MATH 145: Foundations of Mathematics and Introductory Algebra

Instructor: Nickolas Rollick

Fall 2020

Table of Contents

2. Abstract Algebra

Reading 18: Binary Operations and an Introduction to Groups

Reading 19: More on Groups; Cyclic Groups

Reading 20: Subgroups and Symmetric Groups

Reading 21: Homomorphisms and Cosets

Reading 22: Quotient Groups and Lagrange's Theorem

Reading 23: An Introduction to Rings

Reading 24: Subrings and Homomorphisms

Reading 25: Ideals and Quotient Rings

Reading 26: Integral Domains and Divisibility

3. Elementary Number Theory

Reading 27: Elementary Number Theory: Division with Remainder and Greatest Common Divisor

Reading 28: The Euclidean Algorithm

Reading 29: Linear Diophantine Equations and Linear Congruences

Reading 30: Primitive Roots Modulo a Prime and Polynomials over a Field

Reading 31: Chinese Remainder Theorem for Integers

Reading 32: Field of Fractions and Localization

4. Complex Numbers

Reading 33: Complex Numbers: An Introduction

Reading 34: The Fundamental Theorem of Algebra and Solving Equations over the Field of Complex Numbers

Reading 35: Factoring Polynomials with Coefficients in a Field

How to Succeed in Math Class

What do you think is most important for success in math: "talent" or hard work? I'd like to quote at length from one of my favourite books, *Grit*, by Angela Duckworth (p. 16–17):

"Even that first week, it was obvious that some of my students picked up mathematical concepts more easily than their classmates. Teaching the most talented students in the class was a joy. They were, quite literally, 'quick studies.' Without much prompting, they saw the underlying pattern in a series of math problems that less able students struggled to grasp. They'd watch me do a problem once on the board and say 'I get it!' and then work out the next one correctly on their own.

And yet, at the end of the first marking period, I was surprised to find that some of these very able students weren't doing as well as I'd expected. Some did very well, of course. But more than a few of my most talented students were earning lackluster grades or worse.

In contrast, several of the students who initially struggled were faring better than I'd expected. These 'overachievers' would reliably come to class every day with everything they needed. Instead of playing around and looking out the window, they took notes and asked questions. When they didn't get something the first time around, they tried again and again, sometimes coming for extra help during their lunch period or during afternoon electives. Their hard work showed in their grades.

Apparently, aptitude did *not* guarantee achievement. Talent for math was different from excelling in math class.

This came as a surprise. After all, conventional wisdom says that math is a subject in which the more talented students are expected to excel, leaving classmates who are simply 'not math people' behind. To be honest, I began the school year with that very assumption. It seemed a sure bet that those for whom things came easily would continue to outpace their classmates. In fact, I expected that the achievement gap separating the naturals from the rest of the class would only widen over time.

I'd been distracted by talent.

Gradually, I began to ask myself hard questions. When I taught a lesson and the concept failed to gel, could it be that the struggling student needed to struggle just a bit longer? Could it be that I needed to find a different way to explain what I was trying to get across? Before jumping to the conclusion that talent was destiny, should I be considering the importance of effort? And, as a teacher, wasn't it my responsibility to figure out how to sustain effort – both the students' and my own – just a bit longer?

At the same time, I began to reflect on how smart even my weakest students sounded when they talked about things that genuinely interested them. These were conversations I found almost impossible to follow: discourses on basketball statistics, the lyrics to songs they really liked, and complicated plotlines about who was no longer speaking to whom and why. When I got to know my students better, I discovered that all of them had mastered any number of complicated ideas in their very complicated daily lives. Honestly, was getting x all by itself in an algebraic equation all that much harder?"

Granted, Duckworth was talking about a Grade 7 class. But deep down, university math is not much different. I hold a very similar position – effort, patience, and perseverance will get you much further than relying on innate "talent". More importantly, you need a *reason* to put in all that effort, and strategies for success.

It might surprise you to know that I seriously considered quitting my PhD program, despite the fact that everything was going very well. Whenever I hit roadblocks in my research, I took that as evidence that I wasn't good at it. I eventually realized this couldn't be further from the truth. Getting stuck and making mistakes is a completely normal part of doing math, and it happens to *everybody*. It just happens to be

something that isn't talked about as much as it should be... The main difference between someone who is successful in math and someone who isn't is that the successful person never gives up, never stops trying.

Given that you are enrolled in an advanced first-year math course at UWaterloo, you are undoubtedly interested in math, and certainly have some of your own strategies for success. However, it never hurts to have more of them! This is especially true in the current term, where online learning has taken the place of the usual benefits of in-person instruction. With that in mind, here are a few things you can do to be as successful as possible:

- **Complete your readings on time, and engage actively with them.** The course assignments, and the activities we do in the weekly synchronous active learning sessions, depend crucially on the readings posted on Perusall. There are three readings per week, due on Mondays, Wednesdays, and Fridays, closely mimicking the structure of an in-person course, where we would usually meet three times a week. Set aside dedicated time for this reading, and *ask lots of questions* about what you are reading, presenting those questions as annotations on Perusall. Ensure that you are convinced by what you read, and that you follow each deductive step.

- **Participate regularly in the discussions on Perusall and Piazza.** Should you encounter any questions in your reading, or find yourself really curious about extensions to what you are reading about, you are strongly encouraged to post a question on Perusall, or to answer one of your peers' questions. Likewise, you are encouraged to use Piazza to ask and answer questions that are less directly tied to our course readings. Aside from being a part of your course grade, lively dialogue about interesting problems and questions is a key part of doing mathematics, and your classmates may have some great insights to share. Beyond that, by participating in these online conversations, you play a role in building a supportive mathematical community in our class, making your studies much more fun and interesting! This is particularly crucial this term, when online engagement is the *only* form of engagement.

- **Take the reflective responses seriously.** In these responses, you will have frequent opportunities to set goals, propose ways to meet those goals, identify what isn't working, and adapt accordingly. Among other things, you will also be able to reflect on your learning in the course – what topics are most and least important, how you feel about the subject, and so on. By thinking deeply about your own learning, you can take an active role in it, and identify what motivates and excites you. Moreover, these responses are an important communication channel between you and me, which I can use to maintain a virtual dialogue with each of you all the way through the course.

- **If you are able, attend the weekly synchronous active learning sessions, or watch the recordings.** This is especially important in this course, since our in-class activities will be focused on opportunities for mathematical thinking and problem-solving. Even during a regular term of classes, I firmly believe that in-class meetings should be a "value added" experience – if I'm doing something in class that you could get elsewhere just as effectively, then I shouldn't be doing it. The weekly online active learning sessions I'm organizing are conceived of in exactly the same spirit – they are the one opportunity you have to engage with me and large numbers of your peers in real time, benefiting from their perspectives. By participating fully in these sessions, you will extract the greatest benefit.

- **Do not hesitate to contact me, for any reason whatsoever.** If anything you see in this course just isn't making sense, or you are worried about a gap in your background, please reach out to me. I am happy to work with you for as long as it takes. That said, you should definitely make a real effort to understand the topic before approaching me. The more accurately you can identify what you *don't* understand, the more helpful I can be. First and foremost, I care about your *learning* in this course, and improving your mathematical thinking abilities. Grades are entirely secondary to me. If learning is *your* priority, I will do anything I can to help you along.

Of course, this is just a start. Above all, I recognize that you are human beings, not just math students. Your success in this course depends on all kinds of factors above and beyond the math itself. You have the power to guarantee a successful outcome in this course, and I'm here to help.

MATH 145 Course Reading 1: Mathematical Logic, Logical Connectives, and Basic Proof Techniques

September 11, 2020

This course is dedicated to studying the foundations of the branch of mathematics known as "abstract algebra". We do this as a way of illustrating the so-called "mathematical process", which you can view as the mathematical equivalent of the scientific method. Mathematicians start with a set of formal definitions and *axioms*, which form the foundation for the system under investigation. Once these systems are defined, mathematicians will play around with specific examples, and look for patterns or other interesting truths. These are formulated as *conjectures*, and the next step is to try to prove or disprove those conjectures. Once this is done, a mathematician will return to exploring examples again, armed with more information than before.

This process sounds a lot like the scientific method in some ways: a scientist also starts from fundamental assumptions about how a system behaves, experiments with the system, and forms conjectures. But the mathematical process differs from the scientific method in one important respect: a conjecture in mathematics can be proved or disproved beyond a shadow of a doubt, whereas a scientist can only gather evidence for their conjectures.

What makes this possible is the fact that mathematics has formal, logical underpinnings. It comes with its own language, and using this language, you can derive logically certain conclusions from known facts. Therefore, it is important to begin our study with a look at how this mathematical language works. The rules of deduction we discuss here are common to *all* branches of mathematics, making this perhaps the most fundamental object of mathematical study.

Mathematical Statements

The most important feature of a mathematical statement is that it has a well-defined *truth value*. This means a mathematical statement must be either true or false, even if we don't yet know the truth or falsity of the statement. Examples of mathematical statements include:

"All integers are positive."

"The equation $x^4 + x^2 + 3 = 0$ *has a real solution."*

"The real number π *is irrational."*

To stress the point: each of the mathematical sentences above has a consistent truth value assigned to it, even if we don't (yet) know what that truth value is. For instance, you may not know at a glance whether the second statement above is true or false (exercise: which is it?), or how to prove that the third one is true, but that doesn't affect whether or not it counts as a statement.

On the other hand, the sentence "this sentence is false" is not a statement, because there is no way to assign a consistent truth value to it (why not?)

Logical Connectives

Once we've established statements as our mathematical "building blocks", we can put statements together to build new ones. These *compound statements* are built using *logical connectives*. We discuss five of them: conjunction ("and", symbolically denoted $\wedge$), disjunction ("or", denoted $\vee$), negation ("not", denoted $\neg$), implication (denoted $\Rightarrow$), and biconditional (denoted $\Leftrightarrow$).

If P and Q denote mathematical statements, we can form the *conjunction* of the statements, $P \wedge Q$ ("P and Q"), which is true exactly when both P and Q are true. If either P or Q is false, then $P \wedge Q$ is false. For example, the statement

"All integers are positive and the real number π is irrational"

is false, because the statement "all integers are positive" is false. On the other hand, the statement

"The real number π is positive and π is irrational"

is true because both component statements are true.

Similarly, given statements P and Q, we can form the *disjunction* of the statements, $P \vee Q$ ("P or Q"), which is true exactly when at least one of P and Q are true (and false if both P and Q are false). In this case, the statement

"All integers are positive or the real number π is irrational."

is true, because one of the component statements is true. Contrary to some uses of "or" in the English language, this is an *inclusive* or, so that the statement $P \vee Q$ is still true when P and Q are both true.

Given a statement P, the statement $\neg P$ ("not P") has exactly the opposite truth value to P. So if P is true, then $\neg P$ is false, and if P is false, then $\neg P$ is true. For instance, if the statement P is the false statement

"All integers are positive"

one way to translate $\neg P$ into words is "Not all integers are positive", which is true. Given a statement P, the statement $\neg P$ is usually referred to as the *negation* of P.

At this stage, it is useful to introduce the notion of a *truth table*. Given a compound statement, built up from statements and logical connectives, the truth table provides a way of specifying when the compound statement is true, as a function of the truth values of the individual statements that it is built from. So, for instance, we can summarize the truth values of the three types of compound statements we have seen so far in the following truth tables:

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	T	F
F	F	F

P	Q	$P \vee Q$
T	T	T
T	F	T
F	T	T
F	F	F

P	$\neg P$
T	F
F	T

In the table, T denotes that a statement is true, and F denotes that a statement is false. Note that the first columns of these truth tables give all the possible combinations of truth values for the statements that make up the compound statement. So, if we were considering a compound statement built from three simpler statements, P, Q, and R, there would be 8 rows of truth values, covering all possible assignments of true and false to the statements (P, Q, R).

Implications

The next logical connective, implication, is perhaps the most important of all, because the vast majority of mathematical statements are in the form of implications. Given statements P and Q, we can form the compound statement $P \Rightarrow Q$ ("If P, then Q"), with truth values given by the following truth table:

P	Q	$P \Rightarrow Q$
T	T	T
T	F	F
F	T	T
F	F	T

In the implication $P \Rightarrow Q$, P is sometimes called the *hypothesis* of the statement, and Q is sometimes called the *conclusion*. Hence the statement "If P, then Q" is only false if the hypothesis is true and the conclusion is false.

Other (less common) ways mathematicians translate the statement $P \Rightarrow Q$ into words are:

"Q if P."

"P is sufficient for Q."

"Q is necessary for P."

For the moment, it is less important to remember these other formulations, but they do come up from time to time.

Perhaps the most counter-intuitive part of mathematical implications comes in the third line of the truth table above. It can seem strange at first to consider the statement "If P, then Q" to be true when P is false and Q is true. But let's take an every-day example to illustrate. Suppose you're cooking for your young cousin, and they are refusing to eat their vegetables. To help give them some incentive, you make the promise that "If you eat your vegetables, then I will give you some ice cream."

For this statement, the first and last lines of the truth table make complete sense. If your cousin eats their vegetables and you give them some ice cream, you've definitely kept your promise (i.e. your statement was true). If your cousin does not eat their vegetables and you do not give them ice cream, you've still stayed true to your promise (i.e. your statement was true). On the other hand, looking at the second line of the truth table, if your cousin eats their vegetables and you do not give them ice cream, you've broken your promise (i.e. your statement was false).

The third line is the one that might seem a little odd. If your cousin does not eat their vegetables *and you give them ice cream anyway*, have you kept your promise, or broken it? Since doing this *seems* inconsistent with the promise you made, some people might argue that your statement was false in this case. But let's take a closer look – you only promised your cousin that you would do something *if* they ate their vegetables. You have made no such statement about the case where they *don't* eat those vegetables. So if your cousin does not eat their vegetables and you still give them ice cream, you are certainly not lying – at worst, you can be accused of being overly generous! So the statement can't be considered false in this case, and therefore must be true.

Hopefully this interpretation justifies why the statement $P \Rightarrow Q$ is considered true when P is false but Q is true.

Biconditionals

The final logical connective is the biconditional. Given statements P and Q, the compound statement $P \Leftrightarrow Q$ ("P if and only if Q") is the statement with truth values given in the table below:

P	Q	$P \Leftrightarrow Q$
T	T	T
T	F	F
F	T	F
F	F	T

In words, the statement "P if and only if Q" is true exactly when P and Q have the same truth value, and is false otherwise. Another useful way to view these statements uses the notion of *logical equivalence*. Two compound statements R and S are said to be logically equivalent, and we write $R \equiv S$, if they take the same truth values for all truth value assignments of their component statements. In other words, R is logically

equivalent to S when R and S have the same entries in their column in a truth table.

To illustrate, let's prove that $P \Leftrightarrow Q$ is logically equivalent to $(P \Rightarrow Q) \wedge (Q \Rightarrow P)$ by drawing out a truth table. To work out the truth values for the second statement, we break it into smaller component statements, and determine the truth values for those first. This is illustrated below:

P	Q	$P \Rightarrow Q$	$Q \Rightarrow P$	$(P \Rightarrow Q) \wedge (Q \Rightarrow P)$	$P \Leftrightarrow Q$
T	T	T	T	T	T
T	F	F	T	F	F
F	T	T	F	F	F
F	F	T	T	T	T

The truth values in the columns corresponding to $(P \Rightarrow Q) \wedge (Q \Rightarrow P)$ and $P \Leftrightarrow Q$ are identical, and so the two statements are logically equivalent.

For example, the biconditional within the (true) statement

"For all integers n that are perfect squares, n is even if and only if n is a multiple of 4"

can be replaced with the conjunction of two implications:

"For all integers n that are perfect squares, if n is even then n is a multiple of 4, and if n is a multiple of 4 then n is even."

In passing, we should note that the start of the statement, "For all integers n that are perfect squares...", is introducing a *quantifier* into the statement. We will take a closer look at quantified statements in the next reading.

Once you know that two statements are logically equivalent, a proof that a statement in one form is true is the same as a proof that a statement in the other form is true, since the statements are true in exactly the same cases. This will be useful to us in the next section.

Proving Compound Statements

Given the truth tables for the five types of compound statements above, the general strategy for proving each type of compound statement is true can be summarized as follows:

- To prove that a statement of the form $P \wedge Q$ is true, you must independently prove the truth of each of P and Q.
- To prove that a statement of the form $P \vee Q$ is true, it is enough to prove that one of P and Q is true. In fact, since $P \vee Q$ is only false when *both* of P and Q are false, you are actually entitled to assume that P is false, and can make use of that assumption to show that Q is true.
- To prove that a negated statement, $\neg P$, is true, you can prove that the original statement P is false. This is actually more commonly used the other way: to prove that a statement P is false, it is common to instead write a proof that $\neg P$ is true.
- To prove an implication $P \Rightarrow Q$, we see from the truth table that the statement is automatically true whenever P is false. Therefore, we can *assume* that P is true and use this information to show that Q must also be true.
- The usual way to prove a biconditional statement $P \Leftrightarrow Q$ is to use the logical equivalence to the statement $(P \Rightarrow Q) \wedge (Q \Rightarrow P)$, and prove this logically equivalent statement by showing that both $P \Rightarrow Q$ and $Q \Rightarrow P$ are true.

MATH 145 Course Reading 2:
Quantifiers and Proving Quantified Statements

September 14, 2020

Last time, we looked at the foundations of mathematical logic. We introduced the notion of a statement, and looked at different ways to build compound statements using five logical connectives. Already, when formulating those statements, the notion of a quantifier quietly introduced itself.

The vast majority of mathematical statements are quantified: either we wish to prove a statement for *all* objects of a certain type, or else to prove that *at least one* mathematical object of a given type satisfies a given statement. These are formally captured by the notions of *universal* and *existential* quantifiers, respectively. In fact, these two types of quantifiers are the only ones we need to introduce into our mathematical language.

Quantified Statements

As mentioned above, most mathematical statements are quantified. To formulate a quantified statement, we require several ingredients:

- A quantifier (either *universal* or *existential*).
- A *set* of objects being quantified over.
- A *variable* representing an element of the set being quantified over.
- A statement whose truth value may depend on the variable mentioned above.

Let's look at each ingredient in turn. A *universal quantifier* is denoted by the symbol $\forall$, and is usually read as "for all". It makes a claim about *all* the elements in a particular set. An *existential quantifier* is denoted by the symbol $\exists$, and is usually read as "there exists". It makes a claim that there is *at least* one element in a particular set having a given property.

Both types of quantifiers refer back to a *set*. For the moment, we can view these informally as a collection of mathematical objects, enclosed between braces. For example, $\{1, 2, 3\}$ is a set, and so is $\mathbb{Z} = \{\dots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \dots\}$, the set of *integers*. We will be studying the theory of sets in detail in this course, but for now, you can take this informal description of sets as sufficient.

Along with a set, we have a variable to stand in for a member of that set. For instance, if the quantified statement is a claim about all the integers, the symbolic way to start the statement would be something like $\forall n \in \mathbb{Z}$ (read "for all integers n"). The symbol $\in$ denotes *set membership*, which is just a fancy way of saying that n is an element of the set $\mathbb{Z}$. We use the negated symbol $\notin$ to show that an element does not belong to a given set. For example, $1 \in \{1, 2, 3\}$, while $\pi \notin \mathbb{Z}$.

For another example, let's say we're making a claim that a certain equation in one variable x has a real solution. This time, the statement might begin with $\exists x \in \mathbb{R}$ (read "there exists a real number x"). In this course, as is traditional, we will use the blackboard-bold letter $\mathbb{R}$ to denote the set of real numbers, and x is our variable in this case.

The remaining ingredient is a mathematical statement, depending on the variable(s) being quantified over. For each specific value of the variable, that statement will take a truth value, and that truth value can change depending on the value of the variable. For example, let's say we want to express that $x^3 - 2 = 0$ has a real solution, and to do so as a quantified statement. Symbolically, we might write

$$\exists x \in \mathbb{R}, x^3 - 2 = 0.$$

Out loud, we might read this as "there exists a real number x such that $x^3 - 2 = 0$." The "statement" part of this quantified statement, which depends on x, is the claim that $x^3 - 2 = 0$. This statement may be true for some values of x and false for others (and indeed, it is). Because we are using an existential quantifier here, the quantified statement $\exists x \in \mathbb{R}, x^3 - 2 = 0$ is true if we can find *at least one* value for the variable x for which $x^3 - 2 = 0$, and false if there are *no* such values of x.

For another example, let's say we want to assert that every integer has a non-negative square. This is a claim about all integers, so we use a universal quantifier. We might write the claim symbolically as

$$\forall n \in \mathbb{Z}, n^2 \geq 0.$$

Out loud, we might read this as "for all integers n, we have $n^2 \geq 0$." The statement here, depending on n, is the inequality $n^2 \geq 0$. For each value of n, this statement has a truth value. Since we are using a universal quantifier, the quantified statement above is true if the statement $n^2 \geq 0$ is true for *all* integer values of n, and false otherwise.

Proving Quantified Statements

Let's recap what we've said so far.

- A universally quantified statement is of the form $\forall x \in S, P(x)$, where S is a set of objects under consideration, and $P(x)$ is a statement, whose truth value depends on the particular choice of element x from the set S. The quantified statement is true if $P(x)$ is true for every choice of $x \in S$, and false if $P(x)$ is false for at least one $x \in S$.
- An existentially quantified statement is of the form $\exists x \in S, P(x)$, where S and $P(x)$ are as above. The statement is true if $P(x)$ is true for at least one $x \in S$, and false if $P(x)$ is false for all $x \in S$.

In turn, this leads to the following strategies for proving quantified statements are true:

- To prove that a universally quantified statement $\forall x \in S, P(x)$ is true, we let x denote a fixed, but arbitrary element of the set S. Next, using only properties we know x has from belonging to S, we argue that the statement $P(x)$ is true for this arbitrary x.
- To prove that an existentially quantified statement $\exists x \in S, P(x)$ is true, it is enough to give a specific example of an element x in S that satisfies the statement $P(x)$. This example can be obtained in many different ways; examples include trial-and-error, appealing to known facts (theorems), or by using some kind of counting argument.

To illustrate, let's look at proving each of the two main examples of quantified statement given in the previous section. First, let's try to prove that $x^3 - 2 = 0$ has a real solution. This is existentially quantified, so it's enough to give a single value of x. This result is a good illustration of the different levels of rigour that you may need for a proof. Here's one way to write the proof:

Proof. We know that $\sqrt[3]{2} \in \mathbb{R}$, so we let $x = \sqrt[3]{2}$. Then $x^3 - 2 = (\sqrt[3]{2})^3 - 2 = 2 - 2 = 0$. Thus there is a real number x such that $x^3 - 2 = 0$. □

Here, we gave a proof by eyeballing a solution to the equation and verifying that it works. However, what if someone questions the existence of the real number $\sqrt[3]{2}$? After all, by claiming this number exists, doesn't it seem like we're assuming what we're trying to prove?

So, here's an even more careful proof, using tools you might have encountered in calculus class. This illustrates another mathematical technique: appealing to one or more known results as part of your proof (commonly called *lemmas* in such cases). In our particular example, we will use the *intermediate value theorem*:

Lemma 2.1. *Suppose that $f(x)$ is a continuous real-valued function defined on the closed interval $[a, b]$. If $f(a) < 0$ and $f(b) > 0$, or $f(a) > 0$ and $f(b) < 0$, then there is a real number $c \in [a, b]$ such that $f(c) = 0$.*

Using this lemma, we can supply a more careful proof:

Proof. Let $f(x)$ denote the function $x^3 - 2$. We observe that $f(x)$ is a continuous function on the interval $[1, 2]$, and that $f(1) = 1 - 2 = -1 < 0$ and $f(2) = 2^3 - 2 = 6 > 0$. By the Intermediate Value Theorem, there is some real number $c \in [1, 2]$ such that $f(c) = 0$. In particular, c is a solution to the equation $x^3 - 2 = 0$, and so there is a real number x such that $x^3 - 2 = 0$. □

Notice that this proof does not give an explicit value for the real number x satisfying the equation. At best, we know x lies between 1 and 2. Note also that even this more careful proof has gaps we can fill in if we please. For example, why is $x^3 - 2$ a continuous function?

At this stage, you might be wondering how much you have to justify your steps for proofs in this course. The answer is that you should write so that your arguments are convincing to your classmates. If everyone in your class can justify why each of your claims is true, unaided by you, then you have supplied enough detail.

Now let's turn to proving the universally quantified statement $\forall n \in \mathbb{Z}, n^2 \geq 0$. Here, we must let n denote a fixed, but arbitrary integer, and use only general properties of integers to show that $n^2 \geq 0$ is true for this arbitrary choice of n:

Proof. Let n be an arbitrary integer. We consider two cases, according to whether $n \geq 0$ or $n < 0$. If $n \geq 0$, we take this inequality and multiply both sides by n. Since n is nonnegative, multiplying both sides of the inequality by n does not change its direction, and so

$$n \cdot n \geq n \cdot 0 = 0.$$

In other words, $n^2 \geq 0$, proving the statement in this case. Now suppose that $n < 0$. Again, we take the inequality $n < 0$ and multiply both sides by n. However, since n is negative now, multiplication of both sides by n flips the direction of the inequality. Thus $n \cdot n > n \cdot 0 = 0$, and so again $n^2 \geq 0$. Having shown that $n^2 \geq 0$ in both cases, which cover all possibilities, the proof is complete. □

This proof illustrates the common technique of *proof by cases*, where the set S being quantified over is split into smaller pieces, and the statement is proved separately for each piece. We will have more to say on proof strategies like this in the next reading.

Notice also that we took certain facts about inequalities for granted in this proof. You may find it interesting to think about proving them. More specifically, you are invited to think about how to prove that for all real numbers a, b, c, if $a > b$ and $c > 0$, then $ac > bc$, and that if $a > b$ and $c < 0$, then $ac < bc$.

Negating Quantified Statements

Our rules from last time about building compound statements out of simpler ones also apply for statements that involve quantifiers. In particular, negating quantified statements is interesting, because the negations can be simplified somewhat. We discuss each type of quantifier in turn:

- To negate the statement $\forall x \in S, P(x)$, we want a statement with the opposite truth values. The statement $\forall x \in S, P(x)$ is false exactly when some element x in S makes the statement $P(x)$ false. But that's the same as there being an element x of S for which $\neg P(x)$ is true. The negated statement should be true in exactly this case. In other words,

  $$\neg(\forall x \in S, P(x)) \equiv \exists x \in S, \neg P(x).$$

  Note that the negation of a universally quantified statement is existentially quantified.

- To negate the statement $\exists x \in S, P(x)$, we again look for a statement with opposite truth values. The statement $\exists x \in S, P(x)$ is false exactly when $P(x)$ is false for every element of S, i.e. $\neg P(x)$ is true for every element of S. Since the negated statement should be true in exactly these cases, we have

  $$\neg(\exists x \in S, P(x)) \equiv \forall x \in S, \neg P(x).$$

Applying this to our two running examples, the negation of $\exists x \in \mathbb{R}, x^3 - 2 = 0$ is $\forall x \in \mathbb{R}, x^3 - 2 \neq 0$. In words, the negation of the claim that $x^3 - 2 = 0$ has a real solution is the claim that $x^3 - 2$ is never zero for any real number x. The negation of $\forall n \in \mathbb{Z}, n^2 \geq 0$ is $\exists n \in \mathbb{Z}, n^2 < 0$. In words, the negation of the claim that every integer has a nonnegative square is the claim that some integer has a negative square.

Mixed Quantified Statements

In a quantified expression, the statement $P(x)$ can itself be a quantified statement, depending on variables other than x. When this occurs, we get a statement with multiple quantifiers. If a statement includes both types of quantifiers, then it is called a *mixed quantified statement*. Proving or negating such statements can be more complicated, because of the need to deal with all the nested quantifiers. For a famous example from calculus, let's look at the official definition of limit.

Given a real-valued function $f(x)$ and a real number a, we say that $\lim_{x \to a} f(x) = L$ if the following mixed quantified statement is true: for every real number $\epsilon > 0$, there is a real number $\delta > 0$, such that for all real numbers x, if $0 < |x - a| < \delta$, then $|f(x) - L| < \epsilon$. Written more symbolically, if we let $\mathbb{R}^+$ denote the set of positive real numbers, we have

$$\forall \epsilon \in \mathbb{R}^+, \exists \delta \in \mathbb{R}^+, \forall x \in \mathbb{R}, ((0 < |x - a| < \delta) \Rightarrow (|f(x) - L| < \epsilon)).$$

In particular, a proof that the limit of a certain function at a is L would start by taking an arbitrary positive real number ϵ, and giving an explicit choice of positive real number δ that satisfies the implication in the statement above for an arbitrary choice of x. Since the quantifier over δ appears within the statement quantified over ϵ, the chosen value of δ is allowed to depend on ϵ.

At first, we won't be working with a lot of mixed quantified statements, so that is as much as we will say at this point, but we will return to this theme when it becomes more relevant. However, one more thing is worth saying. The ideas we mentioned around negating quantified statements continue to hold for statements with multiple quantifiers. For example,

$$\neg(\exists x \in S, \forall y \in T, P(x, y)) \equiv \forall x \in S, \neg(\forall y \in T, P(x, y)) \equiv \forall x \in S, \exists y \in T, \neg P(x, y).$$

In short, when negating a statement with multiple quantifiers, each existential quantifier is converted to universal, and vice versa, and the innermost statement is negated.

Uniqueness Statements

We conclude this reading with a discussion of a particular type of existentially quantified statement: uniqueness statements. Sometimes, we not only wish to establish that a mathematical object satisfying a given property exists, but also that it is the *only* object in the set with this property. For example, consider an upgraded version of our existence statement above:

*"There is a **unique** real number x such that $x^3 - 2 = 0$."*

Symbolically, we might write this statement as

$$\exists! x \in \mathbb{R}, x^3 - 2 = 0,$$

where we use the symbol $\exists!$ to denote that the statement is true for *exactly* one value of the variable in the specified set. In general, to prove that a statement of the form $\exists! x \in S, P(x)$ is true, we must show two things:

- That the statement $\exists x \in S, P(x)$ is true (to show that *at least one* element of S satisfies $P(x)$).
- If we have *two* elements of S satisfying $P(x)$, then they must be identical. In other words, assume that there are $x, y \in S$ such that $P(x)$ and $P(y)$ are both true, and prove that $x = y$.

Using this template, let's prove the uniqueness statement given above.

Proof. We already showed that $\exists x \in \mathbb{R}, x^3 - 2 = 0$ is true, so now we assume that we have real numbers x and y such that $x^3 - 2 = 0$ and $y^3 - 2 = 0$. This implies that $x^3 - 2 = y^3 - 2$, so $x^3 - y^3 = 0$. Factoring, we get

$$(x - y)(x^2 + xy + y^2) = 0.$$

The product of two real numbers is 0 only if one of the factors in the product is 0, so we deduce that either $x - y = 0$ or $x^2 + xy + y^2 = 0$. If $x - y = 0$, then $x = y$, and we have what we wanted to prove.

Our proof will be complete if we show that the alternative $x^2 + xy + y^2 = 0$ is impossible. Completing the square, we get

$$x^2 + xy + y^2 = (x + (1/2)y)^2 + (3/4)y^2.$$

For any real numbers X and Y, we have $X^2 + Y^2 = 0$ if and only if $X = 0$ and $Y = 0$, so knowing that

$$(x + (1/2)y)^2 + (3/4)y^2 = (x + (1/2)y)^2 + ((\sqrt{3}/2)y)^2 = 0$$

tells us that $(\sqrt{3}/2)y = 0$ and $x + (1/2)y = 0$. The first equation implies $y = 0$, and this together with the second equation implies that $x = 0$. But x and y were solutions to $x^3 - 2 = 0$ and $y^3 - 2 = 0$, and 0 is not a solution to the equation, so we conclude that this case is impossible.

In summary, the assumption that $x^3 - 2 = 0$ and $y^3 - 2 = 0$ led to the conclusion that $x = y$, and so we have finished the proof of the uniqueness statement. □

MATH 145 Course Reading 3: Review of Proof Techniques – Proof by Contrapositive, Contradiction, and Cases

September 16, 2020

By this point, we have now completely reviewed the "mathematical language" we will need to formulate mathematical statements, and have discussed the basics of how to prove them. In particular, we've seen the five common types of logical connectives and the two types of quantifiers. To prepare us further for the types of proofs we will write and encounter in this course, this reading is dedicated to three commonly-occurring proof strategies: proof by contrapositive, proof by contradiction, and proof by cases.

Proof by Contrapositive

When implications were introduced, we mentioned that the vast majority of mathematical statements we will encounter take the form of (possibly quantified) implications. Recall that the statement $P \Rightarrow Q$ is automatically true whenever P is false. On the other hand, if P is true, the statement $P \Rightarrow Q$ is only true when Q is also true. So, we can begin a proof of $P \Rightarrow Q$ by *assuming* the truth of the statement P, and using that to show the truth of the statement Q necessarily follows. A proof of an implication that proceeds in this manner is referred to as a *direct proof*.

However, sometimes it is not very convenient to work with P as an assumption in a proof of the implication $P \Rightarrow Q$. In such cases, we can instead seek to prove the so-called *contrapositive* implication. The contrapositive of the implication $P \Rightarrow Q$ is the implication $\neg Q \Rightarrow \neg P$. The method of proof by contrapositive is justified by the fact that an implication and its contrapositive are logically equivalent, i.e. for any statements P and Q we have

$$(P \Rightarrow Q) \equiv (\neg Q \Rightarrow \neg P).$$

This logical equivalence can be verified by constructing the truth table for the compound statement $\neg Q \Rightarrow \neg P$ and verifying that its truth values align with those of $P \Rightarrow Q$:

P	Q	$\neg P$	$\neg Q$	$\neg Q \Rightarrow \neg P$
T	T	F	F	T
T	F	F	T	F
F	T	T	F	T
F	F	T	T	T

To give a classic example of the usefulness of proof by contrapositive, let's formally introduce the definition of even and odd integers:

Definition 3.1. An integer n is called *even* if there exists an integer k such that $n = 2k$. An integer n is called *odd* if there exists an integer k such that $n = 2k + 1$.

With these definitions in place, consider the following implication, which we couch as a mathematical proposition:

Proposition 3.1. *Let $n \in \mathbb{Z}$. If n^2 is even, then n is even.*

First, a note on the statement is in order. Mathematicians often write "let $n \in \mathbb{Z}$" or "let n be an integer" in place of a universal quantifier $\forall n \in \mathbb{Z}$. Hence the above statement has the symbolic translation

$$\forall n \in \mathbb{Z}, (\exists k \in \mathbb{Z}, n^2 = 2k) \Rightarrow (\exists \ell \in \mathbb{Z}, n = 2\ell).$$

If we tried to prove this implication directly, we would start by assuming $n^2 = 2k$ for some integer k. We would like to conclude something about n, but the only obvious thing to do is to take square roots to get

$|n| = \sqrt{2k}$, and then we cannot do anything immediately useful with $\sqrt{2k}$.

If we take for granted that every integer is either even or odd, and not both, then the contrapositive of the statement "If n^2 is even, then n is even" will be the statement "If n is odd, then n^2 is odd". So, to prove our original statement, it is enough to prove the contrapositive implication: for all $n \in \mathbb{Z}$, if n is odd, then n^2 is odd.

Here's how we might write out the proof by contrapositive formally:

Proof. We prove the implication "If n^2 is even, then n is even" by contrapositive. Thus, assume that n is not even, i.e. that n is odd. Our goal is to prove that n^2 is odd. By definition, since n is odd, this means $n = 2k + 1$ for some integer k. But note that

$$n^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1.$$

Since k is an integer, so is $2k^2 + 2k$. Thus the equation above shows that $n^2 = 2\ell + 1$, where the integer ℓ is equal to $2k^2 + 2k$. This proves that n^2 is odd when n is odd, completing our proof by contrapositive. □

As a rule, an implication $P \Rightarrow Q$ is easier to prove by contrapositive if $\neg Q$ seems like a more useful assumption to work with than P. In the implication above, the assumption that n is odd was more useful in proving something about n^2 than the assumption that n^2 is even was for proving something about n. For one more example where proof by contrapositive is useful, consider the following:

Proposition 3.2. *Let $x \in \mathbb{R}$. If $x^3 + x + 1 < 0$, then $x < 0$.*

Proof. Let x be an arbitrary real number. We prove the implication by contrapositive, by showing that if $x \geq 0$, then $x^3 + x + 1 \geq 0$. Since $x^2 \geq 0$ for all real numbers x, we can multiply both sides of the inequality $x \geq 0$ by x^2 to get that $x^3 \geq x^2 \cdot 0 = 0$. Thus $x^3 \geq 0$, $x \geq 0$, and $1 \geq 0$. Since the sum of non-negative real numbers is non-negative, we conclude that $x^3 + x + 1 \geq 0$, as desired. □

In this proof, notice that the assumption $x \geq 0$ was much more useful for proving something about $x^3 + x + 1$ than the assumption that $x^3 + x + 1 < 0$ would have been for proving something about x.

Proof by Contradiction

Just as the method of proof by contrapositive allows us to assume the negation of part of our statement, proof by contradiction works on a similar philosophy. Sometimes we wish to prove a statement P whose definition is hard to work with directly, but whose negation $\neg P$ is easier to work with. This proof strategy allows us to assume $\neg P$ is true, and make a series of deductions that lead to a *contradiction*, something logically inconsistent. Usually, the contradiction is obtained by deducing the truth of R and $\neg R$, for some statement R. Since a statement and its negation cannot both be true at the same time, our initial assumption that $\neg P$ is true must have been incorrect.

One of the most famous examples of proof by contradiction is the standard proof of the following result:

Theorem 3.1. *The real number $\sqrt{2}$ is irrational.*

Since irrationality of a real number is a difficult notion to work with directly, it is easier to assume that the statement is false, and work with the definition of *rational* number instead. Here's how this goes:

Proof. For a contradiction, assume that $\sqrt{2}$ is in fact a rational number. By definition of rational number, this means we may write $\sqrt{2} = \frac{a}{b}$ for some integers a and b, where $b \neq 0$. Furthermore, we may assume we have written $\frac{a}{b}$ as a fraction in lowest terms, so that a and b do not share any common factors.

Squaring both sides, we get $2 = \frac{a^2}{b^2}$, so that

$$2b^2 = a^2.$$

By definition of even number, we see from the above that a^2 is even. Since a^2 is even, applying Proposition 3.1, which we proved above, we deduce that a is also even. By definition, we may write $a = 2k$ for some integer k. Thus the above equation may be rewritten

$$2b^2 = (2k)^2 = 4k^2,$$

and cancelling a factor of 2 from both sides, we get

$$b^2 = 2k^2.$$

But now this tells us b^2 is even, so that b is even by Proposition 3.1.

At this point, we have now shown that both a and b are even, so that both are divisible by 2. This contradicts our assumption that the fraction $\frac{a}{b}$ had been written in lowest terms. In turn, this contradiction tells us that our initial assumption that $\sqrt{2}$ is rational must be incorrect. So, $\sqrt{2}$ is irrational after all. □

In this proof, we appealed to the fact that every rational number may be written in lowest terms. This is something we have not yet proved, but which we will address in our study of elementary number theory later in this course.

For one more illustration of proof by contradiction, we will prove

Proposition 3.3. *An integer cannot be both even and odd.*

Since it is difficult to directly prove that something is *not* the case, we will proceed by assuming its negation, which is easier to work with, and deducing a contradiction. The negation of the universally quantified statement "an integer cannot be both even and odd" is the existentially quantified statement "there is an integer that is both even and odd". This gives us the starting point for our proof:

Proof. For a contradiction, suppose there is an integer n that is both even and odd. Since n is even, we know we may write $n = 2k$ for some integer k. But since n is odd, we may also write $n = 2\ell + 1$ for some integer ℓ. This gives us the equation $2k = n = 2\ell + 1$. This implies

$$1 = 2k - 2\ell = 2(k - \ell),$$

which shows that 1 is even. This is a contradiction, so our assumption that there is an integer that is both even and odd must be incorrect. In conclusion, an integer cannot be both even and odd. □

In the above, note that the proof would have been invalid if we had used k again instead of ℓ in the definition of odd. (Why?). Thus it's important to introduce new variables to avoid unintended relationships to variables previously defined in the proof!

Proof by Cases

Our final proof strategy is one that we've already encountered, in our proof that every integer has a non-negative square. The idea behind a *proof by cases* is that we wish to prove a universally quantified statement by partitioning the *domain* of the statement (the set being quantified over) into smaller pieces, and proving the statement independently for each of these smaller pieces. When we do this, it's important that the cases, taken together, cover all the possibilities.

In the proof that every integer has a non-negative square, we broke into two cases, according to whether the integer was negative or non-negative. This is a common division into cases, with another common example being cases based on divisibility. This is illustrated in the proof of the following result:

Proposition 3.4. *For every integer n, the integer $n^2 + n + 1$ is odd.*

Proof. We consider two cases, according to whether n is even or n is odd.

Case 1: n is even. In this case, we may write $n = 2k$ for some integer k. It follows that

$$n^2 + n + 1 = (2k)^2 + (2k) + 1 = 4k^2 + 2k + 1 = 2(2k^2 + k) + 1,$$

where $2k^2 + k \in \mathbb{Z}$. This proves that $n^2 + n + 1$ is odd in this case.

Case 2: n is odd. In this case, we may write $n = 2\ell + 1$ for some integer ℓ. It follows that

$$n^2 + n + 1 = (2\ell + 1)^2 + (2\ell + 1) + 1 = (4\ell^2 + 4\ell + 1) + (2\ell + 1) + 1 = 4\ell^2 + 6\ell + 3 = 2(2\ell^2 + 3\ell + 1) + 1,$$

where $2\ell^2 + 3\ell + 1 \in \mathbb{Z}$. This proves that $n^2 + n + 1$ is odd in this case as well.

Since these two cases cover all possibilities for n, and the statement is true in both cases, we have proved the statement for all integers n. □

MATH 145 Course Reading 4: Introduction to Axiomatic Set Theory – The Logic and Language of Sets

September 18, 2020

Now that we have the logic of mathematical statements in place, we turn to studying the most fundamental mathematical structure – the *set*. Over the next several weeks, we will undertake a study of *axiomatic set theory*, which seeks to define and rigorously study sets using only a handful of fundamental assumptions (*axioms*) about sets and their properties.

What's In a Set?

The reason that axiomatic set theory is so fundamental is that many (in fact, most) other types of mathematical objects may be ultimately defined in terms of sets. Certainly, sets of objects occur everywhere in mathematics, and occur as an essential part of quantified statements. For example, we have the following commonly occurring sets of numbers:

$$\begin{aligned} \mathbb{N} &= \{0, 1, 2, 3, \dots\}, && \text{the set of } \textit{natural numbers} \\ \mathbb{Z} &= \{\dots, -2, -1, 0, 1, 2, \dots\}, && \text{the set of } \textit{integers} \\ \mathbb{Q} &= \{\frac{a}{b} : a, b \in \mathbb{Z}, b \neq 0\}, && \text{the set of } \textit{rational numbers} \\ \mathbb{R}, && && \text{the set of } \textit{real numbers} \\ \mathbb{C} &= \{a + bi, a, b \in \mathbb{R}\}, && \text{the set of } \textit{complex numbers}. \end{aligned}$$

At least the first four should be quite familiar to you; we will undertake a more formal study of complex numbers towards the end of the course.

Interestingly, while we consider each of these to be sets of *numbers*, it turns out that we can formulate the elements of each of these sets *as sets*. The basic idea is to reduce each set of numbers to a simpler set of numbers, and to treat the simplest set of numbers, $\mathbb{N}$, as a set of sets.

For instance, you'll notice that the set of complex numbers really amounts to a set of pairs of real numbers. In turn, as you will learn when studying real analysis, each real number can essentially be treated as a sequence of rational numbers (more or less, you can think of a real number as the sequence of its truncated decimal expansions, each of which is a rational number). In turn, a rational number can be treated as a pair of integers, where the second element in the pair is non-zero.

In a similar way, $\mathbb{Z}$ can be built from $\mathbb{N}$: every negative integer $-n$ can be thought of as a difference $0 - n$ of natural numbers, and thus as a pair $(0, n)$ of natural numbers. Each nonnegative integer n can also be thought of in this way, as the difference $n - 0$, or as the pair $(n, 0)$ of natural numbers. And as we will soon see, there is a precise prescription for thinking of each natural number as a set.

Even fundamental mathematical constructions like ordered pairs and functions can be treated as sets: an ordered pair of elements is just a special type of set built from those elements, and a function can be viewed as a set of ordered pairs (and thus as a set of special types of sets).

The moral of this story is two-fold. Firstly, almost every mathematical object can be viewed as a set, once the right translation is made. This justifies studying set theory as one of the foundational subjects of mathematics. Secondly, since all the elements of sets can themselves be treated as sets, the "language" of set theory is somewhat simplified: the only primitive objects we work with are sets, and no distinction needs to be made between "sets" and "elements" in statements like $x \in \mathbb{R}$, since both x and $\mathbb{R}$ are indeed sets.

Properties and Paradoxes

As we will see, the axioms of set theory are carefully constructed in order to precisely define what should count as a set. The reason this is done is that it is very easy to introduce mathematical contradictions into the theory if we are too broad about what constitutes a set.

One of the first attempts to formally define a theory of sets allowed a set to be determined by any property whatsoever. Here, a *property* is a mathematical statement that a set may or may not satisfy. Thus, it was proposed that for any property $P(X)$ that a set X might satisfy, one could construct the set of sets X for which $P(X)$ is true, denoted using the familiar set-builder notation by

$$\{X|P(X)\}.$$

The difficulty is that even a relatively simple property $P(X)$ can lead to contradictory results. This was formally outlined by Bertrand Russell in 1901, in the case where the statement $P(X)$ is $X \notin X$. In other words, the set

$$S = \{X|X \notin X\}$$

we are trying to construct is the set of sets that are not an element of themselves.

The paradox arises when we ask whether or not the set S we've just apparently constructed is an element of S. If $S \in S$, then by definition of S, we have $S \notin S$, a contradiction. On the other hand, if $S \notin S$, then by definition of S, we must have $S \in S$, again a contradiction.

This particular contradiction is often called *Russell's paradox*. Another popular formulation of Russell's paradox is known as the *barber paradox*: consider a hypothetical barber in a small town that cuts the hair of exactly the people who don't cut their own hair. The paradox is: should this barber be cutting their own hair? If they do, then they shouldn't be, and if they don't, then they should be!

In order to avoid paradoxes like these, the aim of axiomatic set theory is to lay out several "self-evident" properties of sets (the axioms), and from there, deduce all the properties of sets that we hope to be true. Most importantly, we will want an axiomatic system that is *sound*: that all the things we can prove from the axioms are true, and in particular, that it's impossible to derive a contradiction from them.

The Language of Sets

We will explore many of the axioms in our next reading, but for now, let's introduce the so-called "language of sets": the symbols that we build statements about sets from. In addition to all the logical connectives and the two quantifiers, the other relationship we will need between sets is *set membership*, denoted by $\in$. We have already used this set membership symbol informally several times; from the point of view of set theory, set membership is the one set-theoretic relationship that is built into the language of sets.

Other relations between sets are then defined using mathematical statements in the language of sets. The subset relation is a major example of this. We define a relation $\subseteq$ between sets by declaring that the statement $X \subseteq Y$ is true exactly when every element of X is an element of Y. Or using the formal language of sets, the statement $X \subseteq Y$ is a shorthand for

$$\forall x \in X, x \in Y.$$

We will formally define other familiar set-theoretic constructions, such as the *empty set* $\varnothing$, the *union* $X \cup Y$, and the *intersection* $X \cap Y$, directly from the axioms in the next reading.

MATH 145 Course Reading 5: Six Fundamental Axioms – Existence, Extensionality, Comprehension, Pair, Union, and Power Set

September 21, 2020

With the language of set theory established, we can now begin stating and investigating the various axioms of set theory. For now, we will content ourselves with six of them: the Axioms of Existence, Extensionality, Comprehension, Pair, Union, and Power Set. While these axioms are not a complete list, they will allow us to establish many of the familiar set-theoretic properties, and keeping only to these axioms will highlight their limitations. This will motivate the introduction of further axioms for overcoming those limitations.

Existence and Exensionality

In order to get started with a study of sets, it's necessary for there to *be* sets to study. Without an axiom saying so, our universe of sets may very well be empty! In fact, we say more than just that a set exists, but in fact that an *empty* set exists. This is our Axiom of Existence:

Axiom 5.1 (Axiom of Existence)**.** There is a set containing no elements. For this set A, the statement $X \in A$ is false for all sets X.

Now that we have a set to work with, the other important thing we need to establish is how to recognize when two sets are the same. Since a set is intuitively a collection of objects, two sets should be considered the same if and only if they have the same elements. The Axiom of Exensionality formalizes this intuition:

Axiom 5.2 (Axiom of Extensionality)**.** If two sets have the same elements, then they are equal. In other words, for all sets X and Y, if every element of X is an element of Y and every element of Y is an element of X, then $X = Y$.

The Axiom of Extensionality is a useful tool for proving that a set defined by a given property is unique. Without it, we wouldn't have much to go on! For instance, let's combine this with the Axiom of Existence to show that there is only one empty set:

Theorem 5.1. *There is a unique set containing no elements, which we denote by* $\varnothing$*.*

Proof. The Axiom of Existence already tells us that a set with no elements exists, so we need only prove uniqueness. Suppose we have two sets A_1 and A_2, each with no elements. Then every element of A_1 is vacuously an element of A_2, because there are no elements of A_1 in the first place! Similarly, every element of A_2 is an element of A_1, simply because there are no elements of A_2. By the Axiom of Extensionality, we conclude that $A_1 = A_2$, so that the empty set is unique. □

Comprehension

The next axiom tries to capture the intuition that we should be able to define a collection of sets satisfying a given property, without succumbing to the difficulties of Russell's paradox. The way around such contradictions is to define the collection only when the sets come from a set already known to exist. This is captured by the Axiom Schema of Comprehension:

Axiom 5.3 (Axiom Schema of Comprehension)**.** Let $P(X)$ be a property of sets (a mathematical statement in the language of sets that is true or false depending on the set X). For any set A, there is a set B defined by the property that $X \in B$ if and only if $X \in A$ and the statement $P(X)$ is true for X.

Just a little note on terminology: the above is referred to as an *axiom schema* rather than an axiom, because it really is a large collection of axioms, one for every possible property $P(X)$ of sets that we might care to define.

Once we have shown that the set B in the axiom schema is uniquely determined by the set A and statement $P(X)$ (which we do immediately below), we will usually adopt set-builder notation to denote this unique set. More precisely, the set B in the axiom schema will often be written

$$\{X \in A : P(X)\}.$$

Let's justify this notation by proving uniqueness:

Theorem 5.2. *For every set A and property $P(X)$ of sets, the set B defined by A and $P(X)$ in the Axiom Schema of Comprehension is unique.*

Proof. Again, the Axiom Schema already tells us that such a set exists, so we only need to worry about uniqueness. Suppose we have sets B_1 and B_2, both determined by the property that a set X belongs to these sets if and only if $X \in A$ and $P(X)$ is true. In particular, for any set X, if $X \in B_1$, then $X \in A$ and $P(X)$ is true, so that $X \in B_2$. Similarly, if $X \in B_2$, then $X \in B_1$. By the Axiom of Extensionality, we conclude that $B_1 = B_2$, so that the set B given in the Axiom Schema of Comprehension is unique. □

The axioms we've outlined so far do not give us a very rich theory of sets, unfortunately. In fact, you can show that assuming only Existence, Extensionality, and Comprehension, the only set guaranteed to exist is the empty set (this is a good exercise to discuss!) Therefore, we will need further axioms in order to capture the full power of the theory of sets.

Pair, Union, and Power Set

The Axiom of Pair that we're about to introduce will finally guarantee us a set other than the empty set. In fact, it will allow us to construct an infinite family of sets!

Axiom 5.4 (Axiom of Pair)**.** Given any sets A and B, there is a set C whose elements are exactly A and B. In other words, for all sets X, we have $X \in C$ if and only if $X = A$ or $X = B$.

As usual, you can use the Axiom of Extensionality to prove that the set C in the Axiom of Pair is uniquely determined by A and B (this is a great exercise!). Thus, we will adopt the notation $\{A, B\}$ for the set containing the elements A and B, and the shorthand $\{A\}$ for the set $\{A, A\}$.

At the moment, the only set we know to exist is the empty set. If we apply the Axiom of Pair with $A = B = \varnothing$, we get the set $\{\varnothing, \varnothing\} = \{\varnothing\}$. Notice that the set $\{\varnothing\}$ is different from the set $\varnothing$, since $\varnothing$ has no elements, while $\{\varnothing\}$ has one element, namely $\varnothing$.

Now that we have two different sets, we can again use the Axiom of Pair to construct a third one. Applying it to $A = \varnothing$ and $B = \{\varnothing\}$, we construct the set $\{\varnothing, \{\varnothing\}\}$, which is different from both $\varnothing$ and $\{\varnothing\}$, because it is a set containing two distinct elements.

The next axiom we introduce also allows us to create larger sets out of smaller ones. It corresponds to the intuitive notion of taking the union of an arbitrary collection of sets, and thus goes by the name Axiom of Union:

Axiom 5.5 (Axiom of Union)**.** For any set S, there exists a set U such that, for any set X, $X \in U$ if and only if $X \in A$ for some set $A \in S$.

Let's parse out this axiom, since the official formulation might be tricky to digest. We begin with a set S, and use it to construct a set U, which will represent the union of all the elements of S (remember that each element of S is a set!). A set belongs to the union U exactly when it belongs to some member of S.

As usual, we can use the Axiom of Extensionality to show that the union of elements in a set S is unique, and we adopt the notation $\bigcup S$ to represent this union of elements in S.

Now, if you're like me, having encountered this axiom, you might be asking: have we gained anything from the Axiom of Union? Can we obtain any set with this axiom that we could not obtain from the four axioms

introduced before it? In particular, why is the Axiom of Union not just a particular special case of the Axiom Schema of Comprehension?

To answer this latter question, note that the set constructed in the Axiom of Union from the set S is a "bigger" set than S is, and the Axiom Schema of Comprehension only allows us to construct sets that are subsets of some known starting set A. You might imagine trying to construct the union of elements in S via the Axiom Schema by declaring a statement $P(X)$ to be $\exists B \in S, X \in B$, and constructing the union as

$$\{X \in A : P(X)\}.$$

But what should the set A be? It has to be a set that contains all the elements of the sets in S already, which would amount to assuming what we're trying to prove. You will investigate what is gained by the Axiom of Union in more detail on the second assignment.

To give a couple examples of how this axiom is applied, if we use it on the set $S = \{\varnothing, \{\varnothing\}\}$, we get $\bigcup S = \{\varnothing\}$ (do you see why?). Similarly, $\bigcup \varnothing = \varnothing$. We also introduce some familiar notation for finite unions: given two sets A and B, we usually write $A \cup B$ instead of $\bigcup\{A, B\}$.

Our final axiom of this section is the Axiom of Power Set. This re-visits the notion of subset formally introduced in the previous reading. Essentially, what we would like to do is take a set, and construct the set of all its subsets, which is called its *power set*.

Axiom 5.6 (Axiom of Power Set)**.** For any set A, there exists a set $\mathcal{P}$ such that, for any set X, $X \in \mathcal{P}$ if and only if $X \subseteq A$.

Once more, you can show that the set $\mathcal{P}$ given in the axiom is unique (exercise!). We adopt the notation $\mathcal{P}(A)$ for the power set of A.

Again, why does this axiom not follow from the other axioms given so far, particularly the Axiom of Comprehension? While we could declare a property $P(X)$ to be the statement $X \subseteq A$, we can't construct the power set through set-builder notation because we don't know there is a set B that contains all the elements of $\mathcal{P}(A)$ in advance. So there's no way to express the power set as

$$\{X \in B : X \subseteq A\}$$

for some set B that we know exists ahead of time.

Again, for some quick examples we have $\mathcal{P}(\varnothing) = \{\varnothing\}$, $\mathcal{P}(\{\varnothing\}) = \{\varnothing, \{\varnothing\}\}$, and for any sets x_1 and x_2, we have $\mathcal{P}(\{x_1, x_2\}) = \{\varnothing, \{x_1\}, \{x_2\}, \{x_1, x_2\}\}$. In general, for any set A, both the empty set and A itself are subsets of A, and so both $\varnothing$ and A are always elements of $\mathcal{P}(A)$.

Other Set Constructions

Given the axioms defined so far, we can now introduce other familiar set-theoretic constructions. While we needed a new axiom to construct the union of sets, this is not necessary for the intersection. Given sets A and B, we can build the intersection of A and B, denoted $A \cap B$, by applying the Axiom Schema of Comprehension. We can take A as our ambient set in the set-builder notation, and let the statement $P(X)$ be the statement $X \in B$. In other words, we define

$$A \cap B = \{X \in A : X \in B\}.$$

This captures exactly the familiar notion of intersection of sets: a set belongs to $A \cap B$ if and only if it belongs to both A and B.

Similarly, we can build the set difference $A \setminus B$ in this way, given any two sets A and B. We want a set of elements that belong to A but not to B, so we take A to be the ambient set, declare the statement $P(X)$ to be $X \notin B$, and define

$$A \setminus B = \{X \in A : X \notin B\}$$

via the Axiom Schema of Comprehension. Further familiar set constructions (such as the symmetric difference of sets) are possible, but we will content ourselves with the two common constructions just discussed.

MATH 145 Course Reading 6:
Construction of the Natural Numbers, and the Axiom of Infinity

September 23, 2020

Given all the axioms for sets we've now introduced, what is still missing? What types of sets are we now able to create, and which ones are we not able to create? One major deficiency is that the six axioms so far only allow us to build *finite* sets – infinite sets do not have to exist without the addition of further axioms. In essence, this is true because all of our axioms so far (at least the ones that allow us to build bigger sets from smaller ones) take the form "for all sets A, there is a set X such that..." where the condition that ends the statement guarantees X is finite if A is finite. And since the only set we assumed to exist through axioms was the empty set (which is finite), we expect that the only sets we can construct with the first six axioms are again finite.

But in turn, this raises another question – how do we officially define finite and infinite sets? We'll turn to this question in a few weeks when we discuss the *cardinality* of a set, but you may want to start thinking now about what might be involved!

Taking the intuitive notion of finite and infinite as sufficient for now, it becomes clear that we'll need to introduce another axiom even to construct a set like $\mathbb{N}$. And to avoid having to define an infinite set at this juncture, we will not directly assume the existence of an infinite set, but rather, something called an *inductive* set. This definition will set us up for a discussion of the famous Principle of Mathematical Induction for the set of natural numbers.

Natural Numbers and Successors

When we begin thinking about defining the natural numbers as a set, one guiding principle we might use is the following: we would like the natural number n to be a set with exactly n elements. The empty set is the unique set with no elements, so this forces a choice on us for the natural number 0:

$$0 = \varnothing.$$

On the other hand, there are lots of choices for sets with one element. For instance, $\{\varnothing\}$, $\{\{\varnothing, \{\varnothing\}\}\}$, and $\{\{\varnothing\}\}$ all have a single element, and more generally, for any set x, the set $\{x\}$ has one element as well. The amount of choice only increases as we look for candidate sets having two, three, or even more elements. So how can we define $1, 2, 3, \ldots$ in the most elegant way?

One natural way to proceed (pun totally intended) is to define each new natural number in terms of the natural numbers already defined. So of all the possible one-element sets, we define 1 to be the set $\{0\}$, which is the set $\{\varnothing\}$. In the same way, it seems natural to take $2 = \{0, 1\} = \{\varnothing, \{\varnothing\}\}$. Carrying on in this fashion, this leads to a recursive (or inductive) construction where we define the natural number n to be the set $\{0, 1, \ldots, n-1\}$, solely in terms of the natural numbers already defined.

In fact, you will notice that another way to define the next natural number $n+1$ (given that the natural numbers up to n have been defined) is to take $n+1 = n \cup \{n\}$, since the elements of $n+1$ are exactly the elements of n, along with n itself. This motivates the following definition, which is valid for an arbitrary set:

Definition 6.1. For any set x, the *successor* of x is the set $S(x) = x \cup \{x\}$.

Roughly speaking then, the natural numbers are exactly the set of elements you would get from applying the successor operation to the empty set a finite number of times.

Inductive Sets and the Axiom of Infinity

However, we do not yet have a way of explicitly defining the natural numbers in terms of the axioms of set theory introduced thus far. In order to resolve this, we define the notion of an *inductive set*:

Definition 6.2. A set I is called *inductive* if it has the following two properties:

1. $0 \in I$.
2. If $n \in I$, then $S(n) \in I$ (i.e. $n + 1 \in I$).

Whatever the natural numbers turn out to be, we certainly want them to be an inductive set! In order to construct the natural numbers as part of our axiomatic set theory, we make it an axiom that an inductive set exists:

Axiom 6.1 (Axiom of Infinity)**.** An inductive set exists.

To use this axiom to construct $\mathbb{N}$, notice that given the way the definition of an inductive set is phrased, it seems that every inductive set should contain the natural numbers as a subset. In turn, this leads us to define $\mathbb{N}$ to be the set of elements that belong to every inductive set. In other words, we let $\mathcal{I}$ be the inductive set that exists by the Axiom of Infinity, and use the Axiom of Comprehension to define

$$\mathbb{N} = \{x \in \mathcal{I} : x \in I \text{ for all inductive sets } I\}.$$

As usual, $\mathbb{N}$ is uniquely determined, thanks to the Axiom of Extensionality.

Now that we have constructed $\mathbb{N}$, we can verify that it is indeed inductive:

Theorem 6.1. *The set $\mathbb{N}$ is inductive.*

Proof. Here, we just need to verify the two properties of an inductive set. First, we show that $0 \in \mathbb{N}$. For every inductive set I, we know that $0 \in I$ by definition. Hence, the definition of $\mathbb{N}$ implies that $0 \in \mathbb{N}$ as well. Now, suppose we know that an element n belongs to $\mathbb{N}$. By definition, this means $n \in I$ for every inductive set I. For each such set I, we know that $n + 1 \in I$ by definition of inductive set. So $n + 1$ belongs to every inductive set, hence belongs to $\mathbb{N}$ as well. In turn, this completes the proof that $\mathbb{N}$ is inductive. □

Finally, to conclude this axiomatic construction of the natural numbers, we use the details of our construction to define a way of ordering the natural numbers (determining when a natural number is less than another):

Definition 6.3. Let $m, n \in \mathbb{N}$. We say that $m < n$ (said "m is less than n") if $m \in n$.

Notice that this agrees with our conventional usage of the less-than symbol. For example, the statement $2 < 5$ is true because $2 \in 5 = \{0, 1, 2, 3, 4\}$. This is an initial example of an *order relation* on a set. In the next couple of weeks, we will look carefully at the notion of a relation on a set, and order relations like this one in particular.

MATH 145 Course Reading 7:
The Principle of Induction and the Well-Ordering Principle

September 25, 2020

For our first goal, we will derive the famous Principle of Mathematical Induction, as a direct consequence of the way we defined the natural numbers via the Axiom of Infinity. In this reading, we state and apply the principle exclusively to study properties of $\mathbb{N}$ itself. However, in the next reading, we will have much more to say about the way induction arguments are applied throughout mathematics, along with some famous examples worth looking at.

The Principle of Induction

Informally speaking, arguments by induction show that a property holds for all natural numbers by showing that it holds for the first natural number (which is 0, under our conventions), and then proving that if the statement is true for a natural number n, then it is true for the next natural number $n + 1$. We formally state and prove this principle below:

Theorem 7.1 (Principle of Induction)**.** *Suppose that $P(x)$ is a statement, depending on the variable x. Assume that the statement $P(0)$ is true, and that for all $n \in \mathbb{N}$, if $P(n)$ holds, then $P(n+1)$ holds. Under these hypotheses, the statement $P(n)$ is true for all $n \in \mathbb{N}$.*

Proof. Using the Axiom Schema of Comprehension, we define the set $A = \{n \in \mathbb{N} : P(n)\}$. In particular, note that $A \subseteq \mathbb{N}$. Since $P(0)$ is true, we know that $0 \in A$. Furthermore, for each natural number n, we know that if $n \in A$, then $P(n)$ is true, so $P(n+1)$ is true, so $n+1 \in A$. Together, these two facts establish that A is an inductive set. Since $\mathbb{N}$ is a subset of every inductive set by definition, we get that $\mathbb{N} \subseteq A$. Combining this with $A \subseteq \mathbb{N}$, we conclude that $A = \mathbb{N}$, so that $P(n)$ is true for all $n \in \mathbb{N}$. □

Let's now apply the principle to rigorously prove an "obvious" fact about the ordering on the natural numbers:

Proposition 7.1. *For all $n \in \mathbb{N}$, we have $0 \leq n$. (Here, as usual, $0 \leq n$ means that either $0 = n$ or $0 < n$.)*

Proof. We proceed by induction, where the statement $P(x)$ is the statement $0 \leq x$. Note that $P(0)$ is true, since $0 = 0$. Now let $n \in \mathbb{N}$ be arbitrary, and assume that $P(n)$ is true, i.e. $0 \leq n$. Our goal is to show that $P(n+1)$ is also true, so that $0 \leq (n+1)$. Now, our assumption that $0 \leq n$ means either that $0 = n$ or $0 < n$. Either way, it is immediate from the definition that $0 \in n \cup \{n\}$, i.e. $0 \in (n+1)$, proving that $0 \leq (n+1)$. By the Principle of Induction, the statement $0 \leq n$ is true for all natural numbers n. □

The Principle of Strong Induction

There's another common version of induction out there, one that is often more convenient to use, but yet logically equivalent to the Principle of Induction. We call it the Principle of Strong Induction, because it apparently allows us to make a stronger assumption than the ordinary Principle of Induction does.

Theorem 7.2 (Principle of Strong Induction)**.** *Suppose that $P(x)$ is a statement, depending on the variable x. Assume that the following implication is true for all $n \in \mathbb{N}$:*

$$\text{If } P(k) \text{ is true for all natural numbers } k < n, \text{ then } P(n) \text{ is true.} \tag{1}$$

Then $P(n)$ is true for all $n \in \mathbb{N}$.

Proof. We apply the Principle of Induction to the statement $Q(n)$, which is the statement

$$P(k) \text{ is true for all natural numbers } k < n.$$

First, note that $Q(0)$ says "$P(k)$ is true for all natural numbers $k < 0$". This statement is automatically true, since there are no natural numbers $k < 0$ by Proposition 7.1 above. Now let n be an arbitrary natural number, and assume that $Q(n)$ is true. Our goal is to show that $Q(n+1)$ is true.

Hence, assuming $P(k)$ is true for all natural numbers $k < n$, we must show $P(k)$ is true for all natural numbers $k < n+1$. Appealing to statement (1) appearing in the theorem, knowing that $P(k)$ is true for all $k < n$ tells us that $P(n)$ is true as well, so that $P(k)$ is true for all natural numbers $k \in n \cup \{n\}$, i.e. for all $k \in n+1$. Thus we have shown $P(k)$ is true for all $k < n+1$, as needed.

By the Principle of Mathematical Induction, $Q(n)$ is true for all natural numbers n, which implies that $P(n)$ is true for all natural numbers n. □

Two comments on strong induction are now in order:

- The proof above shows how the Principle of Induction implies the Principle of Strong Induction. But the Principle of Strong Induction also implies the Principle of Induction, so that the two principles are logically *equivalent*.

  To see how, assume the Principle of Strong Induction, and suppose we have a statement $P(k)$ such that $P(0)$ is true, and whenever $P(n)$ is true, then $P(n+1)$ is true, for each natural number n. In particular, for this statement $P(k)$, we know that if $P(k)$ is true for all natural numbers $k < n$, then either $n = 0$, in which case the conclusion $P(0)$ follows by assumption, or $0 < n$, in which case $P(n-1)$ is assumed true, so that $P((n-1)+1) = P(n)$ is true by assumption.

  Thus, we may apply strong induction to conclude that $P(n)$ is true for all natural numbers n. This completes the proof that strong induction may be used to prove induction.

- In practice, in order to prove the implication "If $P(k)$ is true for all $k < n$, then $P(n)$ is true" in a strong induction proof, it is usually necessary to independently show that a number of cases of the statement are true, say $P(0), P(1), \ldots, P(m)$ for some natural number m. Then in the proof of the implication, we can assume that $n \geq m$ along with the hypothesis that $P(k)$ is true for all $k < n$. The cases $P(0), \ldots, P(m)$ that we verify independently are called *base cases* in a strong induction proof.

The Well-Ordering Principle

Now that we have the two forms of induction under our belt, let's show that they are equivalent to another fundamental property of the natural numbers, known as the *well-ordering principle*. For $\mathbb{N}$, this amounts to the following claim:

Theorem 7.3 (Well-Ordering Principle)**.** *Every non-empty subset of $\mathbb{N}$ has a least element. In other words, if $A \subseteq \mathbb{N}$ and $A \neq \varnothing$, then there is some element $a \in A$ such that $a \leq n$ for all $n \in A$.*

We will have more to say about well-orderings after we have studied order relations in general, and after we have looked at the so-called Axiom of Choice in set theory. But for now, it is enough to say that the name *well-ordered* comes from the fact that every nonempty subset of $\mathbb{N}$ has a least element. Let's prove this principle from the Principle of Strong Induction:

Proof. Suppose that A is a nonempty subset of $\mathbb{N}$, and suppose for a contradiction that A does not have a least element. We thus consider the set $B = \mathbb{N} \setminus A$.

For any natural number n, we claim that if $k \in B$ for all natural numbers $k < n$, then we must have $n \in B$ as well. Indeed, if $k \in B$ for all natural numbers $k < n$, then none of $0, 1, \ldots, n-1$ belong to A, so that if $n \in A$, then n would be the least element of A, contradicting our hypothesis. Thus $n \in B$ if all natural numbers smaller than n are in B. Applying the Principle of Strong Induction to the statement $P(n)$ given by $n \in B$, we conclude that $n \in B$ for all $n \in \mathbb{N}$, i.e. $B = \mathbb{N}$. But since $B = \mathbb{N} \setminus A$, this implies $A = \varnothing$, a

contradiction.

Altogether, our assumption that A does not have a least element must be false, and we conclude that A has a least element after all. □

Logically speaking, the Well-Ordering Principle also implies the ordinary Principle of Induction, making it logically equivalent to both forms of induction discussed so far:

Theorem 7.4. *If the Well-Ordering Principle holds for subsets of $\mathbb{N}$, then the Principle of Induction is true.*

Proof. Suppose we are given a logical statement $P(x)$, depending on a variable x, for which we know $P(0)$ is true, and if $P(n)$ is true for a natural number n, then $P(n+1)$ is also true. We must show that for all $n \in \mathbb{N}$, $P(n)$ is true.

As in the proof of Theorem 7.1, we again consider the set $A = \{n \in \mathbb{N} : P(n)\}$. But now, we also consider its complement $B = \mathbb{N} \setminus A$. Note that $P(n)$ is true for all $n \in \mathbb{N}$ if and only if $B = \varnothing$. So assume for a contradiction that $B \neq \varnothing$. By the Well-Ordering Principle, B has a least element b. Note that $b \neq 0$, since $0 \in A$. In particular, the number $b-1$ is a natural number too, and by definition of b, we know $b-1 \notin B$, so $b-1 \in A$. But then $P(b-1)$ is true, so that $P((b-1)+1) = P(b)$ is true by assumption.

In turn, this shows $b \in A$, so that $b \notin B$, a contradiction. We conclude that our assumption that B is nonempty must be incorrect, which completes the proof. □

MATH 145 Course Reading 8:
Some Examples of Mathematical Induction Proofs

September 28, 2020

In the previous reading, we looked at the Principle of Induction as an abstract set-theoretic property of $\mathbb{N}$. However, that viewpoint does not adequately capture how central proofs by induction are throughout mathematics, nor how the "flow" of an induction proof generally goes. So here, we pause to take a look at proofs by induction outside of the realm of set theory, to give some indication of how they are typically carried out.

Proofs with Recurrence Relations

One common use for induction proofs is in establishing a *closed-form formula* for the terms of a sequence defined by a *recurrence relation*. A sequence $a_0, a_1, a_2, \ldots$ of real numbers is said to be defined via a recurrence relation if the first few terms of the sequence are defined as explicit values, and the remaining terms of the sequence are defined by formulas involving previous terms of the sequence. For example, the sequences $s_0, s_1, s_2, \ldots$ and $f_0, f_1, f_2, \ldots$ defined by

$$\begin{cases} s_0 = 0 \\ s_{n+1} = s_n + (2n+1), \quad n \in \mathbb{N} \end{cases}$$

$$\begin{cases} f_0 = 0, f_1 = 1 \\ f_{n+2} = f_{n+1} + f_n, \quad n \in \mathbb{N} \end{cases}$$

are both examples of sequences defined by recurrence relations. Note that the first sequence uses only one previous value to determine the next, while the second uses two previous values to determine the next.

In both cases, a proof by induction can be used to verify closed-form expressions for the sequences, i.e. functions $g(n)$ and $h(n)$ such that $s_n = g(n)$ and $f_n = h(n)$ for all natural numbers n. In the case of the first sequence, we might conjecture the formula directly from computing a few small values by hand:

$$\begin{aligned} s_0 &= 0 \\ s_1 &= s_0 + (2(0)+1) = 0 + 1 = 1 \\ s_2 &= s_1 + (2(1)+1) = 1 + 3 = 4 \\ s_3 &= s_2 + (2(2)+1) = 4 + 5 = 9. \end{aligned}$$

Already, we seem to have enough evidence to be able to conjecture that $s_n = n^2$ for all $n \in \mathbb{N}$. We verify this through a proof by induction:

Proposition 8.1. *For the sequence s_n defined above, we have $s_n = n^2$ for all $n \in \mathbb{N}$.*

Proof. We proceed by induction, where the statement $P(n)$ is the statement that $s_n = n^2$. The statement $P(0)$ (the *base case*) is true, since $s_0 = 0$ by definition, while $0^2 = 0$, so $s_0 = 0 = 0^2$.

Now, assume that $P(n)$ is true for some fixed but arbitrary $n \in \mathbb{N}$. We show that $P(n+1)$ is true. In other words, we assume that $s_n = n^2$ and prove that $s_{n+1} = (n+1)^2$. Using that $s_n = n^2$ and applying the recursive definition of the sequence,

$$\begin{aligned} s_{n+1} &= s_n + (2n+1) \\ &= n^2 + (2n+1) \\ &= (n+1)^2, \end{aligned}$$

which proves what we needed to show. By the Principle of Induction, we conclude that $P(n)$ is true for all $n \in \mathbb{N}$, i.e. $s_n = n^2$ for all $n \in \mathbb{N}$. □

When encountering proofs by induction in subsequent courses (and even later in this course), it is common practice to omit mentioning the statement $P(n)$ directly in the proof, appealing to the reader to identify implicitly what that statement is. For now, we draw attention to it only to highlight the steps logically required for an airtight proof.

Next, we verify a closed-form formula for the second sequence above. This time, the sequence is the famous *Fibonacci sequence*, and the closed-form formula might not be so obvious from a computation of a few small values. However, a mathematical study of recurrence relations is possible (though outside the scope of this course), and that general theory yields a formula for the Fibonacci sequence that we can verify by induction.

To do this, however, we will have to rely on strong induction, rather than just the ordinary Principle of Induction. Indeed, since the recurrence formula defines each new term via the two previous terms, it is not enough to assume the statement is valid for a single natural number n, but rather for multiple smaller values of n (and at least two such values of n). In order to carry out the argument fully, we will need to verify two base cases, $P(0)$ and $P(1)$, before we can prove the implication required by strong induction.

Proposition 8.2. *Let $\alpha_1 = \frac{1+\sqrt{5}}{2}$ and $\alpha_2 = \frac{1-\sqrt{5}}{2}$ be the two real roots of the quadratic polynomial $x^2 - x - 1$. Then for each natural number n, we have*

$$f_n = \frac{1}{\sqrt{5}} \left(\alpha_1^n - \alpha_2^n \right).$$

Proof. We proceed by strong induction, where the statement $P(n)$ is $f_n = \frac{1}{\sqrt{5}} (\alpha_1^n - \alpha_2^n)$. First, we verify the base cases $n = 0$ and $n = 1$.

When $n = 0$, note that $f_0 = 0$ by definition, while

$$\frac{1}{\sqrt{5}} \left(\alpha_1^0 - \alpha_2^0 \right) = \frac{1}{\sqrt{5}} (1 - 1) = 0.$$

Thus $P(0)$ is true. When $n = 1$, we have $f_1 = 1$ by definition, while

$$\frac{1}{\sqrt{5}} \left(\alpha_1^1 - \alpha_2^1 \right) = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} - \frac{1-\sqrt{5}}{2} \right) = \frac{1}{\sqrt{5}} \left(\frac{2\sqrt{5}}{2} \right) = 1.$$

Thus $P(1)$ is true. Now, assume that $P(k)$ is true for all natural numbers $k < n$ (and assume that $n \geq 2$, courtesy of the two base cases above). We verify that $P(n)$ is true. Since we've assumed $n \geq 2$, the recurrence relation formula may be applied to f_n, and using this, along with the truth of $P(n-1)$ and $P(n-2)$, we get

$$\begin{aligned}
f_n &= f_{n-1} + f_{n-2} \\
&= \frac{1}{\sqrt{5}} (\alpha_1^{n-1} - \alpha_2^{n-1}) + \frac{1}{\sqrt{5}} (\alpha_1^{n-2} - \alpha_2^{n-2}) \\
&= \frac{1}{\sqrt{5}} ((\alpha_1^{n-1} + \alpha_1^{n-2}) - (\alpha_2^{n-1} + \alpha_2^{n-2})) \\
&= \frac{1}{\sqrt{5}} (\alpha_1^{n-2}(\alpha_1 + 1) - \alpha_2^{n-2}(\alpha_2 + 1)) \\
&= \frac{1}{\sqrt{5}} (\alpha_1^{n-2}(\alpha_1^2) - \alpha_2^{n-2}(\alpha_2^2)) \\
&= \frac{1}{\sqrt{5}} (\alpha_1^n - \alpha_2^n).
\end{aligned}$$

In moving from the third-last line to the second-last line, we used the fact that α_1 and α_2 are both roots of the quadratic polynomial $x^2 - x - 1$. This now verifies that $P(n)$ is true, which completes the proof by strong induction. We now conclude that f_n is given by the formula above for all $n \in \mathbb{N}$. □

This result now lets you derive any Fibonacci number in a single computation, straight from a calculator!

Proofs of Inequalities

Another common type of proof that yields to induction arguments are proofs of inequalities, where the inequalities in question are only valid for natural numbers. In fact, many of these inequalities only hold for sufficiently large natural numbers. In other words, for some natural number $a > 0$, we are asked to prove the inequality $f(n) < g(n)$ for all $n \geq a$. Since we're not trying to prove the statement for all natural numbers, it seems at first that the Principle of Induction does not apply.

However, it can be made to apply through the clever trick where we declare the statement $P(n)$ to be $f(n+a) < g(n+a)$. Then the claim that $P(n)$ holds for all natural numbers n is equivalent to the statement that $f(n) < g(n)$ for all natural numbers $n \geq a$.

In practice, what this means when writing down the proof is that our base case $P(0)$ is actually verifying the inequality $f(a) < g(a)$, and then the proof of the induction implication $P(n) \Rightarrow P(n+1)$ amounts to assuming that $f(n) < g(n)$ for some natural number $n \geq a$ and showing that $f(n+1) < g(n+1)$. This is illustrated in the example proof below.

Before reading it, recall that for any positive integer n, $n!$ is defined as the product of all the integers from 1 up to n. For example, $4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$.

Proposition 8.3. *For all natural numbers $n \geq 4$, we have $n! > 2^n$.*

Proof. We proceed by induction. First, we verify that the statement is true for $n = 4$. Here, $n! = 24$ and $2^n = 16$, so $n! > 2^n$ is true for $n = 4$. Now, assume we are given a natural number $n \geq 4$ for which $n! > 2^n$. Our goal is to show that $(n+1)! > 2^{n+1}$. Since $n \geq 4$, certainly $n + 1 > 2$, so notice that

$$(n+1)! = (n+1) \cdot n! > 2 \cdot n! > 2 \cdot 2^n = 2^{n+1}.$$

This verifies the inequality we needed to show. By the Principle of Induction, the inequality $n! > 2^n$ is valid for all natural numbers $n \geq 4$. □

A question you might be asking is: why is the restriction $n \geq 4$ there in the first place? Why is it not just a claim made for all natural numbers n? If you don't see it immediately, it's worth thinking about for a second!

Other Induction Proofs

Induction can also be used to prove other types of claims. We finish with two examples: one from set theory, and one that has a flavour of number theory.

First, a bit of a definition: if a set A has finitely many elements, we use the notation $|A|$ to denote the number of elements of A, and call that the *cardinality* of A. We will study the cardinality of sets properly in a few weeks, but for now, this intuitive notion of cardinality for finite sets will suffice. The result we're about to state connects the size of a set A and the size of its power set $\mathcal{P}(A)$:

Proposition 8.4. *For every $n \in \mathbb{N}$, if A is a set such that $|A| = n$, then $|\mathcal{P}(A)| = 2^n$.*

Proof. We proceed by induction on n. For the base case, we verify the statement when $n = 0$. Here, if A is a set such that $|A| = 0$, it must be true that $A = \varnothing$, in which case $\mathcal{P}(A) = \{\varnothing\}$, so that $|\mathcal{P}(A)| = 1 = 2^0$. This proves the $n = 0$ case.

Now let $n \in \mathbb{N}$ be arbitrary, and suppose that for all sets B with $|B| = n$, we have $|\mathcal{P}(A)| = 2^n$. Next, let A be a set with $n + 1$ elements; we wish to show that $|\mathcal{P}(A)| = 2^{n+1}$. We list out the elements of A, say $A = \{a_1, a_2, \ldots, a_{n+1}\}$. We split the elements of $\mathcal{P}(A)$ into two types: the subsets that contain a_{n+1}, and the subsets that do not.

The subsets of A that do not contain a_{n+1} correspond exactly with the subsets of $\{a_1, a_2, \ldots, a_n\}$, an n-element set. By hypothesis, there are 2^n subsets of $\{a_1, \ldots, a_n\}$, and thus 2^n subsets of A not containing a_{n+1}. On the other hand, the subsets of A containing a_{n+1} correspond exactly with the subsets of $\{a_1, \ldots, a_n\}$, but with an extra element a_{n+1} thrown into each subset. Again, by hypothesis, there are exactly 2^n such subsets. Combining the subsets covered by these two exhaustive cases, A has a total of $2^n + 2^n = 2 \cdot 2^n = 2^{n+1}$ subsets, as we needed to show. □

Finally, here's an example of an induction proof where we seek to find integer solutions to equations. This one combines both strong induction and the need to verify the statement starting at a natural number other than 0:

Proposition 8.5. *For every natural number $n \geq 24$, there are $x, y \in \mathbb{N}$ such that $5x + 7y = n$.*

Proof. We prove this statement by strong induction on n. First, we verify the base cases $n = 24, 25, 26, 27, 28$ by checking that

$$\begin{aligned} 5(2) + 7(2) &= 24 \\ 5(5) + 7(0) &= 25 \\ 5(1) + 7(3) &= 26 \\ 5(4) + 7(1) &= 27 \\ 5(0) + 7(4) &= 28. \end{aligned}$$

Therefore, the statement is true for these five values of n. Now assume that the statement is true for all positive integers k such that $24 \leq k < n$ (and where we may assume $n \geq 29$). We wish to prove the statement is true for n as well. Notice that since we may assume $n \geq 29$, we have that $n - 5$ is a positive integer greater than or equal to 24. In particular, the statement is true for $n - 5$, so there are $x, y \in \mathbb{N}$ such that

$$5x + 7y = n - 5.$$

Adding 5 to both sides, we get

$$5(x + 1) + 7y = 5x + 7y + 5 = n,$$

and where $x + 1, y \in \mathbb{N}$. Thus, we have now shown that the given statement is true for n as well. By strong induction, we conclude that for all integers $n \geq 24$, there are $x, y \in \mathbb{N}$ such that $5x + 7y = n$. □

Here are two questions you may like to ask yourself, to probe your understanding: (1) Why did we need five base cases for the argument to work? (2) Why do we only claim the statement for $n \geq 24$? Could we expand the statement to be valid for even smaller values of n?

MATH 145 Course Reading 9:
Ordered Pairs, Relations, and Cartesian Products

September 30, 2020

Our next goal is to define some common mathematical constructions formally, with the aid of the set theory we have developed so far. In particular, we will look at how to define the idea of an *ordered pair* using only sets, and construct the related notion of the Cartesian product of sets, which is a set of ordered pairs. At the same time, we will formally define the mathematical notion of *relation*, which will be one of the most important definitions of this course. Special cases of relations include *equivalence relations* and *functions*, both of which will be heavily used throughout the coming weeks.

Ordered Pairs

The fundamental information captured by an ordered pair (x, y) is twofold: not only do we care about the particular elements x, y in the pair, but also the ordering of those two elements. For instance, we want to be able to talk about the *first coordinate* and *second coordinate* of these ordered pairs. A consequence of this is that we want an equality $(x, y) = (x', y')$ if and only if $x = x'$ and $y = y'$.

Given this, it becomes clear after a moment's thought that the set $\{x, y\}$ is insufficient for use as the definition of an ordered pair (x, y). (Why is this the case?) There is no unique option for representing the ordered pair (x, y) as a set, but we will adopt the following convention:

Definition 9.1. Given any two sets x, y, the *ordered pair* (x, y) is defined to be the set $\{\{x\}, \{x, y\}\}$.

As a first check on our definition, let's verify that the notions of "first coordinate" and "second coordinate" make sense by proving that two ordered pairs are equal if and only if their first and second coordinates both agree.

Proposition 9.1. *For any sets x, y, x', y', we have $(x, y) = (x', y')$ if and only if $x = x'$ and $y = y'$.*

Proof. First, we prove the "easy" direction of the biconditional. We assume that $x = x'$ and $y = y'$. Given this, it is clear that

$$\begin{aligned}(x, y) &= \{\{x\}, \{x, y\}\} \\ &= \{\{x'\}, \{x', y'\}\} \\ &= (x', y').\end{aligned}$$

Conversely, assume that $(x, y) = (x', y')$. We wish to show that $x = x'$ and $y = y'$. The assumption that $(x, y) = (x', y')$ amounts to assuming the set equality

$$\{\{x\}, \{x, y\}\} = \{\{x'\}, \{x', y'\}\}.$$

By the Axiom of Extensionality, these two sets are equal if and only if they have the same elements. From here, we break into two cases. If $x = y$, then $\{\{x\}, \{x, y\}\} = \{\{x\}\}$, and so we conclude from the set equality above that $\{x'\} = \{x\}$ and that $\{x', y'\} = \{x\}$. The second equality immediately gives us $x' = x$, and also $y' = x$ again by Axiom of Extensionality. In particular, $x = x'$ and $y = x = y'$, as we needed to show.

On the other hand, if $x \neq y$, then the set $\{x, y\}$ has two distinct elements. Given the set equality above, we must have $\{x, y\} = \{x', y'\}$, since $\{x, y\} = \{x'\}$ is impossible on account of the fact that $\{x'\}$ has only one element. As a consequence, we deduce that $\{x\} = \{x'\}$, so that $x = x'$. But then

$$\{x', y'\} = \{x, y'\} = \{x, y\},$$

and from the last equality we deduce that $y = y'$, as needed. □

We should stress again here that this is not the only way to define an ordered pair: you may wish to seek out alternative set-theoretic definitions to see how they work! Certainly, the definition we've adopted allows us to rigorously define the notion of first and second coordinates. If $x \neq y$, then (x, y) has an element $\{x\}$ with one element, and an element $\{x, y\}$ with two elements. The first coordinate is the element of the singleton set $\{x\}$, and the second coordinate is the other element of the two-element set $\{x, y\}$. And if $x = y$, then $(x, x) = \{\{x\}\}$, and the first and second coordinates are both equal to the only element x in the only set $\{x\}$ belonging to the ordered pair.

We can then iterate this construction to form ordered n-tuples for any positive integer n. We take an ordered 1-tuple (x) to be the set $\{x\}$ by convention, and for $n \geq 3$, an n-tuple is defined recursively in terms of an $n-1$ tuple. For example, the triple (x, y, z) is declared to be $((x, y), z)$, and the 4-tuple (x_1, x_2, x_3, x_4) is taken to be $((x_1, x_2, x_3), x_4)$.

Cartesian Products

You may already be familiar with the Cartesian product of sets. Informally, given sets X and Y, we would like to define a set $X \times Y$ that consists of all the ordered pairs with first coordinate from X and second coordinate from Y. This seems like it could be done by a single application of the Axiom Schema of Comprehension, but in order to accomplish this through set-builder notation, we first need to supply a set that contains all the elements of the set $X \times Y$ we hope to define.

In other words, given sets X and Y, we would like to define it as

$$X \times Y = \{w \in Z : w = (x, y) \text{ for some } x \in X,\, y \in Y\},$$

but what is this set Z we should take? For a hint about this, note that ordered pairs with first coordinate in X and second coordinate in Y look like $\{\{x\}, \{x, y\}\}$ for some $x \in X$ and some $y \in Y$.

So, the elements of the ordered pair are in fact subsets of a larger set, which means we should probably be looking for a power set of something. But what are $\{x\}$ and $\{x, y\}$ subsets of? To start, X and Y do not need to have anything to do with each other, which means we first need a set that contains X and Y. This is accomplished through the Axiom of Union. Certainly, $\{x\}$ and $\{x, y\}$ are both subsets of $X \cup Y$, so the elements of the set $\{\{x\}, \{x, y\}\}$ can be taken to belong to the power set $\mathcal{P}(X \cup Y)$.

However, the ordered pair $\{\{x\}, \{x, y\}\}$ itself is *not* a subset of $X \cup Y$, so we can't just take $Z = \mathcal{P}(X \cup Y)$ above. Rather, to get a *set* of subsets of $\mathcal{P}(X \cup Y)$, we will need an element of $\mathcal{P}(\mathcal{P}(X \cup Y))$. So we actually must take $Z = \mathcal{P}(\mathcal{P}(X \cup Y))$ when applying the Axiom Schema of Comprehension above. (Do you follow this explanation to your satisfaction?) Notice how many of the set theory axioms we had to invoke just to define $X \times Y$: we used Comprehension, Union, and Power Set (twice)!

Let's formally define the Cartesian product now:

Definition 9.2. Given any two sets X and Y, the *Cartesian product* of X and Y is the set

$$X \times Y = \{w \in \mathcal{P}(\mathcal{P}(X \cup Y)) : w = (x, y) \text{ for some } x \in X, y \in Y\}.$$

We can then iterate the Cartesian product construction to define the product of more than two sets. For example, we can take

$$X \times Y \times Z = (X \times Y) \times Z,$$

applying the Cartesian product construction to $X \times Y$ and Z. By convention, we often write X^2 for $X \times X$, X^3 for $X \times X \times X$, and so on.

Relations

In particular, all the work you do in calculus class with the Cartesian plane involves working with certain subsets of $\mathbb{R} \times \mathbb{R}$. Often, you work with the graph of real-valued functions (and we will define a *function* more carefully in the next reading). Sometimes, you even work with subsets of the plane that aren't functions, such as the graph of the unit circle:

$$C = \{(x, y) \in \mathbb{R} \times \mathbb{R} : x^2 + y^2 = 1\}.$$

You can see the set C as defining a relationship between the x-coordinate and y-coordinate, where x and y are only in this relationship if (x, y) lies on the unit circle. Other mathematical relationships can be captured in this way. For example, given two integers m and n, we might define the relationship "m is a multiple of n", and capture the set of all pairs of integers related in this way:

$$D = \{(m, n) \in \mathbb{Z} \times \mathbb{Z} : n = mk \text{ for some integer } k\}.$$

So really, to specify a relationship between two types of objects, one from a set A and one from a set B, we are always specifying the set of ordered pairs $(a, b) \in A \times B$ that satisfy the relation.

All of this motivates the official definition of *(binary) relation*:

Definition 9.3. Given two sets A and B, a *binary relation* from A to B is a subset of $A \times B$. More generally, a set R is called a *relation* if all the elements of R are ordered pairs. Rather than writing $(x, y) \in R$, we will often use the notation xRy. If R is a relation from A to A, we often call R a *relation on* A.

What's the reason for the strange notation xRy? It's motivated by some of the most basic examples of relations:

Example 9.1. On any set A, we define a relation on A called *equality* by declaring that $(x, y) \in R$ if and only if $x = y$. In other words, xRy if and only if $x = y$. Hence we could have just called the relation $=$ instead of R.

Example 9.2. Given any set A, we can define a relation on $\mathcal{P}(A)$ by declaring that $(X, Y) \in R$ if and only if $X \subseteq Y$. In other words, XRY if and only if $X \subseteq Y$. Hence we could have just called the relation $\subseteq$ instead of R.

Already, we see that on many occasions where we write a symbol between two elements of a set, we are implicitly defining a relation between pairs of elements. This justifies the use of the notation xRy for an arbitrary relation R. We now wrap up this reading with a couple more examples of relations, leaving the definition of various concepts connected with relations to the next reading.

Example 9.3. The standard ordering of integers gives a relation on $\mathbb{Z}$. We define a relation R as a subset of $\mathbb{Z} \times \mathbb{Z}$ by declaring that $(x, y) \in R$ if and only if $x < y$. So for example $(1, 10) \in R$, while $(1, -2) \notin R$.

Example 9.4. We can define a relation R from $\mathbb{R}$ to $\mathbb{R}$ by declaring that $(x, y) \in R$ if and only if $|x - y| \leq 1$. Given a fixed $x \in \mathbb{R}$, can you draw a picture of all the $y \in \mathbb{R}$ such that xRy? We will shortly define this to be the *image of* x *under* R.

Example 9.5. For an example that we will have much to say about towards the end of our course, we fix a positive integer n and define a relation R called *congruence modulo* n as a subset of $\mathbb{Z} \times \mathbb{Z}$, given by xRy if and only if $x - y$ is a multiple of n. For instance, if $n = 2$, then xRy if and only if x and y are both even or both odd (can you prove this?). We often write $x \equiv y \pmod{n}$ to show that x and y are *congruent* modulo n (i.e. related by congruence modulo n).

MATH 145 Course Reading 10:
Functions: Formal Definition and Terminology

October 2, 2020

Now that we have the basic set-theoretic construction and terminology behind binary relations, we devote this reading to a very important special case: the *functions*. You have probably seen functions extensively in calculus (at least, between subsets of the real numbers), and we would now like to provide an official and general definition that allows us to talk about a function between any pair of sets. Along the way, we will formalize some of the terminology that arises around functions: domain, range, composition and so on.

Definition of a Function

When you work with the graphs of functions in calculus, you often hear about the "vertical line test" for checking that what you have really is the graph of a function. It says that every vertical line should intersect the function's graph exactly once. Translated into more formal language, for every real number x in the function's domain, there should be exactly one ordered pair (x, y) with first coordinate x on the graph of the function.

This perspective motivates the modern definition of a function, where we can consider a function from any set to any other set.

Definition 10.1. A *function* f is a relation, such that for each first coordinate appearing in f, there is only one ordered pair with that first coordinate. In other words, if the ordered pairs (a, b_1) and (a, b_2) both belong to f, then we must have $b_1 = b_2$. If $(a, b) \in f$, we think of this as f uniquely mapping a to b, and will often write $f(a) = b$ instead of $(a, b) \in f$ or the awkward relation notation afb.

If we are given two sets A and B in advance, a *function from A to B* is a function that happens to be a subset of $A \times B$, such that for every $a \in A$, there *is* some $b \in B$ such that $f(a) = b$. Thus f is a function from A to B exactly when it is a relation from A to B for which each $a \in A$ is related to *exactly* one element $b \in B$. If f is a function from A to B, we will often adopt the notation $f : A \to B$ to represent this.

Example 10.1. In calculus, you may have worked with "functions" such as $f(x) = x^2$. Technically speaking, the "true" function f corresponding to this is the set

$$f = \{w \in \mathbb{R} \times \mathbb{R} : w = (x, x^2) \text{ for some } x \in \mathbb{R}\}.$$

Notice that this is indeed a function according to our definition: for each $r \in \mathbb{R}$, there is exactly one ordered pair in f with r as first coordinate, namely (r, r^2). The same recipe applies to any "function" $y = f(x)$ considered in calculus with domain $\mathbb{R}$; the true underlying function is the set of pairs $\{w \in \mathbb{R} \times \mathbb{R} : w = (x, f(x)) \text{ for some } x \in \mathbb{R}\}$.

Example 10.2. Let $A = \{1, 2, 3\}$ and $B = \{w, x, y, z\}$ (with all elements of A and B distinct). The relation

$$f = \{(1, x), (2, x), (3, w)\}$$

gives a function from A to B. Notice that no explicit formula for the function is required!

On the other hand, the relation

$$R_1 = \{(1, w), (2, z)\}$$

is *not* a function *from A to B*, though it *is* a function. Note the distinction is subtle: what's missing is that not every element of A appears as a first coordinate of some element of R_1. On the other hand,

$$R_2 = \{(1, y), (1, z), (2, x), (3, x)\}$$

is *not* a function, since 1 appears as the first coordinate of two distinct ordered pairs in R_2.

Terminology Around Relations and Functions

We've already thrown around the word "domain" a few times on the first page, so it's about time we define what this is, for an arbitrary relation. At the same time, we make a bunch of other related definitions:

Definition 10.2. Let R be a binary relation. The *domain* of R is the set of all x for which $(x, y) \in R$ for some y. On the other hand, the *range* of R is the set of all y for which $(x, y) \in R$ for some x. More informally, the domain of R is the set of first coordinates of elements of R, and the range of R is the set of second coordinates of elements of R. We often use $\operatorname{dom} R$ to denote the domain of R, and $\operatorname{ran} R$ to denote the range. The set $\operatorname{dom} R \cup \operatorname{ran} R$ is called the *field* of R and denoted by $\operatorname{field} R$.

It is, of course, necessary to check that the sets $\operatorname{dom} R$ and $\operatorname{ran} R$ exist for an arbitrary relation R; happily, they always do. (Why? Hint: Assignment 3). Specializing to the case where R is a function, we recover the familiar notions of domain and range of a function.

Similarly, we can define the image and inverse image of a relation, just as you might be familiar with for functions:

Definition 10.3. Let R be a binary relation. The *image* of a set A under R is the set

$$R(A) = \{b \in \operatorname{ran} R : (a, b) \in R \text{ for some } a \in A\}.$$

Similarly, given a set B, the *inverse image* of B under R is the set

$$R^{-1}(B) = \{a \in \operatorname{dom} R : (a, b) \in R \text{ for some } b \in B\}.$$

Finally, we can define the composition and inverse of a relation, just as you might be used to doing with functions:

Definition 10.4. Let R be a binary relation. The *inverse relation* R^{-1} is defined to be

$$R^{-1} = \{z \in \operatorname{ran} R \times \operatorname{dom} R : z = (b, a) \text{ for some } (a, b) \in R\}.$$

Given two relations R_1 and R_2, the *composition* of those relations, $R_2 \circ R_1$, is defined to be

$$R_2 \circ R_1 = \{z \in \operatorname{dom} R_1 \times \operatorname{ran} R_2 : z = (a, c) \text{ where there exists } b \text{ such that } (a, b) \in R_1 \text{ and } (b, c) \in R_2\}.$$

Notice that the inverse of a relation and a composition of two relations is again a relation. Moreover, if R is a relation from A to B, then R^{-1} is a relation from B to A, and is obtained just by reversing all the ordered pairs in R. Similarly, if R_1 is a relation from A to B and R_2 is a relation from B to C, then $R_2 \circ R_1$ is a relation from A to C.

We will almost exclusively apply these definitions to functions, but it may be worth noting a couple quick examples:

Example 10.3. For the relation R given by $<$ on $\mathbb{Z}$, where $(a, b) \in R$ if and only if $a < b$, note that $\operatorname{dom} R = \operatorname{ran} R = \mathbb{Z}$. The inverse relation, R^{-1}, is the set of ordered pairs (b, a) such that $(a, b) \in R$, i.e. $a < b$. Thus $(b, a) \in R^{-1}$ if and only if $b > a$. This justifies our saying that the inverse of the less-than relation $<$ is the greater-than relation $>$.

Example 10.4. Consider the relation R on $\mathbb{R}$ given by xRy if and only if $|x - y| \leq 1$, as discussed in the previous reading. We claim that an ordered pair (x, y) belongs to $R \circ R$ if and only if $|x - y| \leq 2$.

First, suppose that $(x, y) \in R \circ R$. By definition, there is some real number z such that $(x, z) \in R$ and $(z, y) \in R$. Thus $|x - z| \leq 1$ and $|z - y| \leq 1$. Applying the triangle inequality from calculus,

$$|x - y| = |(x - z) + (z - y)| \leq |x - z| + |z - y| \leq 1 + 1 = 2.$$

Thus if $(x, y) \in R \circ R$, then $|x - y| \leq 2$. Conversely, assume that we have real numbers x, y with $|x - y| \leq 2$. We set $z = \frac{x+y}{2}$, the arithmetic mean of these two numbers x and y. We claim that (x, z) and (z, y) both belong to R, so that $(x, y) \in R \circ R$. Indeed, note that

$$|x - z| = \left| x - \frac{x+y}{2} \right| = \left| \frac{x-y}{2} \right| = \frac{|x-y|}{2} \leq \frac{2}{2} = 1$$
$$|z - y| = \left| \frac{x+y}{2} - y \right| = \left| \frac{x-y}{2} \right| = \frac{|x-y|}{2} \leq \frac{2}{2} = 1.$$

This proves that $(x, z) \in R$ and $(z, y) \in R$, so that $(x, y) \in R \circ R$ as claimed. We have now shown that

$$R \circ R = \{(x, y) \in \mathbb{R} \times \mathbb{R} : |x - y| \leq 2\}.$$

Invertibility, Injectivity, and Surjectivity

According to the definitions above, given any function f, we can always define an inverse, f^{-1}, which will be a relation at the very least. We say that f is *invertible* if f^{-1} is also a function. Clearly, not every function is invertible; for instance, the one considered in Example 10.1 is not. There, we have f^{-1} as the set of ordered pairs of the form (r^2, r) for $r \in \mathbb{R}$, and for instance both $(1, 1)$ and $(1, -1)$ belong to the relation f^{-1}.

A moment's thought suggests that a function f will be invertible only when every element of $\operatorname{dom} f$ gets mapped somewhere different from all the other elements of $\operatorname{dom} f$. In the example above, the failure of invertibility popped up because both 1 and -1 map to 1 under f. This motivates the first part of our definition below:

Definition 10.5. A function f is called *injective* (or *one-to-one*) if, for all $x_1, x_2 \in \operatorname{dom} f$ such that $x_1 \neq x_2$, we have $f(x_1) \neq f(x_2)$. Taking the contrapositive, f is injective if for any $x_1, x_2 \in \operatorname{dom} f$, knowing that $f(x_1) = f(x_2)$ implies that $x_1 = x_2$.

If f is a function from A to B, we say that f is *surjective* (or *onto*) if $\operatorname{ran} f = B$. In other words, for each $b \in B$, there is some $a \in A$ such that $f(a) = b$.

A function $f : A \to B$ is *bijective* (or a *bijection*) if f is both injective and surjective.

By our remarks before the definition, we suspect that the invertible functions are exactly the injective functions. We verify this below:

Proposition 10.1. *A function f is invertible if and only if f is injective. Moreover, if $f : A \to B$, then f^{-1} is a function from B to A if and only if f is bijective.*

Proof. First, suppose that f is an invertible function; we verify f is injective. Thus we assume we have two elements a_1, a_2 such that $f(a_1) = f(a_2)$. In other words, there is some element b for which $(a_1, b) \in f$ and $(a_2, b) \in f$. By definition of f^{-1}, we know that $(b, a_1) \in f^{-1}$ and $(b, a_2) \in f^{-1}$. But since f^{-1} is a function, this forces $a_1 = a_2$, proving that f is injective.

Conversely, assume that f is injective; we show that the inverse relation f^{-1} is a function. If we have two ordered pairs $(b, a_1), (b, a_2)$ belonging to f^{-1} with the same first coordinate, then the definition of f^{-1} implies that both (a_1, b) and (a_2, b) belong to f. Since f is injective, this implies $a_1 = a_2$, so that f^{-1} is a function.

Now, assume that f is a function from A to B. First suppose that f^{-1} is a function from B to A. By the first part of this proof, f is injective, since f^{-1} is a function. Now we show that f is surjective. Given an arbitrary $b \in B$, since f^{-1} is a function from B to A, we know that there is some $a \in A$ such that $(b, a) \in f^{-1}$. But then $(a, b) \in f$, so that $f(a) = b$. This proves f is surjective. Combining this with the fact that f is injective, we have shown that $f : A \to B$ is bijective if its inverse is a function from B to A.

Now assume that $f : A \to B$ is bijective. In particular, it is injective, and so its inverse is a function by the first part of this proof. It only remains to show that $\operatorname{dom} f^{-1} = B$ (why?). Certainly, the containment $\operatorname{dom} f^{-1} \subseteq B$ is immediate from the definition of f^{-1}. On the other hand, given $b \in B$, we know that f is surjective, so there is some $a \in A$ for which $(a, b) \in f$. This says $(b, a) \in f^{-1}$, and so $f^{-1}(b)$ is defined, proving that $B \subseteq \operatorname{dom} f^{-1}$. This now completes the proof that f^{-1} is a function from B to A when $f : A \to B$ is bijective. $\square$

MATH 145 Course Reading 11:
Equivalence Relations, Equivalence Classes, and Partitions

October 5, 2020

This time, we focus on equivalence relations, another important special class of relations. Intuitively, you can think of elements related by an equivalence relation as being "the same" in some way. We will see equivalence relations come up many times, even in this course: in particular, when defining the cardinality of a set and when introducing the *quotient* of an algebraic structure in the second half of our course.

Equivalence Relations

In last week's synchronous class session, we defined what it means for a relation to be an equivalence relation. For convenience, we re-state the necessary components of the definition here:

Definition 11.1. Let R be a binary relation on a set A. We say that

- R is *reflexive* if, for all $a \in A$, we have aRa.
- R is *symmetric* if, for all $a, b \in A$, if aRb, then bRa.
- R is *transitive* if, for all $a, b, c \in A$, if aRb and bRc, then aRc.
- R is an *equivalence relation* (on A) if R if reflexive, symmetric, and transitive.

As we will soon see, equivalence relations on a set end up splitting the elements of that set into pieces, where all the elements in a given piece are related to each other under the given relation. Let's give three examples, two mathematical, and one non-mathematical, to get a feel for how this works.

Example 11.1. Suppose $f : A \to B$ is a function. We define a relation R on the set A by declaring that $a_1 R a_2$ if $f(a_1) = f(a_2)$. We can very quickly check that R is an equivalence relation:

- **Reflexive:** For any $a \in A$, clearly $f(a) = f(a)$, so that aRa.
- **Symmetric:** Given any $a_1, a_2 \in A$ such that $a_1 R a_2$, we have $f(a_1) = f(a_2)$, so $f(a_2) = f(a_1)$, which implies $a_2 R a_1$.
- **Transitive:** Given any $a_1, a_2, a_3 \in A$ such that $a_1 R a_2$ and $a_2 R a_3$, we know $f(a_1) = f(a_2)$ and $f(a_2) = f(a_3)$, so $f(a_1) = f(a_3)$. This says $a_1 R a_3$.

Informally, two elements $a_1, a_2 \in A$ are related by R if the function f maps them to the same place in B. In other words, R splits up the elements of A according to where they map to in B under the function f.

Example 11.2. We define a relation R on the set P of people living in the world at 12pm (Eastern time) on September 1st, 2020 by declaring that $p_1 R p_2$ if p_1 and p_2 are in the same country at 12pm (Eastern time) on September 1st, 2020. To make this work, assume that every part of the world belongs to some country, and that you cannot be occupying two countries at the same time. You can check immediately that R defines an equivalence relation, and it splits the members of the set P into the countries in which they find themselves at that exact moment in time.

Example 11.3. We define a relation R on $\mathbb{Z}$ by declaring that mRn if $m - n$ is even. (You might recognize R as giving the "congruence modulo 2" relation.). Again, we can quickly check that R is an equivalence relation:

- **Reflexive:** For any $m \in \mathbb{Z}$, we have $m - m = 0$, which is even, so mRm.
- **Symmetric:** Suppose we have $m, n \in \mathbb{Z}$ such that mRn. Then $m - n$ is even, so $m - n = 2k$ for some $k \in \mathbb{Z}$. It follows that $n - m = 2(-k)$, where $-k \in \mathbb{Z}$, so that nRm.

- **Transitive:** Assume we have $\ell, m, n \in \mathbb{Z}$ such that $\ell R m$ and $m R n$. By definition, this says $\ell - m$ is even and $m - n$ is even, so $\ell - m = 2k_1$ and $m - n = 2k_2$ for some $k_1, k_2 \in \mathbb{Z}$. It follows that

$$\ell - n = (\ell - m) + (m - n) = 2k_1 + 2k_2 = 2(k_1 + k_2),$$

  where $k_1 + k_2 \in \mathbb{Z}$. Thus $\ell R n$, proving that R is transitive.

Note that if m is even, then mRn if and only if n is even, and if m is odd, then mRn if and only if n is odd. In this way, R splits up the integers into the even ones and the odd ones.

Equivalence Classes

In each of the examples above, you will have noticed a "partitioning" of the underlying set into a bunch of smaller pieces. These pieces will be referred to as *equivalence classes*, according to the definition below:

Definition 11.2. Suppose that E is an equivalence relation on some set A. Given an element $a \in A$, the *equivalence class of a modulo E* is the set

$$[a]_E = \{x \in A : aEx\}.$$

When the equivalence relation E is understood, we sometimes omit this from the notation and simply write $[a]$.

An equivalence relation on a set A splits up A into its equivalence classes: for any two elements of A, either their equivalence classes are identical, or else they are disjoint, sharing no elements in common. We capture this formally below:

Lemma 11.1. *Let E be an equivalence relation on a set A.*

(1) We have aEb if and only if $[a] = [b]$.

(2) We have $(a, b) \notin E$ if and only if $[a] \cap [b] = \varnothing$.

Proof. First, we prove (1). Assume that aEb; we prove that $[a] = [b]$. If $x \in [a]$, by definition we have that aEx. Since E is symmetric and aEb, we know that bEa as well. But then E is transitive, and we know that bEa and aEx, so we conclude that bEx. This says $x \in [b]$ as well, showing that $[a] \subseteq [b]$. A similar argument shows that $[b] \subseteq [a]$, proving that $[a] = [b]$ when aEb.

Next, assume that $[a] = [b]$. Since E is reflexive, we know that bEb, which says $b \in [b]$. But $[a] = [b]$, so we get $b \in [a]$ as well. By definition, this tells us that aEb, as needed.

We can now prove (2) immediately from (1) (taking contrapositives of both implications) if we show that $[a] \neq [b]$ is equivalent to $[a] \cap [b] = \varnothing$. Certainly, if $[a] \cap [b] = \varnothing$, then $[a] \neq [b]$, because both $[a]$ and $[b]$ are nonempty sets, while $[a] \cap [b] = \varnothing$ tells us they have no elements in common. On the other hand, suppose $[a] \cap [b] \neq \varnothing$, and say there is some element $x \in [a] \cap [b]$. We then claim that $[a] = [b]$. Indeed, since $x \in [a]$, we have aEx, and since $x \in [b]$, we have bEx. Since E is symmetric, we deduce that xEb. Then since E is transitive, we use that aEx and xEb to get that aEb. By part (1) of the proof, this says $[a] = [b]$, as needed. Knowing that $[a] \neq [b]$ is logically equivalent to $[a] \cap [b] = \varnothing$, (2) immediately follows. □

Partitions

Already in these notes, we've been using the word "partition" a fair bit. Below, we provide the official definition for this term:

Definition 11.3. Given any set A, a *partition* $\mathcal{P}$ of A is a collection of nonempty sets with the following properties:

(1) For any two distinct sets $P_1, P_2 \in \mathcal{P}$, we have $P_1 \cap P_2 = \varnothing$. (We say that any two distinct sets in $\mathcal{P}$ are *disjoint*.)

(2) The union of all the sets in $\mathcal{P}$ is the whole set A; in other words, $\bigcup \mathcal{P} = A$.

Visually, you might think of the partition $\mathcal{P}$ as splitting up A into a bunch of "countries", much like Example 11.2. Every element of A is assigned a "country" (piece of the partition), and all the "countries" taken together cover the set A.

Notice that Lemma 11.1 essentially tells us that every equivalence relation E on a set A gives us a partition of that set. In this case, the relevant partition is the set of all equivalence classes. Here, we adopt the notation A/E for the set $\{[a]_E : a \in A\}$. The notation A/E is in line with the "quotient" constructions we will be considering on algebraic structures later in the course; elements of such a quotient turn out to be equivalence classes of a certain kind.

Let's formally verify that the set A/E gives us a partition of A:

Proposition 11.1. *Let E be an equivalence relation on the set A. Then A/E is a partition of A.*

Proof. Every equivalence class $[a]$ is non-empty, because it contains the element a, and so A/E is a collection of non-empty sets. We showed in the proof of Lemma 11.1 that if two equivalence classes are not equal, then they are disjoint, which verifies condition (1) in the definition of a partition. Condition (2) of the definition follows from the fact that every $a \in A$ belongs to its own equivalence class $[a]$, so that the union of all the classes $[a]$ is the whole set A. □

The moral of the story is: every equivalence relation gives rise to a partition. Our final goal for this reading is to show that the converse is true: that every partition of a set can be used to define an equivalence relation on that set. This is stated in the following theorem:

Theorem 11.1. *Let A be a set, and let $\mathcal{P}$ be a partition on that set. We define a relation E on A by stating that $a_1 E a_2$ if there is some set $P \in \mathcal{P}$ such that $a_1 \in P$ and $a_2 \in P$. Then E is an equivalence relation on A.*

Proof. We check each of the three properties of equivalence relations in turn:

- **Reflexivity:** Let $a \in A$ be arbitrary. Since $\mathcal{P}$ is a partition, there is some $P \in \mathcal{P}$ containing a. Then clearly $a \in P$ and $a \in P$, so that aEa.
- **Symmetry:** Suppose we are given $a_1, a_2 \in A$ such that $a_1 E a_2$. Then there is some $P \in \mathcal{P}$ for which $a_1 \in P$ and $a_2 \in P$. Symmetrically, $a_2 \in P$ and $a_1 \in P$, showing that $a_2 E a_1$.
- **Transitivity:** Suppose we have $a_1, a_2, a_3 \in A$ such that $a_1 E a_2$ and $a_2 E a_3$. Then we have some $P_1 \in \mathcal{P}$ such that $a_1 \in P_1$ and $a_2 \in P_1$, and also some $P_2 \in \mathcal{P}$ such that $a_2 \in P_2$ and $a_3 \in P_2$. Furthermore, since $\mathcal{P}$ is a partition, if P_1 and P_2 were distinct, then we would have $P_1 \cap P_2 = \varnothing$. This does not occur, as P_1 and P_2 both contain the common element a_2. Therefore, $P_1 = P_2$, and so a_1 and a_3 both belong to the element P_1 of $\mathcal{P}$, proving that $a_1 E a_3$.

□

Therefore, we see that this correspondence runs both ways: every equivalence relation gives a partition, and every partition gives an equivalence relation. In fact, you can verify that the correspondence is reversible. For example, if you start with an equivalence relation, take the partition corresponding to it, and then take the equivalence relation corresponding to that partition, you recover the original equivalence relation. (Can you explain why?)

As a final note, we introduce one more definition:

Definition 11.4. Let A be a set, and let E be an equivalence relation on A. A set X is called a *set of representatives* for the equivalence relation E (or the partition A/E) if X contains exactly one element of each equivalence class. In other words, for each $[a] \in A/E$, we have $X \cap [a] = \{\alpha\}$ for some $\alpha \in [a]$.

In Example 11.3, one natural choice for a set of representatives would be $X = \{0, 1\}$, though any set containing exactly one even integer and exactly one odd integer would do. As we will see, the axioms of set theory we've currently laid out are not enough to guarantee that every equivalence relation *has* a set of representatives. This requires a further axiom, the Axiom of Choice, which we will introduce and discuss very soon.

MATH 145 Course Reading 12: Order Relations

October 7, 2020

To wrap up our detailed study of binary relations on a set, we look at one more important example: order relations. These are relations that behave like the familiar orderings on $\mathbb{Z}$ and $\mathbb{R}$ given by the symbol $\leq$. However, we will no longer insist on one key property of these two examples: we will relax the restriction that every two elements of the set are related (i.e. that for all a, b in the set, either $a \leq b$ or $b \leq a$ must hold). We will reserve the name *total ordering* for such special relations. Accordingly, a general order relation is often called a *partial ordering*, and we will look at a couple such examples below.

Definition of an Order Relation

An order relation on a set shares many properties in common with equivalence relations, but with one subtle difference that changes things considerably. For convenience, the official definition is stated below.

Definition 12.1. Let R be a binary relation on a set A. We say that R is *antisymmetric* if, whenever we have $a, b \in A$ such that aRb and bRa, it follows that $a = b$. A relation $\preceq$ on a set A is called an *order relation* (on A) if it is reflexive, antisymmetric, and transitive. We will also call $\preceq$ a *partial ordering* on A.

It is customary to adopt the notation $\preceq$ for an order relation, just as in the definition above. And we commonly read the expression $a \preceq b$ as "a is less than or equal to b", or else "b is greater than or equal to a". Of course, you can verify right away that the familiar relation $\leq$ on both $\mathbb{Z}$ and $\mathbb{R}$ is an order relation. And in fact, if $\mathcal{C}$ is a collection of sets, the subset relation $\subseteq$ on the sets inside of $\mathcal{C}$ is also an order relation, as you can immediately check.

Here are a couple less obvious examples:

Example 12.1. On any set A, we can define the *identity relation* R on A, where R contains only the ordered pairs (a, a), for each $a \in A$. Then R is trivially reflexive. It is antisymmetric, because as soon as we know aRb, we deduce that $a = b$. And R is transitive, because if aRb and bRc, then $a = b$ and $b = c$, so that $a = c$ and thus aRc.

Example 12.2. On the set $\mathbb{N}^+$ of positive integers, the divisibility relation $\mid$ defines an order relation. To be specific, we declare that for any $a, b \in \mathbb{N}^+$, we have $a \mid b$ if there is some integer k for which $b = ak$. Let's verify that $\mid$ satisfies the three conditions on an order relation:

- **Reflexive:** For every $a \in \mathbb{N}^+$, we have $a \mid a$ because we can write $a = a \cdot 1$.
- **Antisymmetric:** Suppose we have positive integers a and b such that $a \mid b$ and $b \mid a$. By definition, this gives us $b = ak$ and $a = b\ell$ for some $k, \ell \in \mathbb{Z}$. Putting these two equations together,

  $$b = ak = (b\ell)k = b(\ell k).$$

  Cancelling the non-zero integer b gives us $1 = \ell k$. Since k and ℓ are integers, the equation $\ell k = 1$ can only occur if $k = \ell = 1$ or $k = \ell = -1$. However, since a and b are both positive integers and $b = ak$, we must have $k = \ell = 1$, and in particular, $b = a$.
- **Transitive:** Suppose a, b, and c are positive integers such that $a \mid b$ and $b \mid c$. By definition, this tells us $b = ak$ and $c = b\ell$ for some $k, \ell \in \mathbb{Z}$. Putting this together,

  $$c = b\ell = (ak)\ell = a(k\ell),$$

  where $k\ell \in \mathbb{Z}$. This tells us that $a \mid c$, and so we are done.

One more note about order relations is worth making here: just as we can use the ordering $\leq$ to define a strict ordering $<$, we can do the same thing for an arbitrary order relation. If $\preceq$ is a partial ordering on a set A, we can define a strict ordering $\prec$ from it by declaring that for all $a, b \in A$, we have $a \prec b$ if $a \preceq b$ and $a \neq b$. The resulting strict order relation $\prec$ will still be transitive, but the reflexivity and antisymmetry properties are gone, replaced by a new property called *asymmetry*. A relation R on a set A is called *asymmetric* if the two relations aRb and bRa cannot both hold at the same time, for any $a, b \in A$.

Chains and Extremal Elements

As we hinted at earlier, not all pairs of elements have to be related under an order relation. Accordingly, the ones that are related are given a special name.

Definition 12.2. If $\preceq$ is a partial ordering on a set A, we say that two elements $a, b \in A$ are *comparable* if either $a \preceq b$ or $b \preceq a$. A partial ordering in which every pair of elements is comparable is called a *total ordering*, or a *linear ordering*.

As mentioned, the familiar ordering $\leq$ on $\mathbb{Z}$ or $\mathbb{R}$ is a total ordering. On the other hand, the divisibility ordering considered in Example 12.2 is not a total ordering; for instance, 2 and 3 are incomparable (meaning: not comparable).

Another important notion that comes along with an order relation is the notion of a *chain*:

Definition 12.3. If $\preceq$ is a partial ordering on a set A, a subset C of A is called a *chain* if every pair of elements in C are comparable.

In particular, if A is totally ordered by $\preceq$, then A itself is a chain in A. In Example 12.2, the set $C = \{1, 2, 4, 8, 16, \dots\}$ of powers of 2 is a chain in $\mathbb{N}^+$ under the divisibility ordering. On the other hand, in Example 12.1, any pair of distinct elements of A is incomparable, so there are no chains in A with more than one element.

Because not every pair of elements in an ordered set must be comparable, we have to be careful in distinguishing between least/greatest elements of a given set, and maximal/minimal elements of a given set. The distinction is given in the following definition:

Definition 12.4. Let A be a set with partial order relation $\preceq$. Given a subset B of A, we say that

- An element $b \in B$ is a *least element* of B if we have $b \preceq b'$ for all elements $b' \in B$. Similarly, an element $b \in B$ is a *greatest element* of B if we have $b' \preceq b$ for all $b' \in B$.
- An element $b \in B$ is a *minimal element* of B if there are no smaller elements of B, i.e. if $b' \preceq b$ for some $b' \in B$, then $b = b'$. Similarly, an element $b \in B$ is a *maximal element* of B if there are no larger elements of B, so that if $b \preceq b'$ for some $b' \in B$, then $b = b'$.

It is a quick exercise to show that if a subset B has a least or greatest element, then it is unique (proving this is encouraged!). It also follows directly from the definition that every least element is also minimal, and that every greatest element is maximal. However, minimal elements do not have to be least elements, nor do maximal elements have to be greatest elements. We will see an example of this first assertion directly below:

Example 12.3. Let's reconsider Example 12.2, where $\mathbb{N}^+$ is ordered by the divisibility relation. In this case, 1 is a least element, since 1 divides every positive integer. However, the subset $\mathbb{N}^+ \setminus \{1\}$ no longer has a least element. For example, there is no positive integer other than 1 that divides both 2 and 3. On the other hand, this subset $\mathbb{N}^+ \setminus \{1\}$ has infinitely many minimal elements: every prime number p is minimal, since it is not divisible by any positive integers other than 1 and itself.

Bounds on Sets, Suprema, and Infima

Next, we introduce a few notions for ordered sets that you will become quite familiar with if you are taking MATH 147:

Definition 12.5. Suppose that A is a set with order relation $\preceq$, and that B is a subset of A.

- An element $a \in A$ is a *lower bound* for B if $a \preceq b$ for all $b \in B$. Similarly, $a \in A$ is an *upper bound* for B if $b \preceq a$ for all $b \in B$.
- An element $a \in A$ is an *infimum*, or *greatest lower bound* for B if a is the greatest element of the set of lower bounds for B. Similarly, $a \in A$ is a *supremum*, or *least upper bound* for B if a is the least element of the set of all upper bounds for B.

If they exist, we use $\sup B$ and $\inf B$ to denote the supremum and infimum of a set B, respectively.

A given set B can have many lower or upper bounds, but the supremum and infimum are both unique if they exist (being the least/greatest elements of a particular set). You may already be familiar with examples in $\mathbb{R}$ with the usual ordering. For example, the open interval $B = (2, 4)$ has infinitely many lower bounds (any real number up to and including 2 would work), but the infimum of the set is 2, since 2 is the greatest possible lower bound.

MATH 145 Course Reading 13: The Axiom of Choice, Zorn's Lemma, and the Well-Ordering Theorem

October 9, 2020

In today's reading, we take a look at another axiom of set theory, the Axiom of Choice, along with two of its most famous equivalent formulations. This axiom differs from all the ones we've looked at so far, because the Axiom of Choice is sometimes taken as "optional" by set theorists. In other words, they study what parts of mathematics remain true *without* the Axiom of Choice, and which mathematical statements really do require this axiom. One reason that some mathematicians do this is because the Axiom of Choice yields some counter-intuitive mathematical consequences.

For instance, much later in your mathematical studies, you will encounter the notion of *Lebesgue measure*, which is a function μ that takes subsets of the real line as input, and outputs a non-negative real number (or ∞) roughly corresponding to the one-dimensional volume of that set (which we call the *measure* of the set). This measure agrees with our mathematical intuition, in the sense that the measure of an open interval (a, b) is $b - a$ for any real numbers $a < b$.

One consequence of the Axiom of Choice is that no matter how we try to define μ, if we insist that it has a couple reasonable properties (the measure of a set not changing if you translate it, and a property called *countable additivity*), then it will always be impossible to assign a measure to every subset of $\mathbb{R}$.

The study of the Axiom of Choice and its equivalent versions would take us too far afield for this course, and so this reading will focus only on defining and explaining them, without proving too much about them.

Infinite Cartesian Products

So far, we have seen how to formally define the Cartesian product $A \times B$ of two sets, as the set of all ordered pairs whose first coordinate belongs to the set A, and second coordinate belongs to the set B. This construction can then be iterated to construct the Cartesian product of any finite number of sets. For instance, we defined $A \times B \times C$ to be $(A \times B) \times C$, and $A \times B \times C \times D$ to be $(A \times B \times C) \times D$, and so on.

But what if we wanted to define a Cartesian product of an infinite number of sets? This calls for a more general definition, which we will now give. Suppose $\mathcal{C}$ is a non-empty collection of sets (finite or infinite). We would like to define the Cartesian product of all the sets in $\mathcal{C}$, for which we will adopt the notation

$$\prod_{C \in \mathcal{C}} C.$$

If $\mathcal{C}$ is an infinite collection, we'd somehow like to talk about an infinite ordered tuple of elements, with one slot in the tuple for each set in $\mathcal{C}$. This suggests we might make use of a function $\alpha : \mathcal{C} \to \bigcup \mathcal{C}$. For each $C \in \mathcal{C}$, the function value $\alpha(C)$ should give the "Cth coordinate" of the element in the Cartesian product. In particular, the value $\alpha(C)$ should lie in the set C. Let's summarize:

Definition 13.1. Let $\mathcal{C}$ denote a nonempty collection of sets. The *Cartesian product* $\prod_{C \in \mathcal{C}} C$ is the set of functions $\alpha : \mathcal{C} \to \bigcup \mathcal{C}$, with the property that for each $C \in \mathcal{C}$, we have $\alpha(C) \in C$.

We can use this new definition of Cartesian product to essentially recover our old definition of the Cartesian product of finitely many sets. For example, to recover $A \times B$ we can take $\mathcal{C} = \{A, B\}$. The elements of $\prod_{C \in \mathcal{C}} C$ are functions $\alpha : \{A, B\} \to A \cup B$, where $\alpha(A) \in A$ and $\alpha(B) \in B$. For each such function α, we can associate it to the ordered pair $(\alpha(A), \alpha(B)) \in A \times B$, and you can in fact check that this mapping defines a bijection between the new notion of Cartesian product and the old one. Of course, this type of argument

generalizes to any finite collection of sets.

One important question that we'd then like to answer is: is the Cartesian product of any non-empty collection of non-empty sets itself non-empty? More symbolically, if $\mathcal{C}$ is a nonempty collection that contains only nonempty sets, is it always true that $\prod_{C \in \mathcal{C}} C \neq \varnothing$? While it might seem obvious to you that this is true, it turns out (in fact) that none of the standard axioms of set theory allow you to prove the answer is yes! That's why we introduce a new axiom, the Axiom of Choice, that claims this:

Axiom 13.1 (Axiom of Choice)**.** The Cartesian product of any non-empty collection of non-empty sets is non-empty.

This is not the only way to formulate the Axiom of Choice. Another popular equivalent way to state the axiom is via the notion of a *choice function*. Formally, given a collection $\mathcal{C}$ of non-empty sets, a *choice function* for $\mathcal{C}$ is a function $f : \mathcal{C} \to \bigcup \mathcal{C}$, such that for each $A \in \mathcal{C}$, we have $f(A) \in A$. In other words, a choice function for $\mathcal{C}$ picks out an element of every set in $\mathcal{C}$, all at once.

The reason for the name is that the function f just described is making a very large number of simultaneous choices. It should be evident that the Axiom of Choice is equivalent to the existence of a choice function for every nonempty collection of sets $\mathcal{C}$.

Zorn's Lemma and the Well-Ordering Theorem

Another, seemingly different statement about partially ordered sets turns out to be logically equivalent to the Axiom of Choice. The statement is known as Zorn's lemma, and it uses the terminology around order relations that we introduced in the previous reading:

Theorem 13.1 (Zorn's Lemma)**.** *Let A be a partially ordered set, with order relation $\preceq$. Suppose further that every chain in A has an upper bound in A. Then A has a maximal element.*

For this course at least, we will not delve into the thicket of logic that shows this is equivalent to the Axiom of Choice, but will simply take it for granted. Zorn's lemma is a powerful tool for showing that maximal elements of various kinds exist, usually with respect to the ordering given by set containment. As one single application, we will give the following corollary of Zorn's lemma to illustrate how it might be used:

Corollary 13.1. *Suppose that A is a partially ordered set with order relation $\preceq$. Then every chain $\mathcal{C}$ in A is contained in a maximal chain $\mathcal{M}$. In other words, there is a chain $\mathcal{M}$ such that $\mathcal{C} \subseteq \mathcal{M}$, and such that there are no chains in A properly containing $\mathcal{M}$.*

Proof. Let $\mathcal{C}$ be an arbitrary chain in A. We will apply Zorn's lemma to the set Γ of all chains in A containing $\mathcal{C}$, with the subset relation $\subseteq$ on $\mathcal{P}(A)$ giving the order relation between elements of Γ. Now suppose we are given a chain $\mathcal{D}$ in the ordered set Γ (note this is different from a chain in A: each *element* of $\mathcal{D}$ is a chain in A containing $\mathcal{C}$). We claim that $\mathcal{D}$ has an upper bound in Γ.

Indeed, we define $\mathcal{C}_0 = \bigcup \mathcal{D}$, the union of all the elements of all the chains in $\mathcal{D}$. Clearly, for each element $\mathcal{C}' \in \mathcal{D}$, we have $\mathcal{C}' \subseteq \mathcal{C}_0$, because every element of $\mathcal{C}'$ belongs to $\mathcal{C}_0$ by construction. Furthermore, since $\mathcal{C}_0$ contains every chain in $\mathcal{D}$, each of which contains $\mathcal{C}$, we know that $\mathcal{C}_0$ contains $\mathcal{C}$. Thus $\mathcal{C}_0$ will be an upper bound on the chain $\mathcal{D}$, as soon as we verify that $\mathcal{C}_0 \in \Gamma$, i.e. that $\mathcal{C}_0$ is a chain in A.

So suppose we are given two elements $a_1, a_2 \in \mathcal{C}_0$. By construction, each of a_1 and a_2 belong to a chain in $\mathcal{D}$, say $a_1 \in \mathcal{C}_1$ and $a_2 \in \mathcal{C}_2$. Since $\mathcal{D}$ is a chain with respect to set containment, either $\mathcal{C}_1 \subseteq \mathcal{C}_2$ or $\mathcal{C}_2 \subseteq \mathcal{C}_1$. Without loss of generality, say $\mathcal{C}_1 \subseteq \mathcal{C}_2$. Thus both a_1 and a_2 belong to $\mathcal{C}_2$, and since $\mathcal{C}_2$ is a chain, the elements a_1 and a_2 are comparable with respect to $\preceq$. This verifies that the union $\mathcal{C}_0$ is also a chain in A.

By Zorn's lemma, applied to the set Γ of all chains containing $\mathcal{C}$ ordered by $\subseteq$, there is a chain $\mathcal{M} \in \Gamma$, maximal with respect to set containment. This is a chain in A with respect to the ordering $\preceq$, containing $\mathcal{C}$ and not properly contained in any other chain in A. □

The Well-Ordering Theorem is the final well-known statement that turns out to be equivalent to the Axiom of Choice. The idea of a *well-ordering* is already familiar to us from the Well-Ordering Principle for $\mathbb{N}$, but we now define a well-ordering in full generality:

Definition 13.2. Suppose A is a set with order relation $\preceq$. The ordered set A is said to be *well-ordered* if every non-empty subset of A has a least element (with respect to the given order relation $\preceq$).

Thus $\mathbb{N}$, with the usual ordering $\leq$, is a well-ordered set by the Well-Ordering Principle. In particular, well-ordered sets exist. What is perhaps surprising is that the Axiom of Choice is equivalent to the following theorem:

Theorem 13.2 (Well-Ordering Theorem)**.** Every *non-empty set has a well-ordering. In other words, if A is a nonempty set, then there is an order relation $\preceq$ on A, such that A is well-ordered with respect to this order relation.*

Even while being equivalent to the intuitive-looking Axiom of Choice, this theorem is distinctly counter-intuitive. For instance, it implies that $\mathbb{R}$ has some kind of order relation $\preceq$ that makes it into a well-ordered set. This ordering would have to be distinctly different from the usual ordering $\leq$ on $\mathbb{R}$, because with respect to this ordering, there are lots of sets that don't have least elements: every open interval, for instance.

There's a somewhat famous joke in the mathematical community, which we can paraphrase as follows: the Axiom of Choice is obviously true, the Well-Ordering Theorem is obviously false, and Zorn's Lemma is too complicated to be either obviously true or obviously false. (The joke being, of course, that all three are logically equivalent!) Again, this partially explains why some mathematicians are resistant to the Axiom of Choice: it can be used to deduce theorems that seem remarkably counter-intuitive, whereas an "axiom" ought to be intuitively true.

MATH 145 Course Reading 14:
Defining the Cardinality of Sets

October 19, 2020

Both in our synchronous class discussions and on Assignment 4, we have laid the groundwork for defining the *cardinality* of an arbitrary set. Not only will we define the notion of *numerical equivalence* (*same cardinality*) and prove that it resembles an equivalence relation, but we will also be able to say when the cardinality of one set is *at most* the cardinality of another set. In turn, this leads to something resembling an order relation we can place on all sets.

Once this formal definition of cardinality is established, we will officially be able to define *finite set* and *infinite set*, which are two notions you have likely used intuitively for a long time! We will also be able to distinguish between different sizes of infinite set; in particular, it will be useful to single out the *countable sets* from the *uncountable sets*. These tasks will be undertaken in the next several readings following this one.

Comparing Cardinalities of Sets

As we mentioned in our synchronous discussion, one of the best ways to decide if two very large collections of objects have the same size is to match each element in one collection uniquely with an object in another. In a related vein, if we have a collection of objects and we simply change the name of each object, the number of objects we have should not change. This suggests that two sets should have the same size if there is an exact matching between the elements of the two sets. In the language of functions we now have, what we are looking for is a *bijection* between the sets. Let's formalize:

Definition 14.1. Two sets A and B are *numerically equivalent*, or have the *same cardinality*, if there is a bijection $f : A \to B$. In this case, we write $|A| = |B|$ to indicate this.

The next result states that the equals sign on set cardinalities behaves like an equivalence relation. The only reason it is *not* an equivalence relation on all sets is because we can't define a set of all sets on which to define the relation (thanks to Russell's paradox).

Proposition 14.1.

(1) For all sets A, we have $|A| = |A|$.

(2) For all sets A and B, if $|A| = |B|$, then $|B| = |A|$.

(3) For all sets A, B, and C, if $|A| = |B|$ and $|B| = |C|$, then $|A| = |C|$.

Proof. The proof of each of these three claims is essentially identical to the proof of question 2(b) on Assignment 4. There, you showed that on any collection $\mathcal{C}$ of sets, the relation $\mathcal{R}$ on $\mathcal{C}$ given by

$$A \mathcal{R} B \text{ if there is a bijection } f : A \to B$$

is actually an equivalence relation. □

Given that bijections intuitively provide a perfect matching of the elements of one set with the elements of another, we can similarly see an injective function $f : A \to B$ as perfectly matching up elements of A with elements of B, but perhaps without exhausting all the elements of B. So we would want to say that A is at most as big as B:

Definition 14.2. Given two sets A and B, we say that the *cardinality of A is less than or equal to the cardinality of B*, and write $|A| \leq |B|$, if there is an injective function $f : A \to B$.

Properties of Cardinality $\leq$

Ultimately, we would like to prove that this symbol $\leq$ behaves essentially like an order relation on the collection of all sets (with the only stumbling block to defining an *actual* order relation being the inability to define a set of all sets). The one property of order relations that will give us some trouble is antisymmetry, so we first focus on proving a lemma that helps it along. While the result may seem intuitively clear, you'll notice that the proof requires some work.

Lemma 14.1. *Suppose we have three sets A_1, B, A such that $A_1 \subseteq B \subseteq A$. If $|A_1| = |A|$, then $|B| = |A|$.*

Proof. Eventually, our goal is to construct a bijection $g : A \to B$. In order to do this, however, we'll need to do some set-up work up-front.

We're given that $|A_1| = |A|$, so let $f : A \to A_1$ be a bijection. We now use f to define two sequences of sets, starting with A and B. We set $A_0 = A$, $B_0 = B$, and for each $n \in \mathbb{N}$, we set

$$A_{n+1} = f(A_n) \quad B_{n+1} = f(B_n).$$

Since f is a bijection from A to A_1, note that $f(A_0) = A_1$, so that no notational clash occurs above (there really is only one set A_1). In fact, we can say more: that for each $n \in \mathbb{N}$, we have $A_{n+1} \subseteq A_n$. This can be established immediately using induction from the known fact that $A_1 \subseteq A_0$.

To define our bijection $g : A \to B$, certainly we'll want to map all the elements of $A \setminus B$ into B. But we'll also need room in B to map them, so we'll need to map the elements of B into a smaller set too. To accomplish this, we define another sequence of sets; for each $n \in \mathbb{N}$, we set

$$C_n = A_n \setminus B_n.$$

Next, set

$$C = \bigcup_{n=0}^{\infty} C_n,$$

so that C is the union of all the sets in the sequence $C_0, C_1, \ldots$.

We claim that $f(C_n) = C_{n+1}$, for all natural numbers n. First, note that if $a \in f(C_n)$, then $a = f(c)$ for some $c \in C_n$. By definition, $c \in A_n$ and $c \notin B_n$. Then $f(c) \in f(A_n) = A_{n+1}$. We also claim $f(c) \notin f(B_n)$. Indeed, if $f(c) = f(b)$ for some $b \in B_n$, then since f is injective, we get $c = b$, so that $c \in B_n$, a contradiction.

We have now established that $f(c) \in A_{n+1} \setminus B_{n+1}$, i.e. $f(c) \in C_{n+1}$. This proves that $f(C_n) \subseteq C_{n+1}$. On the other hand, if $a \in C_{n+1}$, then $a \in A_{n+1}$ and $a \notin B_{n+1}$. Thus $a = f(a')$ for some $a' \in A_n$, and since $a \notin B_{n+1}$, we know that $a' \notin B_n$. This means $a' \in C_n$, so that $a = f(a') \in f(C_n)$. This proves $C_{n+1} \subseteq f(C_n)$, so that $f(C_n) = C_{n+1}$.

Our next claim is that

$$f(C) = \bigcup_{n=1}^{\infty} C_n.$$

Certainly, if $a \in f(C)$, then $a = f(c)$ for some $c \in C$. Then $c \in C_n$ for some $n \in \mathbb{N}$, and this means $a = f(c) \in f(C_n) = C_{n+1}$, proving that $a \in \bigcup_{n=1}^{\infty} C_n$. Conversely, if $a \in \bigcup_{n=1}^{\infty} C_n$, then $a \in C_n$ for some positive integer n. In particular, $n - 1 \in \mathbb{N}$, and $C_n = f(C_{n-1})$. So $a = f(c)$ for some $c \in C_{n-1}$, and in particular $a \in f(C)$. This proves our most recent claim.

Finally, we define another set $D = A \setminus C$. We define our bijection g by defining it separately on the two sets C and D that partition A. The elements of C will be mapped into a smaller set, and the elements of D are left where they are. Explicitly, we define a function $g : A \to B$ by taking

$$g(x) = \begin{cases} f(x), & \text{if } x \in C \\ x, & \text{if } x \in D. \end{cases}$$

First, we verify that for all $x \in A$, we indeed have $g(x) \in B$. If $x \in C$, then $f(x) \in C_n$ for some $n \geq 1$, which implies $f(x) \in A_n$. Now since $A_0 \supseteq A_1 \supseteq A_2 \supseteq \cdots$, we see in particular that $f(x) \in A_1$ and $A_1 \subseteq B$, so $f(x) \in B$. On the other hand, if $x \in D$, then $x \notin C$, and in particular, $x \notin C_0$. So $x \notin A \setminus B$, which means it is *not* the case that $x \notin B$. In other words, $x \in B$ whenever $x \in D$.

Now that we're assured g defines a function from A to B, we check that it is a bijection. To see that g is one-to-one, suppose we have $x_1, x_2 \in A$ such that $g(x_1) = g(x_2)$. If x_1 and x_2 both belong to C, then $f(x_1) = f(x_2)$, and so $x_1 = x_2$ because f is injective. If x_1 and x_2 both belong to D, then $x_1 = x_2$ immediately follows. If x_1 belongs to C and x_2 belongs to D, then $f(x_1) = x_2$. But $f(x_1) \in C$ while $x_2 \notin C$, so we have a contradiction. The same contradiction applies if x_1 is in D and x_2 is in C, so we have now shown g is injective.

To prove that g is surjective, suppose we are given $b \in B$. If $b \in f(C)$, then clearly there is $c \in C$ such that $f(c) = b$, so that $g(c) = f(c) = b$, as needed. Otherwise, if $b \notin f(C)$, then either $b \in C_0$ or $b \in D$. If $b \in D$, then $g(b) = b$, and we're done. If $b \in C_0$, then $b \in A \setminus B$, and this contradicts the assumption that $b \in B$. So we are now assured that g is surjective.

In conclusion, the function $g : A \to B$ just defined is a bijection, and this shows $|A| = |B|$. □

At last, we're ready to prove that the symbol $\leq$ on cardinality behaves like an order relation:

Proposition 14.2.

(1) For all sets A, B, C, if $|A| \leq |B|$ and $|A| = |C|$, then $|C| \leq |B|$.

(2) For all sets A, B, C, if $|A| \leq |B|$ and $|B| = |C|$, then $|A| \leq |C|$.

(3) For all sets A, B, C, if $|A| \leq |B|$ and $|B| \leq |C|$, then $|A| \leq |C|$.

(4) (Cantor-Schröder-Bernstein Theorem) For all sets A and B, if $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.

Proof. The first three of these results all follow from the same general fact: if $f : A \to B$ and $g : B \to C$ are both injective functions, then so is $g \circ f : A \to C$. (Can you see how (1) and (2) follow from this?). You will have proved this result as part of your solution to Question 2(b) on Assignment 4.

To prove the Cantor-Schröder-Bernstein Theorem, suppose we have injective functions $f : A \to B$ and $g : B \to A$ for some sets A and B. We begin by considering the composition $g \circ f : A \to A$, which is also an injective function, by the previous paragraph. For notational ease, set $X = g(B)$ and $Y = g(f(A))$. Clearly $X \subseteq A$, and since $f(A) \subseteq B$, we get that $Y \subseteq X$. On the other hand, since $g \circ f$ is injective, it is also a bijective function from A to $(g \circ f)(A) = Y$. In particular, $|Y| = |A|$.

So we have set containments $Y \subseteq X \subseteq A$, and $|Y| = |A|$. Applying Lemma 14.1 above, we conclude that $|X| = |A|$. But $X = g(B)$, and g is injective, so $g : B \to X$ is a bijection, and $|X| = |B|$. We conclude that $|A| = |B|$, as desired. □

As expected, we will also adopt the notation $|A| < |B|$ to indicate that $|A| \leq |B|$ and that $|A| \neq |B|$, i.e. there is an injective function from A to B, but no bijective function.

While we'd like to be able to say immediately that any finite subset of $\mathbb{N}$ has smaller cardinality than $\mathbb{N}$ itself, a little work is required to show that no bijection between $\mathbb{N}$ and one of its finite subsets can exist (in fact, this is at the root of the distinction between a finite and an infinite set). One thing we *can* do without extra work is the following:

Example 14.1. For all nonempty sets B, we have $|\varnothing| < |B|$. For this, we must verify that there is an injective function $f : \varnothing \to B$, but no such bijective function. Note first that there is trivially one function from the empty set to any set, namely the empty relation from $\varnothing$ to B. This empty relation vacuously has the property that for each $a \in \varnothing$, there is a unique $b \in B$ belonging to the relation. The function is also vacuously injective: there are no elements $a_1, a_2 \in \varnothing$ for which to have the property $f(a_1) = f(a_2)$. However, the empty function is *not* surjective: given an element $b \in B$, there is no $a \in \varnothing$ for which $f(a) = b$.

MATH 145 Course Reading 15:
Finite Sets

October 21, 2020

In this reading, we give a precise mathematical definition to the notion of *finite set*, and study some of the properties that make finite sets unique. In particular, we will be able to prove that the Axiom of Infinity really is necessary to be able to construct an infinite set.

Definition of Finite, and Basic Properties

Our intuitive notion of what it means to be "finite" is that we should be able to list off the elements in a terminating sequence. In other words, we should be able to count the elements, and have that count stop at some point. More formally, this amounts to defining a bijection between our set and some natural number. That notion is captured exactly in our definition below:

Definition 15.1. A set A is called *finite* if A has the same cardinality as n for some $n \in \mathbb{N}$. In such a case, we also write $|A| = n$ instead of $|A| = |n|$, and we say that A *has* n *elements*. A set is called *infinite* if it is not finite.

The first thing we want to be sure of is that the size of a finite set is well-defined. In other words, can it be the case that there is a set A for which $|A| = n$ and $|A| = m$ for distinct natural numbers n and m? Notice this is equivalent to the assertion that two natural numbers n and m have the same cardinality. We officially rule that out in the results that follow:

Lemma 15.1. *For any $n \in \mathbb{N}$, there is no injective mapping from n to a proper subset X of n.*

Proof. If this statement is false, then by the well-ordering principle, we can find some least $n \in \mathbb{N}$ for which there *is* an injective mapping from n to one of its proper subsets X. Clearly, $n \neq 0$, because there are no proper subsets of 0, so certainly no injective mapping from 0 to a proper subset of 0.

We now consider two cases: either $n-1 \in X$ or $n-1 \notin X$. If $n-1 \notin X$, then in fact $X \subseteq (n-1)$, and $n-1 \in \mathbb{N}$ because $n \neq 0$. If $f : n \to X$ is an injective mapping, we can then define a new injective mapping $g : n-1 \to X \setminus \{f(n-1)\}$ by taking $g(k) = f(k)$ for each $k \in n-1$. (We call g the *restriction of f to the set $n-1$*). Note that g is automatically injective since f is, and g maps $n-1$ to a proper subset of $n-1$, since X is a subset of $n-1$ and the range of g misses at least the element $f(n-1) \in n-1$. This contradicts the minimality of n.

Now suppose that $n-1 \in X$. Since f is injective, we know that $n-1 = f(k)$ for some unique $k \in n$. We use this to define a new function $g : n-1 \to X \setminus \{n-1\}$ by taking

$$g(i) = \begin{cases} f(i), & i \neq k \\ f(n-1), & i = k \end{cases}$$

In other words, g is defined the same as f, except that we redefine $f(k)$ to be $f(n-1)$ instead, which will make sure g maps into $X \setminus \{n-1\}$. You can check immediately that g is an injective function, mapping $n-1$ into the proper subset $X \setminus \{n-1\}$ of $n-1$. Again, this contradicts the minimality of our choice of n. Thus our assumption that the statement is false must be invalid. □

This lemma has a couple consequences, one of which is that there is a unique natural number denoting the cardinality of each finite set:

Corollary 15.1.

(1) If A is a finite set such that $|A| = n$ and $|A| = m$ for $n, m \in \mathbb{N}$, then $n = m$.

(2) The set $\mathbb{N}$ is infinite.

Proof. To prove (1), note that the hypothesis implies that n and m have the same cardinality. Now, if $n \neq m$, then either n is a proper subset of m, or m is a proper subset of n, by the construction of $\mathbb{N}$. Either way, the existence of bijections $f : n \to m$ and $g : m \to n$ contradict the conclusion of Lemma 15.1, because either f or g would give an injection from a natural number to one of its proper subsets.

To prove (2), it suffices to show that $\mathbb{N}$ has the same cardinality as one of its proper subsets (why?). There are many candidate choices; it suffices to note that the doubling map $d : \mathbb{N} \to \mathbb{N}$ given by $d(n) = 2n$ for each $n \in \mathbb{N}$ maps $\mathbb{N}$ bijectively onto the proper subset of even natural numbers. (Check this if it's not clear!) □

So, true to its name, the Axiom of Infinity gives us infinite sets. But can we formally prove that we can't get infinite sets without the Axiom of Infinity? That's one of our main projects for the next section.

Properties of Finite Sets

One of the most intuitive facts about finite sets are that they're the "smallest" class of set, in the sense that every subset of a finite set is finite. We prove this below:

Theorem 15.1. *If A is a finite set and $B \subseteq A$, then $|B| \leq |A|$ and B is finite.*

Proof. Given the assumptions of the theorem, first we define an *inclusion mapping* $\iota : B \to A$, given by $\iota(b) = b$ for all $b \in B$. Clearly this mapping is an injective mapping from B to A, showing that $|B| \leq |A|$. Now we bring in the assumption that A is finite, say $|A| = n$ for some $n \in \mathbb{N}$. Furthermore, since the conclusion that B is finite is trivial if $n = 0$, suppose that $n \geq 1$. By way of the bijection from A to n, we can number the elements of A, say $A = \{a_0, a_1, \ldots, a_{n-1}\}$.

Given that B is a subset of A, if B is empty, we're done. Otherwise, there is some least index i for which $a_i \in B$. Call this element b_0. If $B = \{b_0\}$, we have that B is finite, and we're done. Otherwise, there is some least index $i_1 > i$ such that $a_{i_1} \in B$, and we call this element b_1. Again, if $B = \{b_0, b_1\}$, then B is finite and we're done. Otherwise, we just keep repeating the procedure, which cannot go on indefinitely, because the increasing sequence of indices $i < i_1 < i_2 < \cdots$ for which $a_i, a_{i_1}, a_{i_2}, \ldots \in B$ will eventually be larger than $n - 1$ if allowed to increase indefinitely. Thus there is some $m \in \mathbb{N}$ for which $B = \{b_0, b_1, \ldots, b_{m-1}\}$, proving that B is indeed finite. □

Note that every application of the Axiom Schema of Comprehension to a set X results in a subset of X. Thus the above result shows that if only finite sets exist, then the Axiom Schema of Comprehension can only derive more finite sets.

Similarly, let's prove the intuitive fact that the image of a finite set under a function is always finite:

Theorem 15.2. *Suppose A and B are sets, A is finite, and $f : A \to B$ is a function. Then the image $f(A)$ is a finite subset of B. In fact, $|f(A)| \leq |A|$.*

Proof. Since the result is again trivial if $A = \varnothing$, we assume $|A| = n$, where $n \geq 1$, and we write $A = \{a_0, \ldots, a_{n-1}\}$. Clearly, as a set, we have $f(A) = \{f(a_0), \ldots, f(a_{n-1})\}$, where the elements in the set $f(A)$ may now be presented with repetition. To remove the repetition, we can set $b_0 = f(a_0)$, then find the least index $i_1 > 0$ for which $f(a_{i_1}) \neq f(a_0)$ (if it exists) and set $b_1 = f(a_{i_1})$. Again, if $\{b_0, b_1\}$ does not exhaust $f(A)$, we find the least index $i_2 > i_1$ such that $f(a_{i_2}) \notin \{b_0, b_1\}$ and set $b_2 = f(a_{i_2})$. Continuing in this way, and knowing that the increasing sequence $0 < i_1 < i_2 < \cdots$ cannot surpass $n - 1$, this process must eventually stop and yield $f(A) = \{b_0, b_1, \ldots, b_{m-1}\}$ for some $m \in \mathbb{N}$, proving that $f(A)$ is finite.

Then, we can define an injective function $g : f(A) \to A$ by taking $g(b_0) = a_0$, and for $k > 0$, taking $g(b_k) = a_{i_k}$, where i_k is the sequence element constructed above. You can verify right away that g is injective, establishing that $|f(A)| \leq |A|$. □

Next, we build up to showing that the union of a finite collection of finite sets is finite, so that the Axiom of Union alone cannot give us infinite sets. We begin with the simplest case:

Lemma 15.2. *If A and B are both finite sets, then $A \cup B$ is finite and $|A \cup B| \leq |A| + |B|$. If A and B are disjoint, then in fact $|A \cup B| = |A| + |B|$.*

Proof. Suppose that $|A| = n$ and $|B| = m$, and that we label the elements of A and B accordingly, so that $A = \{a_0, \ldots, a_{n-1}\}$ and $B = \{b_0, b_1, \ldots, b_{m-1}\}$. Certainly, $A \cup B = \{a_0, \ldots, a_{n-1}, b_0, b_1, \ldots, b_{m-1}\}$, but there may be some repetitions of elements in $A \cup B$. With repetitions, $A \cup B$ has $m+n$ elements, but removing the repetitions will only drop the number of distinct elements, so that $|A \cup B| \leq m+n = |A|+|B|$, as desired. On the other hand, if A and B are disjoint, then there are no repetitions on the list $a_0, \ldots, a_{n-1}, b_0, \ldots, b_{m-1}$, which means $|A \cup B| = m + n = |A| + |B|$. □

A straightforward extension by induction then shows

Theorem 15.3. *If S is a finite collection of finite sets, then $\bigcup S$ is also finite.*

Proof. We proceed by induction on $|S|$. If $|S| = 0$, then $S = \varnothing$, so $\bigcup S = \varnothing$, and the conclusion follows immediately. Assuming the result for all collections of n finite sets, for some $n \in \mathbb{N}$, now suppose S is a collection of $n + 1$ finite sets, say $S = \{A_0, A_1, \ldots, A_n\}$. The induction hypothesis tells us that the union $\bigcup_{k=0}^{n-1} A_k$ is finite, and since

$$\bigcup_{k=0}^{n} A_k = \left(\bigcup_{k=0}^{n-1} A_k \right) \cup A_n,$$

and both sets in the union on the right are finite, the set $\bigcup_{k=0}^{n} A_k$ is finite by Lemma 15.2. We conclude the statement is true for all values of $|S|$, by induction. □

Looking at the rest of the six starting axioms for set theory, the Axiom of Existence claims that a finite set exists, the Axiom of Pair only constructs sets with at most two elements, and the Axiom of Extensionality does not construct any new sets whatsoever. As for the last remaining axiom, the Power Set Axiom, we will now see that it too gives us only finite sets when we start with finite sets:

Theorem 15.4. *For any finite set A, the set $\mathcal{P}(A)$ is also finite.*

Proof. Once more, we proceed by induction on $|A|$. If $|A| = 0$, then $A = \varnothing$ and $\mathcal{P}(A) = \{\varnothing\}$, which is finite. Now assume we have $n \in \mathbb{N}$ such that for any set A with $|A| = n$, we know that $\mathcal{P}(A)$ is finite. We take an arbitrary set A with $|A| = n+1$ and show that $\mathcal{P}(A)$ is again finite. Because $|A| = n+1$, we can enumerate the elements, say $A = \{a_0, a_1, \ldots, a_n\}$. We now split $\mathcal{P}(A)$ into two disjoint subsets. We set

$$\mathcal{B} = \{B \in \mathcal{P}(A) : a_n \in B\}$$

and

$$\mathcal{C} = \{C \in \mathcal{P}(A) : a_n \notin B\}.$$

It should be clear from the construction that $\mathcal{P}(A) = \mathcal{B} \cup \mathcal{C}$, and that this is a disjoint union. Furthermore, notice that both $\mathcal{B}$ and $\mathcal{C}$ have the same cardinality as the power set of $A \setminus \{a_n\} = \{a_0, a_1, \ldots, a_{n-1}\}$, which we'll call A'. Indeed, a bijection from $\mathcal{P}(A')$ to $\mathcal{B}$ would take an arbitrary subset B of A' and map it to $B \cup \{a_n\}$, and a bijection from $\mathcal{P}(A')$ to $\mathcal{C}$ would take an arbitrary subset C of A' and again map it to C. In particular, both $\mathcal{B}$ and $\mathcal{C}$ are finite, since $\mathcal{P}(A')$ is finite by induction hypothesis.

We now know that $\mathcal{P}(A)$ can be written as the union of the two finite sets $\mathcal{B}$ and $\mathcal{C}$, so that $\mathcal{P}(A)$ is finite by Lemma 15.2. By induction, we conclude that the power set of every finite set is finite. □

Hence, we have now formally proved that the Axiom of Infinity must be added to the first six axioms of set theory we introduced in order to derive any infinite sets. Over the next two readings, we more fully explore the properties of infinite sets, distinguishing in particular between the countable and uncountable infinite sets.

MATH 145 Course Reading 16: Countable Sets

October 23, 2020

Now that we have studied the finite sets, it is time to conduct a more refined analysis of the infinite sets. In this reading, we spend some time studying the "smallest" class of infinite sets, the so-called *countable* sets. We will see that many familiar examples of infinite sets are indeed countable, and also that the property of countability is preserved under taking finite or countable unions, Cartesian products, and more.

Definition of Countable Sets, and Examples

Similarly to how we define finite sets as the ones with the same cardinality as a fixed natural number, the countable sets are the ones with cardinality equal to that of $\mathbb{N}$.

Definition 16.1. A set A is called *countable* if $|A| = |\mathbb{N}|$, and is called *at most countable* if $|A| \leq |\mathbb{N}|$. If A is countable, we may sometimes write $|A| = \aleph_0$ (pronounced "aleph-nought").

Essentially by definition, if A is countable then there is a bijection $f : \mathbb{N} \to A$. In effect, this accomplishes the task of writing out the elements of A as an infinite sequence: $A = \{a_0, a_1, a_2, \dots\}$. Conversely, if we are successful at writing out the elements of a set as an infinite sequence, then we can be assured that the set is countable. We use this immediately to show the somewhat surprising fact that $\mathbb{Z}$ is countable:

Example 16.1. We demonstrate that $|\mathbb{Z}| = |\mathbb{N}|$ by listing off all the integers in an infinite sequence. We have $\mathbb{Z} = \{a_0, a_1, a_2, \dots\}$, where

$$a_k = \begin{cases} -k/2, & \text{if } k \text{ is even} \\ (k+1)/2, & \text{if } k \text{ is odd.} \end{cases}$$

Explicitly, this gives $\mathbb{Z} = \{0, 1, -1, 2, -2, 3, -3, \dots\}$. Clearly, the sequence $a_0, a_1, a_2, \dots$ hits every integer exactly once, so it explicitly gives our desired bijection from $\mathbb{N}$ to $\mathbb{Z}$. (Question to probe your understanding: why couldn't we have listed the integers in the order $0, 1, 2, 3, \dots, -1, -2, -3, \dots$?)

Perhaps even more surprisingly, it turns out that the set $\mathbb{Q}$ of rational numbers is also countable! We will show this as an immediate consequence of a couple basic lemmas. The first captures the intuitive idea that the countable sets are the smallest size of infinite set.

Lemma 16.1. *Every subset of a countable set is either countable or finite. In particular, subsets of countable sets are at most countable.*

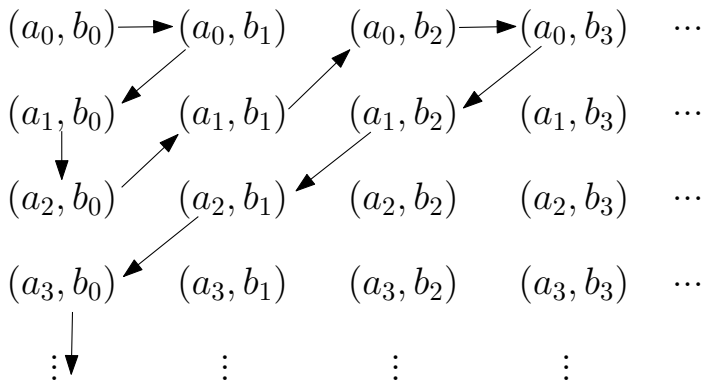
Proof. Suppose that A is a countable set, and let B be a subset of A. If B is finite, we have our conclusion, so suppose B is infinite. If we write out the elements of A in a sequence, say $A = \{a_0, a_1, a_2, \dots\}$, we can define the elements of B in a sequence recursively. First, we choose the smallest index k_0 such that $a_{k_0} \in B$, and set $b_0 = a_{k_0}$. Next, we let k_1 be the smallest index larger than k_0 such that $a_{k_1} \in B$, and again set $b_1 = a_{k_1}$. We continue in this fashion, for each natural number i, letting k_i be the smallest index larger than k_{i-1} such that $a_{k_i} \in B$, and taking $b_i = a_{k_i}$. We can do this at every step, because $B \setminus \{b_0, b_1, \dots, b_{i-1}\}$ will always be nonempty on account of the infiniteness of B. Altogether, this enumerates B as a subsequence of $a_0, a_1, \dots$, namely $a_{k_0}, a_{k_1}, a_{k_2}, \dots$, and proves that B is countable.

In particular, either $|B| = |A|$, or B is finite, so that there is an injection from B to $\mathbb{N}$ given by taking a bijection from B to $|B|$ and considering it as a map from B to $\mathbb{N}$. In both cases, $|B| \leq |A|$, so that B is at most countable. □

Another standard result around countable sets has to do with the Cartesian product of two (or finitely many) countable sets:

Lemma 16.2. *If A and B are both countable sets, then the set $A \times B$ is also countable.*

Proof. Since A and B are countable, they both can be enumerated by infinite sequences, say $A = \{a_0, a_1, a_2, \dots\}$ and $B = \{b_0, b_1, b_2, \dots\}$. The elements of the product $A \times B$ are then ordered pairs of the form (a_i, b_j), where $i, j \in \mathbb{N}$. We can define a bijection from $\mathbb{N}$ to $A \times B$ in visual fashion as follows:



In other words, first we list off all the ordered pairs whose indices sum to zero (in this case, just (a_0, b_0)), then all the ordered pairs whose indices sum to 1 (so (a_1, b_0) and (a_0, b_1)), then all ordered pairs whose indices sum to 2, and so on. Every ordered pair gets hit once and only once in this way, which proves that $A \times B$ is countable. □

This result has a straightforward extension to finite products:

Corollary 16.1. *If $A_0, A_1, \dots, A_n$ are finitely many countable sets (with $n \geq 2$), then the Cartesian product $\prod_{i=0}^{n} A_i$ is countable.*

Proof. We use induction on n, starting with $n = 2$ as our base case. This base case is nothing more than Lemma 16.2. Assuming the statement is true for a collection of $n \geq 2$ sets, now suppose we are given $n + 1$ countable sets, $A_0, A_1, \dots, A_n$. We wish to show that their Cartesian product is countable. By induction hypothesis, $\prod_{i=0}^{n-1} A_i$ is countable, and then by Lemma 16.2 again, we find that

$$\prod_{i=0}^{n} A_i = \left(\prod_{i=0}^{n-1} A_i \right) \times A_n$$

is also countable, completing the proof by induction. □

These two results can be used to fashion together a proof that the set $\mathbb{Q}$ of rational numbers is countable:

Theorem 16.1. *The set of rational numbers is countable.*

Proof. Note that every element $q \in \mathbb{Q}$ may be written in the form $\frac{a}{b}$, where $a, b \in \mathbb{Z}$ and $b \neq 0$. Hence every rational number can be associated to an ordered pair (a, b) of integers, where $b \neq 0$. By Lemma 16.2, we know that $\mathbb{Z} \times \mathbb{Z}$ is countable, so every element of $\mathbb{Q}$ may be identified with an element of the countable set $\mathbb{Z} \times \mathbb{Z}$. While multiple ordered pairs may represent the same rational number, we can always just choose one ordered pair representing each distinct rational number and treat $\mathbb{Q}$ as a subset of $\mathbb{Z} \times \mathbb{Z}$. By Lemma 16.1, every subset of $\mathbb{Z} \times \mathbb{Z}$ is either finite or countable, and so $\mathbb{Q}$ is either finite or countable. However, $\mathbb{Q}$ contains the infinite set $\mathbb{Z}$, so $\mathbb{Q}$ cannot be finite by Theorem 15.1 of the course notes. We conclude that $\mathbb{Q}$ is countable, as we wished to show. □

Building Sets from Countable Sets

We've already seen that any finite Cartesian product of countable sets is countable. Let's now prove a similar result for the union of countable sets:

Lemma 16.3. *If A and B are both countable sets, then $A \cup B$ is countable.*

Proof. Since A and B are both countable, choose enumerations of these sets, say $A = \{a_0, a_1, a_2, \dots\}$ and $B = \{b_0, b_1, b_2, \dots\}$. We can then define a sequence whose range is $A \cup B$, which we'll call $c_0, c_1, c_2, \dots$, by taking $c_{2k} = a_k$ for each $k \in \mathbb{N}$ and $c_{2k+1} = b_k$ for each $k \in \mathbb{N}$. It may so happen that some elements of A also occur as elements of B, but then after removing the duplicates from the sequence $c_0, c_1, \dots$, we end up with an infinite sequence that enumerates $A \cup B$ (why is the sequence necessarily infinite?), proving that $A \cup B$ is countable by Lemma 16.1. □

Induction allows for an immediate extension to any finite union, and the Axiom of Choice will give us the same result for a countable union:

Theorem 16.2. *If $A_0, \dots, A_n$ are finitely many countable sets (with $n \geq 2$), then the union $\bigcup_{i=0}^{n} A_i$ is countable. Furthermore, if $A_0, A_1, A_2, \dots$ is a countable collection of countable sets, then the union $\bigcup_{i=0}^{\infty} A_i$ is also countable.*

Proof. To establish the first part of this statement, we use induction on n, starting from $n = 2$. The base case, where $n = 2$, is just Lemma 16.3. Assuming the result holds for any n countable sets, suppose we are now given $n + 1$ countable sets, $A_0, A_1, \dots, A_n$. By induction hypothesis, the union $\bigcup_{i=0}^{n-1} A_i$ is countable, and thus it follows immediately from Lemma 16.3 that

$$\bigcup_{i=0}^{n} A_i = \left(\bigcup_{i=0}^{n-1} A_i \right) \cup A_n$$

is also countable. By induction, the first part of this theorem is now established.

Suppose now that $A_0, A_1, A_2, \dots$ is a countable collection of countable sets. We wish to show that the union of these sets is again countable. Using the Axiom of Choice, we can enumerate all of the countable sets at once, say $A_i = \{a_{i,0}, a_{i,1}, a_{i,2}, \dots\}$ for each $i \in \mathbb{N}$. We may then express $\bigcup_{i=0}^{\infty} A_i$ as the range of a big sequence, constructed just like in the proof that the Cartesian product of two countable sets is countable. First, we list off $a_{0,0}$, then $a_{1,0}$ and $a_{0,1}$, then $a_{2,0}, a_{1,1}, a_{0,2}$, running through all the elements whose indices sum to 0, then 1, then 2, and so on. While this sequence might contain some duplicates, removing those duplicates and reindexing the sequence still yields the union $\bigcup_{i=0}^{\infty} A_i$ as enumerated by an infinite sequence. Thus, this countable union of countable sets is again countable. □

Given that we needed the Axiom of Infinity to even guarantee the existence of a countable set, you might be forgiven in thinking that all the set-theoretic constructions we've introduced so far applied to countable sets will produce at most countable sets. Certainly, we've seen this to be true for unions, subsets, and finite Cartesian products. In fact, perhaps you may be convinced by now that there are *no* uncountable sets (except for that nagging feeling that we introduced the name "countable" for a reason...). Next time, we will define and briefly study some uncountable sets, which includes the familiar set of real numbers that you are used to working with in calculus.

MATH 145 Course Reading 17: Uncountable Sets – Do They Exist?

October 26, 2020

As you might guess (and as we've already revealed), the answer to the question in the title is "yes"! In particular, we focus on the famous proof (known as *Cantor's diagonal argument*) that leads to the conclusion that $\mathbb{R}$ is uncountable, and we also subsequently show that $\mathcal{P}(\mathbb{N})$ has the same cardinality as $\mathbb{R}$, so that the Axiom of Infinity automatically yields not only the countable set $\mathbb{N}$, but also the uncountable set $\mathcal{P}(\mathbb{N})$, in the presence of the other axioms introduced so far.

The Real Numbers are Uncountable

Our first major step in showing that $\mathbb{R}$ is uncountable is to reduce the task to showing that the open interval $(0, 1)$ is uncountable, with the help of a familiar type of function from calculus:

Lemma 17.1. *The open interval* $(0, 1)$ *has the same cardinality as* $\mathbb{R}$.

Proof. To establish the result, it is enough to construct a bijection $f : (0, 1) \to \mathbb{R}$. Thinking in terms of the graph of such a function, we're looking for the graph to hit every real number exactly once as it crosses from $x = 0$ to $x = 1$. If we define the function to have vertical asymptotes at $x = 0$ and $x = 1$, and to be strictly decreasing or strictly increasing in-between, we will have accomplished our purpose.

The easiest way to control asymptotes is by defining f to be a rational function, and for it to have asymptotes at $x = 0$ and $x = 1$, we can use $x(x - 1) = x^2 - x$ as our candidate denominator. But $f(x) = \frac{1}{x(x-1)}$ is insufficient, because it is strictly negative between $x = 0$ and $x = 1$. To make sure it crosses the x-axis, we throw in a linear factor in the numerator that hits 0 at the midpoint $x = \frac{1}{2}$. Specifically, we choose our candidate function to be $f(x) = \frac{1-2x}{x(x-1)}$. The claim is that $f : (0, 1) \to \mathbb{R}$ is a bijection, as defined.

To show that f is injective, suppose we are given $a, b \in (0, 1)$ such that $f(a) = f(b)$. By definition, this means $\frac{1-2a}{a^2-a} = \frac{1-2b}{b^2-b}$. If we cross-multiply and perform some simplification, we end up with

$$\begin{aligned}(1 - 2a)(b^2 - b) &= (1 - 2b)(a^2 - a)\\ b^2 - b - 2ab^2 + 2ab &= a^2 - a - 2a^2b + 2ab\\ 0 &= a^2 - b^2 - a + b + 2ab^2 - 2a^2b\\ 0 &= (a - b)(a + b) - (a - b) + 2ab(b - a)\\ 0 &= (a - b)(a + b - 2ab - 1).\end{aligned}$$

If we assume that $a \neq b$, we deduce that $a + b - 2ab - 1 = 0$. To see how this leads to a contradiction, note that this implies $ab = a + b - ab - 1 = (1 - a)(b - 1)$. Since both a and b are between 0 and 1, the product ab is positive, while $1 - a > 0$ and $b - 1 < 0$, so that $(1 - a)(b - 1) < 0$. This means the equality $ab = (1 - a)(b - 1)$ is a contradiction! We conclude that $a = b$ after all, so that f is injective.

Next, we show that f is surjective. Given any $r \in \mathbb{R}$, we need to find $x \in (0, 1)$ such that $f(x) = r$. This leads to the equation

$$r = \frac{1 - 2x}{x^2 - x}.$$

Clearing denominators and simplifying, we end up with a quadratic equation in x with the real number r showing up in the coefficients:

$$\begin{aligned}(x^2 - x)r &= 1 - 2x\\ rx^2 + (2 - r)x - 1 &= 0.\end{aligned}$$

If $r = 0$, clearly $x = \frac{1}{2}$ is a solution to the equation, and indeed $f(1/2) = 0$. Otherwise, the quadratic formula tells us that there are two real solutions to the equation:

$$x = \frac{(r-2) \pm \sqrt{(2-r)^2 - 4r(-1)}}{2r} = \frac{(r-2) \pm \sqrt{(4-4r+r^2)+4r}}{2r} = \frac{(r-2) \pm \sqrt{r^2+4}}{2r}.$$

Both solutions are indeed real, since $\sqrt{r^2+4}$ is always defined as a real number. However, our task is to show there is a real number x *from the interval* $(0,1)$ among these two solutions. To see whether we should start by taking the plus or minus sign, note that if we take $r = 1$ as a candidate, the value for x with the minus sign turns out to $\frac{-1-\sqrt{5}}{2}$, which is negative, and outside our range of consideration. On the other hand, if we take the plus sign, we get $\frac{-1+\sqrt{5}}{2}$, which does indeed work out to be between 0 and 1.

So now we verify the claim that for all nonzero $r \in \mathbb{R}$, we have $0 < \frac{r-2+\sqrt{r^2+4}}{2r} < 1$. This will be best handled in two cases: $r > 0$ and $r < 0$. If $r > 0$, note that $2r > 0$. We claim also that $(r-2) + \sqrt{r^2+4} > 0$. This holds because $\sqrt{r^2+4} > \sqrt{0^2+4} = 2$ when $r > 0$, so that $(r-2)+\sqrt{r^2+4} > (r-2)+2 = r > 0$, as needed. Knowing that both numerator and denominator are positive, we conclude that $\frac{(r-2)+\sqrt{r^2+4}}{2r} > 0$ for $r > 0$.

On the other hand, showing that $\frac{(r-2)+\sqrt{r^2+4}}{2r} < 1$ is equivalent (in the case $r > 0$) to showing that $(r-2) + \sqrt{r^2+4} < 2r$, or that $\sqrt{r^2+4} < r+2$. Knowing that $r > 0$, note that

$$\sqrt{r^2+4} < \sqrt{r^2+4r+4} = \sqrt{(r+2)^2} = r+2,$$

as we set out to show. We now know that $0 < \frac{(r-2)+\sqrt{r^2+4}}{2r} < 1$ whenever $r > 0$.

On the other hand, if $r < 0$, then $r^2+4 < r^2-4r+4 = (r-2)^2$, so that $\sqrt{r^2+4} < \sqrt{r^2-4r+4} = \sqrt{(r-2)^2} = -(r-2)$. Thus $(r-2) + \sqrt{r^2+4} < (r-2) + -(r-2) = 0$. Since $2r < 0$ as well, we conclude that $\frac{(r-2)+\sqrt{r^2+4}}{2r} > 0$ for $r < 0$.

Finally, proving that $\frac{(r-2)+\sqrt{r^2+4}}{2r} < 1$ for $r < 0$ is equivalent to showing that $(r-2) + \sqrt{r^2+4} > 2r$. This follows by noting that $(r-2) + \sqrt{r^2+4} > (r-2) + 2 = r$, and $r > 2r$ because $r < 0$.

This completes the verification that $x = \frac{(r-2)+\sqrt{r^2+4}}{2r}$ lies in the interval $(0,1)$ for all nonzero $r \in \mathbb{R}$. Finally, we observe that

$$\begin{aligned}
f(x) &= \frac{1 - 2\left(\frac{(r-2)+\sqrt{r^2+4}}{2r}\right)}{\left(\frac{(r-2)+\sqrt{r^2+4}}{2r}\right)^2 - \left(\frac{(r-2)+\sqrt{r^2+4}}{2r}\right)} \\
&= \frac{(2r)^2 - 4r((r-2)+\sqrt{r^2+4})}{((r-2)+\sqrt{r^2+4})^2 - 2r((r-2)+\sqrt{r^2+4})} \\
&= \frac{4r^2 - 4r^2 + 8r - 4r\sqrt{r^2+4}}{r^2 - 4r + 4 + 2r\sqrt{r^2+4} - 4\sqrt{r^2+4} + (r^2+4) - 2r^2 + 4r - 2r\sqrt{r^2+4}} \\
&= \frac{8r - 4r\sqrt{r^2+4}}{8 - 4\sqrt{r^2+4}} \\
&= r,
\end{aligned}$$

which shows that f is surjective. Now that we know f is a bijection, we conclude that $|(0,1)| = |\mathbb{R}|$, as claimed. □

With this result established, we can now look at the famous proof that $\mathbb{R}$ is uncountable:

Theorem 17.1. *The set of all real numbers is uncountable.*

Proof. By Lemma 17.1, it suffices to show that the open interval $(0, 1)$ is uncountable. We prove this by contradiction, supposing to the contrary that $(0, 1)$ is in fact countable. This means that we can produce an enumeration of $(0, 1)$, say $(0, 1) = \{r_1, r_2, r_3, \dots\}$. Each element of the interval $(0, 1)$ has a decimal expansion, say $r_i = 0.b_{i1}b_{i2}b_{i3}\dots$, where each decimal digit b_{ij} is an integer between 0 and 9. To guarantee that the decimal expansion is unique, we disallow expansions with infinitely repeating 9s, such as $0.129999\dots$, replacing them instead with the equivalent decimal expansion with infinitely repeating 0s (in this case, $0.130000\dots$). Given the enumeration of $(0, 1)$ above, we now list out the decimal expansions of all these numbers:

$$\begin{aligned} r_1 &= 0.b_{11}b_{12}b_{13}b_{14}\dots \\ r_2 &= 0.b_{21}b_{22}b_{23}b_{24}\dots \\ r_3 &= 0.b_{31}b_{32}b_{33}b_{34}\dots \\ r_4 &= 0.b_{41}b_{42}b_{43}b_{44}\dots \end{aligned}$$

Our claim is now this: no matter how we tried to enumerate $(0, 1)$, there is always a real number $r \in (0, 1)$ that does not belong to the set $\{r_1, r_2, r_3, \dots\}$, contradicting the fact that our mapping from the positive integers to $(0, 1)$ is surjective. To construct this real number r, look at the first digit after the decimal point in r_1, the second digit after the decimal point in r_2, and so on (the so-called *diagonal* digits of the list):

$$\begin{aligned} r_1 &= 0.\mathbf{b_{11}}b_{12}b_{13}b_{14}\dots \\ r_2 &= 0.b_{21}\mathbf{b_{22}}b_{23}b_{24}\dots \\ r_3 &= 0.b_{31}b_{32}\mathbf{b_{33}}b_{34}\dots \\ r_4 &= 0.b_{41}b_{42}b_{43}\mathbf{b_{44}}\dots \end{aligned}$$

We then construct r in such a way that it differs from each r_i in the ith decimal place. With this in mind, we give r the decimal expansion $r = 0.c_1c_2c_3c_4\dots$, where for each positive integer i, we have

$$c_i = \begin{cases} 4, & \text{if } b_{ii} \neq 4 \\ 5, & \text{if } b_{ii} = 4. \end{cases}$$

Certainly, $r \in (0, 1)$, but $r \neq r_i$ for any positive integer i, because r differs from r_i in the ith decimal place. Indeed, if $b_{ii} = 4$, then c_i is not 4, and if b_{ii} is not 4, then c_i *is* 4. Since this is true no matter how we tried to enumerate $(0, 1)$, we have a contradiction, and conclude that $(0, 1)$, and hence $\mathbb{R}$, must be uncountable after all. □

Another Uncountable Set

As promised, in this section we show that the power set of $\mathbb{N}$ is uncountable, and in fact show that it has the same cardinality as $\mathbb{R}$. Interestingly, the proof we give makes use of the Cantor-Schröder-Bernstein Theorem, which we stated and proved as part of Proposition 14.2 in an earlier reading.

Theorem 17.2. *The sets $\mathcal{P}(\mathbb{N})$ and $\mathbb{R}$ have the same cardinality, and hence $\mathcal{P}(\mathbb{N})$ is uncountable.*

Proof. Again appealing to Lemma 17.1, it suffices to show that $\mathcal{P}(\mathbb{N})$ has the same cardinality as the open interval $(0, 1)$. For this, we will give injective functions $f : (0, 1) \to \mathcal{P}(\mathbb{N})$ and $g : \mathcal{P}(\mathbb{N}) \to (0, 1)$. This will show that $|(0, 1)| \leq |\mathcal{P}(\mathbb{N})|$ and that $|\mathcal{P}(\mathbb{N})| \leq |(0, 1)|$. By the Cantor-Schröder-Bernstein Theorem, this will yield our desired conclusion $|(0, 1)| = |\mathcal{P}(\mathbb{N})|$.

So, we define a function $f : (0, 1) \to \mathcal{P}(\mathbb{N})$ as follows. Given $r \in (0, 1)$, write out its decimal expansion uniquely as $r = 0.b_1b_2b_3b_4\dots$, where each digit b_i is between 0 and 9, and we avoid infinitely repeating 9s. We define $f(r)$ to be the following subset R of the natural numbers:

$$f(r) = \{10^{n-1}b_n : n \in \mathbb{N}^+\} = R.$$

So, for example, for the rational number $1/3 = 0.3333\ldots$, we have $f(1/3) = \{3, 30, 300, 3000, \ldots\}$. We now verify that f is injective. Suppose we have two real numbers $r, s \in (0, 1)$ such that $f(r) = f(s)$. For convenience, write out the unique decimal expansions of r and s as $r = 0.b_1b_2b_3\ldots$ and $s = 0.c_1c_2c_3\ldots$. If we know that $f(r) = f(s)$, consider the ith digit b_i in the decimal expansion of r. Then $10^{i-1}b_i$ belongs to $R = f(r)$, so it also belongs to $S = f(s)$. But if $b_i \neq 0$, it follows from the construction of $f(r)$ that $10^{i-1}b_i$ is the unique integer between 10^{i-1} and $9 \cdot 10^{i-1}$ belonging to R, and a similar statement holds for S. So $10^{i-1}b_i$ must be the unique integer between 10^{i-1} and $9 \cdot 10^{i-1}$ belonging to S as well, proving that $10^{i-1}b_i = 10^{i-1}c_i$, so that $b_i = c_i$.

On the other hand, if $b_i = 0$, then $0 \in R$, and there is no integer between 10^{i-1} and $9 \cdot 10^{i-1}$ in the set R. Since $R = S$, the same holds true for S, which forces $c_i = 0$ as well. We have now shown that r and s agree in every decimal place, so that $r = s$ and f is injective.

Next, we construct our injective function $g : \mathcal{P}(\mathbb{N}) \to (0, 1)$. For each subset $A \subseteq \mathbb{N}$, we define $g(A) = 0.a_0a_1a_2\ldots$, where

$$a_i = \begin{cases} 4, & \text{if } i \in A \\ 5, & \text{if } i \notin A. \end{cases}$$

Certainly, $g(A)$ is a real number between 0 and 1, and since it does not have any 0s or 9s in its decimal expansion, that decimal expansion is unique. Now, suppose we have two subsets A and B of $\mathbb{N}$ such that $g(A) = g(B)$. Then $g(A) = 0.a_0a_1a_2\ldots = 0.b_0b_1b_2\ldots = g(B)$. If $i \in A$, then by definition $a_i = 4$, so $b_i = 4$ by uniqueness of the decimal expansion, which says that $i \in B$. This proves that $A \subseteq B$, with the proof that $B \subseteq A$ being entirely similar. Thus $A = B$, so that g is injective.

Having constructed injective functions from $(0, 1)$ to $\mathcal{P}(\mathbb{N})$ and from $\mathcal{P}(\mathbb{N})$ to $(0, 1)$, we conclude that they have the same cardinality, which completes the proof of our result. $\square$

MATH 145 Course Reading 18:
Binary Operations and an Introduction to Groups

October 28, 2020

From this point forward, we take a significant shift in direction, away from studying the theory of sets, and towards applying some of the things we've studied in the realm of *abstract algebra*. Simply put, the main focus of modern abstract algebra is to study properties that follow from defining operations on a given set that have a number of familiar properties (like addition or multiplication of real numbers, for instance).

Hence, our initial goal for this reading will be to clarify the definition of a *binary operation*, which will be the template for the various types of operations we define in algebraic structures. Next, we gradually narrow our focus to look particularly at the structure known as a *group*. Groups are an extremely important idea in mathematics, and we will spend almost two weeks exclusively studying various examples and features of groups.

Binary Operations: Definition and Examples

When we add together two real numbers, or two integers, what are we really doing? As input, we take a pair of objects (real numbers, integers), and as output, we get a single object from that same set, the result of the operation. For some operations, the order in which we take our pair of objects might affect the result (consider that $7 - 5$ is different from $5 - 7$), and so the input to our operation really ought to be considered an *ordered* pair of objects. This motivates the formal definition:

Definition 18.1. Let S be a set. A *binary operation* on S is a function from $S \times S$ to S. If $* : S \times S \to S$ is a binary operation, we will usually prefer to write $a * b$ instead of $*(a, b)$ as the output of the function.

The generality of the definition means that there are a whole array of familiar examples:

Example 18.1. On the set $\mathbb{Z}$ of integers, the operations of addition, subtraction, and multiplication all give binary operations on $\mathbb{Z}$. Note, however, that division of two integers does *not* give a binary operation on $\mathbb{Z}$. For instance, it is not the case that the result of every division of integers results in an integer.

Example 18.2. Similarly to the above, the operations of addition, subtraction, and multiplication of real numbers all give binary operations on $\mathbb{R}$. However, division again fails to be a binary operation on $\mathbb{R}$, because there is no way to divide real numbers by 0, so that division does not give us a function $\div : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. But, if we restrict our attention to the set $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$ of non-zero real numbers, then $\div$ is a binary operation on $\mathbb{R}^*$, since the quotient of two nonzero real numbers is a nonzero real number.

Example 18.3. For any set A, the operations of set union and set intersection both define binary operations on $\mathcal{P}(A)$, giving ways to take pairs of subsets of A and define new subsets of A from them.

Example 18.4. For any nonempty set A, the composition of functions $\circ$ defines a binary operation on A^A, the set of functions from A to A. Indeed, if $f : A \to A$ and $g : A \to A$ are functions, then the composition $g \circ f$ is again a function from A to A.

Binary operations can have numerous additional properties that make them more convenient to study (and form parts of the definitions of the more complicated algebraic structures we will soon look at in this course). We give some of the main ones below:

Definition 18.2. Let S be a set, and let $*$ denote a binary operation on S. The operation $*$ is said to be *associative* if, for all $a, b, c \in S$, we have $(a * b) * c = a * (b * c)$. The operation $*$ is said to be *commutative* if, for all $a, b \in S$, we have $a * b = b * a$. An element $e \in S$ is said to be a *unity* (or *identity*) for $*$ if $a * e = e * a = a$ for all elements $a \in S$.

Associativity is a vastly useful property, because it allows us to forget about brackets when writing out a long string of successive operations, such as $a_1 * a_2 * a_3 * \cdots * a_n$. On the face of it, the result might depend on the order we choose the pairs of elements to "multiply" together using $*$ (the choice of bracketing). For example, with four elements, we could take $((a_1 * a_2) * a_3) * a_4$, or $a_1 * ((a_2 * a_3) * a_4)$, among other options, and we might fear that the choice of bracketing might matter. With the help of induction, we can give a formal proof that it does not:

Proposition 18.1. *Let S be a set with associative binary operation $*$. If $a_1, \ldots, a_n$ are arbitrary elements of S, with $n \geq 1$, then the product $a_1 * a_2 * \cdots * a_n$ is well-defined, regardless of the choice of bracketing in the product.*

Proof. We proceed by induction on $n \geq 1$. To standardize our notation, we define a *standard product* $\langle a_1, a_2, \ldots, a_n \rangle$ of the n elements recursively by declaring $\langle a_1 \rangle = a_1$, $\langle a_1, a_2 \rangle = a_1 * a_2$, and for $n \geq 3$, taking $\langle a_1, a_2, \ldots, a_n \rangle = \langle a_1, a_2, \ldots, a_{n-1} \rangle * a_n$. Thus $\langle a_1, a_2, a_3 \rangle = (a_1 * a_2) * a_3$, $\langle a_1, a_2, a_3, a_4 \rangle = ((a_1 * a_2) * a_3) * a_4$, and so on.

What we prove, by strong induction on n, is that every product of the n elements $a_1, a_2, \ldots, a_n$, in that order, is equal to the standard product $\langle a_1, a_2, \ldots, a_n \rangle$. Since there is no choice of bracketing when $n = 1$ or $n = 2$, both of these hold trivially as base cases. Now suppose the result is true for all products of up to n elements, where $n \geq 2$, and suppose we are given $n + 1$ elements of S, say $a_1, a_2, \ldots, a_{n+1}$. Taking any product of these $n + 1$ elements, and looking at the last application of $*$, we know the product can be expressed in the form $b * c$, where b is a product of some elements $a_1, \ldots, a_k$, with $1 \leq k \leq n$, and c the product of the remaining elements $a_{k+1}, \ldots, a_{n+1}$.

If $k = n$, then $c = a_{n+1}$, and by induction hypothesis $b = \langle a_1, \ldots, a_n \rangle$, so that $b * c = \langle a_1, \ldots, a_n \rangle * a_{n+1} = \langle a_1, \ldots, a_{n+1} \rangle$ by definition of the standard product. If $k < n$, then $c = \langle a_{k+1}, \ldots, a_{n+1} \rangle = \langle a_{k+1}, \ldots, a_n \rangle * a_{n+1}$ by induction hypothesis, and $b = \langle a_1, \ldots, a_k \rangle$ by induction hypothesis, so

$$\begin{aligned} b * c &= \langle a_1, \ldots, a_k \rangle * (\langle a_{k+1}, \ldots, a_n \rangle * a_{n+1}) \\ &= (\langle a_1, \ldots, a_k \rangle * \langle a_{k+1}, \ldots, a_n \rangle) * a_{n+1} \\ &= \langle a_1, \ldots, a_n \rangle * a_{n+1} \\ &= \langle a_1, \ldots, a_{n+1} \rangle, \end{aligned}$$

applying associativity of $*$, followed by the induction hypothesis again. This completes the proof that every product of n elements of S is equal to the standard product, and hence is independent of the order of bracketing.

Of the examples given in the previous section, most of the binary operations are associative, but notably subtraction is not. For example, $(3 - 5) - 7 = (-2) - 7 = -9$, while $3 - (5 - 7) = 3 - (-2) = 5$. Of the remaining examples, most are also commutative, though composition of functions gives a notable exception. For example, consider the functions f and g from $\mathbb{R}$ to $\mathbb{R}$ given by $f(x) = x^2$, and $g(x) = -x$. Notice that $g \circ f \neq f \circ g$; in particular, $(g \circ f)(1) = g(1^2) = g(1) = -1$, while $(f \circ g)(1) = f(-1) = (-1)^2 = 1$. □

Monoids, Inverse Elements, and Groups

Algebraists sometimes like to package two of the properties considered in the previous section (associativity and the existence of an identity element) together and give such structures a special name:

Definition 18.3. Let S be a set with binary operation $*$. We call S a *monoid* if the operation $*$ is associative, and there is an identity element $e \in S$ with respect to $*$.

Usually, when studying a general monoid, it is common not to use $*$ for the binary operation, but rather to write products of elements instead, as we do for the familiar multiplication operations we know about. In other words, it is common to write ab in place of $a * b$ as the result of applying the binary operation in a monoid, and to refer to it as the *product* of a and b. This is referred to as *multiplicative notation*. Similarly

then, we would use exponent notation a^n to denote the result of multiplying a by itself n times, for $n \geq 2$, with the conventions that $a^1 = a$ and $a^0 = e$, the identity element of the monoid.

Let's now prove a basic property of monoids: that the identity of such a structure is necessarily unique.

Proposition 18.2. *The identity element of any monoid is unique.*

Proof. Let S be a monoid (which we'll write in multiplicative notation), and suppose S has two identity elements, say e_1 and e_2. Then for any $a \in S$, we know that $ae_1 = e_1a = a$ and $ae_2 = e_2a = a$. Applying the first string of equalities with $a = e_2$, we get $e_2e_1 = e_1e_2 = e_2$, and applying the second string of equalities with $a = e_1$, we get $e_1e_2 = e_2e_1 = e_1$. Together, these imply that $e_1 = e_2$, proving uniqueness. □

It may so happen that we work with a structure where it is possible to "undo" the operation. For example, the addition operation on $\mathbb{Z}$ or $\mathbb{R}$ comes equipped with such a thing, corresponding to the operation of subtraction. The operation of multiplication on the set $\mathbb{R}^*$ of non-zero real numbers also comes equipped with such a thing, corresponding to the operation of division. Let's formally define what we mean:

Definition 18.4. Let S be a monoid, written in multiplicative notation. An element $a \in S$ is called a *unit* of S if there exists some $b \in S$ for which $ab = ba = e$, the identity element of S. In such a case, we call b an *inverse* of a.

Again, we can very quickly prove that inverse elements are unique if they exist:

Lemma 18.1. *In any monoid, the inverse of any unit is unique.*

Proof. Let S be a monoid, and let $a \in S$ be a unit. Suppose b_1 and b_2 are both inverses of a, so that $ab_1 = b_1a = e$ and $ab_2 = b_2a = e$. By definition of the identity element, note that

$$b_1 = b_1e = b_1(ab_2) = (b_1a)b_2 = eb_2 = b_2,$$

giving the desired uniqueness. □

If our monoid is written in multiplicative notation, we often use a^{-1} to denote the unique inverse of the element a, when that inverse exists. Similarly, for any negative integer m, we set a^m to be the result of multiplying a^{-1} together $|m|$ times. Sometimes, algebraists will also use *additive notation* for a monoid, using $+$ for the binary operation. In this case, the inverse of an element a is usually written $-a$, and called the *negative* of a.

Looking back at our examples, both $\mathbb{Z}$ and $\mathbb{R}$ under the addition operation are monoids, with 0 as the identity element, with the property that *every* element is a unit (the inverse of a being the usual number $-a$). Also, both these sets under the multiplication operation are also monoids, with 1 as the identity element. But in $\mathbb{Z}$, only 1 and -1 are units (why?), while in $\mathbb{R}$, *every* real number (except 0) is a unit with respect to the multiplication operation, with the inverse of the real number $r \neq 0$ being $\frac{1}{r}$.

To conclude this reading, we now define a *group*, in terms of the definitions we've built up throughout these last few pages:

Definition 18.5. A set G with a binary operation $*$ is called a *group* if G is a monoid, such that *every* element of G is a unit. If the operation $*$ on G is also commutative, then G is called an *abelian group*.

We can also spell out the definition of a group more traditionally, by isolating each of the conditions. A set G with a binary operation (in multiplicative notation) is called a group if:

(1) For all $a, b, c \in G$, we have $(ab)c = a(bc)$. (Associativity)

(2) There is $e \in G$ such that $ae = ea = a$ for all $a \in G$. (Identity)

(3) For all $a \in G$, there is $b \in G$ such that $ab = ba = e$. (Inverses)

With the optional fourth condition:

(4) For all $a, b \in G$, we have $ab = ba$,

we end up with an abelian group. We already have a couple examples of groups, and we will recap them below, while supplying one more:

Example 18.5. The sets $\mathbb{Z}$ and $\mathbb{R}$, equipped with the binary operation of addition, are both abelian groups. The set $\mathbb{R}^*$ of nonzero real numbers, equipped with the binary operation of multiplication, is also an abelian group. The set $\{1, -1\}$ of units in $\mathbb{Z}$, with respect to the multiplication operation on $\mathbb{Z}$, is a finite abelian group.

Example 18.6. Given a set A, we let G denote the set of all *bijections* from A to A. This set, with function composition as the operation, is a group. Indeed, we've previously verified that the composition of two bijections is a bijection, so composition does form a binary operation on G. You can check right away that function composition is associative. Moreover, the function $e : A \to A$ given by $e(a) = a$ for all $a \in A$ is a bijection, and serves as the identity element of this group. Finally, given a function $f \in G$, we know the inverse relation $f^{-1} : A \to A$ is also a function, and is a bijection. Furthermore, you can check that $f^{-1} \circ f$ and $f \circ f^{-1}$ both give the identity function on A, so f^{-1} is the inverse of f with respect to this given binary operation as well (note the overlap in terminology!). In general, G is not abelian: for example, if $A = \mathbb{R}$, the functions $f(x) = -x$ and $g(x) = x + 1$ are both bijections on $\mathbb{R}$, but $g \circ f \neq f \circ g$ (for example, $(g \circ f)(1) = 0$, while $(f \circ g)(1) = -2$).

MATH 145 Course Reading 19:
More on Groups; Cyclic Groups

October 30, 2020

Now that we've introduced the definition of a group, our next step is to study them more carefully, both through introducing some major types of examples, and also through studying their properties. In this reading, we will first prove a couple results about groups in the abstract, then study *cyclic groups* in more detail. In our next reading, we will then visit the idea of *subgroups* (groups within a group) and also take a look at the *symmetric groups*.

More on Groups

Recall that a *group* G is a set equipped with a binary operation (often denoted by multiplication), such that the operation is associative, G admits an identity element with respect to the operation, and every element of G is a unit (has an inverse). Said more compactly, a group is a monoid in which every element is a unit. In fact, we can derive a group from every monoid as follows:

Theorem 19.1. *Let M be a monoid, and let M^* denote the units of M, the set of elements that have a multiplicative inverse. Then M^* is a group, called the* group of units of M.

Proof. First, we must check that the binary operation on M, restricted to elements of M^*, gives back an element of M^*. In other words, we must show that for all $a, b \in M^*$, we have $ab \in M^*$. (We will then say that M^* is *closed under the binary operation*). By definition, we know there are $a^{-1}, b^{-1} \in M$ such that $aa^{-1} = a^{-1}a = e$ and $bb^{-1} = b^{-1}b = e$. Note then that

$$(ab)(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = aea^{-1} = aa^{-1} = e$$
$$(b^{-1}a^{-1})(ab) = b^{-1}(a^{-1}a)b = b^{-1}eb = b^{-1}b = e.$$

This shows ab is a unit, with inverse $(ab)^{-1} = b^{-1}a^{-1}$. We now know that M^* is closed under the binary operation on M.

Associativity and the existence of an identity element in M^* follow almost immediately from these same properties in M. Indeed, given $a, b, c \in M^*$, we have $(ab)c = a(bc)$ because this equation already holds in the larger set M. Similarly, note that the identity element $e \in M$ belongs to M^* because $ee = e$, showing that $e^{-1} = e$. Then $ea = ae = a$ for all $a \in M^*$, proving that M^* has an identity element. Finally, suppose we have $a \in M^*$. Then it has an inverse a^{-1} in M by construction, and $a^{-1} \in M^*$ because $aa^{-1} = a^{-1}a = e$, showing that $(a^{-1})^{-1} = a$. Thus each $a \in M^*$ has an inverse in M^*, completing the proof that M^* is a group. □

For a quick illustration of this theorem in action, reconsider the monoid A^A introduced in Example 18.4, where A is any nonempty set, and the binary operation is function composition. The group of units $(A^A)^*$ in this case are the set of functions $f : A \to A$ for which there is an inverse function $g : A \to A$ satisfying $f \circ g = g \circ f = e$, where $e : A \to A$ is the identity function taking every element to itself. It is a good exercise to check that this corresponds exactly with the subset of *bijective* functions $f : A \to A$, so that $(A^A)^*$ is the group introduced in Example 18.6 of the previous reading.

The next thing we verify is that exponent notation in a group behaves in the expected way. Recall that for any element g in a group G, we set $g^0 = e$, $g^1 = g$, g^m equal to the product of m copies of g for $m \geq 2$, and g^{-m} equal to the product of m copies of g^{-1} for $m \geq 2$. The following three rules for manipulating exponents are key, and will be used over and over again:

Theorem 19.2. *Let G be a group, and let $g, h \in G$ be elements.*

(1) For all $n, m \in \mathbb{Z}$, we have $g^{n+m} = g^n \cdot g^m$.

(2) For all $n, m \in \mathbb{Z}$, we have $(g^m)^n = g^{mn}$.

(3) If g and h commute, so that $gh = hg$, then $(gh)^n = g^n h^n$ for all $n \in \mathbb{Z}$.

Proof. We establish each of the three statements for all $n \in \mathbb{N}$, $m \in \mathbb{Z}$ by induction on n, then return separately to proving them for all integers n. For (1), we show by induction on n that $g^{n+m} = g^n \cdot g^m$ for all $m \in \mathbb{Z}$. In the base case $n = 0$, $g^{n+m} = g^m$ and $g^n \cdot g^m = e \cdot g^m = g^m$, so this case is true. Assuming the statement for a given $n \in \mathbb{N}$ and all $m \in \mathbb{N}$, we now wish to show that $g^{(n+1)+m} = g^{n+1} \cdot g^m$ for all $m \in \mathbb{Z}$. By definition, and by induction hypothesis,

$$\begin{aligned} g^{n+1} \cdot g^m &= (g \cdot g^n) \cdot g^m \\ &= g \cdot (g^n \cdot g^m) \\ &= g \cdot g^{n+m}. \end{aligned}$$

If $n + m \geq 0$, then this last expression is the result of multiplying $n + m$ copies of g by one more g, which is $g^{(n+m)+1}$ by definition. If $n + m = -1$, then the above is equal to gg^{-1}, which is $e = g^0 = g^{(n+m)+1}$. Finally, if $n + m \leq -2$, then the last expression is multiplying $|n + m|$ copies of g^{-1} by one copy of g, which results in $|n + m| - 1$ copies of g^{-1}, which can again be written $g^{(n+m)+1}$. Thus (1) is established for all $n \in \mathbb{N}$ and all $m \in \mathbb{Z}$ by induction. An exactly similar argument can establish the result for all $n \in \mathbb{Z}$ and all $m \in \mathbb{N}$, proceeding by induction on m instead. Between these two proofs, we have established the truth of the statement for all integers n and m.

For (2), we again show by induction on n that $(g^m)^n = g^{mn}$ for all $m \in \mathbb{Z}$. In the base case $n = 0$, $(g^m)^n = (g^m)^0 = e$ and $g^{mn} = g^0 = e$, so this case is proved. Now, assume the statement for some fixed $n \in \mathbb{N}$. We wish to show that $(g^m)^{n+1} = g^{m(n+1)}$ for all $m \in \mathbb{Z}$. Applying statement (1) of this theorem and the induction hypothesis, we find that

$$\begin{aligned} (g^m)^{n+1} &= (g^m)^n \cdot (g^m)^1 \\ &= g^{mn} \cdot g^m \\ &= g^{mn+m} \\ &= g^{m(n+1)}. \end{aligned}$$

This proves the statement for all $m \in \mathbb{Z}$ and all $n \in \mathbb{N}$ by induction. Finally, if n is a negative integer, say $n = -\ell$ for some $\ell \in \mathbb{N}$, note that $(g^m)^n = (g^m)^{-\ell}$. If we show that for any $h \in G$, $\ell \in \mathbb{N}$, we have $h^{-\ell} = (h^\ell)^{-1}$, we will be done, since then $((g^m)^{-\ell}) = ((g^m)^\ell)^{-1} = (g^{m\ell})^{-1} = g^{-m\ell} = g^{m(-\ell)} = g^{mn}$.

But to prove this final claim, note that $h^{-\ell}$ is the product of ℓ copies of h^{-1}, which is clearly the inverse of h^ℓ, the product of ℓ copies of h. Thus (2) has been proved in general too.

Finally, we prove (3) by induction on n as well. If $n = 0$, then $(gh)^0 = e = ee = g^0 h^0$, so the statement is true for this value of n. If we know the statement to be true for a given $n \in \mathbb{N}$, note that by induction hypothesis and part (1),

$$\begin{aligned} (gh)^{n+1} &= (gh)^n \cdot (gh)^1 \\ &= (g^n h^n) \cdot gh \\ &= g^n (h^n g) h \\ &= g^n (gh^n) h \\ &= g^{n+1} h^{n+1}, \end{aligned}$$

where we use that g commutes with h to get that g commutes with h^n as well. By induction, we have (3) for all $n \in \mathbb{N}$. Finally, if n is negative, say $n = -\ell$ for $\ell \in \mathbb{N}$, then $(gh)^n = (gh)^{-\ell} = ((gh)^\ell)^{-1} = (g^\ell h^\ell)^{-1} = (h^\ell)^{-1}(g^\ell)^{-1} = h^{-\ell} g^{-\ell} = h^n g^n = g^n h^n$ (why does the last equation follow?) □

Orders of Elements and Cyclic Groups

This lengthy discussion about exponent rules features heavily in our first main family of groups, known as the cyclic groups. Our first result is a useful way to find smaller groups within a given group (which we will soon refer to as *subgroups*):

Theorem 19.3. *Let G be a group, and let $g \in G$ be an element. Then the set*

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

is a group contained in G with respect to the operation on G.

Proof. By part (1) of Theorem 19.2, for any elements $g^m, g^n \in \langle g \rangle$, we have $g^m \cdot g^n = g^{m+n} \in \langle g \rangle$, so that $\langle g \rangle$ is closed under the group operation on G. Thus the operation on G restricts to a binary operation on $\langle g \rangle$. Associativity of this operation is automatic from the fact that it is associative in G. Note that $e = g^0$ belongs to $\langle g \rangle$, and for all $g^m \in \langle g \rangle$, we have $eg^m = g^m e = g^m$, so that e is also the identity element of $\langle g \rangle$. Finally, given any $g^m \in \langle g \rangle$, note that $g^{-m} \in \langle g \rangle$, and that $g^m g^{-m} = g^{-m} g^m = g^0 = e$ by exponent rules, so that every element of $\langle g \rangle$ is a unit. Thus $\langle g \rangle$ is a group, as claimed. □

This fundamental result leads to a couple fundamental definitions:

Definition 19.1. If G is a group and $g \in G$ is an element, the set $\langle g \rangle$ is called the *subgroup of G generated by g*. If $G = \langle g \rangle$ for some element $g \in G$, then we say that G is a *cyclic group*, and that g is a *generator* of G.

We now give the main examples of cyclic groups, but presented in additive notation, rather than multiplicative, as is customary for these particular examples.

Example 19.1. Consider the group $\mathbb{Z}$, with addition of integers as the group operation. This group is cyclic, generated by either 1 or -1, since every integer n may be expressed as $n \cdot 1$ and as $(-n) \cdot -1$, i.e. it may be expressed as the sum of copies of either 1 or -1 in this group.

Example 19.2. For a group along the lines of the previous example, but finite, we introduce the set of *integers modulo n*, denoted $\mathbb{Z}/n\mathbb{Z}$, where n is a positive integer. To define this group as a set, we take the set of equivalence classes of $\mathbb{Z}$ under the relation *congruence modulo n*, which is defined as follows:

$$a \equiv b \pmod{n} \text{ if } n \mid a - b.$$

Here, we write $n \mid a - b$ and say n *divides* $a - b$, if there is some $k \in \mathbb{Z}$ such that $a - b = kn$. It is a good exercise to check that this really does give an equivalence relation on $\mathbb{Z}$, and that there are exactly n distinct equivalence classes, namely $[0], [1], \ldots, [n-1]$. (We will also verify a much more general version of this fact later in the course). It turns out we can also define an addition of equivalence classes, via

$$[a] + [b] = [a + b],$$

for any $a, b \in \mathbb{Z}$. Of course, we must check that this is well-defined, i.e. for any integers a, a', b, b' with $[a] = [a']$ and $[b] = [b']$, we have $[a] + [b] = [a'] + [b']$. Let's prove this quickly. If $[a] = [a']$, then $a \equiv a' \pmod{n}$. Similarly, $b \equiv b' \pmod{n}$. Thus $a - a' = kn$ and $b - b' = \ell n$ for some $k, \ell \in \mathbb{Z}$. It follows that

$$(a + b) - (a' + b') = (a - a') + (b - b') = kn + \ell n = (k + \ell)n,$$

showing that $a + b \equiv a' + b' \pmod{n}$. It follows that $[a] + [b] = [a + b] = [a' + b'] = [a'] + [b']$. You can then immediately check that this operation of $+$ turns $\mathbb{Z}/n\mathbb{Z}$ into an abelian group, with identity $[0]$ and the inverse of $[a]$ given by $[-a]$ for any $a \in \mathbb{Z}$.

This is a finite group, with n elements, and again, it is cyclic: for any class $[a] \in \mathbb{Z}/n\mathbb{Z}$, we have $[a] = a \cdot [1]$ (i.e. $[a]$ is the sum of a copies of $[1]$ or $|a|$ copies of $[-1]$ in this group), so that $[1]$ is a generator for the group.

Cyclic groups are tied in with a related notation, applying to any element of any group. We define it below:

Definition 19.2. Given any group G and element $g \in G$, the *order* of the element g, denoted $o(g)$, is the smallest integer $n \geq 1$ such that $g^n = e$, if such an integer exists. If $g^n \neq e$ for all $n \geq 1$, then we say the order of g is infinite and write $o(g) = \infty$.

Let's quickly give some facts about the order of an element, to get a feel for how it works:

Proposition 19.1. *For any element g in a group G, we have $o(g) = o(g^{-1})$.*

Proof. If $o(g) = \infty$, then for all positive integers k, we have $g^k \neq e$. Taking inverses of both sides, we see that for all positive integers k, we also have $(g^{-1})^k \neq e$, so $o(g^{-1}) = \infty$. In the case $o(g) = n$, we know n is the smallest positive integer such that $g^n = e$. Taking inverses of both sides, we get $(g^{-1})^n = e$. If there were a smaller positive integer $m < n$ such that $(g^{-1})^m = e$, again taking inverses of both sides would give $g^m = e$, contradicting our definition of $o(g)$. Thus $o(g^{-1}) = n = o(g)$. □

Here is another example:

Proposition 19.2. *If G is a finite group (a group with finitely many elements), then every element of G has finite order.*

Proof. Let G be a finite group, and let $g \in G$ be an element. Then the sequence $g^0, g^1, g^2, g^3, \ldots$ cannot contain infinitely many elements, so there are distinct $m, n \in \mathbb{N}$ such that $g^m = g^n$, and without loss of generality, suppose $m > n$. Then multiplying both sides by g^{-n} gives $g^{m-n} = g^0 = e$, so that g has finite order, at most $m - n$. □

With the help of division with remainder for integers (which we will study in detail soon), we can prove one final result connecting the subgroup generated by an element to the order of that element:

Theorem 19.4. *Let G be a group, and let $g \in G$ be an element of finite order n. Then:*

(1) $g^k = g^m$ if and only if $k \equiv m \pmod{n}$. In particular, $g^k = e$ if and only if $n \mid k$.

(2) We have $\langle g \rangle = \{e, g, g^2, \ldots, g^{n-1}\}$, where $1, g, \ldots, g^{n-1}$ are all distinct.

Proof. First, suppose that $k \equiv m \pmod{n}$ for some integers k, m. By definition, this means $k - m = \ell n$ for some $\ell \in \mathbb{Z}$. Since $g^n = e$, raising to the power ℓ gives $g^{\ell n} = e^\ell = e$. Thus $g^{k-m} = e$, and multiplying through by g^m gives $g^k = g^m$. Conversely, suppose $g^k = g^m$. This implies $g^{k-m} = e$. To show that $n \mid k-m$, we use division with remainder in $\mathbb{Z}$. Dividing by n with remainder, this says we can find $q, r \in \mathbb{Z}$ with $0 \leq r < n$ such that $k - m = nq + r$. But then

$$e = g^{k-m} = g^{nq+r} = g^{nq} g^r = (g^n)^q g^r = e^q g^r = g^r.$$

Since $r < n$, the definition of $o(g)$ forces $r = 0$, so that $k - m = nq$, and $n \mid (k - m)$. Thus $g^k = g^m$ implies that $k \equiv m \pmod{n}$.

As a special case of (1), we can take $m = 0$, so that $g^k = g^0 = e$ if and only if $k \equiv 0 \pmod{n}$, which holds if and only if $n \mid k$.

To prove (2), note that we clearly have $\{e, g, g^2, \ldots, g^{n-1}\} \subseteq \langle g \rangle$. In the other direction, suppose $k \in \mathbb{Z}$, and use division with remainder by n to write $k = nq + r$, with $q \in \mathbb{Z}$ and $0 \leq r < n$. As above, $g^k = g^r$, which shows $g^k \in \{e, g, g^2, \ldots, g^{n-1}\}$, so that $\langle g \rangle \subseteq \{e, g, g^2, \ldots, g^{n-1}\}$.

Finally, the elements of $\{e, g, \ldots, g^{n-1}\}$ are all distinct, because if we had $g^k = g^m$ for some k, m such that $0 \leq m < k \leq n-1$, then n would divide $k-m$, which is smaller than n and larger than 0. This is impossible, so we conclude that $k = m$, proving the elements of this set are all distinct. □

In particular, note that the size of the group $\langle g \rangle$ is equal to $o(g)$ when g has finite order!

MATH 145 Course Reading 20: Subgroups and Symmetric Groups

November 2, 2020

Much can be learned about a group by studying its *subgroups*: the groups situated inside that group. Here, we formally provide the definition, give a useful test for determining when a subset of a group is a subgroup, and also provide several useful subgroup constructions. To conclude, we briefly introduce the *symmetric groups*, which are highly important in more advanced studies of groups. Here, our treatment will be fairly brief; only enough to touch on how it connects with the various group-theoretic ideas we've discussed so far.

Subgroups

We begin with the official definition:

Definition 20.1. Let G be a group. A subset H of G is called a *subgroup* of G if H itself is a group with respect to the binary operation already defined on G.

For some very immediate examples, if G is a group, then G is always a subgroup of G. Also, you can check right away that the set $\{e\}$ is also a subgroup of G, called the *trivial subgroup* of G.

A little more interestingly, $\mathbb{Z}$ is a subgroup of $\mathbb{Q}$ with respect to the addition operation, and $\mathbb{Q}$ is a subgroup of $\mathbb{R}$ with respect to the addition operation.

In practice, some variation on the following test is usually employed to check when a subset of a group is actually a subgroup:

Theorem 20.1 (Subgroup Test)**.** *Let G be a group, and let H be a nonempty subset of G. Then H is a subgroup of G if and only if, for all $a, b \in H$, we have $ab^{-1} \in H$, where b^{-1} denotes the inverse of b in G. In this case, H has the same identity element as G, and if $h \in H$, its inverse in the subgroup H is the same as its inverse in G.*

Proof. First, suppose H is a subgroup of G. Then for each $a, b \in H$, we know that $b^{-1} \in H$ because (as we argue below) the inverse of b in G will also serve as the inverse of b in H. Then since H is closed with respect to the operation on G, we get that $ab^{-1} \in H$, as desired.

Conversely, suppose that for all $a, b \in H$, we have $ab^{-1} \in H$. In particular, since H is nonempty, taking $b = a$ to be any element of H, we get that $aa^{-1} = e$, the identity of G, belongs to H. Then taking any $b \in H$ and taking $a = e$, we get $eb^{-1} = b^{-1} \in H$ for each $b \in H$. Thus every element of H has an inverse, within H. The operation on H is associative automatically, because the associativity relation $(ab)c = a(bc)$ holds for all $a, b, c \in G$, so certainly for all $a, b, c \in H$. Finally, to see that H is closed under the operation on G, so that we get a binary operation on H, note that for any $a, b \in H$, we know $b^{-1} \in H$ too, so $a(b^{-1})^{-1} = ab \in H$ by hypothesis.

Thus if H is a nonempty subset of G satisfying this one condition of the subgroup test, then H is a subgroup of G.

If we let e_H denote the identity of H and e_G the identity of G, then note that e_G belongs to H and also serves as an identity element of H. By uniqueness of the identity element in H (from Proposition 18.2), we get that $e_G = e_H$. Similarly, given $h \in H$, its inverse in G belongs to H by the Subgroup Test and also serves as the inverse of h in H, so that by uniqueness of inverses (from Lemma 18.1), the inverse of h in G agrees with its inverse in H. □

For examples of the Subgroup Test in action, let's prove that two common constructions involving a group actually yield subgroups.

Example 20.1. Let G be a group. We define the *center* of G to be the set

$$Z(G) = \{z \in G : zg = gz \text{ for all } g \in G\}.$$

Let's now verify that $Z(G)$ is always an abelian subgroup of G. First of all, $Z(G)$ is nonempty, since $e \in Z(G)$ essentially by definition (note that $eg = ge = g$ for all $g \in G$). Now, suppose we have $a, b \in Z(G)$. We must check that $ab^{-1} \in Z(G)$. Suppose we are given $g \in G$. Knowing that $bg = gb$ (since $b \in Z(G)$), multiplying on the left and right by b^{-1} gives $gb^{-1} = b^{-1}g$, so that $b^{-1} \in Z(G)$. Now,

$$(ab^{-1})g = a(b^{-1}g) = a(gb^{-1}) = (ag)b^{-1} = (ga)b^{-1} = g(ab^{-1}),$$

proving that $ab^{-1} \in Z(G)$. This proves that $Z(G)$ is a subgroup of G, by the Subgroup Test. To see why it is abelian, if we are given $a, b \in Z(G)$, by definition we must have $ag = ga$ for all $g \in G$, so in particular $ab = ba$.

Notice that $Z(G) = G$ if and only if G is abelian. Thus, in some way, the center of a group measures "how commutative" the group is.

Example 20.2. Suppose that H is a subgroup of G, and that $g \in G$. We define the *conjugate of H in G* (by g) to be

$$gHg^{-1} = \{ghg^{-1} : h \in H\}.$$

Let's verify that gHg^{-1} is a subgroup of G. Certainly, this set is nonempty, because H is. If we are given two elements $a, b \in gHg^{-1}$, then $a = gh_1g^{-1}$ and $b = gh_2g^{-1}$ for some $h_1, h_2 \in H$. Note that

$$\begin{aligned} ab^{-1} &= (gh_1g^{-1})(gh_2g^{-1})^{-1} \\ &= (gh_1g^{-1})(gh_2^{-1}g^{-1}) \\ &= g(h_1h_2^{-1})g^{-1}, \end{aligned}$$

and $h_1h_2^{-1} \in H$ by the Subgroup Test, so $ab^{-1} \in gHg^{-1}$. Again by the Subgroup Test, gHg^{-1} is a subgroup of G.

These conjugate subgroups of H introduced in the last example are important to study when we turn to the notion of *normal subgroups*. A subgroup H will be called normal if all the conjugate subgroups of H are equal to H. Normal subgroups will turn out to be the "right" type of subgroup for defining *quotient groups* down the road.

Let's take a (very) small group and find all of its subgroups by hand.

Example 20.3. Consider a cyclic group C_4 with four elements, say $C_4 = \{e, a, a^2, a^3\}$, with $a^4 = e$. What are the subgroups of C_4? Certainly, the trivial subgroup and C_4 itself are subgroups. If a belongs to a subgroup of C_4, then all powers of a belong to that subgroup, which means the subgroup is all of C_4. If $a^3 = a^{-1}$ belongs to a subgroup of C_4, then its inverse a belongs to the subgroup, and so again the subgroup is all of C_4. It follows that any proper subgroup of C_4 cannot contain either a or a^3. Aside from the trivial subgroup, then, the only remaining candidate is the subset $H = \{e, a^2\}$. We can check that this is a subgroup using the Subgroup Test. Indeed, the only non-trivial multiplication to check is that $e(a^2)^{-1} = a^{-2} = a^2$ belongs to H, which it does. So H is the only subgroup of C_4 other than the trivial subgroup and C_4 itself.

Symmetric Groups

As we've already noted in Example 18.6, the set of bijections from a nonempty set A to itself provides an example of a group, with composition of functions as the group operation. We can study such groups much more concretely if A is a finite set, say $|A| = n$. Here, we can treat the set of bijections from A to A as the set of bijections from $\{1, 2, \ldots, n\}$ back to itself. In this case, we denote the set by S_n and call it *the symmetric group of degree n*. One way we can represent an element $\sigma \in S_n$ is via a sort of matrix, with the

top row containing the integers 1 through n, and the bottom row containing the images $\sigma(1), \sigma(2), \ldots, \sigma(n)$ in that order. For example, the identity map ε that fixes every element can be written

$$\varepsilon = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ 1 & 2 & 3 & \cdots & n \end{pmatrix}.$$

Similarly, the bijection τ that sends 1 to 2, 2 to 3, ..., n to 1 can be written

$$\tau = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ 2 & 3 & 4 & \cdots & 1 \end{pmatrix}.$$

One fundamental fact that emerges from this representation is that S_n is a group with $n!$ elements. Indeed, at first we have n choices for the number to insert underneath 1 in the second now. Once that choice is made, we have $n-1$ remaining choices for the number to insert underneath 2 and still get an injective function. Then we have $n-3$ choices underneath 3, and so on, down to the number underneath n having only one choice, forced by all the choices made previously. In total, this yields $n(n-1)(n-2)\cdots 1 = n!$ elements. So as you can see, even small values of n can lead to large groups. For example, we already have $|S_6| = 720$.

If $n \geq 3$, then S_n is non-abelian. Indeed, let σ be the function that swaps 1 and 2, leaving all the other elements untouched, and let τ be the function that cycles the set $\{1, 2, 3\}$, sending 1 to 2, 2 to 3, and 3 to 1. These can be represented respectively by the matrices

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ 2 & 1 & 3 & \cdots & n \end{pmatrix} \qquad \tau = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ 3 & 1 & 2 & \cdots & n \end{pmatrix}$$

You can check right away that both $\sigma \circ \tau$ and $\tau \circ \sigma$ leave all the integers $4, 5, \ldots, n$ fixed. On the other hand, $(\sigma \circ \tau)(1) = \sigma(3) = 3$, $(\sigma \circ \tau)(2) = \sigma(1) = 2$, and $(\sigma \circ \tau)(3) = \sigma(2) = 1$, so that

$$\sigma \circ \tau = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ 3 & 2 & 1 & \cdots & n \end{pmatrix}.$$

A similar computation yields that

$$\tau \circ \sigma = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ 1 & 3 & 2 & \cdots & n \end{pmatrix}.$$

In particular, $\sigma \circ \tau \neq \tau \circ \sigma$, and S_n is not abelian.

There is *much* that can be said about symmetric groups, and because of the nature of our course, we can only scratch the surface of it here. So for now, we will leave the topic, and turn our focus toward one of the most important concepts in the study of groups: the notion of cosets and quotient groups, and the notion of a *homomorphism* between two groups (the correct generalization of a function to the setting of groups). These are useful tools both for deriving theoretical results on group structure (like Lagrange's Theorem), and also for learning more about specific instances of groups.

MATH 145 Course Reading 21: Homomorphisms and Cosets

November 4, 2020

After we studied sets as objects in themselves, one of our next steps was to define the functions between them. Among other things, a detailed study of functions allowed us to compare the cardinalities of sets. Now that we've defined groups as objects in themselves, our next step is likewise to study the functions between groups, and the various constructions that arise as a result. We will wish for our functions to preserve the binary operation on a group, and this leads to the definition of a *group homomorphism*.

Group Homomorphisms

Succinctly put, a group homomorphism is a function between two groups that respects the multiplication operation on both:

Definition 21.1. Let G_1 and G_2 be groups. A function $\phi : G_1 \to G_2$ is called a *homomorphism* if, for all elements $a, b \in G_1$, we have

$$\phi(ab) = \phi(a) \cdot \phi(b),$$

where $\cdot$ denotes the binary operation in G_2.

With the definition in play, let's give a bunch of specific examples:

Example 21.1. Given any groups G_1 and G_2, there is always a homomorphism $\phi : G_1 \to G_2$ called the *trivial homomorphism*, given by $\phi(a) = e_{G_2}$ for all $a \in G_1$. Indeed, for any $a, b \in G_1$, note that $\phi(ab) = e_{G_2} = e_{G_2} \cdot e_{G_2} = \phi(a) \cdot \phi(b)$, so the homomorphism condition is satisfied.

Example 21.2. For any group G, the identity map $\iota : G \to G$ given by $\iota(g) = g$ for all $g \in G$ is clearly a group homomorphism.

Example 21.3. For any positive integer n, the "reduction modulo n" map $\phi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ defined by $\phi(m) = [m]$ for all $m \in \mathbb{Z}$ is a homomorphism of groups (in additive notation). Indeed, for any $m_1, m_2 \in \mathbb{Z}$, note that

$$\phi(m_1 + m_2) = [m_1 + m_2] = [m_1] + [m_2] = \phi(m_1) + \phi(m_2).$$

Observe that we use the plus sign to denote two different operations here: one in $\mathbb{Z}$ and one in $\mathbb{Z}/n\mathbb{Z}$. As we will soon see, this homomorphism is one of the prototype examples of a *quotient mapping*.

Example 21.4. For a more interesting example, the natural logarithm function $\ln : \mathbb{R}^+ \to \mathbb{R}$ is a homomorphism from the group of positive real numbers (with multiplication as the group operation) to the group of all real numbers (with addition as the group operation). Indeed, one of the main rules of logarithms,

$$\ln(ab) = \ln(a) + \ln(b),$$

valid for all positive real numbers a and b, is just expressing the fact that ln is a group homomorphism.

Example 21.5. As one more example, given any group G and element $g \in G$, there is a homomorphism $\phi : \mathbb{Z} \to \langle g \rangle$, given by $\phi(m) = g^m$ for each $m \in \mathbb{Z}$. This is called the *exponent map*, and is a homomorphism because of the exponent rules. For all $k, m \in \mathbb{Z}$, we have

$$\phi(k + m) = g^{k+m} = g^k \cdot g^m = \phi(k) \cdot \phi(m).$$

As it turns out, asking a homomorphism to preserve the group operation means that it preserves many other features of a group automatically. We summarize them in the following theorem:

Theorem 21.1. *Let $\phi : G_1 \to G_2$ be a group homomorphism.*

(1) The map ϕ preserves the identity: $\phi(e_{G_1}) = e_{G_2}$.

(2) The map ϕ preserves inverses: for all $g \in G_1$, we have $\phi(g^{-1}) = \phi(g)^{-1}$.

(3) The map ϕ preserves powers: for all $g \in G_1$ and $k \in \mathbb{Z}$, we have $\phi(g^k) = \phi(g)^k$.

(4) The composition of homomorphisms is a homomorphism: if $\psi : G_2 \to G_3$ is another group homomorphism, then the composition $\psi \circ \phi : G_1 \to G_3$ is a group homomorphism.

Proof.

(1) Applying the homomorphism property, note that

$$\phi(e_{G_1}) = \phi(e_{G_1} \cdot e_{G_1}) = \phi(e_{G_1}) \cdot \phi(e_{G_1}).$$

Taking $\phi(e_{G_1}) = \phi(e_{G_1}) \cdot \phi(e_{G_1})$ and multiplying both sides by $\phi(e_{G_1})^{-1}$ on the left, we get $e_{G_2} = \phi(e_{G_1})$, as needed.

(2) To prove this, note that for each $g \in G$, we have

$$\phi(g) \cdot \phi(g^{-1}) = \phi(g \cdot g^{-1}) = \phi(e_{G_1}) = e_{G_2},$$

by item (1). Similarly, $\phi(g^{-1}) \cdot \phi(g) = e_{G_2}$. By uniqueness of inverses in a group, this proves that $\phi(g^{-1}) = \phi(g)^{-1}$.

(3) We prove this for $k \in \mathbb{N}$ first by induction on k. For $k = 0$, note that $\phi(g^0) = \phi(e_{G_1}) = e_{G_2} = \phi(g)^0$. Now assuming the result holds for some $k \in \mathbb{N}$, note that

$$\phi(g^{k+1}) = \phi(g^k \cdot g) = \phi(g^k) \cdot \phi(g) = \phi(g)^k \cdot \phi(g) = \phi(g)^{k+1},$$

completing the proof by induction. If k is negative, say $k = -m$ for some positive integer m, we now know that $\phi(g^k) = \phi(g^{-m}) = \phi((g^{-1})^m) = \phi(g^{-1})^m = (\phi(g)^{-1})^m = \phi(g)^{-m} = \phi(g)^k$, completing this proof.

(4) This one is left as an exercise for you!

□

One of the most important structures associated to a homomorphism is its *kernel*:

Definition 21.2. If $\phi : G_1 \to G_2$ is a homomorphism of groups, the *kernel* of ϕ, denoted $\ker \phi$, is the set

$$\ker \phi = \{g \in G_1 : \phi(g) = e_{G_2}\}.$$

Two of the most useful facts about kernels are stated and proved below:

Theorem 21.2. *Let $\phi : G_1 \to G_2$ be a homomorphism of groups.*

(1) The set $\ker \phi$ is a subgroup of G_1.

(2) The function ϕ is injective if and only if $\ker \phi = \{e_{G_1}\}$.

Proof.

(1) We apply the Subgroup Test. First, $\ker \phi$ is nonempty because $\phi(e_{G_1}) = e_{G_2}$, so that e_{G_1} always belongs to $\ker \phi$. Now, if $a, b \in \ker \phi$, we show that $ab^{-1} \in \ker \phi$. This we can check directly with properties of homomorphisms:

$$\phi(ab^{-1}) = \phi(a)\phi(b^{-1}) = \phi(a)\phi(b)^{-1} = e_{G_2} e_{G_2}^{-1} = e_{G_2} e_{G_2} = e_{G_2}.$$

By the Subgroup Test, $\ker \phi$ is indeed a subgroup of G_1.

(2) Suppose first that ϕ is injective. The containment $\{e_{G_1}\} \subseteq \ker\phi$ always holds, as noted above, so now suppose that $g \in \ker\phi$. Then $\phi(g) = e_{G_2} = \phi(e_{G_1})$, so by injectivity of ϕ, we get $g = e_{G_1}$, proving that $\ker\phi \subseteq \{e_{G_1}\}$. We have now proved that if ϕ is injective, then $\ker\phi = \{e_{G_1}\}$.

Now assume that $\ker\phi = \{e_{G_1}\}$, and suppose we have $a, b \in G_1$ such that $\phi(a) = \phi(b)$. Multiplying both sides by $\phi(b)^{-1}$ on the right and applying homomorphism properties, we get

$$e_{G_1} = \phi(a)\phi(b)^{-1} = \phi(a)\phi(b^{-1}) = \phi(ab^{-1}),$$

so that $ab^{-1} \in \ker\phi$. Since $\ker\phi = \{e_{G_1}\}$, we know that $ab^{-1} = e_{G_1}$, and multiplying both sides by b on the right yields $a = b$. Thus if $\ker\phi = \{e_{G_1}\}$, then ϕ is injective.

□

Cosets

Given any function $f : A \to B$ between sets, we were able to define an equivalence relation $\sim$ on A by saying that $a_1 \sim a_2$ if $f(a_1) = f(a_2)$. If we apply this relation to a group homomorphism $\phi : G_1 \to G_2$, can we describe the equivalence relation and the equivalence classes another way?

For notational convenience, let's set $H = \ker\phi$. Given elements $a, b \in G$, we know that $a \sim b$ if and only if $\phi(a) = \phi(b)$, which holds if and only if $\phi(ab^{-1}) = e_{G_2}$, which holds if and only if $ab^{-1} \in H$.

This last condition can be generalized to subgroups H other than $\ker\phi$, and we can check right away that this does give an equivalence relation:

Theorem 21.3. *Let G be a group, and let H be a subgroup of G. We define a relation $\sim$ on G by declaring that for $a, b \in G$, we have $a \sim b$ if and only if $ab^{-1} \in H$. Then $\sim$ is an equivalence relation. Moreover, the equivalence class of an element $a \in G$ is the set $Ha = \{ha : h \in H\}$, which is called the* (right) coset of H generated by a.

Proof. We check each of the equivalence relation properties:

- **Reflexive:** For any $a \in G$, we have $a \sim a$ because $e_G = aa^{-1} \in H$, since H is a subgroup.
- **Symmetric:** Given $a, b \in G$ such that $a \sim b$, we know $ab^{-1} \in H$. Since H is closed under taking inverses, we get that $(ab^{-1})^{-1} = ba^{-1} \in H$ as well, which says $b \sim a$.
- **Transitive:** If we have $a, b, c \in G$ for which $a \sim b$ and $b \sim c$, we know $ab^{-1} \in H$ and $bc^{-1} \in H$. Since H is closed under multiplication, we get

  $$(ab^{-1})(bc^{-1}) = ac^{-1} \in H,$$

  which says $a \sim c$.

Now that we know $\sim$ is an equivalence relation, choose $a \in G$. By definition,

$$\begin{aligned}
[a] &= \{g \in G : g \sim a\} \\
&= \{g \in G : ga^{-1} \in H\} \\
&= \{g \in G : ga^{-1} = h \text{ for some } h \in H\} \\
&= \{g \in G : g = ha \text{ for some } h \in H\} \\
&= Ha,
\end{aligned}$$

proving that the equivalence class of a is the right coset generated by a. □

We can define an analogous equivalence relation $\sim_L$ on G by declaring that $a \sim_L b$ if and only if $b^{-1}a \in H$, and the resulting equivalence classes are sets of the form $aH = \{ah : h \in H\}$, which are called *left cosets of H*. Note that if G is abelian, then these two equivalence relations are the same, and $aH = Ha$ for any element $a \in G$ and subgroup H of G.

Example 21.6. It is instructive to see what happens when we take $G = \mathbb{Z}$. We let n be a positive integer, and note that $H = n\mathbb{Z} = \{m \in \mathbb{Z} : m = nk \text{ for some } k \in \mathbb{Z}\}$ is a subgroup of G (which you can quickly check). Since the group operation is addition here, we use additive notation for the cosets too, writing $n\mathbb{Z}+a$ for the right coset of $n\mathbb{Z}$ generated by a.

Can we describe these cosets in a more familiar way? Note that $b \in n\mathbb{Z} + a$ if and only if $b \sim a$ according to the equivalence relation above, which holds if and only if $b - a \in n\mathbb{Z}$ (using the additive notation, so that $-a$ is the inverse of a). Now $b - a \in n\mathbb{Z}$ if and only if $n \mid (b - a)$, so that b belongs to the equivalence class $n\mathbb{Z} + a$ if and only if $b \equiv a \pmod{n}$. Thus the equivalence relation just defined here is the equivalence relation of congruence modulo n in this particular case!

One extra thing we were able to do with equivalence classes modulo n was to take the addition operation on $\mathbb{Z}$ and define it on the equivalence classes in $\mathbb{Z}/n\mathbb{Z}$ in exactly the same way, taking $[a] + [b] = [a + b]$. In similar fashion, we might hope to then define a group operation on the right cosets of H in G by taking $(Ha)(Hb) = Hab$. Unfortunately, this prescription does not always result in a group operation; extra conditions on H are required in order to make this work. Once those extra conditions are found, the resulting set of cosets with this group operation will be called a *quotient group* and be denoted by G/H.

MATH 145 Course Reading 22:
Quotient Groups and Lagrange's Theorem

November 6, 2020

At the end of the previous reading, we introduced the concept of *cosets* of a subgroup, which were nothing more than equivalence classes of group elements under a certain equivalence relation. In particular, the left or right cosets of a subgroup partition a group G, and a careful analysis of this fact will lead immediately to the proof of Lagrange's Theorem, a powerful result in group theory. Before we get there, however, we investigate the conditions necessary to ensure that a natural group structure may be imposed on the (right) cosets of a subgroup within a group.

Normal Subgroups and Quotient Groups

As we investigated in our most recent synchronous session, the rule $(Ha)(Hb) = Hab$ for multiplying right cosets of a subgroup will be well-defined if and only if that subgroup is a *normal* subgroup, according to the definition below:

Definition 22.1. Let G be a group. A subgroup H is called a *normal subgroup* of G if $gHg^{-1} = H$ for all elements $g \in G$. (Here, gHg^{-1} is the notation for the conjugate subgroup of H introduced in Example 20.2). In this case, we adopt the notation $H \lhd G$ to denote that H is a normal subgroup of G.

Let's now prove that the normal subgroups are exactly the ones for which the set G/H of right cosets of H has a natural group structure:

Theorem 22.1. *Let G be a group. For every subgroup H of G, the formula $(Ha)(Hb) = Hab$ gives a well-defined multiplication of right cosets if and only if $H \lhd G$.*

Proof. First, suppose that $(Ha)(Hb) = Hab$ is well-defined for all right cosets of H in G. Letting $g \in G$ be arbitrary, and $h \in H$ be arbitrary, clearly $Hg = Hg$ and $Hh = He$. Thus $(Hg)(Hh) = (Hg)(He)$, or in other words, $Hgh = Hge$. By definition, this says $(gh)(ge)^{-1} = ghg^{-1} \in H$. This proves that $gHg^{-1} \subseteq H$. Taking g^{-1} in place of g in the argument above, we end up with $g^{-1}Hg \subseteq H$ for all $g \in G$, and you can check right away that this implies $H \subseteq gHg^{-1}$ for all $g \in G$. Thus $H = gHg^{-1}$ for all $g \in G$, so that H is normal in G.

Conversely, suppose that H is normal in G. We prove that multiplication of right cosets is well-defined. Suppose we have $a, b, a_1, b_1 \in G$ such that $Ha = Ha_1$ and $Hb = Hb_1$. Then by definition, $aa_1^{-1} \in H$ and $bb_1^{-1} \in H$. We wish to show that $(Ha)(Hb) = (Ha_1)(Hb_1)$, which is equivalent to $(ab)(a_1b_1)^{-1} = abb_1^{-1}a_1^{-1} \in H$. But notice that

$$abb_1^{-1}a_1^{-1} = a(bb_1^{-1})a_1^{-1} = (a(bb_1^{-1})a^{-1})(aa_1^{-1}).$$

Now, $aHa^{-1} = H$ and $bb_1^{-1} \in H$, so $a(bb_1^{-1})a^{-1} \in aHa^{-1} = H$. Also, $aa_1^{-1} \in H$ by assumption. By closure of H under multiplication, we get that $(a(bb_1^{-1})a^{-1})(aa_1^{-1}) = abb_1^{-1}a_1^{-1} \in H$, as needed. Thus multiplication of right cosets is well-defined. □

With this foundational result out of the way, let's prove a bunch of facts about the collection of right cosets of a normal subgroup:

Theorem 22.2. *Let G be a group, and let H be a normal subgroup of G.*

(1) The set G/H of right cosets of H is a group under the operation $(Ha)(Hb) = Hab$, called the quotient group *of G by H.*

(2) The function $\phi : G \to G/H$ given by $\phi(g) = Hg$ is a surjective group homomorphism, called the quotient mapping.

(3) If G is abelian, then G/H is abelian.

(4) If G is a cyclic group, then G/H is a cyclic group.

Proof.

(1) The previous theorem, Theorem 22.1, tells us that the binary operation on G/H is well-defined. For associativity, note that for any cosets Ha, Hb, Hc, we have

$$((Ha)(Hb))(Hc) = (Hab)(Hc) = H(ab)c = Ha(bc) = (Ha)(Hbc) = (Ha)((Hb)(Hc)),$$

applying associativity in G. You can check right away that the identity element of G/H is the coset $H = He$, and the inverse of the coset Ha is the coset Ha^{-1}. Thus G/H is indeed a group.

(2) The fact that ϕ is a homomorphism is immediate from the definitions: for $a, b \in G$, we have

$$\phi(ab) = Hab = (Ha)(Hb) = \phi(a)\phi(b).$$

To see that it is surjective, note that for any coset $Ha \in G/H$, we have $\phi(a) = Ha$.

(3) If G is abelian, then for any cosets $Ha, Hb \in G/H$, note that $(Ha)(Hb) = Hab = Hba = (Hb)(Ha)$, so G/H is abelian too.

(4) If $G = \langle g \rangle$ for some $g \in G$, then every element of G is of the form g^k for some $k \in \mathbb{Z}$. Thus, given $Ha \in G/H$, we know $a = g^k$ for some integer k, and then $Ha = Hg^k = \phi(g^k) = \phi(g)^k = (Hg)^k$, where ϕ denotes the quotient homomorphism. This shows $G/H = \langle Hg \rangle$, so that G/H is cyclic.

□

At this point, a couple remarks are in order. If G is an abelian group, then for any subgroup H and element $g \in G$, note that $gHg^{-1} = \{ghg^{-1} : h \in H\} = \{gg^{-1}h : h \in H\} = \{h : h \in H\} = H$, so *every* subgroup of G is normal. In particular, in the group $\mathbb{Z}$ with addition as the operation, each of the subgroups $n\mathbb{Z} = \{nk : k \in \mathbb{Z}\}$ is normal (for each positive integer n), and so we may form the quotient groups $\mathbb{Z}/n\mathbb{Z}$ with elements $n\mathbb{Z}+a$, which are none other than equivalence classes under the relation of congruence modulo n. So the quotient group notation $\mathbb{Z}/n\mathbb{Z}$ aligns with our existing use of this notation for the group of integers modulo n.

Let's look quickly at another example of a quotient group, just to get a sense for how these might behave:

Example 22.1. Consider the group $\mathbb{Q}$, with addition as the binary operation. The subset $\mathbb{Z}$ is then a subgroup of $\mathbb{Q}$ under addition, and since $\mathbb{Q}$ is abelian, $\mathbb{Z}$ is automatically a normal subgroup. Thus the quotient group $\mathbb{Q}/\mathbb{Z}$ exists. The elements of this group are cosets of the form $\mathbb{Z} + q$, where $q \in \mathbb{Q}$.

We claim in fact that every element of $\mathbb{Q}/\mathbb{Z}$ has a unique representative of the form $\mathbb{Z} + \delta$, where $\delta \in \mathbb{Q}$ satisfies $0 \leq \delta < 1$. Indeed, for any $q \in \mathbb{Q}$, we can always "round down" q to an integer $\lfloor q \rfloor$ (called the *floor* of q), for which $0 \leq q - \lfloor q \rfloor < 1$. For instance, $\lfloor 3/2 \rfloor = 1$, and $\lfloor -8/5 \rfloor = -2$. Note that $\delta = q - \lfloor q \rfloor \in \mathbb{Q}$, because δ is the difference of two rational numbers. Thus $q - \delta = \lfloor q \rfloor \in \mathbb{Z}$, so that $\mathbb{Z} + q = \mathbb{Z} + \delta$, and $0 \leq \delta < 1$.

If δ' was another rational number with $0 \leq \delta' < 1$ and $\mathbb{Z} + \delta' = \mathbb{Z} + q$, then we would have $\mathbb{Z} + \delta' = \mathbb{Z} + \delta$, and so $\delta' - \delta \in \mathbb{Z}$. But given the constraints on δ and δ', we have $-1 < \delta' - \delta < 1$, so if this quantity is an integer, it must be 0, which implies $\delta = \delta'$.

In summary, every element of $\mathbb{Q}/\mathbb{Z}$ is uniquely represented by a coset of the form $\mathbb{Z} + \delta$, where $0 \leq \delta < 1$.

We can say one more thing about this group: despite the fact that it's countably infinite, every element has finite order (such a group is called *torsion*). Indeed, given any coset $\mathbb{Z} + q$, write $q = \frac{a}{b}$, where a and b are integers, and $b \neq 0$. In fact, without loss of generality, we may assume $b > 0$, since if b is negative, then $\frac{a}{b} = \frac{-a}{-b}$, and $-b$ is positive.

Note then that $b(\mathbb{Z} + q) = \mathbb{Z} + bq = \mathbb{Z} + a = \mathbb{Z} + 0$, since $a \in \mathbb{Z}$. Thus the order of the coset $\mathbb{Z} + (a/b)$ is at most b, and in particular is finite.

Lagrange's Theorem

Turning back to looking at cosets in general, we can use the facts we've derived thus far to prove a far-reaching theorem about finite groups. We begin with a definition:

Definition 22.2. Let G be a group, and let H be a subgroup of G. The *index* of H in G, denoted $|G : H|$, is equal to the number of distinct right cosets of H in G (this can be either finite or infinite).

In particular, if H is normal in G, then $|G : H|$ gives the number of elements of the group G/H. It is also worth noting that $|G : H|$ is equal to the number of distinct *left* cosets of H in G: you can check as an exercise that the mapping sending Ha to $a^{-1}H$ is a bijection between the set of right cosets of H and the set of left cosets of H.

We may now state and prove the theorem:

Theorem 22.3 (Lagrange's Theorem)**.** *Let G be a finite group, and let H be a subgroup of G. Then $|H|$ divides $|G|$, and in fact $|G : H| = \frac{|G|}{|H|}$.*

Proof. Since the right cosets of H in G form a partition of G, let $Ha_1, Ha_2, \ldots, Ha_n$ denote the collection of distinct right cosets of H in G (there are finitely many because G is finite). In particular, the union of $Ha_1, \ldots, Ha_n$ is G, and $Ha_i \cap Ha_j = \varnothing$ if $i \neq j$. Note that for any index i, we have $|H| = |Ha_i|$, because the mapping sending h to ha_i is a bijection from H onto Ha_i. The map is clearly surjective, and to check injectivity, suppose $h_1 a_i = h_2 a_i$. Multiplying by a_i^{-1} on the right yields $h_1 = h_2$.

Thus $|Ha_1| = |Ha_2| = \cdots = |Ha_n|$. Thus G is a disjoint union of $n = |G : H|$ cosets, each with size $|H|$. Putting all this together yields $|G : H| \cdot |H| = |G|$, so that $\frac{|G|}{|H|} = |G : H|$ is an integer. In particular, $|H|$ must divide $|G|$. □

This theorem can be immediately applied to derive far-reaching consequences:

Corollary 22.1. *If G is a finite group and $g \in G$, then $o(g)$ divides $|G|$.*

Proof. For this, note that $H = \langle g \rangle$ is a subgroup of G, and that $|H| = o(g)$. The conclusion then follows immediately from Lagrange's Theorem. □

Corollary 22.2. *If G is a finite group with n elements, then for all $g \in G$, we have $g^n = e$.*

Proof. Given an arbitrary element $g \in G$, we know that $k = o(g)$ divides n by Corollary 22.1. Thus $n = k\ell$ for some integer ℓ, and $g^n = g^{k\ell} = (g^k)^\ell = e^\ell = e$, as needed. □

Corollary 22.3. *If p is a prime number, then every group with p elements is cyclic. In fact, if G is a group with p elements, then for any non-identity element $g \in G$, we have $G = \langle g \rangle$.*

Proof. Let G be a group with p elements, where p is a prime. Since $p \geq 2$, there is a non-identity element $g \in G$. We set $H = \langle g \rangle$. Then $|H| > 1$, because the generator g of H has order larger than 1. But $|H|$ must divide $|G|$ by Lagrange's Theorem, and $|G| = p$ has only 1 and p as positive divisors. Thus we must have $|H| = p = |G|$, so that $G = H = \langle g \rangle$, showing also that G is cyclic. □

MATH 145 Course Reading 23:
An Introduction to Rings

November 9, 2020

As we've seen, a group is a set with only a single binary operation defined on it. However, many of the algebraic structures we are used to working with have more than one binary operation (think of the addition and multiplication operations on $\mathbb{Z}$ or $\mathbb{R}$, for example). The next algebraic structure we're about to define, called a *ring*, captures the features of these two examples we just mentioned. A ring will come equipped with two binary operations (one usually denoted by addition, the other usually denoted by multiplication), and these two operations will satisfy several properties, much like a group does.

Definition of a Ring and Basic Examples

One curious feature about the definition of a ring is that not all authors agree on what parts must belong to the definition. That said, here is the definition we will use in this course:

Definition 23.1. A *ring* is a set R, equipped with two binary operations, usually denoted by addition and multiplication, satisfying the following conditions:

(1) The set R, along with the binary operation $+$, is an abelian group. (In this case, the identity of the group is usually denoted 0).

(2) The set R, along with the binary operation $\cdot$, is a monoid. (In this case, the identity of the monoid is usually denoted 1).

(3) Left and right distributive laws hold: for all $a, b, c \in R$, we have $a(b+c) = ab+ac$, and $(a+b)c = ac+bc$.

To spell out these conditions explicitly, the binary operations $+$ and $\cdot$ must satisfy all of the following, for all $a, b, c \in R$:

- $(a + b) + c = a + (b + c)$
- There is $0 \in R$ such that $a + 0 = 0 + a = a$ for all $a \in R$.
- For each $a \in R$, there is $-a \in R$ such that $a + (-a) = (-a) + a = 0$.
- $a + b = b + a$.
- $(ab)c = a(bc)$
- There is $1 \in R$ such that $a \cdot 1 = 1 \cdot a = a$ for all $a \in R$.
- $a(b + c) = ab + ac$
- $(a + b)c = ac + bc$

If multiplication in R is commutative, so that $ab = ba$ for all $a, b \in R$, then we call R a *commutative ring*.

The one point of disagreement in this definition is the condition that R should have an identity (unity) element with respect to multiplication. Some authors do not insist on this in their definition of ring, and would call a ring such as defined above a *ring with unity*. This is something to keep in mind if you ever do some reading on your own in ring theory!

As you might expect, this definition is crafted so that all of our familiar examples satisfy it:

Example 23.1. The sets $\mathbb{Z}, \mathbb{Q}$, and $\mathbb{R}$, with the standard definitions of addition and multiplication, are all commutative rings.

Example 23.2. For any positive integer n, the set of integers modulo n, $\mathbb{Z}/n\mathbb{Z}$, can also be given the structure of a ring. For the addition operation, we use the one we have already previously defined. For the multiplication operation, we will set $[a] \cdot [b] = [ab]$ for all $a, b \in \mathbb{Z}$. Of course, we must first check that this operation is well-defined.

Supposing that we have integers a, a', b, b' such that $[a] = [a']$ and $[b] = [b']$, this means that $a \equiv a' \pmod{n}$ and $b \equiv b' \pmod{n}$, so $a - a'$ and $b - b'$ are multiples of n, say $a - a' = kn$ and $b - b' = \ell n$. We wish to show that $[ab] = [a'b']$, so that $ab - a'b'$ is a multiple of n. Here, we use a clever trick:

$$ab - a'b' = ab - ab' + ab' - a'b' = a(b - b') + (a - a')b' = a(\ell n) + (kn)b' = (a\ell + kb')n,$$

where $a\ell + kb' \in \mathbb{Z}$. This proves that $ab \equiv a'b' \pmod{n}$, so $[ab] = [a'b']$.

Now that we know multiplication is well-defined as a binary operation on $\mathbb{Z}/n\mathbb{Z}$, only a moment's checking shows that the operation is associative and has identity $[1]$ (exercise!), so that $\mathbb{Z}/n\mathbb{Z}$ is a monoid with respect to this operation. Furthermore, you can check right away that multiplication is commutative.

Finally, we check the distributive laws: given $[a], [b], [c] \in \mathbb{Z}/n\mathbb{Z}$, note that

$$[a]([b] + [c]) = [a] \cdot [b + c] = [a(b + c)] = [ab + ac] = [ab] + [ac] = [a][b] + [a][c],$$

and a similar check verifies that $([a] + [b])[c] = [a][c] + [b][c]$. Thus $(\mathbb{Z}/n\mathbb{Z})$ is a commutative ring.

Example 23.3. For a trivial example, let $R = \{0\}$, with addition and multiplication given by $0+0 = 0 \cdot 0 = 0$. You can check right away that these binary operations make R into a commutative ring, with 0 as the identity with respect to both addition and multiplication.

Example 23.4. For an example that sometimes yields non-commutative rings, let G be an abelian group, with group operation denoted by addition. We let $\mathrm{End}(G)$ denote the set of group homomorphisms from G to G, called the set of *endomorphisms of* G. We can define an addition on $\mathrm{End}(G)$ by pointwise addition, taking $(\phi + \psi)(g) = \phi(g) + \psi(g)$ for all $\phi, \psi \in \mathrm{End}(G)$. Note that $\phi + \psi$ is indeed a group homomorphism from G to G, because for any $g, h \in G$, we have

$$\begin{aligned}(\phi + \psi)(g + h) &= \phi(g + h) + \psi(g + h) \\ &= \phi(g) + \phi(h) + \psi(g) + \psi(h) \\ &= (\phi(g) + \psi(g)) + (\phi(h) + \psi(h)) \\ &= (\phi + \psi)(g) + (\phi + \psi)(h).\end{aligned}$$

Note that this verification uses the fact G is abelian!

We can now check that addition on $\mathrm{End}(G)$ makes $\mathrm{End}(G)$ into an abelian group, which I leave as an exercise! I will say only that the identity element of the group is the zero homomorphism $\mathbf{0} : G \to G$ given by $\mathbf{0}(g) = 0$ for all $g \in G$. For the multiplication operation in $\mathrm{End}(G)$, we use function composition, so that $\phi\psi$ stands for the composition $\phi \circ \psi$. Since the composition of homomorphisms is a homomorphism, note that $\phi\psi \in \mathrm{End}(G)$, as needed.

Associativity of multiplication in $\mathrm{End}(G)$ follows from the associativity of function composition. The identity element with respect to multiplication is the identity function $\iota : G \to G$ given by $\iota(g) = g$ for all $g \in G$. Left distributivity and right distributivity are a bit more interesting. Given $\phi, \psi, \pi \in \mathrm{End}(G)$, and $g \in G$, we have

$$\begin{aligned}\phi(\psi + \pi)(g) &= \phi(\psi(g) + \pi(g)) \\ &= \phi(\psi(g)) + \phi(\pi(g)) \\ &= (\phi\psi)(g) + (\phi\pi)(g) \\ &= (\phi\psi + \phi\pi)(g),\end{aligned}$$

so $\phi(\psi + \pi) = \phi\psi + \phi\pi$. (Note that the key here is that ϕ is a group homomorphism). On the other hand, we have

$$\begin{aligned}((\phi + \psi)\pi)(g) &= (\phi + \psi)(\pi(g)) \\ &= \phi(\pi(g)) + \psi(\pi(g)) \\ &= (\phi\pi)(g) + (\psi\pi)(g) \\ &= (\phi\pi + \psi\pi)(g),\end{aligned}$$

so $(\phi+\psi)\pi = \phi\pi+\psi\pi$. (Note that this did not need the fact that any of the maps are group homomorphisms!)

As for why this ring need not be commutative, we will investigate this further on Assignment 7.

Properties of Rings and Definitions

Because the additive structure of a ring forms an abelian group, and the multiplicative structure of a ring is a monoid, we can deduce several properties about rings directly from our previous studies. In the points that follow, let R be an arbitrary ring.

- The additive and multiplicative identities of R are unique (justifying the unique notation 0 and 1 for these identities).
- For any $a \in R$, its additive inverse is unique, and usually denoted $-a$.

These properties follow directly from Proposition 18.2 and Lemma 18.1 in our discussion of monoids.

Some additional properties of rings follow from the distributive law, telling us how the additive and multiplicative structure of a ring interact:

Theorem 23.1. *Let R be a ring, and let 0 be its additive identity. Then for all elements $a \in R$, we have $a \cdot 0 = 0 \cdot a = 0$.*

Proof. Using the fact that 0 is an additive identity, we know that $0 + 0 = 0$. Thus, by distributivity,

$$a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0.$$

If we add $-(a \cdot 0)$ to both sides, this results in $0 = a \cdot 0$, as needed. The proof that $0 \cdot a = 0$ is similar. □

Building off this fact, we can say some familiar things about how additive inverses interact with multiplication:

Theorem 23.2. *Let R be a ring, and let $a, b \in R$ be arbitrary. Then $(-a)b = a(-b) = -(ab)$, and $(-a)(-b) = ab$.*

Proof. Since the additive inverse of an element is unique, we show that both $(-a)b$ and $a(-b)$ are additive inverses of ab. This will immediately establish that both are equal to $-(ab)$. By distributivity,

$$(-a)b + ab = (-a + a)b = 0b = 0,$$

applying Theorem 23.1 in the last equality. By commutativity of addition in R, $ab+(-a)b = 0$ as well. Thus $(-a)b = -(ab)$ by uniqueness of inverses. The proof that $a(-b) = -(ab)$ is similar.

Finally, applying the first part of our result, we can establish the second: we have $(-a)(-b) = -(a(-b)) = -(-(ab)) = ab$, using the fact that ab is the unique additive inverse of $-ab$. □

One more useful definition to have about rings is the following:

Definition 23.2. For any ring R, the *characteristic* of R, denoted $\operatorname{char} R$, is the order of the multiplicative identity 1 in the group R under addition, if this order is finite. If the order of 1 is not finite, then we declare the characteristic of R to be 0.

Said differently, the characteristic of R is the least positive integer n such that $n \cdot 1$ (1 added together n times) is equal to 0, if such an integer exists. Otherwise, $\operatorname{char} R = 0$.

Looking back at our examples, all of $\mathbb{Z}, \mathbb{Q}$, and $\mathbb{R}$ are rings with characteristic 0, while for any positive integer n, $\mathbb{Z}/n\mathbb{Z}$ has characteristic n, since $n \cdot [1] = [n] = [0]$, while for any positive integer $m < n$, we have $m \cdot [1] = [m] \neq [0]$. Similarly, the zero ring has characteristic 1, since in this ring, $1 = 0$. (Do all of these examples make sense to you?)

Using facts we've already encountered about cyclic groups, the following properties of ring characteristic are almost immediate:

Theorem 23.3. *If R is a ring of characteristic $n > 0$, then we have $k \cdot r$ (the result of adding r to itself k times) equal to 0 for all $r \in R$ if and only if $n \mid k$. If R is a ring of characteristic 0, then $k \cdot r = 0$ holds for all $r \in R$ if and only if $k = 0$.*

Proof. First, suppose R is a ring of characteristic n, let k be an integer such that $n \mid k$, and let $r \in R$ be arbitrary. Then $k = mn$ for some integer m. Note that $k \cdot r = (mn) \cdot r = ((mn) \cdot 1) \cdot r$ by distributivity. Furthermore, $n \cdot 1 = 0$, so $(mn) \cdot 1 = m \cdot (n \cdot 1) = m \cdot 0 = 0$ by exponent rules (in additive notation). This implies $k \cdot r = 0 \cdot r = 0$, as needed.

Now suppose that k is an integer such that for any $r \in R$, we have $k \cdot r = 0$. In particular, $k \cdot 1 = 0$, and so k must be a multiple of the order of 1 in the group R under addition. By Theorem 19.4, part (1), but translated to additive notation, we conclude that $n \mid k$.

Now suppose that $\operatorname{char} R = 0$. Certainly, if $k = 0$, then $k \cdot r = 0 \cdot r = 0$ for all $r \in R$. On the other hand, if $k \cdot r = 0$ for all $r \in R$, then $k \cdot 1 = 0$, and since 1 has infinite order, this happens only when $k = 0$. □

MATH 145 Course Reading 24:
Subrings and Homomorphisms

November 11, 2020

By now, you might be starting to see a pattern emerging in our studies. When we looked at sets, two important constructions included subsets and functions between sets. By analogy, when we studied groups, we looked at subgroups and the group homomorphisms. Now that we've introduced rings, we have opportunity to study both the rings contained in a given ring (called the *subrings* of that ring), and also the functions between rings that preserve the ring structure (known as the *ring homomorphisms*). This reading will be devoted to introducing and studying these two concepts.

Subrings

Now that we're familiar with the definition of a subgroup, the definition of a subring will probably feel quite predictable:

Definition 24.1. Let R be a ring. A subset S of R is called a *subring* if the addition and multiplication operations on R restrict to binary operations on S, and S is a ring with respect to these restricted operations from R. Furthermore, we insist that $1_R = 1_S$, i.e. that the multiplicative identity of the rings R and S agree.

The condition $1_R = 1_S$ represents something new, and perhaps unexpected. As we saw when studying groups, it was not necessary to specify that the identity of a subgroup be the same as the identity of the larger group; that property followed automatically from the definition of a subgroup. The following example shows why the condition $1_R = 1_S$ does not automatically follow for subrings:

Example 24.1. Consider the ring $\mathbb{Z}/6\mathbb{Z}$ of integers modulo 6, and the subset $S = \{[0], [2], [4]\}$ of "even" equivalence classes. This subset is a subgroup of the abelian group $\mathbb{Z}/6\mathbb{Z}$ under addition; in fact, it is the cyclic subgroup generated by $[2]$. Furthermore, this set S is closed under multiplication of congruence classes, since any class times $[0]$ gives $[0]$, and $[2] \cdot [2] = [4]$, $[2] \cdot [4] = [2]$, and $[4] \cdot [4] = [4]$. It should also be clear from these multiplications that $[4]$ is an identity element for the multiplication operation on this subset, since $[4] \cdot [a] = [a]$ for $a \in \{0, 2, 4\}$. Thus S is a subset of $\mathbb{Z}/6\mathbb{Z}$ that is a ring under the operations on $\mathbb{Z}/6\mathbb{Z}$, but it's not a subring according to our definition, since the identity $[4]$ for S is different from the identity $[1]$ for $\mathbb{Z}/6\mathbb{Z}$.

On the other hand, the expected examples of subrings do work: $\mathbb{Z}$ is a subring of both $\mathbb{Q}$ and $\mathbb{R}$, for instance. Just like we have the Subgroup Test for groups, we now give a practical test for identifying when a subset of a ring is a subring, known as the Subring Test:

Theorem 24.1 (Subring Test)**.** *Let R be a ring, and let S be a nonempty subset of R. Then S is a subring if and only if the following conditions hold:*

- *$1_R \in S$, where 1_R denotes the multiplicative identity of R (and in this case, 1_R is also the multiplicative identity for S).*
- *For all $a, b \in S$, we have $a - b \in S$.*
- *For all $a, b \in S$, we have $ab \in S$.*

Proof. First, suppose that S is a subset of R satisfying the three conditions laid out above. Then by the second bullet and the Subgroup Test, S is a subgroup of R with respect to addition. Given the third bullet, S is also closed under multiplication, and associativity of that multiplication in S follows from the fact that it holds true in R. Given the first bullet, S has a multiplicative identity, given by 1_R, since $1_R \cdot s = s \cdot 1_R = s$ for all $s \in S$. In particular, by uniqueness of the identity element in a monoid, we know that $1_R = 1_S$, as required. Finally, the distributive laws hold in S because they already hold in the larger set R. Thus if S

satisfies the three conditions above, then it is a subring of R.

Conversely, suppose that S is a subring of R; we show that it satisfies each of the three conditions above. Firstly, S is a subgroup of R with respect to addition, and so the Subgroup Test tells us that $a-b \in S$ for all $a, b \in S$. Secondly, since S is closed under the multiplication on R, we know that $ab \in S$ whenever $a, b \in S$. Finally, it is part of our definition of subring that $1_R = 1_S \in S$, so the first bullet holds true as well. $\square$

Just as we defined the centre of a group, we can also define the centre of a ring, and apply the Subring Test to verify that the centre really is a ring:

Example 24.2. Given any ring R, we define the *centre* of R, $Z(R)$, to be

$$Z(R) = \{z \in R : zr = rz \text{ for all } r \in R\}.$$

Certainly, $Z(R)$ is not empty, since $1 \in Z(R)$ more or less by definition of 1. This also verifies the first condition in the Subring Test. Now, suppose we are given $a, b \in Z(R)$; we wish to show that $a - b \in Z(R)$. We check this argument carefully. Given any $r \in R$, note that

$$(a-b)r = (a+(-b))r = ar + (-b)r = ar + -(br) = ra + -(rb) = ra + r(-b) = r(a+(-b)) = r(a-b).$$

This proves that $a - b \in Z(R)$. Note that this argument also gives a careful proof that $(a-b)r = ar - br$ and $r(a-b) = ra - rb$ for any elements a, b, r in a ring R. This fact will be used often without comment.

Finally, supposing that $a, b \in Z(R)$, given $r \in R$ we have that

$$(ab)r = a(br) = a(rb) = (ar)b = (ra)b = r(ab),$$

proving that $ab \in Z(R)$. By the Subring Test, we now know that $Z(R)$ is a subring of R.

Note also that if R is a commutative ring, then $Z(R) = R$, so the centre is only interesting to study for a non-commutative ring.

Ring Homomorphisms

In order for a function between groups to preserve all the group structure, we asked such a function to preserve the binary operation in the group. When defining a homomorphism between rings, we will similarly require that both binary operations are preserved, and also that the multiplicative identity is preserved by the map:

Definition 24.2. Let R and S be rings. A function $\phi : R \to S$ is called a *ring homomorphism* if the following three conditions hold:

- For all $a, b \in R$, we have $\phi(a+b) = \phi(a) + \phi(b)$.
- For all $a, b \in R$, we have $\phi(ab) = \phi(a)\phi(b)$.
- $\phi(1_R) = 1_S$.

Just as the condition $1_R = 1_S$ in the definition of a subring did not follow from the other parts of the definition, it turns out that $\phi(1_R) = 1_S$ also does not follow from the other parts of a homomorphism definition, and so it must be included. Needless to say, authors who do not assume rings have multiplicative identities will also choose to leave the third bullet point out of their definition of homomorphisms.

We'll give some basic examples of homomorphisms here, and explore others during this week's synchronous session.

Example 24.3. Just as with groups, given any ring R, the *identity mapping* $\iota : R \to R$ given by $\iota(r) = r$ for all $r \in R$ is always a ring homomorphism. There is also a homomorphism from any ring R to the zero ring, called the *zero morphism*, which is the map $\mathbf{0} : R \to \{0\}$ given by $\mathbf{0}(r) = 0$ for all $r \in R$. (Question: if the zero ring is replaced by an arbitrary ring S, is this map necessarily still a homomorphism?)

Example 24.4. We can also consider the reduction modulo n map $\phi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$, given by $\phi(m) = [m]$ for all $m \in \mathbb{Z}$, to be a homomorphism of rings. We previously checked that ϕ is a homomorphism on the additive groups, and it's a quick check to see that this map preserves multiplication and the multiplicative identity.

We conclude this reading by stating several basic properties of ring homomorphisms, mirroring Theorem 21.1 for groups. However, we will not prove any of these statements here, since they can all be directly deduced from, or else adapted from, the corresponding result for groups.

Theorem 24.2. *Suppose $\phi : R_1 \to R_2$ is a homomorphism of rings, and let $r \in R_1$ be arbitrary.*

(1) $\phi(0) = 0$.

(2) $\phi(-r) = -\phi(r)$.

(3) For all $k \in \mathbb{Z}$, we have $\phi(kr) = k\phi(r)$.

(4) For all $n \in \mathbb{N}$, we have $\phi(r^n) = \phi(r)^n$.

(5) If $u \in R$ has a multiplicative inverse, then $\phi(u^k) = \phi(u)^k$ for all $k \in \mathbb{Z}$.

MATH 145 Course Reading 25: Ideals and Quotient Rings

November 13, 2020

Continuing the parallels with our study of groups, we now investigate the quotients of a ring. Since the additive structure of every ring forms an abelian group, it will automatically be possible to form a quotient group for every additive subgroup of the ring. But when do these quotients inherit a ring structure? That's the question we will investigate in this reading. In studying this topic, we'll be quickly led to the notion of an *ideal* of a ring, which plays the same role for rings as normal subgroups do for groups.

Quotient Rings – First Attempts

Let R be a ring. Looking only at the addition operation on R, we have that R is an abelian group. Thus if A is any subgroup of R under addition, a quotient group R/A can be defined. Addition of right cosets is given by $(A+r)+(A+s)=A+(r+s)$ for all $r,s\in R$, and our previous work with group quotients shows that this operation is well-defined and turns R/A into an abelian group.

Naturally, we would wish to make a similar definition for multiplication. We would *like* to set

$$(A+r)(A+s)=A+rs$$

for all cosets $A+r, A+s\in R/A$. But is this well-defined? Suppose we know that $A+r=A+r_1$ and $A+s=A+s_1$ for some $r,s,r_1,s_1\in R$. We would like to say that $(A+r)(A+s)=(A+r_1)(A+s_1)$, or in other words, that $A+rs=A+r_1s_1$. This is equivalent to asking that $rs-r_1s_1\in A$. Thus we need A to have the property that, for any $r,s,r_1,s_1\in R$ for which $r-r_1\in A$ and $s-s_1\in A$, we have $rs-r_1s_1\in A$.

It is not at all clear that all additive subgroups of R will carry this property automatically, and in fact this is not the case in general. For instance, consider the subgroup $\mathbb{Z}$ of the ring $\mathbb{Q}$. Taking $r=\frac{3}{2}$, $r_1=\frac{1}{2}$, $s=\frac{4}{3}$, $s_1=\frac{1}{3}$, notice $r-r_1=s-s_1=1\in\mathbb{Z}$, while

$$rs-r_1s_1=\frac{3}{2}\cdot\frac{4}{3}-\frac{1}{2}\cdot\frac{1}{3}=2-\frac{1}{6}\notin\mathbb{Z}.$$

Perhaps it should not be surprising that not every group quotient R/A can be given a ring structure; after all, A is defined without any reference to the multiplicative structure of R. Furthermore, even in the group context, it was not possible to get a group structure when quotienting by an arbitrary subgroup. There, we needed to work with *normal* subgroups. So how can we find a sub-structure of R for which taking a quotient ring *is* possible?

For inspiration, let's look back at groups for a moment, and prove a fundamental result about the relationship between kernels and normal subgroups:

Theorem 25.1. *Let G be a group.*

(1) If G_1 is any group and $\phi: G\to G_1$ is a group homomorphism, then $\ker\phi$ is a normal subgroup of G.

(2) If H is a normal subgroup of G, then there is a group homomorphism $\phi: G\to G_1$ (for some group G_1) such that $H=\ker\phi$.

Proof. (1) Suppose $\phi: G\to G_1$ is a group homomorphism, and set $K=\ker\phi$. By Theorem 21.2, we already know that K is a subgroup of G; it only remains to prove that K is normal in G. For each $g\in G$, we must show that $gKg^{-1}=K$. So first suppose that $h\in gKg^{-1}$. Then $h=gkg^{-1}$ for some $k\in K$. Observe that

$$\phi(h)=\phi(gkg^{-1})=\phi(g)\phi(k)\phi(g)^{-1}=\phi(g)e\phi(g)^{-1}=\phi(g)\phi(g)^{-1}=e.$$

This shows $h \in \ker\phi = K$, so that $gKg^{-1} \subseteq K$. As usual, taking g^{-1} in place of g shows that $g^{-1}Kg \subseteq K$ for each $g \in G$, and this immediately implies $K \subseteq gKg^{-1}$, so that $K = gKg^{-1}$. This holds for each $g \in G$, so $K = \ker\phi$ is normal in G.

(2) Now suppose H is a normal subgroup of G. Set $G_1 = G/H$, and consider the quotient homomorphism $q : G \to G/H$ given by $q(g) = Hg$ for each $g \in G$. Observe that $g \in \ker q$ if and only if $Hg = He$, if and only if $e^{-1}g \in H$, if and only if $g \in H$. Thus $H = \ker q$, and the proof is complete.

□

Courtesy of Theorem 25.1, we see that the normal subgroups of a group G coincide exactly with the kernels of group homomorphisms with domain G. Since normal subgroups are the "right" object to quotient by when forming a quotient group, it makes sense to define and study the kernels of ring homomorphisms in order to find the "right" object to quotient by when forming a quotient ring.

Definition 25.1. Let R and S be rings, and $\phi : R \to S$ be a group homomorphism. The *kernel* of ϕ, denoted $\ker\phi$, is the kernel of the homomorphism of abelian groups R and S given by ϕ:

$$\ker\phi = \{r \in R : \phi(r) = 0_S\}.$$

By its very construction, the kernel of a ring homomorphism with domain R is automatically an additive subgroup of R, since this is true already of group homomorphisms. But we know we need more structure than an additive subgroup in order to define quotient rings... Do kernels of ring homomorphisms have extra structure?

Ideals

A first guess might be that kernels have to be subrings, but we can quickly see this is not the case. If $\phi : R \to S$ is a ring homomorphism, and S is not the zero ring, then $\phi(1_R) = 1_S \neq 0_S$ (why is $1_S \neq 0_S$?) This shows $1_R \notin \ker\phi$, and so by our definition of subring, $\ker\phi$ cannot be a subring.

However, $\ker\phi$ is indeed closed under multiplication: given a homomorphism $\phi : R \to S$ and $r_1, r_2 \in R$, we see that $\phi(r_1r_2) = \phi(r_1)\phi(r_2) = 0_S \cdot 0_S = 0_S$, so $r_1r_2 \in \ker\phi$. But in fact something stronger is at work: if $r \in \ker\phi$ and $a \in R$ is arbitrary, notice that

$$\phi(ra) = \phi(r)\phi(a) = 0_S \cdot \phi(a) = 0_S,$$

so $ra \in \ker\phi$. Similarly, $ar \in \ker\phi$. Thus $\ker\phi$ is not just closed under multiplication, it actually *absorbs* multiplication, in the sense that multiplying a kernel element by any element of R gives another kernel element.

This motivates the definition of an *ideal* of a ring:

Definition 25.2. Let R be a ring. A subset I of R is called an *ideal* of R if:

- I is a subgroup of the additive group R.
- I absorbs multiplication, meaning that if $r \in I$ and $a \in R$, then ra and ar both belong to I.

Before going any further, we will give a few examples of ideals in rings:

Example 25.1. If R is any ring, then both R and the subset $\{0\}$ are ideals of R. The fact that R is an ideal of R should be clear. To check that $\{0\}$ is an ideal of R, note that it corresponds to the trivial subgroup of R under addition, so it is certainly a subgroup of the additive group R. Furthermore, for any $a \in R$, we have $a \cdot 0 = 0 \cdot a = 0$, so the set $\{0\}$ absorbs multiplication as well. The ideal $\{0\}$ is usually referred to as the *zero ideal* of R.

Example 25.2. For any $n \in \mathbb{N}$, the additive subgroups $n\mathbb{Z} = \{m \in \mathbb{Z} : m = nk \text{ for some } k \in \mathbb{Z}\}$ are ideals of $\mathbb{Z}$. It was noted in Example 21.6 that these are additive subgroups. Now, we check that they absorb multiplication. If $m \in n\mathbb{Z}$ and $\ell \in \mathbb{Z}$, we have $m = nk$ for some integer k. It follows that

$$\ell m = m\ell = (nk)\ell = n(k\ell),$$

where $k\ell \in \mathbb{Z}$. Thus ℓm and $m\ell$ belong to $n\mathbb{Z}$, proving that these sets absorb multiplication.

Example 25.3. More generally, for any commutative ring R and element $a \in R$, we can define the set $Ra = aR = \{s \in R : s = ar \text{ for some } r \in R\}$. It is a good exercise to check that Ra is an ideal of R, called the *principal ideal generated by* a.

As we hoped, if we quotient a ring by an ideal, then we are able to give a natural ring structure to the quotient:

Theorem 25.2. *Let R be a ring, and let I be an ideal of R. Then the set of (right) cosets R/I can be given the structure of a ring, with addition given by $(I + a) + (I + b) = I + (a + b)$ for all $I + a, I + b \in R/I$, and with multiplication given by $(I + a)(I + b) = I + ab$ for all $I + a, I + b \in R/I$.*

Proof. Given the set-up of the theorem, we already know that I is an additive subgroup of R, and since R is abelian, the quotient group R/I is well-defined, with addition given by $(I + a) + (I + b) = I + (a + b)$ for all $I + a, I + b \in R/I$, by Theorem 22.2 of our course readings. We now check that the multiplication operation

$$(I + a)(I + b) = I + ab$$

is well-defined and satisfies the conditions imposed on multiplication in a ring. Suppose we have $a', b' \in R$ such that $I + a = I + a'$ and $I + b = I + b'$. By definition, this means $a - a' \in I$ and $b - b' \in I$. We wish to show that $I + ab = I + a'b'$, which requires that we show $ab - a'b' \in I$. Note that

$$ab - a'b' = ab - a'b + a'b - a'b' = (a - a')b + a'(b - b').$$

Since $a - a' \in I$ and I absorbs multiplication, we have $(a - a')b \in I$. Since $b - b' \in I$ and I absorbs multiplication, $a'(b - b') \in I$. Finally, since I is closed under addition, $(a - a')b + a'(b - b') \in I$. In other words, $ab - a'b' \in I$, proving that multiplication is well-defined.

Certainly, the coset $I + 1$ is the identity with respect to multiplication, since $(I + a)(I + 1) = I + a = (I + 1)(I + a)$ for all $I + a \in R/I$. Also, multiplication in R/I is associative, since

$$((I + a)(I + b))(I + c) = (I + ab)(I + c) = I + (ab)c = I + a(bc) = (I + a)(I + bc) = (I + a)((I + b)(I + c))$$

for all $I + a, I + b, I + c \in R/I$. I leave checking the distributive laws as an exercise for you! □

Another sign that we're on the right track comes from the fact that a version of Theorem 25.1 holds for ideals of ring homomorphisms:

Theorem 25.3. *Let R be a ring.*

(1) If S is any ring and $\phi : R \to S$ is a ring homomorphism, then $\ker \phi$ is an ideal of R.

(2) If I is an ideal of R, then there is a ring R_1 and a ring homomorphism $\phi : R \to R_1$ such that $I = \ker \phi$.

Proof. (1) We already verified in our work above that $\ker \phi$ satisfies all the defining properties of an ideal, for any homomorphism $\phi : R \to S$.

(2) Let I be an ideal of R. We set $R_1 = R/I$, and consider the *quotient mapping* $q : R \to R_1$ given by $q(a) = I + a$ for all $a \in R$. You can check immediately from the definition of the operations in R/I that q is a ring homomorphism, and we claim that $I = \ker q$. Note that $a \in \ker q$ if and only if $q(a) = I + a = I + 0$, which holds if and only if $a \in I$, so that $\ker q = I$ as claimed.

□

Thus ideals of a ring R correspond exactly to kernels of ring homomorphisms with domain R.

We conclude now with some quick examples of quotient rings:

Example 25.4. Let R be an arbitrary ring. If we take $I = R$, then there is only one coset in the quotient R/I, namely the coset $I + 0 = R + 0$. Thus R/R can be identified with the zero ring. On the other hand, if we take $I = \{0\}$, note that we have $I + a = I + b$ if and only if $-b + a \in I = \{0\}$, which is true if and only if $a = b$. Thus coset equality corresponds to actual equality in R, and so $R/\{0\}$ can be identified with R.

Example 25.5. As we saw in Example 25.2, the sets $n\mathbb{Z}$ are ideals of the ring $\mathbb{Z}$ for every $n \in \mathbb{N}$. When $n \geq 1$, note that the coset $n\mathbb{Z} + m$ corresponds to the equivalence class $[m]$ under congruence modulo n, and our multiplication rule $(n\mathbb{Z} + m_1)(n\mathbb{Z} + m_2) = n\mathbb{Z} + m_1 m_2$ corresponds exactly to the multiplication $[m_1][m_2] = [m_1 m_2]$ of equivalence classes we've seen previously. Hence the quotient ring $\mathbb{Z}/n\mathbb{Z}$ coincides exactly with the ring we were calling $\mathbb{Z}/n\mathbb{Z}$ previously in this course, the *ring of integers modulo n*.

MATH 145 Course Reading 26:
Integral Domains and Divisibility

November 16, 2020

Now that we've been introduced to the study of rings, we transition towards applying this knowledge in the realm of elementary number theory. Simply put, elementary number theory seeks to understand algebraic properties of the ring $\mathbb{Z}$, as well as related rings sharing similar properties. In particular, we will be interested only in *integral domains*, which are commutative rings having an additional cancellation property. In this reading, we introduce the definition and basic properties of integral domains, particularly those properties that relate to divisibility. As we go, we will see how those properties apply in $\mathbb{Z}$, while introducing other integral domains for comparison.

Integral Domains – Definition and Examples

Since elementary number theory is all about studying integral domains, we jump straight into the definition:

Definition 26.1. Let R be a commutative ring. An element $a \in R$ is called a *zero divisor* if there is some $b \in R$, $b \neq 0$, for which $ab = 0$. An *integral domain* is a commutative ring R, different from the zero ring, such that 0 is the only zero divisor. In other words, if $ab = 0$ holds in R, then either $a = 0$ or $b = 0$.

Another common, equivalent way to define an integral domain is as a commutative ring in which "cancellation" of elements is possible:

Theorem 26.1. *A commutative ring $R \neq 0$ is an integral domain if and only if, for all $a, b, c \in R$, if $ab = ac$ and $a \neq 0$, then $b = c$.*

Proof. First, suppose R is an integral domain, and that we have elements $a, b, c \in R$ such that $ab = ac$ and $a \neq 0$. Then $ab - ac = 0$, so $a(b - c) = 0$. Since R is an integral domain, this implies $a = 0$ or $b - c = 0$. But $a \neq 0$ by assumption, so $b - c = 0$. In other words, $b = c$ as desired.

Now, suppose R is a commutative ring for which the cancellation property holds, and suppose we have elements $a, b \in R$ such that $ab = 0$. If $a = 0$, we have what we need to show, so suppose $a \neq 0$. Then $ab = 0 = a \cdot 0$ and $a \neq 0$, so the cancellation property tells us $b = 0$, as needed. □

Now, let's take a look at some well-known examples (and a non-example):

Example 26.1. The ring $\mathbb{Z}$ is an integral domain (and in fact could be considered the "prototype" integral domain).

Example 26.2. The commutative rings $\mathbb{Q}$ and $\mathbb{R}$ are integral domains.

In fact, $\mathbb{Q}$ and $\mathbb{R}$ are something *more* than integral domains: they are examples of fields. Formally:

Definition 26.2. A ring F is called a *field* if it is commutative, and if every non-zero element of F has a multiplicative inverse in F. In other words, for each $a \in F$ with $a \neq 0$, there is $b \in F$ such that $ab = 1$ (and as with groups, we write a^{-1} for the element b, the unique multiplicative inverse of a).

Certainly $\mathbb{Q}$ is a field according to this definition, with the multiplicative inverse of $\frac{a}{b}$ being $\frac{b}{a}$ if $a \neq 0$. Likewise, for every nonzero $r \in \mathbb{R}$, the real number $\frac{1}{r}$ is also defined, showing that $\mathbb{R}$ is a field. On the other hand, $\mathbb{Z}$ is not a field, since the only integers with multiplicative inverses in $\mathbb{Z}$ are 1 and -1. The following result shows that fields are excellent sources of integral domains:

Theorem 26.2. *Every subring of a field is an integral domain. In particular, every field is an integral domain.*

Proof. Let F be a field, and let R be a subring of F. Since multiplication in R is defined the same as in F, and multiplication in F is commutative, we know that R is a commutative ring. Now, suppose we have $a, b \in R$ such that $ab = 0$ in R. Then this equation also holds in F. If $a \neq 0$, then a^{-1} exists in F by definition of a field, and if we multiply both sides of $ab = 0$ by a^{-1}, we get $b = 0$. This proves that if $ab = 0$ in R, then $a = 0$ or $b = 0$. □

In particular, since $\mathbb{Z}$ is a subring of $\mathbb{Q}$, the ring $\mathbb{Z}$ is an integral domain, as claimed. Now, we proceed to give an important non-example of an integral domain:

Example 26.3. If $n \geq 2$ is a composite integer, then $\mathbb{Z}/n\mathbb{Z}$ is not an integral domain. Indeed, if n admits a factorization $n = ab$, where $1 < a, b < n$, then $[a]$ and $[b]$ are nonzero elements of $\mathbb{Z}/n\mathbb{Z}$, such that $[a][b] = [ab] = [0]$.

We will later see that if n is prime, then $\mathbb{Z}/n\mathbb{Z}$ *is* an integral domain (in fact, a field). However, in order to get there, we will need to build up some additional number theory first.

For the moment, we introduce one further example of an integral domain, one which parallels the properties of $\mathbb{Z}$ very closely:

Example 26.4. Define the ring of *Gaussian integers* to be $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$, with addition given by $(a+bi)+(c+di) = (a+c)+(b+d)i$, and multiplication given by $(a+bi)(c+di) = (ac-bd)+(ad+bc)i$. Those who have seen the complex numbers $\mathbb{C}$ before will recognize the addition and multiplication operations here are exactly the same as in $\mathbb{C}$, and in fact, $\mathbb{Z}[i]$ could be described simply as the subring of $\mathbb{C}$ consisting of all complex numbers with integer real part and integer imaginary part. If you haven't seen complex numbers before, don't worry: we'll define and study them more carefully later in the course.

For calculation purposes though, it's enough to know that Gaussian integers can be added and multiplied like binomials $a + bx$ which you are used to working with in high school, but subject to the additional rule that $i^2 = -1$. So for example:

$$(2 + i) + (3 + 2i) = 5 + 3i$$
$$(2 + i)(3 + 2i) = 6 + 3i + 4i + 2i^2 = 6 + 7i - 2 = 4 + 7i.$$

The quickest way to prove that $\mathbb{Z}[i]$ is an integral domain is to prove that $\mathbb{C}$ is a field (which we'll do later in the course), and then argue that $\mathbb{Z}[i]$ is a subring of $\mathbb{C}$. Since we haven't worked with $\mathbb{C}$ yet, we'll just take this fact for granted for now.

Basic Properties of Integral Domains

Armed with some basic examples of integral domains, let's now prove some important ring-theoretic facts about them. The first result restricts the possible characteristics of an integral domain, providing an efficient way to prove that certain rings are *not* integral domains:

Theorem 26.3. *If R is an integral domain, then* $\operatorname{char} R = 0$ *or* $\operatorname{char} R$ *is prime.*

Proof. Suppose that R is an integral domain, and suppose to the contrary that $\operatorname{char} R \neq 0$ and $\operatorname{char} R$ is not prime. Then either $\operatorname{char} R = 1$ or $\operatorname{char} R$ is a positive composite integer. If $\operatorname{char} R = 1$, then R is the zero ring (why?), contrary to the definition of integral domain. So suppose $\operatorname{char} R = n$, where n is composite. Then we can write $n = ab$, where $1 < a, b < n$. Then the elements $r = a \cdot 1_R$ and $s = b \cdot 1_R$ are non-zero elements of R, but $rs = n \cdot 1_R = 0$, contradicting that R is an integral domain.

Thus, if R is an integral domain, then either $\operatorname{char} R = 0$ or $\operatorname{char} R$ is prime. □

Note that this theorem re-proves that $\mathbb{Z}/n\mathbb{Z}$ is not an integral domain if $n \geq 2$ is composite, since $\operatorname{char} \mathbb{Z}/n\mathbb{Z} = n$. In fact, the proof of this theorem is just a generalization of the argument in Example 26.3.

Here is one other major ring-theoretic result, which makes use of some of the facts about injective functions, surjective functions, and set cardinality we discussed earlier in the course:

Theorem 26.4. *Every finite integral domain is a field.*

Proof. Let R be an integral domain, and suppose R has finitely many elements, say $|R| = n$ for some positive integer n. Let a be a non-zero element of R, and consider the multiplication map $\phi : R \to R$ given by $\phi(r) = ar$ for each $r \in R$. Note that ϕ is an injective function: if $\phi(r) = \phi(s)$ for some $r, s \in R$, then $ar = as$. Since $a \neq 0$ and R is an integral domain, we can use the cancellation property to get $r = s$, proving injectivity.

Thus we have constructed an injective function $\phi : R \to R$. Since R is a finite set, this implies ϕ is surjective as well. Indeed, given that ϕ is injective, the cardinality of $\phi(R)$ is equal to n, and $\phi(R)$ is a subset of the n-element set R, so $\phi(R) = R$. In particular, by surjectivity, there must be some $b \in R$ for which $\phi(b) = 1$, which says $ab = ba = 1$. Thus a has a multiplicative inverse in R. Given that $a \neq 0$ was arbitrary, we conclude that R is a field. □

Divisibility and Associates

At last, we are ready to define what divisibility means in an arbitrary integral domain, generalizing the definition we've already given for $\mathbb{Z}$:

Definition 26.3. Let R be an integral domain, and let $a, b \in R$. We say that a *divides* b, and write $a \mid b$, if there is some $c \in R$ such that $b = ac$.

For example, 5 divides both 10 and -10 in $\mathbb{Z}$, since $10 = 5 \cdot 2$ and $-10 = 5 \cdot (-2)$. On the other hand, 3 does not divide 5, which we can write as $3 \nmid 5$, since there is no integer b for which $5 = 3b$. For a Gaussian integer example, $2 + i$ divides 5 in $\mathbb{Z}[i]$, since $5 = (2 + i)(2 - i)$. Also, in any field F, if $a \neq 0$, then $a \mid b$ for all $b \in F$, since $b = a(a^{-1}b)$.

Next, we verify that the divisibility relation is reflexive and transitive, while also proving a further useful property of divisibility:

Proposition 26.1. *Let R be an integral domain. Then:*

(1) For all $a \in R$, we have $a \mid a$.

(2) If $a, b, c \in R$ are such that $a \mid b$ and $b \mid c$, then $a \mid c$.

(3) If $a, b, c \in R$ are such that $a \mid b$ and $a \mid c$, then $a \mid (bx + cy)$ for all $x, y \in R$.

Proof. (1) Since $a = a \cdot 1$ for all $a \in R$, we have $a \mid a$ for all $a \in R$.

(2) Given that $a \mid b$, we know $b = ak$ for some $k \in R$. Similarly, since $b \mid c$, we know $c = b\ell$ for some $\ell \in R$. Thus

$$c = b\ell = (ak)\ell = a(k\ell),$$

where $k\ell \in R$. Thus $a \mid c$, as needed.

(3) Given that $a \mid b$ and $a \mid c$, we have $b = ak$ and $c = a\ell$ for some $k, \ell \in R$. Then for any $x, y \in R$ we have

$$bx + cy = (ak)x + (a\ell)y = a(kx + \ell y),$$

where $kx + \ell y \in R$. Thus $a \mid (bx + cy)$, as needed.

□

As a consequence of the first two points in the theorem, we can define a reflexive, transitive relation $\mid$ on any integral domain R, according to the definition above. Now recall the result of Question 3 on Assignment 4, which allows us to take the relation $\mid$ and define an equivalence relation $\sim$ on R by declaring that $a \sim b$ if $a \mid b$ and $b \mid a$.

If $a \sim b$, then we say that a and b are *associate* in R. Furthermore, that assignment problem tells us that we can define an order relation on the set of equivalence classes under $\sim$ by declaring that $[a]_\sim \mid [b]_\sim$ if and

only if $a \mid b$. This relation is well-defined, no matter what representatives we choose for each equivalence class, by the work we did on Assignment 4.

What do the equivalence classes look like under this associate relation? The next theorem answers that question. First, we make a definition:

Definition 26.4. Let R be a ring. An element $r \in R$ is called a *unit* of R if r has a multiplicative inverse in R. The set of all units of R is denoted R^*, and coincides with the group of units of the monoid R under multiplication (see Theorem 19.1).

Now, here is the result:

Theorem 26.5. *Let R be an integral domain. Given elements $a, b \in R$, we have $a \sim b$ if and only if $a = ub$ for some unit $u \in R^*$.*

Proof. First, suppose that $a \sim b$ in R. Then $a \mid b$ and $b \mid a$, so there are $k, \ell \in R$ such that $b = ak$ and $a = b\ell$. Putting these together, we find

$$b = ak = (b\ell)k = b(\ell k).$$

If $b = 0$, then $a = b\ell = 0\ell = 0$, and so $a = b = 1 \cdot b$, where $1 \in R^*$. Otherwise, if $b \neq 0$, then $b \cdot 1 = b(\ell k)$ from the equation above, so $1 = \ell k$ by the cancellation property. This proves that $\ell \in R^*$, so that $a = \ell b$ where $\ell \in R^*$.

Conversely, suppose that $a = ub$ for some $u \in R^*$. From this equation, $b \mid a$ immediately follows. Multiplying both sides by u^{-1} gives $b = u^{-1}a$, so that $a \mid b$. This shows $a \sim b$, completing the proof. □

Applying this theorem to the integral domain $\mathbb{Z}$, since $\mathbb{Z}^* = \{1, -1\}$, we see that $a \sim b$ in $\mathbb{Z}$ if and only if $a = b$ or $a = -b$.

MATH 145 Course Reading 27: Elementary Number Theory – Division with Remainder and Greatest Common Divisor

November 18, 2020

In the previous reading, we introduced a special class of rings known as integral domains. Ultimately, our goal in number theory is to study the integers and other "integer-like" rings, and so we now proceed toward a more detailed study of $\mathbb{Z}$. As we will see, the integral domain $\mathbb{Z}[i]$ will share essentially all the same properties, and so we will develop the theory for $\mathbb{Z}[i]$ in tandem with the theory for $\mathbb{Z}$.

Division with Remainder in $\mathbb{Z}$

Even though exact division of one integer by another is not always possible, we can always perform such a division if we allow ourselves a small remainder. Dividing integers with remainder is something you may have done when learning long division for the first time. We now provide a rigorous proof that this procedure works, and that the quotient and remainder of the division are unique:

Theorem 27.1. *Let a and b be integers, with $b > 0$. Then there exist unique integers q and r, with $0 \leq r < b$, such that $a = bq + r$. The integer q is called the* quotient *of the division, and the integer r is called the* remainder *of the division.*

Proof. First, we treat the case where $a \geq 0$. This is a direct application of the well-ordering principle for $\mathbb{N}$. To show that q and r exist, consider the set $S = \{n \in \mathbb{N} : n = a - bq \text{ for some } q \in \mathbb{Z}\}$. Note that S is non-empty, since $a = a - b(0) \in \mathbb{N}$. By the Well-Ordering Principle, S has a least element, which we call r. Then $r = a - bq$ for some integer q by construction, meaning $a = bq + r$. We also claim that $0 \leq r < b$. Certainly $r \geq 0$, since r is a natural number. If we suppose to the contrary that $r \geq b$, then $r - b \geq 0$, and $r - b = a - bq - b = a - b(q + 1)$, so that $r - b \in S$. This contradicts minimality of r. Thus $0 \leq r < b$ as claimed. This completes the proof of existence if $a \geq 0$.

If a is negative, then $-a > 0$, so the first part of the proof gives us $q_0, r_0 \in \mathbb{Z}$ such that $-a = bq_0 + r_0$, and $0 \leq r_0 < b$. If $r_0 = 0$, then $a = b(-q_0) + 0$, so that $q = -q_0$ and $r = 0$ satisfy the conditions of the theorem. If $r_0 \neq 0$, we write $a = b(-q_0) - r_0 = b(-q_0 - 1) + (b - r_0)$, where $-q_0 - 1, b - r_0 \in \mathbb{Z}$ and $0 < b - r_0 < b$, so that $q = -q_0 - 1$ and $r = b - r_0$ satisfy the conditions of the theorem. This completes the proof of existence if $a < 0$.

To show uniqueness, suppose we also have integers q', r' such that $a = bq' + r'$ and $0 \leq r' < b$. Equating the two expressions for a, we get

$$bq + r = bq' + r'.$$

Thus $r' - r = bq - bq' = b(q - q')$. If $q = q'$, this equation forces $r = r'$, and the uniqueness proof is complete. If $q \neq q'$, then taking absolute values of both sides and using that $|q - q'| \geq 1$, we get

$$|r' - r| = |b||q - q'| \geq b.$$

But r' and r are both natural numbers strictly less than b, so $|r' - r| \geq b$ is impossible. We conclude that $q = q'$ after all, and therefore the integers q and r constructed above are unique. □

How do we compute the integers q and r in the theorem, in practice? For simplicity, let's assume that $a \geq 0$. If working by hand, any of the old familiar methods for performing long division will work. But if you have a calculator handy, simply punch in $\frac{a}{b}$. The output will have an integer part and a part after the decimal point, say $\frac{a}{b} = q + x$, where $q \in \mathbb{N}$ and $0 \leq x < 1$. Then $a = bq + bx$ with $0 \leq bx < b$, so by uniqueness in division with remainder, $r = bx$ is the remainder and q is the quotient. So q can be obtained as the integer before the decimal point in $\frac{a}{b}$, and r is obtained by multiplying the quantity after the decimal point by b.

To illustrate, suppose we wish to divide 1009 by 33 with remainder. Pulling out a calculator, we get $\frac{1009}{33} = 30.5757575\ldots$. Thus $q = 30$, and $r = 33 \cdot (0.57575\ldots) = 19$. This yields the decomposition $1009 = 33 \cdot 30 + 19$ that we seek.

Division with Remainder in $\mathbb{Z}[i]$

It turns out that a similar process can be carried out for Gaussian integers too, though we have to give up any claims of uniqueness of the quotient and remainder. Before we can do this, we have to decide on a notion of "size" for our Gaussian integers, since we would like the remainder of the division to be smaller than the number we are dividing by. This motivates the following definition:

Definition 27.1. For any Gaussian integer $a + bi$, we define the *norm* of $a + bi$, written $N(a + bi)$, to be the quantity $a^2 + b^2 \in \mathbb{N}$.

For those familiar with complex absolute value, $N(a + bi)$ is nothing more than $|a + bi|^2$ in the ring $\mathbb{C}$. We will study complex absolute value in this course, and one property we will show is the familiar fact from real absolute values that $|zw| = |z| \cdot |w|$ for all $z, w \in \mathbb{C}$. This immediately implies that $N(\alpha\beta) = N(\alpha)N(\beta)$ for all $\alpha, \beta \in \mathbb{Z}[i]$.

To motivate how division with remainder will work in $\mathbb{Z}[i]$, let's take a concrete example. Suppose we wished to divide $2 + i$ by $1 + i$ with remainder. In other words, we want to find a quotient γ and remainder δ in $\mathbb{Z}[i]$ such that $(2 + i) = (1 + i)\gamma + \delta$, and with $0 \leq N(\delta) < N(1 + i)$, to capture the fact that the remainder is smaller than the number $1 + i$ that we are dividing by.

Let's take our cue from the "calculator" solution to division with remainder in $\mathbb{Z}$. Let's first calculate the quotient $\frac{2+i}{1+i}$ in the complex numbers, and then "round" the result to a quotient in $\mathbb{Z}[i]$. The way to simplify the quotient $\frac{2+i}{1+i}$ is to multiply both the numerator and denominator by the *conjugate* $1 - i$ of the denominator, obtained by flipping the sign on the coefficient of i. This leads to

$$\frac{2+i}{1+i} = \frac{(2+i)(1-i)}{(1+i)(1-i)} = \frac{2-i-i^2}{1-i^2} = \frac{3-i}{2} = \frac{3}{2} - \frac{1}{2}i.$$

To round this quotient to the nearest Gaussian integer, we now have four choices. We can round $\frac{3}{2}$ up to 2, or down to 1. Similarly, we can round $-\frac{1}{2}$ up to 0, or down to -1. To make a convenient choice, let's round to $\gamma = 2 + 0i = 2$. Then we calculate our remainder δ as

$$(2 + i) - (1 + i)\gamma = (2 + i) - (1 + i)(2) = (2 + i) - (2 + 2i) = -i.$$

This gives the decomposition $(2 + i) = (1 + i)(2) + (-i)$, with $\gamma = 2$ and $\delta = -i$. Note that $N(-i) = 0^2 + (-1)^2 = 1$, and this is smaller than $N(1 + i) = 1^2 + 1^2 = 2$, as needed. It only remains to show that this procedure works in general:

Theorem 27.2. *Let α and β be Gaussian integers, with $\beta \neq 0$. Then there exist $\gamma, \delta \in \mathbb{Z}[i]$ such that $\alpha = \beta\gamma + \delta$, and $0 \leq N(\delta) < N(\beta)$.*

Proof. Write $\alpha = a + bi$ and $\beta = c + di$, for some integers a, b, c, d, with c and d not both zero. Performing division in $\mathbb{C}$, we get

$$\frac{\alpha}{\beta} = \frac{a+bi}{c+di} = \frac{(a+bi)(c-di)}{(c+di)(c-di)} = \frac{(ac+bd)+(bc-ad)i}{c^2+d^2} = r + si,$$

where $r = \frac{ac+bd}{c^2+d^2} \in \mathbb{Q}$ and $s = \frac{bc-ad}{c^2+d^2} \in \mathbb{Q}$. Now choose integers m, n such that $|m - r| \leq \frac{1}{2}$ and $|n - s| \leq \frac{1}{2}$. (Why do such integers exist?) Set $\gamma = m + ni \in \mathbb{Z}[i]$, and set $\delta = \alpha - \beta\gamma$. Note $\delta \in \mathbb{Z}[i]$ since $\mathbb{Z}[i]$ is a ring, and thus closed under multiplication, addition, and additive inverses.

Certainly, we have $\alpha = \beta\gamma + \delta$, and it only remains to show that $0 \leq N(\delta) < N(\beta)$. The inequality $0 \leq N(\delta)$ follows automatically from the definition of norms. For the other inequality, note that

$$\delta = \alpha - \beta\gamma = \beta(r + si) - \beta(m + ni) = \beta((r - m) + (s - n)i).$$

Taking complex absolute values and squaring, we get

$$N(\delta) = |\delta|^2 = |\beta((r-m)+(s-n)i)|^2 = |\beta|^2((r-m)^2+(s-n)^2) = N(\beta)((r-m)^2+(s-n)^2).$$

Now, since $|r-m| \leq \frac{1}{2}$ and $|s-n| \leq \frac{1}{2}$, we know that $(r-m)^2 \leq \frac{1}{4}$ and $(s-n)^2 \leq \frac{1}{4}$. Adding up, we get $(r-m)^2+(s-n)^2 \leq \frac{1}{4}+\frac{1}{4} = \frac{1}{2}$, and so the equation above implies $N(\delta) \leq N(\beta) \cdot \frac{1}{2} < N(\beta)$, as we needed to show. □

As mentioned earlier, we lose uniqueness in our division with remainder now. For instance, in the example given before Theorem 27.2, we could have chosen any one of four possible values of γ, each of which would have led to a valid choice of remainder δ.

What if we wanted to study the general situation of integral domains having division with remainder? What should that mean? We now give the formal definition:

Definition 27.2. Let R be an integral domain. We say that R *has a division algorithm* if there exists a function $d : R \setminus \{0\} \to \mathbb{N}$, called a *divisor function*, such that for any $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ such that

$$a = bq + r,$$

and either $d(r) < d(b)$, or else $r = 0$.

For instance, Theorem 27.2 proves that $\mathbb{Z}[i]$ has a division algorithm with divisor function $d(\alpha) = N(\alpha)$. Our work in Theorem 27.1 essentially shows that $\mathbb{Z}$ has a division algorithm, with divisor function $d(a) = |a|$. The only case not covered by the theorem in the generality required by Definition 27.2 is the case $b < 0$ (can you work out this missing case for yourself?)

We will explore an additional example of an integral domain with a division algorithm in Assignment 9, but you will find it has a very similar feel to $\mathbb{Z}$ and $\mathbb{Z}[i]$.

Greatest Common Divisor

In the ring $\mathbb{Z}$, the notion of greatest common divisor is fairly intuitive once we have a definition of divisibility. Given a pair of integers a and b, we seek to find an integer dividing both a and b, and to choose the *largest* such integer with this property. But what if we wanted to give a definition that works in more general integral domains, like $\mathbb{Z}[i]$? In $\mathbb{Z}$, we are lucky enough to have a total ordering on the ring elements, but in general, there may be no ordering at all, so that the idea of *greatest* common divisor is harder to specify.

However, the existence of a divisibility relation on our integral domains means there *is* an order relation handy: the one that comes from divisibility. As mentioned in the previous reading, divisibility puts a partial order relation on the equivalence classes of an integral domain with respect to the associate relation $\sim$. Once we have this order relation, the idea of *greatest* common divisor now makes sense. This motivates our general definition:

Definition 27.3. Let R be an integral domain, and let $a, b \in R$ be arbitrary elements, not both zero. An element $d \in R$ is called a *greatest common divisor* (gcd) of a and b if d satisfies the following two properties:

(1) $d \mid a$ and $d \mid b$.

(2) If $e \in R$ is another common divisor of a and b, so that $e \mid a$ and $e \mid b$, then $e \mid d$.

Your intuitive experience of working with greatest common divisors in $\mathbb{Z}$ may suggest to you that the greatest common divisor of two elements should be unique. However, we will prove in a moment that this is not quite correct: the gcd of two elements picks out a unique equivalence class with respect to the associate relation $\sim$. Furthermore, there are integral domains with two nonzero elements having no greatest common divisor, as we will explore on the assignments. What we *can* say about gcds in general is captured in the following theorem:

Theorem 27.3. *Let R be an integral domain, and let a and b be elements of R, not both zero. If d_1 and d_2 are both greatest common divisors of a and b, then $d_1 \sim d_2$. Conversely, if d_1 is a greatest common divisor of a and b and $d_2 \in R$ is such that $d_2 \sim d_1$, then d_2 is a gcd of a and b.*

Proof. Given the set-up of the theorem, suppose d_1 and d_2 are both gcds of a and b. Since d_2 is a common divisor of a and b, and d_1 is a *greatest* common divisor, we must have $d_2 \mid d_1$ by definition. By symmetry, since d_1 is a common divisor of a and b, and d_2 is a *greatest* common divisor, we get $d_1 \mid d_2$. By definition of the associate relation on R, we conclude that $d_1 \sim d_2$.

Now assume that d_1 is a gcd of a and b, and that $d_2 \sim d_1$. Then $d_1 \mid d_2$ and $d_2 \mid d_1$. Since $d_1 \mid a$ and $d_1 \mid b$ and $d_2 \mid d_1$, the transitivity of divisibility gives us $d_2 \mid a$ and $d_2 \mid b$, so that d_2 is a common divisor of a and b. Furthermore, if e is a common divisor of a and b, then by definition $e \mid d_1$. Since $d_1 \mid d_2$, transitivity gives us $e \mid d_2$. Thus d_2 satisfies the definition of a gcd of a and b. □

So, as we stated, even though the gcd of two elements is not unique in an integral domain R (if a gcd exists at all), it at least picks out a single equivalence class in $R/\sim$. By Theorem 26.5, this means that if d is a gcd of a and b, then all gcds of a and b take the form ud for some suitably chosen unit $u \in R^*$. Thus we can choose to adopt $\gcd(a, b)$ as notation for this unique equivalence class of gcds.

For instance, in $\mathbb{Z}$, if d is one gcd of a and b, then $-d$ is the only other gcd. By insisting on taking a *positive* gcd when possible, we can recover the familiar uniqueness of gcds in $\mathbb{Z}$. We will explore the notion of gcd in $\mathbb{Z}$ more carefully in the next reading, when we introduce the Euclidean algorithm for calculating gcds.

MATH 145 Course Reading 28:
The Euclidean Algorithm

November 20, 2020

Now that we've defined what it means for an arbitrary integral domain to possess a division algorithm, we explore one of the main functions of a division algorithm: efficiently calculating gcds. As you will see on Assignment 8, the gcd of two arbitrary elements of an integral domain does not necessarily exist. However, when a division algorithm is available, it can be iterated to yield a computationally efficient procedure for calculating gcds of two elements. In particular, if an integral domain has a division algorithm, then the gcd of two elements (not both zero) necessarily exists.

Euclidean Algorithm: The Procedure

The key lemma that makes the Euclidean algorithm work can be stated as follows:

Lemma 28.1. *Let R be an integral domain, and suppose we have elements $a, b, q, r \in R$ such that $a = bq + r$. Then an element $d \in R$ is a gcd of a and b if and only if it is a gcd of b and r. In other words,* $\gcd(a, b) \sim \gcd(b, r)$.

Proof. Given the set-up of the lemma, suppose d is a gcd of a and b. In particular, $d \mid a$ and $d \mid b$. It follows right away that $d \mid a \cdot 1 + b \cdot (-q)$, by part (3) of Proposition 26.1. In other words, $d \mid r$, so that d is a common divisor of b and r. Now suppose e is a common divisor of b and r. Thus $e \mid b$ and $e \mid r$. Again by Proposition 26.1, we get that $e \mid b \cdot q + r \cdot 1$, i.e. $e \mid a$. So e is a common divisor of a and b, which implies $e \mid d$ by definition of d as a gcd of a and b. This proves that d satisfies the conditions for being a gcd of b and r. The proof that a gcd of b and r is also a gcd of a and b is analogous. □

Now let R be an integral domain with a division algorithm, and suppose D denotes the divisor function in this case. Let's see how to use Lemma 28.1 to calculate gcds of any two elements $a, b \in R$, not both zero.

Since the definition of $\gcd(a, b)$ is symmetric in a and b, we may swap a and b if needed in order to arrange that $b \neq 0$. Then we carry out a division with remainder:

$$a = bq_0 + r_1,$$

where $q_0, r_1 \in R$ and either $r_1 = 0$ or $D(r_1) < D(b)$. If $r_1 = 0$, then $b \mid a$, and from this you can quickly check that b is a gcd of a and b (do this!). Otherwise, we can apply Lemma 28.1 to deduce that $\gcd(a, b) \sim \gcd(b, r_1)$. Hence, we are reduced to the task of calculating a gcd of b and r_1. The advantage here is that $D(r_1) < D(b)$, so that the divisor function is decreasing.

Next, we perform division with remainder of b by r_1, to get

$$b = r_1 q_1 + r_2,$$

where $q_1, r_2 \in R$ and either $r_2 = 0$ or $D(r_2) < D(r_1)$. Again, if $r_2 = 0$, then r_1 is a gcd of b and r_1 (and hence a gcd of a and b). Otherwise, Lemma 28.1 tells us that $\gcd(b, r_1) \sim \gcd(r_1, r_2)$, and we are reduced to calculating a gcd of r_1 and r_2. Again, since $D(r_2) < D(r_1)$, our divisor function is decreasing. From here, we continue to repeat the process until arriving at a zero remainder.

So to summarize, given $a, b \in R$, we perform a succession of divisions with remainder:

$$\begin{aligned} a &= bq_0 + r_1 \\ b &= r_1 q_1 + r_2 \\ r_1 &= r_2 q_2 + r_3 \\ &\vdots \\ r_{n-2} &= r_{n-1} q_{n-1} + r_n \\ r_{n-1} &= r_n q_n + 0, \end{aligned}$$

where r_n is the last non-zero remainder obtained. We must eventually get a remainder of 0, for we have $D(b) > D(r_1) > D(r_2) > D(r_3) > \cdots$, which is a strictly decreasing sequence of natural numbers. This sequence cannot go on forever, so we must eventually arrive at a zero remainder.

Furthermore, repeated applications of Lemma 28.1 give us $\gcd(a, b) \sim \gcd(b, r_1) \sim \gcd(r_1, r_2) \sim \cdots \sim \gcd(r_{n-1}, r_n)$. And since the last division tells us $r_n \mid r_{n-1}$, we know that r_n is a gcd of r_n and r_{n-1}, and thus a gcd of a and b.

This discussion provides a constructive proof of the following theorem:

Theorem 28.1. *Let R be an integral domain with a division algorithm. Given any two elements $a, b \in R$, not both zero, $\gcd(a, b)$ exists.*

The procedure outlined above is called the *Euclidean algorithm* (as you might have guessed).

Euclidean Algorithm: Examples

Let's give a couple examples of this algorithm in practice: one in $\mathbb{Z}$, and one in $\mathbb{Z}[i]$:

Example 28.1. Let's compute $\gcd(1009, 33)$ in $\mathbb{Z}$. Repeatedly performing divisions with remainder, we end up with

$$\begin{aligned} 1009 &= 33 \cdot 30 + 19 \\ 33 &= 19 \cdot 1 + 14 \\ 19 &= 14 \cdot 1 + 5 \\ 14 &= 5 \cdot 2 + 4 \\ 5 &= 4 \cdot 1 + 1 \\ 4 &= 1 \cdot 4 + 0. \end{aligned}$$

Thus the last non-zero remainder, 1, is a gcd of 1009 and 33 (in fact, the unique positive such remainder).

Example 28.2. Let's compute $\gcd(17, 2 + 9i)$ in $\mathbb{Z}[i]$. This time, we'll show our work in each division with remainder step.

First, we must divide 17 by $2 + 9i$ with remainder. We get

$$\frac{17}{2+9i} = \frac{17(2-9i)}{(2+9i)(2-9i)} = \frac{34-153i}{2^2+9^2} = \frac{34-153i}{85} = \frac{34}{85} - \frac{153}{85}i.$$

Rounding both coefficients of the final result to the nearest integer, we get a quotient of $-2i$, so the remainder is $17 - (2+9i)(-2i) = 17 - (18 - 4i) = -1 + 4i$. Thus our first division with remainder is

$$17 = (2+9i)(-2i) + (-1+4i).$$

As a sanity check, you can verify that $N(-1+4i) = 17 < 85 = N(2+9i)$, as required.

Now, we must divide $2+9i$ by $-1+4i$ with remainder. This leads to

$$\frac{2+9i}{-1+4i} = \frac{(2+9i)(-1-4i)}{(-1+4i)(-1-4i)} = \frac{34-17i}{17} = 2-i.$$

In particular, the division is exact, and the second division with remainder is thus

$$2+9i = (-1+4i)(2-i)+0.$$

Since we've hit a remainder of 0, the last non-zero remainder, $-1+4i$, is a gcd of 17 and $2+9i$.

Extended Euclidean Algorithm

It turns out that the Euclidean algorithm is also useful for solving linear equations over the integral domain we're working with. To be precise, suppose we are given an integral domain R with a division algorithm, and two nonzero elements a and b in R. Suppose the Euclidean algorithm gives us d as a gcd of a and b. We now wish to find elements $x, y \in R$ such that $ax + by = d$. We can accomplish this by a *back-substitution process*, as described below.

Suppose that after running the Euclidean algorithm on a and b, we generate divisions with remainder

$$\begin{aligned} a &= bq_0 + r_1 \\ b &= r_1q_1 + r_2 \\ r_1 &= r_2q_2 + r_3 \\ &\vdots \\ r_{n-2} &= r_{n-1}q_{n-1} + r_n, \end{aligned}$$

where we've left out the last step, the one with a zero remainder. In particular, r_n is the gcd d mentioned in the setup above.

We now reverse the order of the equations, and isolate the remainder in each one:

$$\begin{aligned} r_n &= r_{n-2} - r_{n-1}q_{n-1} \\ r_{n-1} &= r_{n-3} - r_{n-2}q_{n-2} \\ &\vdots \\ r_2 &= b - r_1q_1 \\ r_1 &= a - bq_0. \end{aligned}$$

By substituting the right-hand side of the second equation in place of r_{n-1} in the first equation, we can write r_n in the form $r_{n-2}x + r_{n-3}y$ for some $x, y \in R$. Then using the right-hand side of the third equation, we can replace r_{n-2} with a combination of r_{n-3} and r_{n-4} in the expression $r_{n-2}x + r_{n-3}y$, resulting in an expression for r_n in the form $r_{n-3}x + r_{n-4}y$ for some $x, y \in R$. We keep repeating this procedure until we get to the last equation, at which point we will have written r_n in the form $ax + by$ for some $x, y \in R$, as desired.

This description is made much clearer with examples, so let's build on the two that we started earlier:

Example 28.3. We found that 1 is a gcd of 1009 and 33 in $\mathbb{Z}$, and now we wish to find $x, y \in \mathbb{Z}$ such that $1009x + 33y = 1$. We take the divisions with remainder that we used as intermediate computations, reverse the order of the equations, and solve for each non-zero remainder:

$$1 = 5 - 4 \cdot 1 \tag{1}$$
$$4 = 14 - 5 \cdot 2 \tag{2}$$
$$5 = 19 - 14 \cdot 1 \tag{3}$$
$$14 = 33 - 19 \cdot 1 \tag{4}$$
$$19 = 1009 - 33 \cdot 30. \tag{5}$$

Now we take equation (2) and substitute its right-hand side in place of 4 in equation (1):

$$1 = 5 - (14 - 5 \cdot 2) \cdot 1 = 5 - (14 - 5 \cdot 2) = 5 \cdot 3 - 14 \cdot 1.$$

This expresses 1 as a combination of the remainders 5 and 14. Next, we take equation (3) and substitute its right-hand side in place of 5 above:

$$1 = 5 \cdot 3 - 14 \cdot 1 = (19 - 14 \cdot 1) \cdot 3 - 14 \cdot 1 = 19 \cdot 3 - 14 \cdot 3 - 14 \cdot 1 = 19 \cdot 3 - 14 \cdot 4.$$

This now expresses 1 as a combination of the remainders 19 and 14. Now we take equation (4) and replace 14 with its right-hand side:

$$1 = 19 \cdot 3 - 14 \cdot 4 = 19 \cdot 3 - (33 - 19 \cdot 1) \cdot 4 = 19 \cdot 3 - 33 \cdot 4 + 19 \cdot 4 = 19 \cdot 7 - 33 \cdot 4.$$

Now we have 1 as a combination of 19 and 33. Finally, we use equation (5) to replace 19 with an expression in terms of 1009 and 33:

$$1 = 19 \cdot 7 - 33 \cdot 4 = (1009 - 33 \cdot 30) \cdot 7 - 33 \cdot 4 = 1009 \cdot 7 - 33 \cdot 210 - 33 \cdot 4 = 1009 \cdot 7 - 33 \cdot 214.$$

Thus the equation $1009x + 33y = 1$ has a solution $x = 7$ and $y = -214$ over the integers.

Example 28.4. In our Gaussian integer example, the relevant equation asks us to find $x, y \in \mathbb{Z}[i]$ such that $17x + (2 + 9i)y = (-1 + 4i)$, since $-1 + 4i$ is the gcd we obtained. Since our Euclidean algorithm work only had one non-zero remainder, the solution follows immediately from re-writing that equation with the remainder isolated:

$$(-1 + 4i) = 17 - (2 + 9i)(-2i) = 17 \cdot 1 + (2 + 9i) \cdot (2i).$$

In particular, $x = 1$ and $y = 2i$ gives a solution.

Our discussion above can be formalized into a proof of the following theorem:

Theorem 28.2. *Let R be an integral domain with a division algorithm, and let a and b be non-zero elements of R. If d is any gcd of a and b, then there are $x, y \in R$ such that $ax + by = d$.*

Writing up a proof of this is highly recommended as an exercise! (What is left to do in the proof, on top of what we already discussed?)

For additional practice with these computational techniques, you are highly encouraged to make up your own examples of Euclidean algorithm and Extended Euclidean algorithm computations, choosing your own values of a and b. You are then welcome to share your questions and solutions on Piazza!

MATH 145 Course Reading 29:
Linear Diophantine Equations and Linear Congruences

November 23, 2020

One of the major topics studied in number theory is the solution of *Diophantine equations*. Simply put, a Diophantine equation is an equation in one or more variables, with integer coefficients, for which we search for integer solutions. For example, all of the following are important types of Diophantine equations:

$$7x + 19y = 4$$
$$x^3 + y^3 = z^3$$
$$x^2 - 2y^2 = 1.$$

The first equation is a *linear Diophantine equation*, so-called because all the variables in the equation occur with degree 1. This is the type of equation we will solve in today's reading. The other two are examples of a *Fermat equation* and *Pell equation*, respectively. These types of Diophantine equations are explored much more carefully in a course on elementary or algebraic number theory (such as PMATH 340 or PMATH 441 here at Waterloo).

Linear Diophantine Equations in Two Variables

Suppose R is an integral domain with a division algorithm. Given $a, b, c \in R$, we would like to study solutions to the equation

$$ax + by = c,$$

with $x, y \in R$. The two main questions of interest are:

(1) Does a solution to the equation exist?

(2) If yes, can we find *all* the solutions?

Our first result helps narrow down situations where the solution *does* exist:

Theorem 29.1. *Let R be an integral domain with a division algorithm, and let a, b, and c be arbitrary elements of R, with a and b not both zero. If there is a solution to the equation*

$$ax + by = c$$

with $x, y \in R$, then $\gcd(a, b) \mid c$.

Proof. Let d denote any gcd of a and b. Then $d \mid a$ and $d \mid b$. Given the elements $x, y \in R$ mentioned above, we apply part (3) of Proposition 26.1. This tells us that $d \mid (ax + by)$, i.e. $d \mid c$, as needed. □

So, for instance, the equation $4x + 6y = 5$ in $\mathbb{Z}$ cannot have an integer solution, because $\gcd(4, 6) = 2$ does not divide 5. Put another way, the left-hand side $4x + 6y$ of the equation is always going to be a multiple of 2 for any $x, y \in \mathbb{Z}$, so it can never equal 5.

A perhaps more interesting result is that the converse of the previous theorem holds:

Theorem 29.2. *Suppose R is an integral domain with a division algorithm. Suppose we have $a, b, c \in R$ with a, b not both zero, and where* $\gcd(a, b) \mid c$. *Then the equation*

$$ax + by = c$$

has a solution with $x, y \in R$.

Proof. Let d denote any gcd of a and b. (Such a gcd exists by Theorem 28.1). Furthermore, by Theorem 28.2, there are elements $x_0, y_0 \in R$ for which $ax_0 + by_0 = d$. Since $d \mid c$, by definition there is some $k \in \mathbb{Z}$ for which $c = kd$. Therefore,

$$c = kd = k(ax_0 + by_0) = a(kx_0) + b(ky_0),$$

so that the equation $ax + by = c$ has a solution $x = kx_0$ and $y = ky_0$. □

Some Divisibility Results

In order to proceed with solving these linear Diophantine equations, it is necessary to establish a few results on divisibility. These follow from the theory we have developed so far in a very slick way.

Theorem 29.3. *Suppose R is an integral domain with a division algorithm, and suppose we are given $a, b, c \in R$. If $a \mid bc$ and $1 \sim \gcd(a, b)$, then $a \mid c$.*

Proof. Given that $a \mid bc$, we know that $bc = ak$ for some $k \in R$. Furthermore, since 1 is a gcd of a and b, Theorem 29.2 tells us there are $x, y \in R$ such that

$$ax + by = 1.$$

Multiplying both sides of the above by c, we get

$$acx + bcy = c.$$

Now $bc = ak$, and so

$$c = acx + aky = a(cx + ky),$$

proving that $a \mid c$, as required. □

This next result backs the intuition that if two elements are "divided" by their gcd, then the resulting elements have no common factors left:

Lemma 29.1. *Suppose R is an integral domain with a division algorithm, and suppose we are given $a, b \in R$, not both zero. Suppose d is a gcd of a and b, so that we may write $a = da_0$ and $b = db_0$, for some $a_0, b_0 \in R$. Then $\gcd(a_0, b_0) \sim 1$.*

Proof. We begin by applying Theorem 29.2, with $c = d$. This tells us that there are $x, y \in R$ such that

$$ax + by = d.$$

Equivalently, we have

$$(da_0)x + (db_0)y = d(a_0x + b_0y) = d,$$

and we may cancel the non-zero element d to get

$$a_0x + b_0y = 1.$$

In particular, applying Theorem 29.1, we get that $\gcd(a_0, b_0) \mid 1$. But $1 \mid \gcd(a_0, b_0)$ always holds trivially, so this shows $\gcd(a_0, b_0) \sim 1$. □

The General Solution of a Linear Diophantine Equation

Our results in the first section tell us that when R is an integral domain with a division algorithm, the equation $ax + by = c$ has a solution if and only if $\gcd(a, b) \mid c$. When this condition holds, how do we find *all* the solutions? The next theorem spells it out for us:

Theorem 29.4. *Let R be an integral domain with a division algorithm. Let $a, b, c \in R$ be such that a and b are not zero, and let d be a gcd of a and b. Furthermore, assume that $d \mid c$. Also, write $a = da_0$ and $b = db_0$ for some $a_0, b_0 \in R$. Then the complete set of solutions to the equation*

$$ax + by = c$$

with $x, y \in R$ is given by

$$(x, y) = (x_0 + kb_0, y_0 - ka_0),$$

where $k \in R$ is arbitrary, and (x_0, y_0) is a particular solution to $ax + by = c$ (which exists by Theorem 29.2).

Proof. Suppose a, b, c, d, a_0, b_0 are as given in the setup of the theorem. In particular, the equation $ax+by = c$ *has* a solution (x_0, y_0) by Theorem 29.2. Given such a solution, suppose that (x_1, y_1) is another solution to the equation. Then we know that both

$$ax_1 + by_1 = c$$
$$ax_0 + by_0 = c$$

hold. If we subtract the two equations, we end up with

$$a(x_1 - x_0) + b(y_1 - y_0) = 0.$$

Writing $a = da_0$ and $b = db_0$, the above becomes

$$da_0(x_1 - x_0) + db_0(y_1 - y_0) = 0.$$

Cancelling away the common factor of d in this integral domain and re-arranging, we end up with

$$a_0(x_1 - x_0) = -b_0(y_1 - y_0).$$

From here, we see that $b_0 \mid a_0(x_1 - x_0)$. But $\gcd(a_0, b_0) \sim 1$ by Lemma 29.1, and so $b_0 \mid (x_1 - x_0)$ by Theorem 29.3. By definition, this means $x_1 - x_0 = kb_0$ for some $k \in R$, so that $x_1 = x_0 + kb_0$. Substituting this back into the displayed equation above,

$$a_0(kb_0) = -b_0(y_1 - y_0).$$

Cancelling away the common factor of b_0, we get $ka_0 = -(y_1 - y_0)$, and solving for y_1 gives $y_1 = y_0 - ka_0$.

We have now shown that if (x_1, y_1) is another solution to the equation, then $(x_1, y_1) = (x_0 + kb_0, y_0 - ka_0)$ for some $k \in R$. Conversely, we can check that every ordered pair $(x_1, y_1) = (x_0 + kb_0, y_0 - ka_0)$ is a solution to the equation:

$$ax_1 + by_1 = a(x_0 + kb_0) + b(y_0 - ka_0) = (ax_0 + by_0) + k(ab_0 - ba_0) = c + k((da_0)b_0 - (db_0)a_0) = c.$$

This completes the proof of the theorem. □

To illustrate the theorem in action, let's pick up the two examples from the previous reading and solve a corresponding linear Diophantine equation.

Example 29.1. Suppose we wish to find all integer solutions to the equation

$$1009x + 33y = 5.$$

By running the Extended Euclidean algorithm on 1009 and 33 (as described in our previous reading), we find that $\gcd(1009, 33) = 1$ and that $1009(7) + 33(-214) = 1$. In particular, since the gcd of 1009 and 33 divides 5, the equation above has a solution by Theorem 29.2. Following the proof of that theorem, we multiply the equation

$$1009(7) + 33(-214) = 1$$

through by 5 to make the right-hand side equal to 5, giving us

$$1009(35) + 33(-1070) = 5.$$

This means our equation has the particular solution $(x_0, y_0) = (35, -1070)$. To find all solutions, we observe that in the notation of Theorem 29.4, we have $a_0 = a = 1009$ and $b_0 = b = 33$, because the gcd is 1 in this case. Thus the general solution is

$$(x, y) = (35 + 33k, -1070 - 1009k),$$

where k is an arbitrary integer parameter. So for example, we could take $k = -1$ to get another solution $(x, y) = (2, -61)$.

Example 29.2. Suppose we wish to find all solutions in $\mathbb{Z}[i]$ to the equation

$$17x + (2+9i)y = (-5+3i).$$

By running the Extended Euclidean algorithm on 17 and $2+9i$ (as described in our previous reading), we found that $-1+4i$ is a gcd of these two numbers in $\mathbb{Z}[i]$. We know that the equation above only has a solution if $\gcd(17, 2+9i)$ divides the constant term, $-5+3i$. To see if this is the case, we carry out complex number division:

$$\frac{-5+3i}{-1+4i} = \frac{(-5+3i)(-1-4i)}{(-1+4i)(-1-4i)} = \frac{17+17i}{17} = 1+i.$$

Thus $(-5+3i) = (-1+4i)(1+i)$. In particular, we can find a single solution to the equation above by taking the end result of our Extended Euclidean algorithm computation (from Example 28.4),

$$(-1+4i) = 17 \cdot 1 + (2+9i) \cdot (2i),$$

and multiplying through by $1+i$:

$$-5+3i = 17 \cdot (1+i) + (2+9i) \cdot (-2+2i),$$

giving us the particular solution $(x_0, y_0) = (1+i, -2+2i)$. In the notation of Theorem 29.4, we have $a_0 = \frac{17}{-1+4i} = -1-4i$, and $b_0 = \frac{2+9i}{-1+4i} = 2-i$. So the general solution to the equation is

$$(x, y) = ((1+i) + (2-i)k, (-2+2i) - (-1-4i)k),$$

where k is an arbitrary parameter in $\mathbb{Z}[i]$.

Multiplicative Inverses in $\mathbb{Z}/n\mathbb{Z}$

One useful by-product of this Diophantine equation discussion is that we now have a constructive procedure for calculating a multiplicative inverse of an element in $\mathbb{Z}/n\mathbb{Z}$ (when it exists). Suppose we take $[a] \in \mathbb{Z}/n\mathbb{Z}$. This congruence class has an inverse if and only if there is $[x] \in \mathbb{Z}/n\mathbb{Z}$ such that $[a][x] = [1]$. This is equivalent to the assertion that $ax \equiv 1 \pmod{n}$, which is the same as saying that $n \mid (1-ax)$. In turn, this is equivalent to $1 - ax = ny$ for some integer y, or $ax + ny = 1$ for some integers x and y.

Thus finding a multiplicative inverse for $[a] \in \mathbb{Z}/n\mathbb{Z}$ is reduced to solving the linear Diophantine equation $ax + ny = 1$ for some integers x and y. Our results from earlier in this reading say this is possible if and only if $\gcd(a, n) = 1$.

In particular, if $n = p$, a prime, and $[a] \in \mathbb{Z}/p\mathbb{Z}$ is such that $[a] \neq [0]$, then $\gcd(a, p) = 1$ (why?). This implies $[a]^{-1}$ exists in $\mathbb{Z}/p\mathbb{Z}$, and proves that $\mathbb{Z}/p\mathbb{Z}$ is a field (given that we already knew it was a commutative ring).

Let's give a very quick example of a multiplicative inverse computation:

Example 29.3. Suppose we want to find the multiplicative inverse of $[5] \in \mathbb{Z}/13\mathbb{Z}$. This reduces to finding a single solution to the Diophantine equation

$$5x + 13y = 1,$$

and $[x]$ will give a multiplicative inverse. So we run the Extended Euclidean algorithm on 13 and 5. First we perform the divisions with remainder:

$$\begin{aligned} 13 &= 5 \cdot 2 + 3 \\ 5 &= 3 \cdot 1 + 2 \\ 3 &= 2 \cdot 1 + 1 \\ 2 &= 1 \cdot 2 + 0. \end{aligned}$$

Since the last non-zero remainder is 1, $\gcd(13, 5) = 1$ and $[5]^{-1}$ exists. Now we use back-substitution to find a solution to the given Diophantine equation. Writing each remainder by itself, we get

$$\begin{aligned} 1 &= 3 - 2 \cdot 1 \\ 2 &= 5 - 3 \cdot 1 \\ 3 &= 13 - 5 \cdot 2. \end{aligned}$$

Now making the substitutions, we find

$$\begin{aligned} 1 &= 3 - 2 \cdot 1 \\ &= 3 - (5 - 3 \cdot 1) \cdot 1 \\ &= 3 \cdot 2 - 5 \cdot 1 \\ &= (13 - 5 \cdot 2) \cdot 2 - 5 \cdot 1 \\ &= 13 \cdot 2 - 5 \cdot 4 - 5 \cdot 1 \\ &= 13 \cdot 2 - 5 \cdot 5. \end{aligned}$$

Hence $5x + 13y = 1$ has a solution $x = -5$ and $y = 2$. In particular, $[5]^{-1} = [-5] = [8]$, as you can verify directly by computing $[5] \cdot [8]$.

This method for finding multiplicative inverses in $\mathbb{Z}/n\mathbb{Z}$ by solving a corresponding linear Diophantine equation extends in a similar way to solving any linear congruence of the form $ax \equiv c \pmod{n}$. We will explore this in more detail on the next assignment.

MATH 145 Course Reading 30: Primitive Roots Modulo a Prime and Polynomials over a Field

November 25, 2020

At this stage of our journey through abstract algebra, we are now in a position to state and prove a famous result, one that has real-life implications in *cryptography* (the science of making communications secure from unauthorized parties). We just saw in the previous reading that for any prime p, the ring $\mathbb{Z}/p\mathbb{Z}$ is actually a field. In particular, its set of units, $(\mathbb{Z}/p\mathbb{Z})^*$, is a group with respect to multiplication. What we will see is that $(\mathbb{Z}/p\mathbb{Z})^*$ is actually a *cyclic* group, which implies that every nonzero element of $\mathbb{Z}/p\mathbb{Z}$ is a power of some fixed nonzero element of $\mathbb{Z}/p\mathbb{Z}$.

Along the way to proving this, we will introduce a new type of ring, the *ring of polynomials with coefficients in a field*, and prove some familiar properties about these rings, including showing that these rings have a division algorithm.

Polynomials over a Field

Let's begin with a fundamental definition:

Definition 30.1. Let F be a field. We define the set $F[x]$ of *polynomials with coefficients in* F to be

$$F[x] = \{a_0 + a_1x + a_2x^2 + \cdots + a_nx^n : n \geq 0, a_i \in F \text{ for each } i.\}$$

The *degree* of a polynomial $f = a_0 + a_1x + \cdots + a_nx^n \in F[x]$, denoted $\deg(f)$, is the largest index i such that $a_i \neq 0$. So for instance, $\deg(x^2 + x^4) = 4$. We do not assign a degree to the zero polynomial (the polynomial with all coefficients zero).

We can turn $F[x]$ into a commutative ring by defining addition and multiplication of polynomials in the way you are familiar with for polynomials over $\mathbb{R}$. If $f = \sum_{i=0}^{m} a_ix^i$ and $g = \sum_{i=0}^{n} b_ix^i$, then we define

$$f + g = \sum_{i=0}^{\max(m,n)} (a_i + b_i)x^i$$

$$fg = \sum_{i=0}^{m+n} c_ix^i,$$

where for each i, $c_i = \sum_{j=0}^{i} a_jb_{i-j}$. (Do you see why this definition of multiplication really is the "usual" one?)

The verification that these operations make $F[x]$ into a commutative ring will be omitted here, but you are encouraged to work through at least some of the verification for yourself! (The associativity of multiplication is particularly involved).

The following result lays the groundwork to prove that $F[x]$ is an integral domain, and also to establish that $F[x]$ has a division algorithm:

Theorem 30.1. *Let F be an arbitrary field, and let $f, g \in F$ be non-zero polynomials. We have the following:*

(1) If $f + g \neq 0$, then $\deg(f + g) \leq \max(\deg(f), \deg(g))$

(2) $\deg(fg) = \deg(f) + \deg(g)$

(3) $F[x]$ is an integral domain.

Proof. Suppose we write $f = \sum_{i=0}^{m} a_i x^i$ and $g = \sum_{i=0}^{n} b_i x^i$, with $a_m \neq 0$ and $b_n \neq 0$, so that $m = \deg(f)$ and $n = \deg(g)$.

(1) By definition, $f+g = \sum_{i=0}^{\max(m,n)} (a_i + b_i)x^i$. In particular, all coefficients of powers of x after $\max(m,n)$ are equal to 0, and so if $f+g \neq 0$, then $\deg(f+g) \leq \max(m,n) = \max(\deg(f), \deg(g))$.

(2) Again by definition, $fg = \sum_{i=0}^{m+n} c_i x^i$, with each coefficient c_i as given in the definition. All coefficients of powers of x after $m+n$ are equal to 0, so that $\deg(fg) \leq m+n = \deg(f) + \deg(g)$. Now, let's look at the coefficient c_{m+n} of x^{m+n}. By definition, it is equal to

$$\sum_{j=0}^{m+n} a_j b_{m+n-j} = a_0 b_{m+n} + a_1 b_{m+n-1} + \cdots + a_m b_n + \cdots + a_{m+n} b_0.$$

Note that $a_j = 0$ whenever $j > m$, so all terms after $a_m b_n$ in the last sum above are zero. Likewise, $b_{m+n-j} = 0$ whenever $j < m$, so all terms before $a_m b_n$ in the last sum above are equal to zero. Finally, since $a_m \neq 0$ and $b_n \neq 0$, the fact that F is an integral domain tells us $a_m b_n \neq 0$. Thus $c_{m+n} = a_m b_n \neq 0$, proving that $\deg(fg) = m+n = \deg(f) + \deg(g)$.

(3) Given part (2) of this theorem, note that if $f \neq 0$ and $g \neq 0$ are polynomials in $F[x]$, then f and g both have a degree, and $\deg(fg) = \deg(f) + \deg(g)$ exists. So in particular, $fg \neq 0$. Combined with the fact that $F[x]$ is a commutative ring, this shows $F[x]$ is an integral domain.

□

Polynomial Division with Remainder, and Consequences

Now, we add to the above result by showing that $F[x]$ has a division algorithm:

Theorem 30.2. *For any field F, the integral domain $F[x]$ admits a division algorithm with divisor function $\deg(\cdot)$. In other words, for any polynomials $f, g \in F[x]$ with $g \neq 0$, there are polynomials $q, r \in F[x]$ such that $f = gq + r$, and either $r = 0$ or $\deg(r) < \deg(g)$.*

Proof. Just like the proof that worked over $\mathbb{Z}$, the main trick is to study the set $S = \{f - gq : q \in F[x]\}$, which we can see as the set of candidate remainders. If S contains the zero polynomial, then there is $q \in F[x]$ such that $f - gq = 0$, i.e. $f = gq + 0$, and so the desired polynomials exist.

Otherwise, we can look at the degrees of all the polynomials in the set S, and take one such polynomial r of lowest degree. By construction, $r = f - gq$ for some polynomial $q \in F[x]$, and so $f = gq + r$ as needed. It only remains to check that $\deg r < \deg g$. So suppose to the contrary that $\deg r \geq \deg g$, and express the polynomials r and g as $r = a_n x^n + a_{n-1}x^{n-1} + \cdots$ and $g = b_m x^m + b_{m-1}x^{m-1} + \cdots$, where $b_m \neq 0$ since g is the zero polynomial. Note that b_m is a nonzero element of the field F, so b_m^{-1} exists. Since we have assumed that $\deg r \geq \deg g$, we know that $n \geq m$. Thus, we consider the new polynomial

$$\begin{aligned} r_1 &= r - a_n b_m^{-1} x^{n-m} g \\ &= (a_n x^n + a_{n-1}x^{n-1} + \cdots) - (a_n x^n + a_n b_m^{-1} b_{m-1} x^{n-1} + \cdots) \\ &= (a_{n-1} - a_n b_m^{-1} b_{m-1}) x^{n-1} + \cdots, \end{aligned}$$

so that $\deg r_1 < \deg r$. On the other hand, $r_1 = r - a_n b_m^{-1} x^{n-m} g = (f - qg) - a_n b_m^{-1} x^{n-m} g = f - (q + a_n b_m^{-1} x^{n-m}) g$, so $r_1 \in S$, contradicting our choice of r as an element of S of smallest degree. We deduce that we must indeed have $\deg r < \deg g$ after all, and so a decomposition $f = gq + r$ where $r = 0$ or $\deg r < \deg g$ does exist. □

To calculate the result of a division with remainder in $F[x]$ practically, the approach is to use the polynomial long division technique that you've likely encountered in high school and in calculus for polynomials with real coefficients. The only thing to keep in mind when working over arbitrary fields is that any instance of "division" that you perform in long division should be treated as multiplying by the multiplicative inverse

of the given element. This long division process will be explored further on Assignment 9.

This long division procedure leads directly to some important results around roots of a polynomial. If $f = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in F[x]$ is a polynomial and $c \in F$, we define the *evaluation of f at c* to be

$$f(c) = a_n c^n + a_{n-1}c^{n-1} + \cdots + a_0,$$

obtained by replacing every indeterminate x with the field element c. We say that c is a *root* of f if $f(c) = 0$. You can verify that for each $c \in F$, the mapping $\phi_c : F[x] \to F$ given by $\phi_c(f) = f(c)$ is a ring homomorphism, called the *evaluation homomorphism*.

Theorem 30.3. *Suppose F is a field, $f \in F[x]$ is an arbitrary polynomial, and $c \in F$ is an arbitrary element.*

(1) The element c is a root of f if and only if $(x - c) \mid f$ in the integral domain $F[x]$. Equivalently, $\ker \phi_c = F[x](x - c)$, the ideal generated by $x - c$. (This is often called the Factor Theorem.*)*

(2) If f is a nonzero polynomial, then f has at most $\deg(f)$ distinct roots in F.

Proof. (1) Given the polynomial f and element $c \in F$, first suppose that $(x - c) \mid f$. Then by definition there is some polynomial $q \in F[x]$ such that $f = (x-c)q$. Note that $\phi_c(x-c) = c-c = 0$. Thus applying the evaluation homomorphism to $f = (x - c)q$, we get $f(c) = \phi_c(f) = \phi_c((x - c)q) = \phi_c(x - c)\phi_c(q) = 0 \cdot q(c) = 0$. So, c is a root of f.

Conversely, suppose that c is a root of f. Applying division with remainder of f by $(x - c)$, we find polynomials $q, r \in F[x]$ such that

$$f = (x - c)q + r,$$

where $r = 0$ or $\deg r < \deg(x - c) = 1$. Either way, r is a *constant polynomial*, of the form $r = r_0$ for some $r_0 \in F$. Applying ϕ_c to the equation above and using that $f(c) = 0$, we find that $0 = f(c) = 0 \cdot q(c) + r(c) = r(c) = r_0$. Thus $r = 0$, and $f = (x - c)q$, proving that $(x - c) \mid f$.

Putting this together, we find that $f(c) = 0$ exactly when f is a multiple of $(x - c)$, which exactly describes the principal ideal $F[x](x - c)$, so that $\ker \phi_c$ is given by this ideal.

(2) We prove this result by induction on $\deg(f)$. We first check the base cases $\deg(f) = 0$ and $\deg(f) = 1$. If $\deg(f) = 0$, then $f = f_0$, where f_0 is a non-zero element of f, and so $f(c) = f_0 \neq 0$ for any $c \in F$. Thus f has no roots in F, and the proof is complete in this case. If $\deg(f) = 1$, then $f = ax+b$ for some $a, b \in F$, with $a \neq 0$. Then note that $f(c) = 0$ if and only if $ac+b = 0$, if and only if $ac = -b$, if and only if $c = -a^{-1}b$. This proves that f has exactly one root in F, completing the proof in this base case as well.

Now suppose $\deg(f) = n+1$ for some integer $n \geq 1$, and assume that all polynomials of degree $k < n+1$ have at most k roots in F. If f has no roots in F, then the result is trivially true, so suppose $c \in F$ is a root of f. Applying the Factor Theorem (part (1) of this result), we find that $f = (x - c)f_n$ for some polynomial $f_n \in F$. Taking degrees of both sides, we deduce that $\deg(f_n) = n$. Note that for any $a \in F$, we have $f(a) = 0$ if and only if $(a - c)f_n(a) = 0$, which holds if and only if $a = c$ or $f_n(a) = 0$. But $f_n(a) = 0$ for at most n distinct values $a \in F$, so altogether, $f(a) = 0$ for at most $n + 1$ distinct values a (all the roots of f_n, along with c). By induction, the proof is complete.

□

Primitive Roots Modulo p

We have now built up the theory required to show that $(\mathbb{Z}/p\mathbb{Z})^*$ is a cyclic group for any prime p. At this stage, what we do know is that this group is a finite abelian group with exactly $p - 1$ elements (since every nonzero element of $\mathbb{Z}/p\mathbb{Z}$ is a unit). To break up the proof a little, we present the first half of the proof as a general result about finite abelian groups:

Lemma 30.1. *Let G be a finite abelian group, and suppose $g \in G$ is an element for which $o(g)$ is maximal. If $o(g) = k$, then $h^k = e$ for all elements $h \in G$.*

Proof. You will establish this result on Assignment 9. □

Now, let's prove the theorem we're after, but even more generally:

Theorem 30.4. *Let F be a finite field. Then the group F^* of units of F is cyclic. (We refer to any generator of F^* as a* primitive element for F*.)*

Proof. Since F^* is a finite abelian group, we can choose an element $c \in F^*$ that has maximal order, say k. By Lemma 30.1, applied to the group F^*, we know that $a^k = 1$ for all $a \in F^*$. In particular, the polynomial $x^k - 1 \in F[x]$ has at least $|F^*|$ distinct roots. On the other hand, part (2) of Theorem 30.3 tells us that $x^k - 1$ has at most k distinct roots, so $|F^*| \leq k$. But $k \leq |F^*|$, because Lagrange's Theorem tells us that $k = o(c)$ must divide the size of the group, which is $|F^*|$. We conclude that $k = |F^*|$, and in particular, there is an element of F^* having order $|F^*|$. In turn, this implies F^* is cyclic, as desired. □

Corollary 30.1. *For any prime p, the group $(\mathbb{Z}/p\mathbb{Z})^*$ is cyclic.*

Proof. For any prime p, the ring $\mathbb{Z}/p\mathbb{Z}$ is a finite field. Now apply Theorem 30.4. □

In general, there is no systematic way of *locating* a primitive element of $\mathbb{Z}/p\mathbb{Z}$, even though we know such an element must exist. In practice, the search usually boils down to educated trial-and-error, as in the next example:

Example 30.1. Let's find a primitive element for $\mathbb{Z}/13\mathbb{Z}$. Thus, we're looking for a congruence class $[a]$ such that $o([a]) = 12$. Note that since $o([a]) \mid 12$ in any case, if we show that $[a]^6 \neq [1]$ and $[a]^4 \neq 1$, it will immediately follow that $o([a]) = 12$ (why?). We know $[1]$ cannot be a primitive element, so let's try $[2]$:

$$[2]^4 = [16] = [3] \neq [1]$$
$$[2]^6 = [2]^4 \cdot [2]^2 = [3] \cdot [4] = [12] \neq [1].$$

Based on these two computations, we must have $o([2]) = 12$, and so $[2]$ is a primitive element. (Exercise: can you find all the other primitive elements of $\mathbb{Z}/13\mathbb{Z}$, given that we know $[2]$ is one of them?)

Some of the protocols of modern cryptography make use of the fact that $(\mathbb{Z}/p\mathbb{Z})^*$ is a cyclic group. They take advantage of the fact that when p is large, if you are given a primitive element $[a] \in \mathbb{Z}/p\mathbb{Z}$, and some element $[b] \in (\mathbb{Z}/p\mathbb{Z})^*$, it is very difficult to find an exponent e such that $[a]^e = [b]$. One cryptographic application, the *Diffie-Hellman protocol*, will be explored on Assignment 10.

MATH 145 Course Reading 31:
Chinese Remainder Theorem for Integers

November 27, 2020

In the reading before last, we took up the question of solving linear congruences modulo n, as a natural outgrowth of our work on solving linear Diophantine equations over the integers. But what if we wish to find integers satisfying multiple congruences at the same time? One tool for accomplishing this is the Chinese Remainder Theorem, which we will now investigate. We will take two different looks at this theorem: one is the "classical" perspective, and the other takes a look at the result through the lens of ring theory, as giving a certain type of bijective homomorphism (aka *isomorphism*) between two different rings. This second look at the theorem will give us more opportunity to discuss the notion of isomorphism in general.

Chinese Remainder Theorem, Classical Version

The starting point for the result known as the Chinese Remainder Theorem is the following general question: given a whole bunch of congruence constraints,

$$\begin{cases} x \equiv a_1 \pmod{m_1} \\ x \equiv a_2 \pmod{m_2} \\ \vdots \\ x \equiv a_k \pmod{m_k}, \end{cases}$$

where $m_1, \ldots, m_k$ are arbitrary positive integers and $a_1, \ldots, a_k$ are arbitrary integers, will there always be an integer x satisfying them all?

It is easy to see that the answer is no, at least not in general. For instance, it is easy to argue that the two congruences

$$\begin{cases} x \equiv 1 \pmod{2} \\ x \equiv 0 \pmod{4} \end{cases}$$

do not have a simultaneous solution $x \in \mathbb{Z}$, since the first congruence would require x to be odd, while the second clearly requires x to be even. (Can you come up with other counterexamples?)

The theorem we are about to present shows that this difficulty can be avoided, provided we make sure the moduli $m_1, m_2, \ldots, m_k$ don't share any factors in common. Before we state the theorem, we first need to introduce some terminology. Two integers a and b are often called *coprime* or *relatively prime* if $\gcd(a, b) = 1$, and a list $m_1, \ldots, m_k$ of integers are called *pairwise coprime* if every pair of distinct elements is coprime, meaning that $\gcd(m_i, m_j) = 1$ for $i \neq j$.

Theorem 31.1 (Chinese Remainder Theorem)**.** *Let $a_1, \ldots, a_k$ denote arbitrary integers, and let $m_1, \ldots, m_k$ denote pairwise coprime positive integers. The system of congruences*

$$\begin{cases} x \equiv a_1 \pmod{m_1} \\ x \equiv a_2 \pmod{m_2} \\ \vdots \\ x \equiv a_k \pmod{m_k} \end{cases}$$

has a solution $x \in \mathbb{Z}$, and this solution is unique modulo $m_1 m_2 \ldots m_k$. In other words, if $y \in \mathbb{Z}$ is also a solution to the system above, then $x \equiv y \pmod{m_1 m_2 \cdots m_k}$.

Proof. First, we show that a solution exists. One useful way to conceptualize the proof is to think of each of the k congruences in the system as specifying a coordinate in some ordered k-tuple. Our first goal will be to track down some "standard coordinates" with respect to this system.

More specifically, we first seek to find integers $b_1, \ldots, b_k$ for which

$$\begin{cases} b_1 \equiv 1 \pmod{m_1} \\ b_1 \equiv 0 \pmod{m_2} \\ \vdots \\ b_1 \equiv 0 \pmod{m_k} \end{cases} \qquad \begin{cases} b_2 \equiv 0 \pmod{m_1} \\ b_2 \equiv 1 \pmod{m_2} \\ \vdots \\ b_2 \equiv 0 \pmod{m_k} \end{cases} \qquad \begin{cases} b_k \equiv 0 \pmod{m_1} \\ b_k \equiv 0 \pmod{m_2} \\ \vdots \\ b_k \equiv 1 \pmod{m_k}. \end{cases}$$

So you can think of b_1 as being "1 in the first coordinate and 0 in the rest", b_2 as being "1 in the second coordinate and 0 in the rest", and so on.

Let's take for granted for a moment that these integers $b_1, \ldots, b_k$ can be found. We then claim that

$$x = \sum_{i=1}^{k} a_i b_i$$

is a solution to the system of congruences. To verify this, let's first reduce modulo m_1. This gives

$$\begin{aligned} x &= a_1 b_1 + a_2 b_2 + \cdots + a_k b_k \\ &\equiv a_1(1) + a_2(0) + \cdots + a_k(0) \\ &\equiv a_1 \pmod{m_1}. \end{aligned}$$

A similar computation when reducing modulo m_i for each i shows that $x \equiv a_i \pmod{m_i}$. Thus, if these integers $b_1, \ldots, b_k$ can be found, then a solution to the system of congruences exists.

Let's first argue why b_1 exists. In order for b_1 to satisfy the congruences that it does, the conditions $b_1 \equiv 0 \pmod{m_i}$ for $2 \leq i \leq k$ show that b_1 is divisible by each of $m_2, m_3, \ldots, m_k$. So let's set $M_1 = m_2 m_3 \cdots m_k$. If we take b_1 in the form $c_1 M_1$ for some integer c_1, then b_1 automatically satisfies $b_1 \equiv 0 \pmod{m_i}$ for $2 \leq i \leq k$.

To make sure b_1 satisfies the first congruence, we will then need $c_1 M_1 \equiv 1 \pmod{m_1}$. We already have a value for M_1, so we see that c_1 must be the multiplicative inverse of M_1 modulo m_1. Does this inverse exist? We know from previous work that this holds if and only if $\gcd(M_1, m_1) = 1$. But suppose this were not the case. Then M_1 and m_1 would share a prime factor. This factor would divide both m_1 and one of the integers in the product $M_1 = m_2 m_3 \cdots m_k$. Then this prime would be a common factor of m_1 and some m_i for $i \neq 1$, contradicting the pairwise coprime hypothesis!

So M_1 has a multiplicative inverse in $\mathbb{Z}/m_1\mathbb{Z}$, and so we can indeed find an integer c_1 satisfying $c_1 M_1 \equiv 1 \pmod{m_1}$. Taking $b_1 = c_1 M_1$ gives us the integer we were looking for.

Similarly, for $2 \leq i \leq k$, set $M_i = \prod_{j \neq i} m_j$, the product of all the moduli except for m_i. A similar argument to the above shows that $\gcd(M_i, m_i) = 1$, and so there is an integer c_i such that $c_i M_i \equiv 1 \pmod{m_i}$. Setting $b_i = c_i M_i$, we immediately get $b_i \equiv 0 \pmod{m_j}$ for $j \neq i$, and $b_i \equiv 1 \pmod{m_i}$. This completes the proof of existence of a solution to the original system of congruences!

Now, let's argue why the solution is unique modulo $m_1 m_2 \cdots m_k$. Suppose that x and y are integers both satisfying the system of congruences. Then in particular, $x \equiv y \pmod{m_i}$ for $1 \leq i \leq k$, so that $m_i \mid (x - y)$ for each i. Thus, since $m_1 \mid (x - y)$ and $m_2 \mid (x - y)$ and $\gcd(m_1, m_2) = 1$, Question 4 of Assignment 9 yields that $m_1 m_2 \mid (x - y)$. It is easy to check that the pairwise coprime hypothesis also guarantees that $\gcd(m_1 m_2, m_3) = 1$, so iterating the result of this assignment question, together with the fact that $m_3 \mid (x - y)$, tells us that $m_1 m_2 m_3 \mid (x - y)$. Continuing to apply this result, we eventually conclude that $m_1 m_2 \cdots m_k \mid (x - y)$, so that $x \equiv y \pmod{m_1 m_2 \cdots m_k}$. □

Let's now see this procedure in action on a small example:

Example 31.1. Suppose we wish to find the solutions $x \in \mathbb{Z}$ to the system

$$\begin{cases} x \equiv 2 \pmod 3 \\ x \equiv 3 \pmod 5 \\ x \equiv 2 \pmod 7. \end{cases}$$

In the notation of the theorem's proof, we have $M_1 = 5 \cdot 7 = 35$, $M_2 = 3 \cdot 7 = 21$, and $M_3 = 3 \cdot 5 = 15$. Then c_1, c_2, c_3 are determined by solving $35c_1 \equiv 1 \pmod 3$, $21c_2 \equiv 1 \pmod 5$, and $15c_3 \equiv 1 \pmod 7$. Reducing the coefficients, these drop to $-c_1 \equiv 1 \pmod 3$, $c_2 \equiv 1 \pmod 5$, and $c_3 \equiv 1 \pmod 7$. So taking $(c_1, c_2, c_3) = (-1, 1, 1)$ is acceptable. Then

$$\begin{aligned} b_1 &= c_1 M_1 = (-1)(35) = -35 \\ b_2 &= c_2 M_2 = (1)(21) = 21 \\ b_3 &= c_3 M_3 = (1)(15) = 15, \end{aligned}$$

and a solution to our system (given that $(a_1, a_2, a_3) = (2, 3, 2)$) is

$$x = a_1 b_1 + a_2 b_2 + a_3 b_3 = 2 \cdot (-35) + 3 \cdot 21 + 2 \cdot 15 = -70 + 63 + 30 = 23.$$

This solution is unique modulo $3 \cdot 5 \cdot 7 = 105$, so that if y is another integer solution to the system, then $y \equiv 23 \pmod{105}$.

Chinese Remainder Theorem, Ring Theory Version

Taking a careful look at the proof we just gave, the proof strategy of thinking about each congruence constraint as giving a "coordinate" can be fleshed out in terms of ring theory. Since the solution to the given system is unique modulo $m_1 m_2 \cdots m_k$, it looks like we are establishing some kind of correspondence between the ring $(\mathbb{Z}/m_1 m_2 \cdots m_k \mathbb{Z})$ and the Cartesian product $(\mathbb{Z}/m_1\mathbb{Z}) \times (\mathbb{Z}/m_2\mathbb{Z}) \times (\mathbb{Z}/m_k\mathbb{Z})$. And in fact, this correspondence can be lifted to the level of rings, taking the direct product (introduced on Assignment 7) as the ring structure on the Cartesian product.

To fully explain what it means for two rings to have the "same structure", we now introduce the notion of isomorphism, for both groups and rings. A homomorphism $\phi : R \to S$ of rings is called an *isomorphism* if ϕ is also a bijection. Equivalently, ϕ is an isomorphism if there is a homomorphism $\psi : S \to R$ such that $\psi \circ \phi = \mathrm{id}_R$ and $\phi \circ \psi = \mathrm{id}_S$, where id_R and id_S are the identity maps on R and S respectively. (Proving this equivalence is a good exercise!)

If there is an isomorphism from a ring R to a ring S, then we say R and S are *isomorphic*, and write $R \cong S$. The relation of isomorphism on rings is easily checked to be an equivalence relation, and for all practical purposes, isomorphic rings behave in the same way in all ring-theoretic respects.

Similarly, a homomorphism of groups $\phi : G_1 \to G_2$ is an *isomorphism* if ϕ is also a bijection, or equivalently, if there is a homomorphism $\psi : G_2 \to G_1$ such that $\psi \circ \phi$ is the identity map on G_1 and $\phi \circ \psi$ is the identity map on G_2. Two groups are *isomorphic* if there is an isomorphism between them. Again, this gives an equivalence relation on groups.

The ring-theoretic version of the Chinese Remainder Theorem boils down to showing that two particular rings are isomorphic:

Theorem 31.2 (Chinese Remainder Theorem, Ring Version)**.** *Suppose $m_1, \ldots, m_k$ are pairwise coprime positive integers. Then there is a ring isomorphism*

$$(\mathbb{Z}/m_1 m_2 \cdots m_k \mathbb{Z}) \cong (\mathbb{Z}/m_1\mathbb{Z}) \times (\mathbb{Z}/m_2\mathbb{Z}) \times \cdots \times (\mathbb{Z}/m_k\mathbb{Z}).$$

Proof. We begin by defining a ring homomorphism $\phi : (\mathbb{Z}/m_1m_2\cdots m_k\mathbb{Z}) \to (\mathbb{Z}/m_1\mathbb{Z}) \times (\mathbb{Z}/m_2\mathbb{Z}) \times \cdots \times (\mathbb{Z}/m_k\mathbb{Z})$, and then showing that it is a bijection.

Using coset notation in the quotient rings for clarity, our mapping is given by

$$\phi(m_1m_2\cdots m_k\mathbb{Z} + x) = (m_1\mathbb{Z} + x, m_2\mathbb{Z} + x, \ldots, m_k\mathbb{Z} + x),$$

for all cosets $m_1m_2\cdots m_k\mathbb{Z} + x \in (\mathbb{Z}/m_1m_2\cdots m_k\mathbb{Z})$. First, we argue why this map is well-defined. If $m_1m_2\cdots m_k\mathbb{Z} + x = m_1m_2\cdots m_k\mathbb{Z} + y$, then $m_1m_2\cdots m_k \mid (x - y)$. In particular, for each i between 1 and k, we have $m_i \mid (x - y)$, so $m_i\mathbb{Z} + x = m_i\mathbb{Z} + y$. This implies $\phi(m_1m_2\cdots m_k\mathbb{Z} + x) = \phi(m_1m_2\cdots m_k\mathbb{Z} + y)$, proving that the map is well-defined.

Checking that ϕ preserves addition, multiplication, and the unity is routine, and thus we omit the details here. To show that ϕ is injective, we prove that $\ker\phi$ is the zero ideal; just like for groups, this is equivalent to the injectivity of ϕ. So suppose we have a coset $m_1m_2\cdots m_k\mathbb{Z} + x$ such that $\phi(m_1m_2\cdots m_k\mathbb{Z} + x) = (m_1\mathbb{Z} + 0, m_2\mathbb{Z} + 0, \ldots, m_k\mathbb{Z} + 0)$. Then x satisfies the system of congruences

$$\begin{cases} x \equiv 0 \pmod{m_1} \\ x \equiv 0 \pmod{m_2} \\ \vdots \\ x \equiv 0 \pmod{m_k}. \end{cases}$$

Clearly 0 is also a solution to the system, and by the uniqueness claim in Theorem 31.1, we deduce that $x \equiv 0 \pmod{m_1m_2\cdots m_k}$, so that $m_1m_2\cdots m_k\mathbb{Z} + x = m_1m_2\cdots m_k\mathbb{Z} + 0$. This completes the proof that $\ker\phi$ is the zero ideal, so that ϕ is injective.

To prove that ϕ is surjective, suppose we are given an arbitrary element

$$(m_1\mathbb{Z} + a_1, m_2\mathbb{Z} + a_2, \ldots, m_k\mathbb{Z} + a_k) \in (\mathbb{Z}/m_1\mathbb{Z}) \times (\mathbb{Z}/m_2\mathbb{Z}) \times \cdots \times (\mathbb{Z}/m_k\mathbb{Z}).$$

By Theorem 31.1, there is an integer x such that

$$\begin{cases} x \equiv a_1 \pmod{m_1} \\ x \equiv a_2 \pmod{m_2} \\ \vdots \\ x \equiv a_k \pmod{m_k}, \end{cases}$$

and you can check right away that

$$\phi(m_1m_2\cdots m_k\mathbb{Z} + x) = (m_1\mathbb{Z} + a_1, m_2\mathbb{Z} + a_2, \ldots, m_k\mathbb{Z} + a_k).$$

We now know that ϕ is a bijective homomorphism, hence an isomorphism. □

The existence of this isomorphism can then be used to tell us ring-theoretic facts about $(\mathbb{Z}/m_1m_2\cdots m_k\mathbb{Z})$, by reducing such questions to analyzing the same ring-theoretic facts for the direct product $(\mathbb{Z}/m_1\mathbb{Z}) \times (\mathbb{Z}/m_2\mathbb{Z}) \times \cdots \times (\mathbb{Z}/m_k\mathbb{Z})$. Again, this will be explored a little further on Assignment 10.

MATH 145 Course Reading 32:
Field of Fractions and Localization

November 30, 2020

Up to this point in the course, we have seen fields only in passing, and mostly in connection with our detailed study of integral domains. The subject of *field theory* is a rich and interesting subject in its own right, but even the most foundational results about fields require a good working knowledge of the subject of linear algebra (something you will not study until MATH 146). Thus, while we will indeed look at fields for the rest of the course, our discussion in subsequent readings will be restrained mostly to introducing and studying the field of complex numbers in particular. Before that, this reading will be spent looking at the way fields are intimately linked to integral domains by the so-called "field of fractions" construction.

The Field of Fractions of an Integral Domain

Back in Theorem 26.2, we saw that every subring of a field is an integral domain (and indeed, this has been the main way that we've been able to verify that rings are integral domains). It turns out that a converse holds as well: every integral domain is actually a subring of some field. We will prove this by explicit construction, taking the way $\mathbb{Q}$ is built out of $\mathbb{Z}$ as the motivating example.

Let's pretend for a moment that the ring $\mathbb{Q}$ has not yet been constructed, and we wish to do so using $\mathbb{Z}$ as our starting point. To specify a fraction $\frac{a}{b} \in \mathbb{Q}$, what are we really doing? Certainly, we're giving an (ordered) pair of integers, so we might initially start by using the ordered pair (a, b) to stand in for the fraction $\frac{a}{b}$. But this isn't quite right... First of all, 0 should not be allowed as the second coordinate of any ordered pair (we can't have any zero denominators!) Also, the notion of equality between ordered pairs is too tight. For instance, we want the fractions $\frac{1}{2}$ and $\frac{3}{6}$ to be the same, even though $(1, 2) \neq (3, 6)$ in $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$.

Thus, it seems like we're asking for a relation on this set of ordered pairs, specifying when two such pairs should be equal. In order to do this "from scratch", within the integers, we would like to define the relation using only operations available in $\mathbb{Z}$. Clearing denominators, we know we ought to have $\frac{a}{b} = \frac{c}{d}$ when $ad = bc$. As such, we proceed to define a relation $\sim$ on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ by declaring that $(a, b) \sim (c, d)$ if $ad = bc$.

The verification that this is an equivalence relation does not depend crucially on special properties of $\mathbb{Z}$, so we define it in more generality:

Proposition 32.1. *Let R be an integral domain. We define a relation $\sim$ on $R \times (R \setminus \{0\})$ by declaring that $(a, b) \sim (c, d)$ if $ad = bc$. Then $\sim$ is an equivalence relation, and we denote the set of equivalence classes by $Q(R)$.*

Proof. We check each of the three equivalence relation properties in turn:

- **Reflexivity:** For any ordered pair $(a, b) \in R \times (R \setminus \{0\})$, we have $(a, b) \sim (a, b)$ because $ab = ba$.
- **Symmetry:** Suppose we have $(a, b), (c, d) \in R \times (R \setminus \{0\})$ such that $(a, b) \sim (c, d)$. This tells us that $ad = bc$. This implies $cb = da$, so that $(c, d) \sim (a, b)$.
- **Transitivity:** Suppose we are given $(a, b), (c, d), (e, f) \in R \times (R \setminus \{0\})$ such that $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$. By definition, this tells us $ad = bc$ and $cf = de$. We need to show that $(a, b) \sim (e, f)$, which is equivalent to $af = be$. If we take $ad = bc$ and multiply both sides by f, we end up with

$$\begin{aligned} adf &= bcf \\ adf &= b(de) \\ (af)d &= (be)d. \end{aligned}$$

  Now, since R is an integral domain and $d \neq 0$, we can cancel it away to get $af = be$, implying that $(a, b) \sim (e, f)$, as required.

□

Going back to our "case study" of constructing $\mathbb{Q}$ from $\mathbb{Z}$, our candidate set for $\mathbb{Q}$ is the set of equivalence classes $Q(\mathbb{Z})$, with $[(a,b)]$ standing in for our fraction $\frac{a}{b}$. We would now like to define an addition and multiplication on this set, inspired by how it "ought" to work in $\mathbb{Q}$. Since we have $\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$, we set $[(a,b)] + [(c,d)] = [(ad+bc, bd)]$, and since $\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}$, we again set $[(a,b)] \cdot [(c,d)] = [(ac,bd)]$. Notice that our denominator bd in both cases is non-zero because $\mathbb{Z}$ is an integral domain and b and d are non-zero.

This procedure again generalizes to arbitrary integral domains, so we couch the result in general terms:

Proposition 32.2. *Let R be an integral domain. We define two binary operations, addition and multiplication, on $Q(R)$ by taking*

$$[(a,b)] + [(c,d)] = [(ad+bc, bd)]$$
$$[(a,b)] \cdot [(c,d)] = [(ac,bd)]$$

for all $[(a,b)], [(c,d)] \in Q(R)$. Then $Q(R)$ is a field with respect to these two operations, called the field of fractions of R.

Proof. Most of the computational details of this proof will be left as an exercise for you, but we will demonstrate that addition is well-defined, and verify the existence of an additive identity, as well as the multiplicative inverse of each non-zero element.

To show that addition is well-defined, suppose $[(a_1,b_1)] = [(a_2,b_2)]$ and $[(c_1,d_1)] = [(c_2,d_2)]$. We need to show that $[(a_1,b_1)] + [(c_1,d_1)] = [(a_2,b_2)] + [(c_2,d_2)]$. First, note from the definitions of equivalence classes that $a_1b_2 = b_1a_2$ and $c_1d_2 = d_1c_2$. Our goal is to show that $[(a_1d_1 + b_1c_1, b_1d_1)] = [(a_2d_2 + b_2c_2, b_2d_2)]$. We verify this as follows:

$$\begin{aligned}(a_1d_1 + b_1c_1)(b_2d_2) &= (a_1b_2)d_1d_2 + (c_1d_2)b_1b_2 \\ &= (b_1a_2)d_1d_2 + (d_1c_2)b_1b_2 \\ &= (a_2d_2 + b_2c_2)(b_1d_1).\end{aligned}$$

This string of equalities proves that $[(a_1d_1 + b_1c_1, b_1d_1)] = [(a_2d_2 + b_2c_2, b_2d_2)]$, as needed.

Next, we show that for any $b \in R \setminus \{0\}$, the element $[(0,b)] \in Q(R)$ is the additive identity. Indeed, for any $[(c,d)] \in Q(R)$, note that $[(0,b)] + [(c,d)] = [(0 \cdot d + bc, bd)] = [(bc,bd)] = [(c,d)]$, where the last equality holds since $bc(d) = bd(c)$. A similar check shows that $[(c,d)] + [(0,b)] = [(c,d)]$. Thus $[(0,b)]$ does indeed act as the additive identity.

Similarly, one can check that $[(1,1)]$ is the multiplicative identity of $Q(R)$. Then, given any element $[(a,b)] \in Q(R)$ with $[(a,b)] \neq [(0,c)]$ for any nonzero $c \in R$, we claim that $[(b,a)]$ is its multiplicative inverse. First, note that $a \neq 0$, since the condition $[(a,b)] \neq [(0,c)]$ reduces to $ac \neq 0$, which implies $a \neq 0$ since c is nonzero. Thus $[(b,a)] \in Q(R)$, and

$$[(b,a)] \cdot [(a,b)] = [(ba,ab)] = [(1,1)],$$

where the last equality follows immediately from the definition of the equivalence relation. Similarly, $[(a,b)] \cdot [(b,a)] = [(1,1)]$, so $[(a,b)]^{-1} = [(b,a)]$.

As mentioned, all the other verifications required to show that $Q(R)$ is a field are left as an exercise! □

In our toy case where $R = \mathbb{Z}$, we can see that $Q(\mathbb{Z})$ behaves exactly as we're used to manipulating $\mathbb{Q}$, except that we're using the notation $[(a,b)]$ for $\frac{a}{b}$. This completes the formal mathematical construction of $\mathbb{Q}$!

But we can say a little bit more about the field $Q(R)$ for arbitrary integral domains as well. We can verify that R is isomorphic to a natural-looking subring of $Q(R)$, and that every element of $Q(R)$ is indeed a quotient of two elements of this subring isomorphic to R. For notational ease, we will start writing $\frac{a}{b}$ in place of the equivalence class $[(a,b)]$, just as we're used to doing for $\mathbb{Q}$.

Theorem 32.1. *Let R be an arbitrary integral domain. Then R is isomorphic to the subring $R_0 = \left\{\frac{r}{1} : r \in R\right\}$ of its field of fractions $Q(R)$. Identifying R with the subring R_0, every nonzero element of R has an inverse in $Q(R)$, and every element of $Q(R)$ may be written $\frac{a}{b} = ab^{-1}$, where a and b belong to the subring R.*

Proof. First, it is necessary to check that R_0 is a subring of $Q(R)$, for which we may apply the Subring Test. Clearly R_0 contains the unity of $Q(R)$, since $\frac{1}{1}$ belongs to R_0. Now, if we are given $\frac{a}{1}, \frac{b}{1} \in R_0$, we can check that the difference and product of these two elements belongs to $Q(R)$ as well. Indeed,

$$\frac{a}{1} - \frac{b}{1} = \frac{a \cdot 1 - b \cdot 1}{1 \cdot 1} = \frac{a-b}{1} \in R_0$$
$$\frac{a}{1} \cdot \frac{b}{1} = \frac{ab}{1 \cdot 1} = \frac{ab}{1} \in R_0.$$

Thus R_0 is indeed a subring of R. Next, we define a function $\sigma : R \to R_0$ given by $\sigma(r) = \frac{r}{1}$ for each $r \in R$. We can check right away that this is a ring homomorphism, since for any $r, s \in R$, we have

$$\sigma(r+s) = \frac{r+s}{1} = \frac{r}{1} + \frac{s}{1} = \sigma(r) + \sigma(s)$$
$$\sigma(rs) = \frac{rs}{1} = \frac{r}{1} \cdot \frac{s}{1} = \sigma(r) \cdot \sigma(s)$$
$$\sigma(1) = \frac{1}{1}.$$

Note that σ is also plainly surjective, by definition of R_0: for any $\frac{r}{1} \in R_0$, we have $\sigma(r) = \frac{r}{1}$. Finally, to check that σ is injective, suppose we have $r \in R$ such that $\sigma(r) = \frac{r}{1} = \frac{0}{1}$. By definition of the equivalence relation, this says $r \cdot 1 = 0 \cdot 1$, or $r = 0$. Thus $\ker \sigma$ is trivial, and this implies σ is injective.

Assured now that σ is an isomorphism, we see that R is isomorphic to R_0, and identifying R with R_0 via this isomorphism, clearly every nonzero element of R has an inverse in $Q(R)$, since $Q(R)$ is a field. Also, it is clear that for any $a, b \in R$, we have $\frac{a}{b} = \frac{a}{1} \cdot \frac{1}{b} = \frac{a}{1} \cdot \left(\frac{b}{1}\right)^{-1} = ab^{-1}$, under the abuse of notation where we write a in place of the element $\frac{a}{1}$. □

It is now possible to apply the field of fractions construction to the integral domains we have studied thus far. We've already seen that $Q(\mathbb{Z})$ can be identified with $\mathbb{Q}$, and on Assignment 10, you will apply this construction to $\mathbb{Z}[i]$. But at least on the surface level, the field of fractions of an integral domain R can be manipulated in familiar ways, and treated as formal fractions with numerators and denominators from R.

Localization

We now wrap up this reading by briefly introducing a construction related to the field of fractions, known as *localization*. It begins with a definition:

Definition 32.1. Let R be an integral domain. A nonempty subset S of R is called a *multiplicative set* if $1 \in S$ and S is closed under multiplication. In other words, if $a \in S$ and $b \in S$, then $ab \in S$.

When we constructed the field of fractions of an integral domain R, we artificially constructed a ring built out of "fractions" from R, with any nonzero element of R as an allowable denominator. This construction can be generalized, where we allow only the elements of a fixed multiplicative set as our denominators. It's a good exercise to quickly convince yourself why the non-zero elements of R form a multiplicative set!

Given an integral domain R and multiplicative set S in R, we can form a new ring called the *localization of R at S*, denoted $S^{-1}R$. The elements are equivalence classes of ordered pairs in $R \times S$, where the equivalence relation is $(a, b) \sim (c, d)$ if $ad = bc$ in R, as before. Rules for addition and multiplication are the same in $S^{-1}R$ as they are in $Q(R)$ (why is $S^{-1}R$ closed under addition and multiplication?). One thing that changes is that $S^{-1}R$ is not necessarily a field. However, R still is isomorphic to a subring of $S^{-1}R$, as above, and every element of S has a multiplicative inverse in $S^{-1}R$.

On Assignment 10, you will get the chance to experiment with one particularly useful example of localization, applied to the integers.

MATH 145 Course Reading 33:
Complex Numbers: An Introduction

December 2, 2020

Here, we continue our study of fields by looking specifically at the field of complex numbers. This field may already be familiar to you, and we've been using complex number calculations all along when manipulating Gaussian integers. However, here we officially construct the field $\mathbb{C}$, given the existence of the field $\mathbb{R}$ of real numbers. We do it from scratch, taking a careful, rigorous approach.

Complex Numbers: Definition

To some extent, the construction of $\mathbb{C}$ from $\mathbb{R}$ looks a bit like the construction of $\mathbb{Q}$ from $\mathbb{Z}$, though without the headache of having to deal with an equivalence relation.

For those who have seen complex numbers before, you treat them informally as sums $a + bi$, where a and b are real numbers. This suggests our formal construction should begin by declaring the elements of $\mathbb{C}$ to be ordered pairs (a, b) in $\mathbb{R} \times \mathbb{R}$. Complex numbers are added component-wise, just like polynomials, so we wish to define a binary operation of addition on $\mathbb{C}$ to be the ordinary one on $\mathbb{R} \times \mathbb{R}$:

$$(a, b) + (c, d) = (a + c, b + d),$$

for all $(a, b), (c, d) \in \mathbb{R} \times \mathbb{R}$. Since this agrees with the ordinary additive structure on $\mathbb{R} \times \mathbb{R}$, we can already conclude that $\mathbb{C}$ is an abelian group under addition.

As for multiplication, we do things a little differently, deviating from the usual multiplicative structure on $\mathbb{R} \times \mathbb{R}$. Informally, complex number multiplication of $(a + bi)$ by $(c + di)$ uses a binomial expansion, subject to the rule that $i^2 = -1$. This informal manipulation tells us the product *ought* to be

$$(a + bi)(c + di) = ac + (ad + bc)i + bdi^2 = (ac - bd) + (ad + bc)i.$$

Thus, we *define* the product of (a, b) and (c, d) in $\mathbb{C}$ to be

$$(a, b) \cdot (c, d) = (ac - bd, ad + bc).$$

Now, we must check that with this binary operation of multiplication on $\mathbb{C}$, this set becomes a field. We begin with the simplest verifications, checking that $(1, 0)$ is a multiplicative identity and that multiplication is commutative. These are done as follows:

$$\begin{aligned}
(a, b) \cdot (1, 0) &= (a \cdot 1 - b \cdot 0, a \cdot 0 + b \cdot 1) = (a, b) \\
(1, 0) \cdot (a, b) &= (1 \cdot a - 0 \cdot b, 1 \cdot b + 0 \cdot a) = (a, b) \\
(a, b) \cdot (c, d) &= (ac - bd, ad + bc) = (ca - db, cb + da) = (c, d) \cdot (a, b).
\end{aligned}$$

Next, let's verify associativity of multiplication, where $(a, b), (c, d), (e, f)$ are arbitrary elements of $\mathbb{C}$:

$$\begin{aligned}
((a, b) \cdot (c, d)) \cdot (e, f) &= (ac - bd, ad + bc) \cdot (e, f) \\
&= ((ac - bd)e - (ad + bc)f, (ac - bd)f + (ad + bc)e) \\
&= (ace - bde - adf - bcf, acf - bdf + ade + bce).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
(a, b) \cdot ((c, d) \cdot (e, f)) &= (a, b) \cdot (ce - df, cf + de) \\
&= (a(ce - df) - b(cf + de), a(cf + de) + b(ce - df)) \\
&= (ace - adf - bcf - bde, acf + ade + bce - bdf).
\end{aligned}$$

Upon comparison of the results, we can verify immediately that $((a,b)\cdot(c,d))\cdot(e,f) = (a,b)\cdot((c,d)\cdot(e,f))$, proving that multiplication in $\mathbb{C}$ is associative.

To check that every non-zero element has a multiplicative inverse, we start with informal manipulations, then verify them formally. So given $(a,b) \in \mathbb{C}$, corresponding to $a+bi$, we would like to calculate the ordered pair corresponding to $\frac{1}{a+bi}$. We do this by multiplying numerator and denominator by the *complex conjugate* $a-bi$. The result is

$$\frac{1}{a+bi} = \frac{a-bi}{(a+bi)(a-bi)} = \frac{a-bi}{a^2+b^2} = \frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i.$$

Thus we take the ordered pair $(a/(a^2+b^2), -b/(a^2+b^2))$ as our candidate inverse to (a,b). Note that $a^2+b^2 \neq 0$ if $(a,b) \neq (0,0)$, so this candidate inverse is well-defined.

The following computation officially verifies that $(a,b)^{-1} = (a/(a^2+b^2), -b/(a^2+b^2))$:

$$\begin{aligned}(a,b)\cdot(a/(a^2+b^2), -b/(a^2+b^2)) &= (a^2/(a^2+b^2) + b^2/(a^2+b^2), -ab/(a^2+b^2) + ab/(a^2+b^2)) \\ &= ((a^2+b^2)/(a^2+b^2), 0) \\ &= (1,0).\end{aligned}$$

To complete the verification that $\mathbb{C}$ is a field, we would also have to check the distributive law, which I leave as an exercise for you!

Now, if we define i to be the ordered pair $(0,1) \in \mathbb{C}$, note that $i^2 = (0,1)\cdot(0,1) = (0\cdot 0 - 1\cdot 1, 0\cdot 1 + 1\cdot 0) = (-1,0) = -(1,0)$, which verifies the well-known complex number relation $i^2 = -1$.

Now that the ordered pair definition is complete, we will (almost always) abuse notation and write a in place of the complex number $(a,0)$. You can then immediately check that for any $a,b \in \mathbb{R}$, we have $(a,b) = (a,0) + (b,0)(0,1)$, which justifies our writing $a+bi$ in place of (a,b).

Complex Number Constructions and Properties

Now that we've defined the field $\mathbb{C}$ carefully and justified our use of the $a+bi$ notation for complex numbers, it's time to introduce and state a number of familiar properties of these numbers.

Definition 33.1. Let $z \in \mathbb{C}$, and write $z = a+bi$ for some $a,b \in \mathbb{R}$.

- The form $a+bi$ is often called the *standard form* for z.
- The real number a is called the *real part* of z and is denoted $\operatorname{Re} z$.
- The real number b is called the *imaginary part* of z and is denoted $\operatorname{Im} z$.
- The *complex conjugate* of z is the complex number $\overline{z} = a-bi$.
- The *modulus*, or *absolute value* of z is the quantity $|z| = \sqrt{a^2+b^2}$.

For example, we have $\operatorname{Re}(2+3i) = 3$, and $|4+i| = \sqrt{4^2+1^2} = \sqrt{17}$. The following are fundamental properties often used when working with complex numbers, which are left as exercises for you!

- For all $z,w \in \mathbb{C}$, we have $\overline{z+w} = \overline{z} + \overline{w}$, $\overline{z-w} = \overline{z} - \overline{w}$, $\overline{zw} = \overline{z}\cdot\overline{w}$, and $\overline{1/z} = 1/\overline{z}$. (In our fancy ring-theory language, we might say that the function $\phi : \mathbb{C} \to \mathbb{C}$ given by $\phi(z) = \overline{z}$ is a ring homomorphism!)
- For all $z \in \mathbb{C}$, we have $\overline{\overline{z}} = z$.
- For all $z \in \mathbb{C}$, we have $z + \overline{z} = 2\operatorname{Re} z$.
- For all $z \in \mathbb{C}$, we have $z - \overline{z} = 2i\operatorname{Im} z$.

- For all $z \in \mathbb{C}$, we have $z \cdot \overline{z} = |z|^2$.
- For all $z \in \mathbb{C}$, we have $\operatorname{Re} z \leq |\operatorname{Re} z| \leq |z|$ and $\operatorname{Im} z \leq |\operatorname{Im} z| \leq |z|$.
- For all $z, w \in \mathbb{C}$, we have $|zw| = |z||w|$ and $|\overline{z}| = |z|$.

One consequence of these elementary properties is the *triangle inequality*, which you might be familiar with over the real numbers:

Theorem 33.1. *For all $z, w \in \mathbb{C}$, we have $|z + w| \leq |z| + |w|$.*

Proof. To begin with, we write $|z + w|^2$ in another way and simplify:

$$\begin{aligned} |z + w|^2 &= (z + w) \cdot \overline{(z + w)} \\ &= (z + w)(\overline{z} + \overline{w}) \\ &= z\overline{z} + z\overline{w} + w\overline{z} + w\overline{w} \\ &= |z|^2 + |w|^2 + (z\overline{w} + \overline{z\overline{w}}) \\ &= |z|^2 + |w|^2 + 2\operatorname{Re}(z\overline{w}). \end{aligned}$$

Now, note that $\operatorname{Re}(z\overline{w}) \leq |z\overline{w}| = |z| \cdot |\overline{w}| = |z||w|$. Hence, from the above we find that

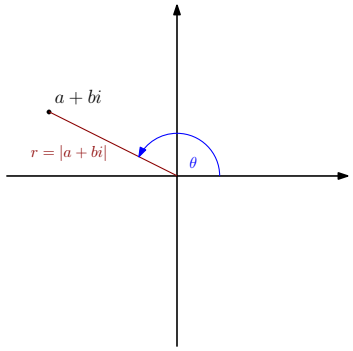
$$|z + w|^2 = |z|^2 + |w|^2 + 2\operatorname{Re}(z\overline{w}) \leq |z|^2 + |w|^2 + 2|z||w| = (|z| + |w|)^2.$$

Taking positive square roots of the inequality $|z + w|^2 \leq (|z| + |w|)^2$, we get the triangle inequality, as desired. □

Polar Form of a Complex Number

We have already seen the standard form representation $z = a + bi$ of a complex number. Another popular notation for complex numbers (particularly when extracting nth roots) is the *polar form*. To begin, we introduce the complex exponential function, taking $e^{i\theta} = (\cos(\theta) + i\sin(\theta))$ for any $\theta \in \mathbb{R}$. Note that for any $\theta \in \mathbb{R}$, we have $|e^{i\theta}| = \sqrt{\cos^2(\theta) + \sin^2(\theta)} = \sqrt{1} = 1$.

Using the natural identification of $\mathbb{C}$ with $\mathbb{R} \times \mathbb{R}$, we can think of every complex number $a + bi$ in standard form as picking out a point (a, b) in the Cartesian plane. But if we use polar coordinates for that same point, we are led to the polar form of a complex number. We write $a + bi = re^{i\theta}$, where $r = |a + bi|$ is the distance from the origin to the point (a, b), and θ is any value (in radian measure) marking the angle from the positive x-axis to the line segment from (a, b) to the origin. This is illustrated in the image below:



The angle θ is often called the *argument* of $a + bi$. Notice that there is not a unique value for the argument: if θ is one possible value for the argument, then so is $\theta + 2k\pi$ for any integer k. When given a complex number in standard form, converting it to polar form can be accomplished using standard techniques from high school trigonometry. For an explicit example:

Example 33.1. Let's convert $-1-i$ to polar form. First, we compute $r = |-1-i| = \sqrt{(-1)^2 + (-1)^2} = \sqrt{2}$. Now, we wish to find $\theta \in \mathbb{R}$ for which $-1 - i = re^{i\theta} = \sqrt{2}(\cos\theta + i\sin\theta)$. Comparing real and imaginary parts, we see that we must have

$$-1 = \sqrt{2}\cos(\theta)$$
$$-1 = \sqrt{2}\sin(\theta).$$

Thus we need to choose θ such that $\cos\theta = \sin\theta = -\frac{1}{\sqrt{2}}$. Using special triangles, the unit circle, or whatever your favourite trigonometry technique is, we see that $\theta = \frac{5\pi}{4}$ will do the trick. Thus we may write $-1 - i = \sqrt{2}e^{i(5\pi/4)}$.

While standard form makes it easy to add complex numbers, polar form makes multiplication particularly easy. This is explained in the following theorem:

Theorem 33.2. *Suppose $z_1, z_2 \in \mathbb{C}$ are given in polar form, say*

$$z_1 = r_1 e^{i\theta_1}$$
$$z_2 = r_2 e^{i\theta_2}$$

for some $\theta_1, \theta_2 \in \mathbb{R}$ and nonnegative real numbers r_1 and r_2. Then a polar form for the product is

$$z_1 z_2 = (r_1 r_2) e^{i(\theta_1 + \theta_2)}.$$

Proof. This result comes down to some well-known trigonometric identities. We know

$$z_1 = r_1 e^{i\theta_1} = r_1(\cos\theta_1 + i\sin\theta_1)$$
$$z_2 = r_2 e^{i\theta_2} = r_2(\cos\theta_2 + i\sin\theta_2),$$

and multiplying together the complex numbers in standard form gives

$$\begin{aligned} z_1 z_2 &= r_1 r_2(\cos\theta_1 + i\sin\theta_1)(\cos\theta_2 + i\sin\theta_2) \\ &= r_1 r_2(\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2 + i(\sin\theta_1\cos\theta_2 + \cos\theta_1\sin\theta_2)) \\ &= r_1 r_2(\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)). \end{aligned}$$

Here, trigonometric identities were applied in the last line. This verifies that $z_1 z_2 = (r_1 r_2)e^{i(\theta_1+\theta_2)}$, as claimed. □

This has an immediate corollary for taking powers of a complex number:

Corollary 33.1 (de Moivre's Theorem)**.** *Let $z \neq 0$ be a complex number, and write z in polar form, say $z = re^{i\theta}$ for some real number θ and positive real number r. For all $n \in \mathbb{Z}$, we have*

$$z^n = r^n e^{i(n\theta)}.$$

Proof. First, we show the result for all $n \in \mathbb{N}$ by induction on n. For the base case $n = 0$, we have $z^0 = 1$ by definition, while $r^0 e^{i(0\theta)} = e^{i\cdot 0} = \cos 0 + i\sin 0 = 1$. Thus the base case holds, as needed.

Now, assume the result is true for a given $n \in \mathbb{N}$, so that $z^n = r^n e^{i(n\theta)}$. Applying Theorem 33.2, we find that

$$z^{n+1} = z^n \cdot z = (r^n e^{i(n\theta)})(re^{i\theta}) = (r^n \cdot r)(e^{i(n\theta+\theta)}) = r^{n+1}e^{i((n+1)\theta)}.$$

This completes the proof by induction, so that the corollary is valid for all $n \in \mathbb{N}$.

Finally, we prove that for any positive integer n, $z^{-n} = r^{-n}e^{i(-n\theta)}$. By uniqueness of inverses, it is enough to show that $r^{-n}e^{i(-n\theta)}z^n = 1$. This follows directly from Theorem 33.2 and the first part of this proof, since $r^{-n}e^{i(-n\theta)}z^n = (r^{-n}e^{i(-n\theta)})(r^n e^{i(n\theta)}) = r^0 e^{i(0)} = 1$. □

This theorem is particularly useful when we wish to extract the nth root of a complex number, a procedure that we will carry out in the next reading.

MATH 145 Course Reading 34: The Fundamental Theorem of Algebra and Solving Equations over $\mathbb{C}$

December 4, 2020

Now that we have properly introduced the complex numbers, we discuss procedures for solving polynomial equations over the complex numbers. One of the main motivations for constructing $\mathbb{C}$ in the first place is to have the ability to solve any polynomial equation. This is encapsulated in the Fundamental Theorem of Algebra. In essence, it tells us that every non-trivial polynomial equation with complex number coefficients has a complex solution, which has numerous theoretical advantages. Unfortunately, we will not be able to *prove* this theorem here (this is traditionally done in a course on complex analysis), but we will at least have the opportunity to explore the theorem to some degree.

After we've introduced the Fundamental Theorem of Algebra, we narrow in to a discussion of the solution of two particular types of complex number equations: equations of the form $z^n - a = 0$, and quadratic equations, of the form $az^2 + bz + c = 0$, where $a, b, c \in \mathbb{C}$.

Algebraically Closed Fields and the Fundamental Theorem of Algebra

One of the main historical motivations for enlargements to our number system is the ability to describe solutions to an increasingly large collection of polynomial equations. For instance, if we only had $\mathbb{N}$, not all equations of the form $x + a = b$ (for $a, b \in \mathbb{N}$) have a solution within $\mathbb{N}$. If we enlarge our system to $\mathbb{Z}$, all linear equations $x + a = b$ (where $a, b \in \mathbb{Z}$) *do* have solutions in $\mathbb{Z}$. But then equations of the form $ax + b = c$ (for $a, b, c \in \mathbb{Z}$) do not always have solutions until we enlarge $\mathbb{Z}$ to the field $\mathbb{Q}$. But $\mathbb{Q}$ is missing solutions to equations like $x^2 = 2$, which is remedied by enlarging the system to $\mathbb{R}$. But even $\mathbb{R}$ does not contain the solution to all polynomial equations over $\mathbb{R}$, since the equation $x^2 + 1 = 0$ does not have a real solution.

What is somewhat remarkable is that after we enlarge our number system to $\mathbb{C}$, there is no need to expand any further: we now have solutions to every possible polynomial equation. This is encapsulated in the following theorem:

Theorem 34.1 (Fundamental Theorem of Algebra)**.** *Let $f \in \mathbb{C}[x]$ be a non-constant polynomial. Then f has a root in $\mathbb{C}$. In other words, there is some $c \in \mathbb{C}$ such that $f(c) = 0$.*

Somewhat interestingly, there is no completely algebraic proof of this algebraic result; every proof appeals to the subject of *analysis* in some way. Perhaps the simplest, clearest proofs are in the subject area of *complex analysis*, which you would likely encounter in PMATH 352 here at Waterloo.

The Fundamental Theorem of Algebra dovetails nicely with the notion of an *algebraically closed field*. We have the following definition:

Definition 34.1. A field F is called *algebraically closed* if every non-constant polynomial $f \in F[x]$ has a root in F.

So the Fundamental Theorem of Algebra amounts to the assertion that $\mathbb{C}$ is algebraically closed.

Another common way of formulating the definition of algebraically closed field is given in the following result:

Theorem 34.2. *A field F is algebraically closed if and only if every non-constant polynomial $f \in F[x]$ may be factored as a product of linear polynomials*

$$f = c(x - a_1)(x - a_2) \cdots (x - a_n),$$

where $c, a_1, \ldots, a_n \in F$, and $n = \deg f$.

Proof. First, suppose that F is algebraically closed, according to our definition. We prove that every non-constant polynomial $f \in F[x]$ may be factored as a product of linear factors by induction on $n = \deg f$. In the base case $n = 1$, we have $f = ax + b$ for some $a, b \in F$ with $a \neq 0$, and so $f = a(x + a^{-1}b)$ gives the factorization of f in the required form.

Now assume the result is true for all polynomials of degree n in $F[x]$, where $n \geq 1$. Suppose $f \in F[x]$ has degree $n + 1$. Since F is algebraically closed, f has a root in F, which we will call a_{n+1}. By the Factor Theorem, Theorem 30.3, the polynomial $(x - a_{n+1})$ divides f, so that we may write $f = g(x - a_{n+1})$ for some polynomial $g \in F[x]$. Notice that $\deg g = n$, and so by induction hypothesis, g may be factored

$$g = c(x - a_1)(x - a_2) \cdots (x - a_n),$$

where $c, a_1, \ldots, a_n \in F$. In turn, this means we have found the desired factorization of f:

$$f = g(x - a_{n+1}) = c(x - a_1)(x - a_2) \cdots (x - a_{n+1}).$$

For the converse, suppose F is a field for which every non-constant polynomial in $F[x]$ factors as a product of linear polynomials. Now let $f \in F[x]$ be a non-constant polynomial, of degree $n \geq 1$. By assumption, we have

$$f = c(x - a_1)(x - a_2) \cdots (x - a_n),$$

where $c, a_1, \ldots, a_n \in F$. Note now that f certainly has a root in F; indeed, all of $a_1, a_2, \ldots, a_n$ are roots. □

Working in an algebraically closed field has a number of technical advantages. These fields are very useful in linear algebra (studied in MATH 146), and are at the heart of classical algebraic geometry (which you would study in PMATH 464 here at Waterloo). The details of *why* these fields are useful in these subjects is beyond the scope of our course, but it's worth at least noting their value!

Solving Equations in $\mathbb{C}$

The Fundamental Theorem of Algebra asserts that it is theoretically possible to find a complex root of every non-constant complex polynomial. But how do we do this in practice? From a strictly algebraic standpoint, there are unfortunately other obstacles standing in the way. Even for polynomials $f \in \mathbb{Q}[x]$, there are such polynomials of degree 5 for which there is no formula for the roots purely in terms of the field operations on $\mathbb{Q}$ and the extraction of nth roots. This is often more colloquially known as the *insolubility of the quintic*.

However, there are still certain types of commonly occurring polynomial equations over $\mathbb{C}$ that we *can* write down a satisfying answer to. One of which is the extraction of complex nth roots, corresponding to solving the equation $z^n = a$ over $\mathbb{C}$.

To solve $z^n = a$, where $a \neq 0$, suppose we write $a = re^{i\theta}$, where $r > 0$ and $\theta \in \mathbb{R}$. Assuming a solution z exists to this equation, we can write it in polar form, say $z = se^{i\phi}$, where $s > 0$ and $\phi \in \mathbb{R}$. Taking a power of z in polar form, we get

$$z^n = s^n e^{i(n\phi)} = a = re^{i\theta}.$$

Comparing these two polar forms, we see that we must have $s^n = r$ (the two sides of the equation must have the same absolute value), and we see that $n\phi = \theta + 2k\pi$ for some integer k (since the argument of a complex number is only well-defined up to a multiple of 2π). We thus conclude that we must have $s = \sqrt[n]{r}$, the unique positive real nth root of r, and $\phi = \frac{\theta}{n} + \frac{2k\pi}{n}$ for some integer k. Notice that only taking $k = 0, 1, 2, \ldots, n-1$ yield distinct values for the argument of z, so there are exactly n possible nth roots. And we can check immediately that for $k \in \{0, 1, \ldots, n-1\}$, we have

$$(\sqrt[n]{r}e^{i(\theta/n + 2k\pi/n)})^n = re^{i(\theta + 2k\pi)} = re^{i\theta} = a.$$

We summarize our findings below:

Proposition 34.1. *Let a be a nonzero complex number, and write a in polar form, say*

$$a = re^{i\theta}$$

for some real number $r > 0$ and real number θ. Then for any positive integer n, there are exactly n distinct solutions z to the equation $z^n = a$, given by

$$z = \sqrt[n]{r}e^{i(\theta/n + 2k\pi/n)},$$

where $k \in \{0, 1, \ldots, n-1\}$.

Let's do one numerical example:

Example 34.1. Let's find all complex solutions z to $z^4 = (1+i)$. The first step is to convert $1+i$ to polar form. Note that $r = |1+i| = \sqrt{1^2+1^2} = \sqrt{2}$. To find θ, we want to ensure that $\sqrt{2}(\cos\theta + i\sin\theta) = \sqrt{2}e^{i\theta} = 1+i$. This tells us $\sqrt{2}\cos\theta = 1$ and $\sqrt{2}\sin\theta = 1$, so $\cos\theta = \sin\theta = \frac{1}{\sqrt{2}}$. In turn, we see from here that $\theta = \pi/4$ is one solution.

Thus the solutions z to $z^4 = (1+i)$ will take the form

$$z = \sqrt[4]{\sqrt{2}}e^{i[(\pi/4)/4 + (2k\pi)/4]} = \sqrt[8]{2}e^{i(\pi/16 + k\pi/2)},$$

where $k \in \{0, 1, 2, 3\}$. You are encouraged to plot these four solutions as points in the plane, to see what is going on geometrically – the result is actually quite aesthetically pleasing!

Finally, let's turn to solving quadratic equations over $\mathbb{C}$. In fact, we will do something more general and show that the quadratic formula remains valid in any field, provided we can extract the appropriate square root.

Theorem 34.3. *Let F be a field with $\operatorname{char} F \neq 2$, and suppose we are given $a, b, c \in F$ with $a \neq 0$. Suppose further that there is an element $y \in F$ such that $y^2 = b^2 - 4ac$. Then the quadratic equation $ax^2 + bx + c = 0$ has solutions given exactly by*

$$x = (-b \pm y) \cdot (2a)^{-1}.$$

Proof. First, we verify that both values for x given above really are solutions to the equation. This is a direct computation:

$$\begin{aligned} a((-b \pm y)\cdot(2a)^{-1})^2 + b((-b \pm y)\cdot(2a)^{-1}) + c &= (4a)^{-1}(b^2 \mp 2by + y^2) + (2a)^{-1}(-b^2 \pm by) + c \\ &= (4a)^{-1}(b^2 \mp 2by + (b^2 - 4ac) - 2b^2 \pm 2by + 4ac) \\ &= 0. \end{aligned}$$

Conversely, suppose that $x \in F$ satisfies $ax^2 + bx + c = 0$. We use the familiar technique of "completing the square" to get

$$\begin{aligned} a(x^2 + a^{-1}bx) + c &= 0 \\ a(x^2 + a^{-1}bx + (2a)^{-2}b^2) - (4a)^{-1}b^2 + c &= 0 \\ a(x + (2a)^{-1}b)^2 + (4a)^{-1}(4ac - b^2) &= 0 \\ a(x + (2a)^{-1}b)^2 &= (b^2 - 4ac)\cdot(4a)^{-1} \\ (x + (2a)^{-1}b)^2 &= (b^2 - 4ac)\cdot(2a)^{-2} \end{aligned}$$

From here, we see that the two square roots of the right-hand side are exactly $y\cdot(2a)^{-1}$ and $-y\cdot(2a)^{-1}$, since $y^2 = b^2 - 4ac$. Thus

$$\begin{aligned} x + (2a)^{-1}b &= \pm y\cdot(2a)^{-1} \\ x &= (-b \pm y)\cdot(2a)^{-1}, \end{aligned}$$

showing that x must be of the form presented in the theorem. □

In particular, since we can always extract square roots in $\mathbb{C}$, we find that we can always solve a quadratic equation over $\mathbb{C}$ explicitly (which we knew we could theoretically do already by the Fundamental Theorem of Algebra).

Example 34.2. Let's find all complex solutions to

$$(2+i)z^2 - 2z + (2-i) = 0.$$

We begin by finding solutions y to $y^2 = (-2)^2 - 4(2+i)(2-i) = 4 - 4(5) = -16$. Solving, we get $y = 4i$ as one square root. Thus the two solutions to the quadratic equation are

$$z_1 = \frac{-(-2)+y}{2(2+i)} = \frac{2+4i}{2(2+i)} = \frac{1+2i}{2+i} = \frac{(1+2i)(2-i)}{(2+i)(2-i)} = \frac{4+3i}{5}$$
$$z_2 = \frac{-(-2)-y}{2(2+i)} = \frac{2-4i}{2(2+i)} = \frac{1-2i}{2+i} = \frac{(1-2i)(2-i)}{(2+i)(2-i)} = \frac{-5i}{5} = -i.$$

MATH 145 Course Reading 35:
Factoring Polynomials with Coefficients in a Field

December 7, 2020

In our final reading, we now return to studying the polynomial rings $F[x]$, where F is a field. In particular, we introduce a factorization theory for polynomials, mimicking how prime factorization works over $\mathbb{Z}$, and spend some time discussing how to split up polynomials over various fields as products of *irreducible* polynomials. As we will see, factorization behaviour works very differently depending on the particular field F we happen to be dealing with.

Irreducible Polynomials

Over the integers, the prime numbers form the "factorization building blocks", in the sense that every integer can be broken up as a product of prime factors. The same phenomenon applies to polynomial rings over a field. We make the following definition, in analogy with the definition of prime number:

Definition 35.1. Let F be a field, and let $f \in F[x]$ be a non-constant polynomial. We say that f is *reducible* if f admits a proper factorization $f = gh$, where $g, h \in F[x]$ and $\deg(g), \deg(h) \geq 1$. Otherwise, we say that f is *irreducible*. In other words, f is irreducible if whenever we have a factorization $f = gh$ with $g, h \in F[x]$, it must hold that either g or h is a constant polynomial.

One question you might immediately have is: how can you tell if a polynomial is irreducible? This question is just as subtle in general as the question "how can you tell if an integer is prime?" But there are a few things we can say immediately, which apply no matter what the field F happens to be. Firstly, every linear polynomial in $F[x]$ (i.e. every polynomial of degree 1) is automatically irreducible (why?)

For polynomials of larger degree, the following result can be useful:

Proposition 35.1. *Let F be an arbitrary field, and let $f \in F[x]$. If $\deg f \geq 2$ and f is irreducible, then f has no roots in F. Conversely, if $\deg f = 2$ or $\deg f = 3$ and f has no roots in F, then f is irreducible.*

Proof. Suppose that $\deg f \geq 2$ and that f is irreducible. Suppose now to the contrary that f has a root $c \in F$. By the Factor Theorem, this means we can write $f = (x - c)h$ for some polynomial $h \in F[x]$. Note also that $\deg h = \deg f - 1 \geq 1$. This means we have exhibited a factorization of f as a product of two non-constant polynomials, contradicting irreducibility of f. We conclude that f must have no roots in F after all.

Conversely, suppose $\deg f = 2$ or $\deg f = 3$, and that f has no roots in F. Suppose also to the contrary that f is reducible. Then we can write $f = gh$, where $g, h \in F[x]$ are non-constant polynomials. Taking degrees, we have $\deg f = \deg g + \deg h$, and our constraint on $\deg f$ forces either $\deg g = 1$ or $\deg h = 1$. Either way, f has a linear factor, and this linear factor must have a root in F. In turn, this means f has a root in F, contradicting our hypothesis. Thus f must be irreducible after all. □

You might immediately ask: can we extend the converse to polynomials of higher degree? The answer is no: there are already polynomials of degree 4 over certain fields that have no roots, but are still reducible (can you find examples?)

The result we just proved can be immediately applied to show that $x^2 + 1 \in \mathbb{R}[x]$ is irreducible, since this polynomial is of degree 2 and has no real roots. We can also derive the following corollary, applicable to all algebraically closed fields:

Corollary 35.1. *Suppose F is an algebraically closed field. Then a non-constant polynomial $f \in F[x]$ is irreducible if and only if $\deg f = 1$.*

Proof. We already noted that a linear polynomial over any field is irreducible. Conversely, if $f \in F[x]$ is irreducible and $\deg f > 1$, then f has no roots in F by Proposition 35.1. But since F is algebraically closed, this cannot occur by definition, so there are no irreducible polynomials of degree larger than 1 in $F[x]$. □

In particular, the above corollary applies to $\mathbb{C}[x]$, courtesy of the Fundamental Theorem of Algebra.

For a little more interesting application of Proposition 35.1, consider the following example:

Example 35.1. The polynomial $f = x^3 + x + [1] \in (\mathbb{Z}/2\mathbb{Z})[x]$ is irreducible. By Proposition 35.1, it is enough to check that this degree 3 polynomial has no roots in $(\mathbb{Z}/2\mathbb{Z})$. But this field has only two elements, so we can very quickly check this to be the case:

$$f([0]) = [0]^3 + [0] + [1] = [1]$$
$$f([1]) = [1]^3 + [1] + [1] = [1].$$

Thus f has no roots in $(\mathbb{Z}/2\mathbb{Z})[x]$, proving that f is irreducible.

Irreducible Polynomials in $\mathbb{R}[x]$

We have now said about as much as we can say in general about polynomial factorization over a field. More can be said when we specialize our choice of the field F. If we take $\mathbb{R}$ as our field, it turns out the only irreducible polynomials are linear and quadratic. Along with the Fundamental Theorem of Algebra, we will need the following well-known result:

Lemma 35.1. *Suppose $f \in \mathbb{R}[x]$ is an arbitrary polynomial. If $c \in \mathbb{C}$ is a complex root of f, then so is its complex conjugate $\overline{c}$.*

Proof. First, we write out an expression for f:

$$f = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

where $a_0, a_1, \ldots, a_n \in \mathbb{R}$ and $a_n \neq 0$. Knowing that c is a root of f tells us that

$$f(c) = a_n c^n + a_{n-1}c^{n-1} + \cdots + a_1 c + a_0 = 0.$$

If we apply the complex conjugation homomorphism to the above equation, and use the fact that the complex conjugate of a real number is that same real number, we end up with

$$\begin{aligned} 0 &= \overline{0} \\ &= \overline{a_n c^n + a_{n-1}c^{n-1} + \cdots + a_1 c + a_0} \\ &= \overline{a_n} \cdot \overline{c}^n + \overline{a_{n-1}} \cdot \overline{c}^{n-1} + \cdots + \overline{a_1} \cdot \overline{c} + \overline{a_0} \\ &= a_n \overline{c}^n + a_{n-1}\overline{c}^{n-1} + \cdots + a_1 \overline{c} + a_0 \\ &= f(\overline{c}). \end{aligned}$$

This verifies that $\overline{c}$ is a root of f as well. □

We are now ready to prove our result on irreducible polynomials over $\mathbb{R}[x]$:

Theorem 35.1. *Let $f \in \mathbb{R}[x]$ be a non-constant polynomial. Then f is irreducible in $\mathbb{R}[x]$ if and only if $\deg f = 1$, or $\deg f = 2$ and f has no real roots.*

Proof. If $\deg f = 1$ or $\deg f = 2$ and f has no real roots, then our work previously in this reading already tells us that f is irreducible. Conversely, if $f \in \mathbb{R}[x]$ is an irreducible, non-constant polynomial, we may also consider f as a polynomial in $\mathbb{C}[x]$ (possibly no longer irreducible). By the Fundamental Theorem of Algebra, f has a complex root c. If $c \in \mathbb{R}$, then f has a real root, and irreducibility of f forces $\deg f = 1$. Otherwise, $c \notin \mathbb{R}$, and so its complex conjugate $\overline{c}$ is different from c and also a root of f.

Applying the Factor Theorem with the roots c and $\overline{c}$, we find that we may write $f = [(x - c)(x - \overline{c})]h$ for some polynomial $h \in \mathbb{C}[x]$. If we set $g = (x - c)(x - \overline{c})$, notice that

$$g = x^2 - (c + \overline{c})x + (c\overline{c}) = x^2 - (2 \operatorname{Re} c)x + |c|^2 \in \mathbb{R}[x].$$

Thus f is divisible by the polynomial $g \in \mathbb{R}[x]$. If we carry out division with remainder of f by g in $\mathbb{R}[x]$, we get $f = gh' + r$ for some $h', r \in \mathbb{R}[x]$ and $\deg r < 2$. But if we do the same division with remainder over $\mathbb{C}[x]$, we know we get $f = gh + 0$. It is not hard to argue that the quotient and remainder are unique when the division is done over $\mathbb{C}[x]$, and so we conclude by uniqueness that $h = h' \in \mathbb{R}[x]$.

Thus we have a factorization $f = gh$ in $\mathbb{R}[x]$, where $\deg g = 2$. By irreducibility of f, h must be a constant polynomial. In turn, up to a constant factor, f is equal to g, which is of degree 2 with no real roots, so the same is true of f. □

Irreducible Polynomials in $\mathbb{Q}[x]$

When it comes to determining whether a polynomial in $\mathbb{Q}[x]$ is irreducible, the story is not quite so neat as it is in $\mathbb{R}[x]$. One elementary result we can use is the *Rational Roots Theorem*:

Theorem 35.2 (Rational Roots Theorem)**.** *Let $f \in \mathbb{Q}[x]$ be a non-constant polynomial, and suppose $r \in \mathbb{Q}$ is a root of f. Suppose further that we have*

$$f = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

where $a_0, a_1, \ldots, a_n \in \mathbb{Z}$ and $a_n \neq 0$. If we write $r = \frac{p}{q}$, where $p, q \in \mathbb{Z}$ and $\gcd(p, q) = 1$, then we must have $q \mid a_n$ and $p \mid a_0$ in $\mathbb{Z}$.

Proof. Plugging in the fact that $\frac{p}{q}$ is a root of f, we get

$$a_n \left(\frac{p}{q}\right)^n + a_{n-1} \left(\frac{p}{q}\right)^{n-1} + \cdots + a_1 \left(\frac{p}{q}\right) + a_0 = 0.$$

Multiplying through by q^n and re-arranging, we end up with

$$a_0 q^n = -(a_n p^n + a_{n-1} p^{n-1} q + \cdots + a_1 p q^{n-1}) = -p(a_n p^{n-1} + a_{n-1} p^{n-2} q + \cdots + a_1 q^{n-1}),$$

which shows $p \mid a_0 q^n$. Since $\gcd(p, q) = 1$, it follows that $\gcd(p, q^n) = 1$, and so $p \mid a_0$ by Theorem 29.3.

Similarly, isolating for $a_n p^n$ instead of $a_0 q^n$, we deduce that $q \mid a_n p^n$, and since $\gcd(q, p^n) = 1$, we get $q \mid a_n$ again by Theorem 29.3. □

Let's put this theorem to use to show that a cubic polynomial in $\mathbb{Q}[x]$ is irreducible:

Example 35.2. The polynomial $f = x^3 - 2x + 5 \in \mathbb{Q}[x]$ is irreducible. Since f has degree 3, it is enough to show that f has no roots in $\mathbb{Q}$. By the Rational Roots Theorem, if $\frac{p}{q}$ is a rational root of f with $\gcd(p, q) = 1$, then $q \mid 1$ and $p \mid 5$. This implies $q \in \{-1, 1\}$ and $p \in \{-1, 1, -5, 5\}$. This leads to four distinct possibilities for $\frac{p}{q}$, namely $1, -1, 5, -5$. We try each to check whether it is a rational root of f:

$$\begin{aligned} f(1) &= 4 \neq 0 \\ f(-1) &= 6 \neq 0 \\ f(5) &= 120 \neq 0 \\ f(-5) &= -110 \neq 0. \end{aligned}$$

We conclude that f does not in fact have any rational roots, and so f is irreducible in $\mathbb{Q}[x]$.

One further application of the Rational Roots Theorem lies in showing that certain real numbers are irrational. Given a candidate $\alpha \in \mathbb{R}$ that we wish to show is irrational, we can try to do two things:

(1) Find a non-zero polynomial f with integer coefficients for which α is a root.

(2) Use the Rational Roots Theorem to show that f has no rational roots.

These two steps combined show that α must be irrational. To give a quick example:

Example 35.3. Let's show that $\alpha = \sqrt{2} + \sqrt{3}$ is irrational. First, we track down a polynomial with integer coefficients having α as a root. Note that

$$\begin{aligned} \alpha^2 &= 2 + 2\sqrt{6} + 3 \\ \alpha^2 &= 5 + 2\sqrt{6} \\ \alpha^2 - 5 &= 2\sqrt{6} \\ (\alpha^2 - 5)^2 &= (2\sqrt{6})^2 = 24 \\ \alpha^4 - 10\alpha^2 + 25 &= 24 \\ \alpha^4 - 10\alpha^2 + 1 &= 0. \end{aligned}$$

Thus, α is a root of the integer polynomial $f = x^4 - 10x^2 + 1$. We now show that this polynomial has no rational roots. By the Rational Roots Theorem, the only candidate rational roots of f are 1 and -1, and we can check right away that $f(1) = f(-1) = -8 \neq 0$. We conclude that α, being a root of f, must not be rational!

Another great standard exercise is to apply the Rational Roots Theorem to show that for any prime p and integer $n \geq 2$, the number $\sqrt[n]{p}$ is irrational.

There are other irreducibility tests out there for polynomials in $\mathbb{Q}[x]$, including the *modular irreducibility test* and *Eisenstein's criterion*, but both of these results hinge on a key algebraic fact called *Gauss' lemma*. Proving this lemma is somewhat involved, and therefore, you are encouraged to explore on your own if you would like to find out more!

And with that, our MATH 145 course is complete! This course took you on a whirlwind tour through axiomatic set theory and abstract algebra. While we have scratched the surface of both of these topics, there is much more to learn here, even at the undergraduate level! But you should now have a solid mathematical foundation for further studies in algebraic subjects, particularly linear algebra in MATH 146.