MUSA 74: Transition to Upper Division Mathematics

Mathematics Undergraduate Student Association

Spring 2023

Contents

How strange to actually have to see the path of your journey in order to make it.

—Neal Shusterman, [Shu16]

Contents 2

0 Preface 5
- 0.1 For the Reader . . . 5
- 0.2 For the Teacher . . . 6
- 0.3 Transition to Upper Division . . . 6
- 0.4 Acknowledgements . . . 8

I Proofs 10

1 Logic and Proofs 11
- 1.1 Week 1: Propositional Logic . . . 11
  - 1.1.1 Compound Propositions . . . 11
  - 1.1.2 Truth Tables . . . 13
  - 1.1.3 Problems . . . 14
- 1.2 Week 2: Equivalences, Predicates, and Quantifiers . . . 14
  - 1.2.1 Logical Equivalences . . . 14
  - 1.2.2 Propositional Functions . . . 17
  - 1.2.3 Quantifiers . . . 18
  - 1.2.4 Problems . . . 19
- 1.3 Week 3: Introduction to Proofs . . . 21
  - 1.3.1 Proof Basics . . . 21
  - 1.3.2 Contradiction and Contraposition . . . 24
  - 1.3.3 Uniqueness Proofs . . . 25
  - 1.3.4 Problems . . . 26

2 Introduction to Sets 28
- 2.1 Week 4: Sets and Set Operations . . . 28
  - 2.1.1 Sets . . . 28

2.1.2 Set Operations . . . 29
2.1.3 Problems . . . 32
2.2 Week 5: Functions and Relations . . . 32
2.2.1 Functions . . . 32
2.2.2 Relations and Orders . . . 35
2.2.3 Equivalence Relations . . . 36
2.2.4 Problems . . . 39
2.3 Week 6: Cardinality . . . 40
2.3.1 Cardinalities . . . 40
2.3.2 Finite Sets . . . 42
2.3.3 Countable Sets . . . 44
2.3.4 Diagonalization and Uncountable Sets . . . 46
2.3.5 Problems . . . 49

3 Induction and Recursion **51**
3.1 Week 7: Mathematical Induction . . . 51
3.1.1 Induction . . . 51
3.1.2 Problems . . . 55
3.2 Week 8: Strong Induction and Well Ordering . . . 55
3.2.1 Strong Induction . . . 55
3.2.2 Well-Ordering . . . 59
3.2.3 Well-Ordering for $\mathbb{N}$ . . . 60
3.2.4 Problems . . . 62

II Concepts **63**

4 Introduction to Topology **64**
4.1 Week 9: Metric Spaces . . . 64
4.1.1 Metric Spaces . . . 64
4.1.2 Open Sets . . . 68
4.1.3 Building Open Sets . . . 70
4.1.4 Problems . . . 72
4.2 Week 10: Closed Sets . . . 73
4.2.1 Convergence . . . 73
4.2.2 Closures . . . 75
4.2.3 Closed Sets . . . 76
4.2.4 Building Closed Sets . . . 77
4.2.5 Complements of Open Sets . . . 80
4.2.6 Problems . . . 81
4.3 Week 11: Continuity . . . 82
4.3.1 Continuous Functions . . . 82
4.3.2 Working with Abstract Metric Spaces . . . 85
4.3.3 Continuity by Open Sets . . . 86
4.3.4 Primer on Point-Set Topology . . . 87
4.3.5 Problems . . . 91

5 Introduction to Group Theory **93**
5.1 Week 12: Groups . . . 93
5.1.1 Symmetries of the Square . . . 93
5.1.2 Modular Arithmetic . . . 96
5.1.3 Defining Groups . . . 98
5.1.4 Basic Group Theory . . . 100
5.1.5 Subgroups . . . 101

5.1.6 Problems . . . 105
5.2 Week 13: Cosets . . . 106
5.2.1 Cosets . . . 106
5.2.2 Cosets by Equivalence Relation . . . 108
5.2.3 How to Think About Cosets . . . 109
5.2.4 Lagrange's Theorem . . . 111
5.2.5 Quotient Groups . . . 113
5.2.6 Problems . . . 115
5.3 Week 14: Homomorphisms . . . 116
5.3.1 Isomorphisms . . . 116
5.3.2 Homomorphisms . . . 119
5.3.3 Kernels and Images . . . 122
5.3.4 Groups of Prime Order . . . 126
5.3.5 Problems . . . 128

Bibliography **130**

Index **131**

CHAPTER 0
PREFACE

I was never this afraid, but I swear to God I was never this determined.

—Winston Rowntree, [Row17]

Stepping into your first upper division math course can be a scary thing. Unlike other subjects, the difference between lower and upper division courses in math can be quite overwhelming, the two main culprits being writing proofs and abstract concepts.

In this course we will address these issues head-on. In particular, we will learn how to write proofs and develop good mathematical style and we will give students more familiarity with the mathematical objects appearing in upper division mathematics.

0.1 For the Reader

We briefly explain some of the writing conventions and notations in these notes.

This text has many examples and exercises contained in the text of each section. Some already have solutions, but some do not. The reader is encouraged to do as many of these unsolved exercises as they can stomach, for one learns mathematics by doing. There are also problems at the end of each chapter intended to solidify understanding. If you do no exercises or problems, Nir Elber will find you and smack you over the head with [Ros19].

We use the notation $:=$ for definitions. For example, the statement $x := 2$ means that we are defining x to equal 2. We hope this will help the reader distinguish equalities which are definitions (for which we use $:=$) from equalities which might require explanation (such as $3^2 + 4^2 = 5^2$).

A "theorem" is a proven result which is a main attraction of the section, chapter, or even of the entire course. For example, the following is a theorem.

Theorem 0.1 (Wiles)**.** The real number $\sqrt{2}$ is irrational.

Proof. Omitted until later in the course! ■

The name "theorem" should be used reverently. For example, the following is not a theorem.

Theorem 0.2. We have $1 + 1 = 2$.

Instead, a proven result which is not a main attraction is called a "proposition."

Proposition 0.3. We have $1 + 1 = 2$.

In contrast to a theorem or proposition, a "corollary" is a result which quickly follows from a theorem or proposition. For example, here is a corollary to Theorem 0.1.

Corollary 0.4. There do not exist positive integers a and b such that $a^2 + a^2 = b^2$.

Proof. Rearranging $a^2+a^2 = b^2$, one sees $\sqrt{2} = \frac{a}{b}$. However, $\sqrt{2}$ is irrational by Theorem 0.1, so this doesn't make any sense! ■

A "lemma" is a result which used to help prove a result. (Usually, the result is a theorem or proposition.) For example, the following result could be a lemma for Theorem 0.1.

Lemma 0.5. The real number $\sqrt{2}$ is not an integer.

Proof. Note that $1 < 2 < 4$, so $1 < \sqrt{2} < 2$. However, there are no integers strictly between 1 and 2, so $\sqrt{2}$ cannot be an integer. ■

One might scoff at the above naming conventions and think that it is easier to just call everything a "theorem" and not have to worry about these extra words. However, it is nonetheless helpful to tell the reader explicitly how important various results we prove are and how they fit into the bigger picture. Calling every a result a "theorem" is the mathematical equivalent of screaming every sentence you speak.

As a final note, occasionally in these notes we will want to warn against some bad reasoning but still state the claim we are making. We do so by labeling the result as a "bad theorem."

Bad Theorem 0.6. We have $1 + 1 = 3$.

0.2 For the Teacher

The exposition in these notes tends to take the point of view that the written word should be precise and correct. As such, proofs which are a little incorrect but perhaps still containing the correct intuition are not favored compared to proofs which are more technically correct. Nonetheless, these notes do make an effort to include the intuitive ideas, just not in the course of an argument.

All of this is to say that a teacher may wish to modify some of the exposition presented here while lecturing to a class. For example, these notes discuss induction after some more difficult set theory in order to more properly be able to state the well-ordering principle. It might be preferable in a class to introduce induction earlier on but wait to discuss the well-ordering principle.

0.3 Transition to Upper Division

The bulk of these notes concerns mathematics, but let us say a few words first about the transition to upper division classes more broadly. Of course, these tips will not all be the ideal solution for all students, but as you read, think about whether such study habits might help you, and if not, how else you might achieve the same goals.

A big difference between lower and upper division math classes is that upper division classes focus much less on learning how to follow a particular procedure to achieve a computational result. In other words, you will be asked to construct arguments or lines of reasoning that you have never seen fully-formed before. Rather, you must learn to put together individual parts of the material you have learned in a clever way to prove novel results.

In order to adjust to these changes, it is important to engage with class material in a truly deep way. This can take the form of asking yourself questions about the objects and concepts you learn about, for example. It can also entail bringing in outside sources to get additional perspectives on the material. We propose a few big and small ways to help you make the most of the time you spend studying.

Proposition 0.7. Start early.

Perhaps, not much needs to be said about this theorem, because we all know procrastination is bad, and we all still do it. However, passively contemplating a problem throughout the week is so much more enjoyable, and so much better for learning, than trying to slap a solution together at the last minute, that it really does bear repeating. Know when to take breaks, and do not overwork yourself, but if you can add one model-student habit to your repertoire, make it reading over your assignments as early as you can.

Proposition 0.8. Find a study group.

Ideally early on in your courses, chat with people sitting near you and offer to establish communication with them to work on homework or reviewing material. If you are brave, make a class Discord server (or similar) and ask the professor to send out the invite to all students. Talking about math with other people will help you pick out what's important and what's challenging. It will also make the learning process a lot more fun!

Corollary 0.9. Ask questions!

What makes study groups so helpful is that they establish a two-way conversation that makes you think harder. Asking questions during lecture, in the MUSA office, or elsewhere achieves the same goal. While it can be nerve-wracking to ask questions in class, especially if they feel elementary, keep in mind that teaching is also more fun as a dialogue than as a monologue. Get yourself in the habit of speaking up in class, and you will find that it becomes easier.

Proposition 0.10. Take notes, but don't get lost in taking notes.

Taking notes is a delicate art form that everyone must master for themselves. Spend real time thinking about how notes can help you learn, and try different styles. A few things to consider: Does taking notes keep you from zoning out during lectures, or does it distract you from doing the mental work of understanding what is being taught? Does it help you complete assignments, or study for exams? Does it help you keep track of important definitions and theorems from lecture that you can refer back to once a relevant board has been erased? Whatever purpose you choose to prioritize, make sure you take the right kind of notes for that purpose.

Some professors lecture very fast, and you may not always be able to keep up as much as you would like. Some people have trouble writing and thinking at the same time, so it is not always wise to try to write everything down. If you must make trade-offs with your note-taking, it is better to write definitions and theorems, then think about examples and applications. Long proofs might be better reviewed from the textbook, but if you can jot down the main steps, you will thank yourself later.

Corollary 0.11. Read ahead if you can.

Although it may feel redundant, reading the content of a lecture prior to attending it can transform your lecture experience from one of scrambling to pick up information to one of getting a fresh explanation of the material that helps you remember it by adding intuition. When reading, you can pause at confusing parts and skim over the obvious things, making it a worthwhile way for many students to learn. It also makes taking notes a lot easier—just write down what didn't stick the first time around, or what feels significant after two passes.

If you're apprehensive about investing the extra time, know that having a little extra familiarity with class content pays dividends later on. Having a working knowledge of the material makes homework and exams

much less of an ordeal, and prevents big gaps in your understanding, which might require reading the textbook later on anyway.

> **Corollary 0.12.** Takes notes while reading.

Mathematics is a large and intricate machine, and it is incredibly difficult to fit the entire picture in your head at once. Making matter worse, mathematical language is created to be precise and unambiguous, so authors will often say important points exactly once. This is in contrast to most other English writing, where communicating ideas is an imprecise problem and requires much repetition. This lack of repetition makes reading mathematics much slower and requiring more effort.

Taking notes while reading alleviates many of these difficulties. For example, noting down important definitions or results forces you to repeat important points, as you might when reading any other text. However, be careful to not just copy the text you are reading verbatim onto your notes—at that point you are just rewriting the text! As with other note-taking, finding a medium which works for you will pay back in spades.

> **Proposition 0.13.** Focus on the big picture, but make sure to think about examples.

Upper division math involves a large amount of detail. Trying to stomach everything at once isn't always feasible. Instead, try to summarize bite-sized chunks, such as one lecture, one chapter, or even one exercise, in your mind, and think about patterns or the ways in which concepts connect to each other.

On the other hand, doing math in constant abstraction doesn't work for most students, either. Examples are your friend, and most professors will present some to you. This text itself and has examples scattered throughout, whether labeled examples, exercises, or problems, and you are encouraged to think through as many of them as you can stomach. Solidify in your mind what they demonstrate about the abstract concepts you are studying, and return to them when you are presented with new questions or new objects. This is a good way to build intuition.

> **Proposition 0.14.** Use office hours for your own benefit.

There are many good reasons to go to office hours—either your professor's or your GSI's—and none of them are that professors and grad students are gods who must occasionally receive an offering. Don't be afraid to come with homework or lecture questions, be they specific or general, but also feel free to ask what your instructor feels you should focus on. Seeing new definitions and ideas for the first time can make it hard to see the forest for the trees, and talking to someone with more experience can make a big difference in the intuition you gain. In particular, professors are great sources of interesting and relevant examples. Of course, you can also talk to professors about the big life questions for after graduation.

On the other hand, don't try to force a relationship with a professor you don't vibe with. Just like anyone else, not all professors are fun to talk to, and that's OK. Don't let the expectations of networking and getting good letters of recommendation blind you to the importance of your own enjoyment of a good mentoring relationship.

> **Theorem 0.15.** Hold on to the things that matter to you.

Whatever the reason you are reading this text, make sure you keep it in mind as you forage ahead. You may not find joy in all aspects of your study, but make a point to observe and remember the aspects that brought you to this point. More than any extrinsic success metric, these are what will keep you in math.

0.4 Acknowledgements

A brief history of these notes follows.

- The first draft of these notes was written by Cailan Li and the first class of students taking MUSA 74, in Spring 2018.

- In preparation for Spring 2019, Aidan Backus, Andrew DeLapo, and Java Villano edited these notes.
- In preparation for Spring 2021, Aidan Backus, Katie Lamar, Audrey Litvak, Chris Randall, Bryce Goldman and Tina Li edited these notes.
- In preparation for Spring 2023, Nir Elber and Rhea Kommerell edited these notes.

It would be nice to turn these notes into a textbook some day, but that day is far away.

Part I

Proofs

Chapter 1
Logic and Proofs

So the man gave him the bricks, and he built his house with them.

—Joseph Jacobs, "The Story of the Three Little Pigs" [Jac90]

1.1 Week 1: Propositional Logic

Before we jump into proof techniques, let us first introduce propositional logic. Propositional logic is the foundation of all mathematics. It allows us to construct correct mathematical arguments and with a strong understanding of logic, one's ability to break down and solve problems will improve immensely.

Definition 1.1 (proposition)**.** A *proposition* is a declarative sentence, which declares a fact. Note that this fact can either be true or false, but not both.

Let us go over a couple examples.

Example 1.2. Which of the following are propositions and which are not?

(a) I like peanut butter and jelly sandwiches.

(b) Pigs can fly and talk.

(c) Can you see the pineapple wearing sunglasses?

(d) $9 + 10 = 21$.

(e) $5 + 5 = 10$.

Proof. Sentences (a), (b), (d), and (e) are propositions. Sentence (c) is not a proposition. Can you think of why sentence (c) is not a proposition? Note that although (b) and (d) are not true, they are still propositions. ■

1.1.1 Compound Propositions

Here are a few definitions.

Definition 1.3 (propositional variable)**.** A *propositional variable* (also called a *sentential variable*) is a variable that is assigned to a *truth value* (true or false).

Definition 1.4 (primitive proposition)**.** A *primitive proposition* is a proposition that can't be further broken down.

Example 1.5. The following are examples of primitive proposition examples.

(a) It is sunny outside.

(b) It is raining.

(c) I am tired.

Definition 1.6 (compound proposition)**.** A *compound proposition* is a proposition that can be further broken down into primitive propositions. Compound propositions consist of one or more primitive (or compound) propositions joined by *logical operators*.

Negation (NOT), conjunction (AND), disjunction (NOR), exclusive-OR (XOR), implication (conditional statement: if ... then ...), and equivalence (... if and only if ...) are logical operators. We will explore these operators in more detail over the next few lectures. Note that propositional variables can also be used to represent compound propositions.

Exercise 1.7. Which of the following are compound propositions?

(a) Bob likes PB&J sandwiches.

(b) Bob likes peanut butter sandwiches and he likes jelly sandwiches.

(c) Alice has at least 20 pigs.

(d) Alice likes ice-cream or cookies, but not both.

(e) Tom failed his math test.

(f) If it is sunny outside then we will go to the beach.

Here are our first logical operators.

Definition 1.8 (negation)**.** Let p be a proposition. The *negation* of p is the statement "not p" or "it is not the case that p". We denote the negation of p using $\neg p$, $-p$, p', $\sim p$, $!p$, ... but they all mean the same and you can use any symbol.

Exercise 1.9. What is the negation of each of the following propositions?

(a) Alice likes ice-cream or cookies

(b) Bob has at least 20 pigs.

(c) The pineapple is not wearing sunglasses.

(d) $9 + 10 = 21$.

Definition 1.10 (logical conjunction)**.** Let p and q be propositional variables. The *conjunction* of p and q is the statement "p and q". We denote the conjunction of p and q by $p \wedge q$. Note that the conjunction $p \wedge q$ has a truth value of true when both are true; otherwise, the truth value is false.

Definition 1.11 (logical disjunction)**.** Let p and q be propositional variables. The *disjunction* of p and q is the statement "p or q". We denote the disjunction of p and q by $p \vee q$. Note that the conjunction $p \vee q$ has a truth value of false when both are false; otherwise, the truth value is true.

Definition 1.12 (exclusive disjunction)**.** Let p and q be propositional variables. The *exclusive disjunction* of p and q is the statement "p or q, but not both". We denote the exclusive disjunction of p and q by $p \oplus q$.

Definition 1.13 (conditional statement)**.** Let p and q be propositions. A *conditional statement*, symbolized by $p \to q$, is the statement "if p, then q". Note that $p \to q$ is false when p is true and q is false; otherwise, the truth value is true.

There are many different ways to say something like $p \to q$. Here is a quick table, for reference.

"if p, then q"	"p implies q"
"if p, then q"	"p only if q"
"p is sufficient for q"	"a sufficient condition for q is p"
"q if p"	"q whenever p"
"q when p"	"q is necessary for p"
"a necessary condition for p is q"	"q follows from p"
"q unless $\neg p$"	"q provided that p"

1.1.2 Truth Tables

Truth tables are utilized to keep track of all possible truth values of a compound proposition. Let's see some examples.

Example 1.14. Let p be a proposition. Here is the truth table for $\neg p$.

p	$\neg p$
T	F
F	T

Exercise 1.15. Construct a truth table for each of the following compound propositions.

(a) $p \wedge q$

(b) $p \vee q$

(c) $p \oplus q$

(d) $p \wedge \neg q$

(e) $p \wedge q \wedge r$

Let's do some more examples!

Exercise 1.16. Write these statements using p and q. Remember to define the propositions p and q.

(a) If I study really hard for this course, then I will do well in other upper division math courses.

(b) The dog is cute unless it is a lion.

(c) A necessary condition to do well in math is to practice many problems.

(d) If it rains today, I should bring an umbrella.

Solve the following puzzle by translating statements into logical expressions and then make conclusions from these expressions with the help of truth tables.

Exercise 1.17 ([Ros19, Exercise 1.2.23–27])**.** Suppose that knights always tell the truth and knaves always lie. Now suppose you encounter two people A and B. If A is either a knight or a knave, and B is either a knight or a knave, what conclusions can you draw from each of the following (indepedent) scenarios? When possible, determine the identity (knave or knight) of A and B.

(a) A says "At least one of us is a knave" and B says nothing.

(b) A says "The two of us are both knights," and B says "A is a knave."

(c) A says "I am a knave or B is a knight" and B says nothing.

(d) Both A and B say "I'm a knight."

(e) A says "We are both knaves," and B says nothing.

1.1.3 Problems

Solve the following puzzle by translating statements into logical expressions and then make conclusions from these expressions with the help of truth tables. Show your work!

Problem 1.1 ([Ros19, Exercise 1.2.36])**.** The police have three suspects for the murder of Mr. Cooper: Mr. Smith, Mr. Jones, and Mr. Williams. Smith, Jones, and Williams each declare that they did not kill Cooper. Smith also states that Cooper was a friend of Jones and that Williams disliked him. Jones also states that he did not know Cooper and that he was out of town the day Cooper was killed. Williams also states that he saw both Smith and Jones with Cooper the day of the killing and that either Smith or Jones must have killed him.

(a) Can you determine who the murderer was if we know that one of the three men is guilty, the two innocent men are telling the truth, but the statements of the guilty man may or may not be true?

(b) Can you determine who the murderer was if we know that innocent men do not lie?

1.2 Week 2: Equivalences, Predicates, and Quantifiers

In this section, we discuss a few more aspects of propositional logic.

1.2.1 Logical Equivalences

Some propositions are particularly simple because they are either always true or always false.

Definition 1.18 (tautology)**.** A *tautology* is a compound proposition that is always true, regardless of whether the truth values of the proposition variables are true or false.

Definition 1.19 (contradiction)**.** A *contradiction* is a compound proposition that is always false regardless of whether the truth values of the proposition variables used to construct the compound proposition are true or false.

Example 1.20. Let p and q be propositional variables. For each compound proposition below, determine whether the proposition is a tautology, a contradiction, or neither.

(a) $p \wedge \neg q$

(b) $p \wedge \neg p$

(c) The temperature is 32 degrees Fahrenheit and 0 degrees Celsius.

(d) The temperature is 32 degrees Fahrenheit or not 0 degrees Celsius.

Proof. Here, (d) is a tautology, (b) is a contradiction, and (a) and (c) are neither. ■

Even if two propositions are not both always true, they might still "mean the same thing." In propositional logic, this notion is called logical equivalence.

Definition 1.21 (logical equivalence)**.** Let p and q be compound propositions. We say that p and q are *logically equivalent*, denoted as $p \equiv q$, if and only if $p \to q$ and $q \to p$ are both always true. Equivalently, $p \equiv q$ if and only if p is true exactly when q is true.

Example 1.22. Let p and q be proposition variables. Show that $p \to q \equiv \neg p \vee q$ using a truth table.

Proof. We build a truth table, as follows.

p	q	$p \to q$	$\neg p \vee q$
T	T	T	T
T	F	F	F
F	T	T	T
F	F	T	T

Note the $p \to q$ column is true exactly when the $\neg p \vee q$ column is true. ■

Example 1.23. Let p, q, and r be proposition variables. Show that $(p \vee q) \vee r \equiv p \vee (q \vee r)$ using a truth table.

Proof. We build a truth table, as follows.

p	q	r	$p \vee q$	$q \vee r$	$(p \vee q) \vee r$	$p \vee (q \vee r)$
T	T	T	T	T	T	T
T	T	F	T	T	T	T
T	F	T	T	T	T	T
T	F	F	T	F	T	T
F	T	T	T	T	T	T
F	T	F	T	T	T	T
F	F	T	F	T	T	T
F	F	F	F	F	F	F

Note the $(p \vee q) \vee r$ column is true exactly when the $p \vee (q \vee r)$ column is true. ■

We take a moment to record the most important equivalences and their names. Let p, q, and r be proposition variables.

Logical Equivalence	Name
$p \wedge T \equiv p$ $p \vee F \equiv p$	Identity Laws
$p \vee T \equiv T$ $p \wedge F \equiv F$	Domination Laws
$p \vee p \equiv p$ $p \wedge p \equiv p$	Idempotent Laws
$\neg(\neg p) \equiv p$	Double Negation Law
$p \vee q \equiv q \vee p$ $p \wedge q \equiv q \wedge p$	Commutative Laws
$(p \vee q) \vee r \equiv p \vee (q \vee r)$ $(p \wedge q) \wedge r \equiv p \wedge (q \wedge r)$	Associative Laws
$p \vee (q \wedge r) \equiv (p \vee q) \wedge (p \vee r)$ $p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r)$	Distributive Laws
$\neg(p \wedge q) \equiv \neg p \vee \neg q$ $\neg(p \vee q) \equiv \neg p \wedge \neg q$	De Morgan's Laws
$p \vee (p \wedge q) \equiv p$ $p \wedge (p \vee q) \equiv p$	Absorption Laws
$p \vee \neg p \equiv \mathsf{T}$ $p \wedge \neg p \equiv \mathsf{F}$	Negation Laws

Let's see these in action.

Example 1.24. Let p and q be proposition variables. Show that $\neg(p \to q) \equiv p \wedge \neg q$ using a series of logical equivalences.

Proof. By De Morgan's Laws, we have that $\neg(\neg p \vee q) \equiv \neg(\neg p) \wedge \neg q$. By the Double Negation Law, we know that $\neg(\neg p) \wedge \neg q \equiv p \wedge \neg q$. Considering that $p \to q \equiv \neg p \vee q$ by Example 1.22, we can say that $\neg(p \to q) \equiv \neg(\neg p \vee q) \equiv p \wedge \neg q$. ■

Example 1.25. Let p, q, and r be proposition variables. Show that $(p \wedge q) \to (p \vee q)$ is a tautology using a series of logical equivalences.

Proof. Let $u \equiv p \wedge q$ and $w \equiv p \vee q$. We want to show that $u \to w$. By Example 1.24, we know that $u \to w \equiv \neg u \vee w$. So we have that $(p \wedge q) \to (p \vee q) \equiv \neg(p \wedge q) \vee (p \vee q)$. We now have the equivalences

$$\begin{aligned}
(p \wedge q) \to (p \vee q) &\equiv (\neg p \vee \neg q) \vee (p \vee q) && \text{(De Morgan's Laws)} \\
&\equiv \neg p \vee (\neg q \vee (p \vee q)) && \text{(Associative Laws)} \\
&\equiv \neg p \vee ((p \vee q) \vee \neg q) && \text{(Commutative Laws)} \\
&\equiv \neg p \vee (p \vee (q \vee \neg q)) && \text{(Associative Laws)} \\
&\equiv \neg p \vee (p \vee \mathsf{T}) && \text{(Negation Laws)} \\
&\equiv \neg p \vee \mathsf{T} && \text{(Negation Laws)} \\
&\equiv \mathsf{T}, && \text{(Negation Laws)}
\end{aligned}$$

which complete the proof. ■

Example 1.26 (contraposition)**.** Let p and q be proposition variables. Show that $(p \to q) \equiv (\neg q \to \neg p)$ using a series of logical equivalences. The proposition $\neg q \to \neg p$ is called the *contraposition* of the proposition $p \to q$.

Proof. The main point is to use Example 1.22 to see $(p \to q) \equiv (\neg p \vee q)$ and $(\neg q \to \neg p) \equiv (\neg\neg q \vee \neg p)$. Thus, we compute

$$\begin{aligned} \neg q \to \neg p &\equiv \neg\neg q \vee \neg p \\ &\equiv q \vee \neg p && \text{(Double Negation Law)} \\ &\equiv \neg p \vee q && \text{(Commutative Laws)} \\ &\equiv p \to q, \end{aligned}$$

which completes the proof. ■

1.2.2 Propositional Functions

Thus far we have built an intuition that a proposition like $p \wedge q$ takes in truth values according to p and q and then gives back the truth value of $p \wedge q$. This notion can be generalized, as follows.

Definition 1.27 (propositional function)**.** An *n-ary propositional function* P is a function that maps the n-tuple $(x_1, x_2, \ldots, x_n)$ to either true or false (exclusive). Note P is often called the *predicate* and the evaluation of P at $(x_1, \ldots, x_n)$ is $P(x_1, \ldots, x_n)$.

Here are some examples.

Example 1.28. Let $\mathbb{R}$ denote the set of real numbers. Define $P\colon \mathbb{R} \to \{\mathsf{T}, \mathsf{F}\}$ by

$$P(x) \coloneqq \begin{cases} \mathsf{T} & \text{if } x > 3, \\ \mathsf{F} & \text{if } x \leq 3. \end{cases}$$

Example 1.29. Let $\mathbb{R}$ denote the set of ordered pairs of real numbers. Define $P\colon \mathbb{R} \times \mathbb{R} \to \{\mathsf{T}, \mathsf{F}\}$ by

$$P(x, y) \coloneqq \begin{cases} \mathsf{T} & \text{if } x > y, \\ \mathsf{F} & \text{if } x \leq y. \end{cases}$$

Example 1.30. Define $P\colon \{\text{UC Berkeley Students}\} \to \{\mathsf{T}, \mathsf{F}\}$ by

$$P(x) \coloneqq \begin{cases} \mathsf{T} & \text{if } x \text{ is enrolled in MUSA 74}, \\ \mathsf{F} & \text{if } x \text{ is NOT enrolled in MUSA 74}. \end{cases}$$

As we can see from the above examples, a propositional function maps to a proposition, but the propositional function itself is not a proposition. For example, consider Example 1.30. Let Bob be a UC Berkeley student. The truth table below shows us that $P(\text{Bob})$ is equivalent to the proposition: Bob is enrolled in MUSA 74.

Bob is enrolled in MUSA 74	$P(\text{Bob})$
T	T
F	F

1.2.3 Quantifiers

What if we want to create a proposition from our propositional function without an argument? Quantifiers allow us to do exactly that!

Definition 1.31 (universal quantification)**.** Let $P\colon S \to \{\mathsf{T}, \mathsf{F}\}$ be a propositional function. The *universal quantification* of P is the proposition "$P(x)$ for all x in S". The notation $\forall x P(x)$ denotes the universal quantification of P where $\forall$ represents "for all". Note $\forall x P(x)$ is true when $P(x)$ is true for all $x \in S$; otherwise, $\forall x P(x)$ is false.

Definition 1.32 (existential quantification)**.** Let $P\colon S \to \{\mathsf{T}, \mathsf{F}\}$ be a propositional function. The *existential quantification* of P is the proposition "there exists an element x in S such that $P(x)$". The notation $\exists x P(x)$ denotes the existential quantification of P where $\exists$ represents "there exists". Note $\exists x P(x)$ is true when $P(x)$ is true for some x in S. If there doesn't exist an x in S such that $P(x)$ is true then $\exists x P(x)$ is false.

Remark 1.33. Considering that a combination of quantifiers and propositional functions are utilized to represent a wide range of statements found in both mathematics and in the English language (as well as all other languages for that matter), most propositions constructed using quantifiers and propositional functions have somewhat "less formal" language to express quantifiers. For example, instead of saying "for all the marbles x in the bag, x is blue", we can say "all of the marbles in the bag are blue".

It is important to remember that $\forall x P(x)$ and $\exists x P(x)$ are propositions, so we can treat them as such.

Example 1.34. Express the following statements using propositional functions and quantifiers.

(a) Some cats have richly colored fur.

(b) All cats are domesticated.

(c) No felines that are larger than the average human are domesticated.

(d) Felines that are not cats are not domesticated.

(e) All felines that are larger than the average human are not cats.

Proof. Let $P, Q, R, S\colon \{\text{all felines}\} \to \{\mathsf{T}, \mathsf{F}\}$ be propositional functions. Let $P(x)$ be the statement "x is a cat" and $Q(x)$ be the statement "x is larger than the average human". Let $R(x)$ be the statement "x is domesticated" and $S(x)$ be the statement "x has richly colored fur".

(a) $\exists x(P(x) \wedge S(x))$.

(b) $\forall x(P(x) \to R(x))$.

(c) $\neg\exists x(Q(x) \wedge R(x))$.

(d) $\forall x(\neg P(x) \to \neg R(x))$.

(e) $\forall x(Q(x) \to \neg P(x))$. ■

Let P and Q be propositional functions. The following tables include commonly used propositions with quantifiers (and nested quantifiers) along with information on how to determine their truth values.

Proposition	Equivalent	True when:	False when:
$\forall x P(x)$		$P(x)$ is true for all x	There exists an x for which $P(x)$ is false
$\exists x P(x)$		There exists an x for which $P(x)$ is true	$P(x)$ is false for all x
$\neg\exists x P(x)$	$\forall x \neg P(x)$	For every x, $P(x)$ is false	There is some x for which $P(x)$ is true
$\neg\forall x P(x)$	$\exists x \neg P(x)$	There is some x for which $P(x)$ is false	$P(x)$ is true for all x

We note that we can even nest our quantifiers! Here is the corresponding table.

Proposition	True when:	False when:
$\forall x \forall y Q(x,y)$ $\forall y \forall x Q(x,y)$	$Q(x,y)$ is true for every pair x, y	There is a pair x, y for which $Q(x,y)$ is false
$\forall x \exists y Q(x,y)$	For every x, there is some y for which $Q(x,y)$ is true	There is some x for which $Q(x,y)$ is false for every y
$\exists x \forall y Q(x,y)$	There is some x for which $Q(x,y)$ is true for every y	For every x there is some y for which $Q(x,y)$ is false
$\exists x \exists y Q(x,y)$ $\exists y \exists x Q(x,y)$	There is a pair x, y for which $Q(x,y)$ is true	$Q(x,y)$ is false for every pair x, y

1.2.4 Problems

The following problems can be found in [Ros19].

Problem 1.2. Let p and q be propositional variables. Show that $(p \to q) \equiv (\neg q \to \neg p)$ by

(a) using a truth table, and

(b) using rules of logical equivalence.

Problem 1.3. Let p, q, and r be propositional variables. Prove that the following statements are tautologies by using (a) truth tables and (b) without using truth tables.

(a) $\neg p \wedge (p \vee q) \to q$

(b) $((p \to q) \wedge (q \to r)) \to (p \to r)$

(c) $(p \wedge (p \to q)) \to q$

(d) $((p \vee q) \wedge (p \to r) \wedge (q \to r)) \to r$

Problem 1.4. Let p and q be propositional variables. Determine whether $\neg(p \oplus q)$ is logically equivalent to $p \leftrightarrow q$. If these compound propositions are logically equivalent then provide a proof. If they are not equivalent, provide an explanation.

Problem 1.5. Let $L\colon \{\text{all humans in the world}\} \times \{\text{all humans in the world}\} \to \{\mathsf{T}, \mathsf{F}\}$ be a propositional function. Let $L(x, y)$ be the statement "x loves y". Use quantifiers to express each of the following statements.

(a) Everybody loves Jerry.

(b) Everybody loves somebody.

(c) There is somebody whom everybody loves.

(d) Nobody loves everybody.

(e) There is somebody whom Lydia does not love.

(f) There is somebody whom no one loves.

(g) There is exactly one person whom everyone loves.

(h) There are exactly two people whom Lynn loves.

(i) Everyone loves himself or herself.

(j) There is someone who loves no one besides himself or herself.

Problem 1.6. Let $L\colon \{\text{all humans in the world}\} \times \{\text{all humans in the world}\} \to \{\mathsf{T}, \mathsf{F}\}$ be a propositional function. Let $L(x, y)$ be the statement "x loves y". Translate the following statements from propositional logic to English.

(a) $\forall x L(x, b)$

(b) $\forall x(L(b, x) \to (x = m))$; here, "$=$" means equality.

Problem 1.7. Find a counterexample, if possible, to the following universally quantified statements, where the domain for all variables consists of all the integers. If the statement is true, provide a proof.

(a) $\forall x \exists y \left(x = \frac{1}{y}\right)$

(b) $\forall x \exists y \left(y^2 - x < 100\right)$

(c) $\forall x \forall y \left(x^2 \neq y^3\right)$

Problem 1.8. Express each of the following statements using quantifiers and then form the negation of the statement so that no negation is to the left of a quantifier. Finally, express the negation in simple English. (Do not simply use the phrase "It is not the case that.")

(a) Some student has solved every exercise in this book.

(b) No student has solved at least one exercise in every section of this book.

(c) No one has lost more than one thousand dollars playing the lottery.

(d) There is a student in this class who has chatted with exactly one other student.

(e) No student in this class has sent e-mail to exactly two other students in this class.

1.3 Week 3: Introduction to Proofs

In this section, we discuss proofs.

1.3.1 Proof Basics

The best way to learn what a proof is to see some examples. Here's an example from high-school algebra.

Example 1.35. Set $p(x) \coloneqq x^2 + bx + c$ for some real numbers b and c. If r_1, r_2 are the zeroes of p, then $r_1 + r_2 = -b$, and $r_1 r_2 = c$.

Proof. We want to start any proof by writing down the basic definitions and properties. We know that r_1 and r_2 are zeroes, so $p(r_1) = p(r_2) = 0$. It follows that we can factor

$$p(x) = (x - r_1)(x - r_2)$$

because the leading coefficient on the x^2 term of $p(x)$ is 1. Expanding both sides,

$$x^2 + bx + c = x - (r_1 + r_2)x + r_1 r_2,$$

so $r_1 + r_2 = -b$ and $r_1 r_2 = c$. ■

That proof should convince you beyond a shadow of a doubt that the claim is true, assuming that you know basic facts about quadratics: above all, a proof is an argument, meant to persuade the reader. If it didn't, think about it and figure out where you lost the line of reasoning, and ask around. Often a proof might not make sense when we read it ourselves, but it becomes clearer when someone else explains it.

Example 1.36. Let n be an integer. If n is even, then n^2 is even.

Proof. Again, we start by writing down the definition. If n is even, then there exists another natural number k such that $n = 2k$. Then

$$\begin{aligned} n^2 &= (2k)^2 \\ &= 4k^2 \\ &= 2 \cdot 2k^2. \end{aligned}$$

We have shown that n^2 is 2 times the natural number $2k^2$, so n^2 is an even number by definition. ■

Let's try a more complicated example, which you might've seen in calculus, and it is important in its own right.

Example 1.37 (Euler's formula)**.** For all real numbers x,

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

Proof. At first this might seem a bit hopeless, because it's not clear that $\exp$ and the trigonometric functions have anything to do with each other. The first clue to consider is that the derivative $\frac{d}{d}x \exp(x) = \exp(x)$, while $\frac{d}{dx} \sin(x) = \cos(x)$, and $\frac{d}{dx} \cos(x) = -\sin(x)$. So it's very easy to compute the higher derivatives of these functions, which means we can reason using Taylor series. This is as good a place to start as any, so we'll try this and see what happens.

Recall the Taylor series

$$e^x = \sum_{n=0}^{\infty} \frac{\exp^{(n)}(0)}{n!} x^n = \sum_{n=0}^{\infty} \frac{e^0}{n!} x^n = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Similarly,

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$
$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} - \cdots.$$

Plugging in ix to e^x and using that $i^2 = -1$, we have

$$\begin{aligned} e^{ix} &= 1 + ix + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \frac{(ix)^4}{4!} + \dots \\ &= 1 + ix + \frac{-x^2}{2!} - i\frac{x^3}{3!} + \frac{x^4}{4!} + \dots \\ &= \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \dots\right) + i\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots\right) \\ &= \cos x + i \sin x, \end{aligned}$$

so $e^{ix} = \cos x + i \sin x$. ■

Notice that in all of the above proofs, we needed to use every assumption we made. This will be the case on the homework. In other words, if you finished a proof and didn't use some assumption that the theorem made, then something went wrong. Explicitly, one of the following happened.

- You assumed too much. In this case, the statement of the theorem you were trying to prove should be rephrased without the unnecessary assumptions.
- You used the assumption tacitly in part of the proof, without realizing it. In this case, realize where you used the assumption, and note it explicitly.
- You made an error elsewhere in the proof. In this case, fix your proof!

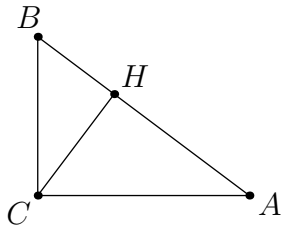
In general, if you prove a theorem but didn't use one of your assumptions, then you have proven a stronger theorem!

Let's see another example. Keep track of where we use the fact that $\triangle ABC$ is a right triangle in the following proof!

Example 1.38 (Pythagorean theorem)**.** Let $\triangle ABC$ be a right triangle with right angle C. Setting $a \coloneqq BC$ and $b \coloneqq CA$ and $c \coloneqq AB$, then

$$a^2 + b^2 = c^2.$$

Proof. We must show $BC^2 + CA^2 = AB^2$. Let H be the point on the hypotenuse such that $\angle CHA$ is a right angle, which gives the following diagram.



We claim that $\triangle ACH$ is similar to $\triangle ABC$. Indeed, $\angle AHC = \angle ACB$ because both are right angles, and $\angle BAC = \angle CAH$ because those are literally the same angle. Lastly, because the angles in both triangles must add up to $180°$ degrees, we must have

$$\angle ABC + \angle BCA + \angle CAB = 180° = \angle ACH + \angle CHA + \angle HAC,$$

so $\angle ACH = \angle ABC = 90° - \angle BAC$. Thus, the triangles are similar. A similar proof shows $\triangle CBH$ is similar to $\triangle ABC$. (Show this yourself, if you'd like.)

Using our similar triangles, we see

$$\frac{CA}{AB} = \frac{HA}{AC} \qquad \text{and} \qquad \frac{BC}{AB} = \frac{BH}{CB}.$$

Thus, $BC^2 = AB \cdot BH$ and $AC^2 = AB \cdot AH$, so

$$BC^2 + AC^2 = AB \cdot BH + AB \cdot AH = AB(AH + BH) = AB^2.$$

This proves the claim. ■

Remark 1.39. If the proof of Example 1.38 did not use the condition that $\triangle ABC$ was a right triangle, then we would have proven the stronger claim that any triangle $\triangle ABC$ has

$$BC^2 + CA^2 = AB^2.$$

However, this claim is false! For example, there is an equilateral triangle with $AB = BC = CA = 1$, and $1^2 + 1^2 \neq 1^2$. This equilateral triangle is called a "counterexample" to our statement.

We close this subsection with an example from linear algebra.

Example 1.40. Let $T\colon V \to W$ be a linear transformation between vector spaces over the scalar field F. Then the kernel $\ker(T)$ of T is a vector subspace of V.

Proof. By definition, the kernel is

$$\ker(T) \coloneqq \{v \in V : T(v) = 0\},$$

where 0 is the zero vector of W.

Now, recall the definition of subspace of a vector space: a set S is a subspace of V if S is a subset of V that contains the zero vector 0_V and is closed under vector addition and scalar multiplication. Therefore, we must check that $\ker(T)$ fulfills all three of the necessary conditions to be a subspace of V.

- By definition, we see $\ker(T)$ is a subset of V.
- It is a property of linear transformations that $T(0_V) = 0_W$. To see this, note $0_V + 0_V = 0_V$, so because T is a linear transformation,

  $$T(0_V) + T(0_V) = T(0_V + 0_V) = T(0_V).$$

  Rearranging gives $T(0_V) = 0_W$, so 0_V is in $\ker(T)$.
- We check that $\ker(T)$ is closed under addition. Well, given two arbitrary vectors v and w in $\ker(T)$, we must show their sum $v + w$ is also in $\ker(T)$. Because T is a linear transformation,

  $$T(v + w) = T(v) + T(w).$$

  Now, we use our assumption that v and w were in $\ker(T)$. This means we know $T(v) = 0_W$ and $T(w) = 0_W$, so

  $$T(v + w) = 0_W + 0_W = 0_W.$$

  It follows $v + w$ is in $\ker(T)$.
- We check that $\ker(T)$ is closed under scalar multiplication. Well, given a vector v in $\ker(T)$ and a scalar k in F, we must show $k \cdot v$ is also in $\ker(T)$. Because T is a linear transformation, we compute

  $$T(k \cdot v) = k \cdot T(v).$$

  However, because v is in $\ker(T)$, we see $T(v) = 0_V$. To finish, we use properties of the zero vector to give

  $$T(k \cdot v) = k \cdot 0_V = 0_V.$$

  It follows $k \cdot v$ is in $\ker(T)$. ■

1.3.2 Contradiction and Contraposition

In order to prove that a statement is true, it is sometimes easier to do so when an extra assumption, let's say P, is true. If another proof proves the statement when P is false, then together the two proofs imply that the statement is true. This is called "proof by cases."

Example 1.41. There exist irrational real numbers x and y such that x^y is rational.

Proof. You will prove in the homework that $\sqrt{2}$ is irrational. For now, assume that $\sqrt{2}$ is irrational. We know $\sqrt{2}^{\sqrt{2}}$ must be either rational or irrational. So, we divide our proof into two cases.

- Suppose $\sqrt{2}^{\sqrt{2}}$ is rational. Then we have found irrational numbers x and y, with $x = y = \sqrt{2}$, such that x^y is rational.
- Suppose $\sqrt{2}^{\sqrt{2}}$ is irrational. Let $x = \sqrt{2}^{\sqrt{2}}$ and $y = \sqrt{2}$. Then

$$x^y = \left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = \sqrt{2}^{\sqrt{2}\cdot\sqrt{2}} = \sqrt{2}^2 = 2.$$

  Because $x^y = 2$ is rational, we have found irrational numbers x and y, with $x = \sqrt{2}^{\sqrt{2}}$ and $y = \sqrt{2}$, such that x^y is rational.

Because one of the above cases must hold, and in both cases such x and y exist, the statement must be true. ■

This proof is "non-constructive." Namely, the proof does not tell us the explicit x and y such that the statement holds; rather, the proof only verifies that such x and y exist. You will encounter plenty of non-constructive proofs in your upper division math classes.

Remark 1.42. It turns out that $\sqrt{2}^{\sqrt{2}}$ is irrational, as in the second case above, but proving this is non-trivial. For the interested, it is a consequence of the Gelfond–Schneider theorem.

Now let's see an example that we'll come back to later: Russell's paradox. It's a silly, but very important, example of a "proof by contradiction."

Example 1.43. It is impossible for a barber to say that he will shave people if and only if they do not shave themselves.

Proof. For a proof by contradiction, we are going to suppose that such a barber exists, and then show that the existence of the barber implies a contradiction. It will then follow that the barber could not exist!

Indeed, suppose that such a barber exists. Either the barber shaves himself or he does not shave himself, which gives the following cases.

- Suppose that the barber shaves himself. If so, then he does not shave himself, so he both shaves himself and does not shave himself. This is impossible.
- Suppose the barber does not shave himself. But then he shaves himself, which is still impossible.

All cases have led to impossibility, so the barber does not exist. ■

Here is another proof by contradiction.

Example 1.44. Let n be an integer. If n^2 is odd, then n is odd.

Proof. Suppose for the sake of contradiction that n is not odd, so n is even. But then n^2 is even by Example 1.36! This is a contradiction because we know n^2 is supposed to be odd. Thus, we must instead have n being odd. ■

A proof technique similar to contradiction is contraposition. To understand contraposition, recall that

$$(p \to q) \equiv (\neg q \to \neg p)$$

by Example 1.26. As such, sometimes when we want to show a statement of the form $p \to q$, we can try to prove the contraposition $\neg q \to \neg p$ instead. Our first example is revamping the proof of Example 1.44.

Another proof. The contraposition of "if n^2 is odd, then n is odd" is the statement "if n is not odd, then n^2 is not odd." However, a positive integer is "not odd" if and only if it is even, so we're actually proving "if n is even, then n^2 is even," which is exactly Example 1.36. ■

Here's a similar example for you to try.

Exercise 1.45. Let n be an integer. If n^2 is even, then n is even.

1.3.3 Uniqueness Proofs

Sometimes we will want to prove that some object is unique. The usual way to do this is to suppose that any two such objects are equal. Let's see some examples.

Example 1.46. Let f be a strictly increasing function from the real numbers to the real numbers. Then for each real number y, there is at most real number x such that $f(x) = y$.

Proof. Suppose that we have real numbers x and x' such that $f(x) = y$ and $f(x') = y$. We show $x = x'$. There are three cases.

- If $x < x'$, then $f(x) < f(x')$, so $y < y$, which is impossible.
- If $x = x'$, then we are done.
- If $x > x'$, then $f(x) > f(x')$, so $y > y$, which is impossible.

All cases give impossibility or $x = x'$, so we conclude $x = x'$. ■

Note that third case of the previous proof was identical to the first case with some letters changed. In the future, we might say "without loss of generality, we have $x > x'$ or $x = x'$ because the case of $x' < x$ is similar."

Continuing with our examples, here is one from calculus.

Example 1.47. Say that a function f from the real numbers to the real numbers is "strictly convex" if f is twice-differentiable and $f''(t) > 0$ for all real numbers t. If f is strictly convex, then there exists at most one real number x such that $f(x)$ is the (global) minimum value of f.

Proof. Assume that f achieves its minimum at x and x'. By some calculus, we can compute the derivatives $f'(x) = 0$ and $f'(x') = 0$. On the other hand, since $f''(t) > 0$ for all t, we see f' is strictly increasing, so $f'(x) = f'(x')$ implies $x = y$ by Example 1.46. ■

Our last example of this subsection will be relevant to us when we study group theory in Chapter 5.

Example 1.48. There exists exactly one real number z such that $z + a = a + z = a$ for all real numbers a.

Proof. Before jumping into the uniqueness part of this proof, we see that there are at least one real number z because $z = 0$ will work: $0 + a = a + 0 = a$ for all real numbers a.

We now show uniqueness. Suppose there are two such real numbers z and z'. The trick, now, is to compute $z + z'$ in two ways: we see

$$z + z' = z \qquad \text{and} \qquad z + z' = z'.$$

Thus, $z = z'$. ■

If you made it this far, then the discussion problems below should not be conceptually too difficult. Focus on writing neat, complete, and rigorous proofs.

Exercise 1.49. We know that if a natural number n is even, then there is another natural number k such that $n = 2k$. Prove that this k is unique.

Exercise 1.50. Let q be a non-zero rational number. Prove that q has a unique multiplicative inverse; that is, there exists a unique rational number r such that $qr = 1$.

Exercise 1.51. Show that the set of rational numbers $\mathbb{Q}$ is closed under addition and multiplication. Is the set $\mathbb{R} \setminus \mathbb{Q}$ of irrational numbers also closed under addition and multiplication? If so, prove it, and if not, find a counterexample.

1.3.4 Problems

Problem 1.9 (triangle inequality). Prove the following.

(a) For any real numbers a and b,
$$|a + b| \leq |a| + |b|.$$

(b) For any real numbers x, y, and z,
$$|x - z| \leq |x - y| + |y - z|.$$

(c) For any vectors u, v, and w in $\mathbb{R}^3$,
$$\|u - w\| \leq \|u - v\| + \|v - w\|,$$
where $\|z\|$ denotes the length of a vector z in $\mathbb{R}^3$.

Problem 1.10. Prove that if n is an integer, then $3n^2 + n + 10$ is even.

Problem 1.11. Prove the following.

(a) If x is even and y is odd then $x + y$ is odd.

(b) If x and y are both even then $x + y$ is even.

(c) If x and y are both odd then $x + y$ is even.

(d) If x is even and y is even then xy is even.

(e) If x is odd and y is even then xy is even.

(f) If x and y are both odd then xy is odd.

Problem 1.12. Show that $\sqrt[3]{2}$ is irrational.

CHAPTER 2
INTRODUCTION TO SETS

Set theorists are the fun police of math.

—Bryce Goldman

2.1 Week 4: Sets and Set Operations

Together with propositional logic, set theory underpins the vast majority of modern mathematics, and is a crucial component of almost any proof you will encounter in any upper division course. The language of set theory provides us with a framework to formalize many mathematical concepts that we are intuitively familiar with—namely, collections of data and functions between them. Throughout the remainder of this course, and almost certainly beyond it, set theory will be used extensively.

2.1.1 Sets

We begin with the definition of a set.

Definition 2.1 (set, element)**.** A *set* X is a collection of objects. An object x in a set X is called an *element* or *member*. If x is an element of X, then we write $x \in X$. Similarly, if x is not an element of X, then we write $x \notin X$. Two sets are equal if and only if they have the same elements.

Example 2.2 (empty set)**.** There is a set, denoted $\varnothing$, which contains no elements. We call $\varnothing$ the *empty set*.

Note that $X = \{1, 1, 1, 1\}$ is the same set as $Y = \{1\}$. After all, $1 \in X$, and 1 is the only number with this property. So sets don't recognize multiple "copies" of their elements. Sets also do not respect order; for example, $\{1, 2, 3\} = \{3, 2, 1\}$.

We will also often want to talk about a set contained within some other, larger set.

Definition 2.3 (subset)**.** Let X and Y be sets. If $y \in Y$ implies $y \in X$ for all y, then we say that Y is a *subset* of X, and we write $Y \subseteq X$. If also $Y \neq X$, we write $Y \subsetneq X$, and we say that Y is a *proper subset* of X.

Warning 2.4. Some authors will use $Y \subset X$ to mean either that Y is a subset or a proper subset of X! While both conventions are acceptable, it's best to choose one of the two and be as consistent with this choice as possible in your writing; to avoid ambiguity, it also helps to explicitly state when a subset is proper.

For example, Barack Obama (let's denote him O) is an element of the set P of all presidents of the United States, so we can write $O \in P$. To write down all the elements of P, we can say

$$P = \{\text{Joe Biden}, \text{Donald Trump}, \text{Barack Obama}, \text{George W. Bush}, \dots\}.$$

If Q denotes the set of all world leaders, then $P \subset Q$. For example, $O \in Q$. Is $P \in Q$? No, because the set of all presidents is not a world leader.

Next, let's define some special sets, written in blackboard font to emphasize their importance.

Definition 2.5. The following sets will be used throughout your mathematical career.

- $\mathbb{N}$ is the set of all natural numbers: $\mathbb{N} \coloneqq \{0, 1, 2, \dots\}$.
- $\mathbb{Z}$ is the set of all integers: $\mathbb{Z} \coloneqq \mathbb{N} \cup \{-1, -2, \dots\}$.
- $\mathbb{Q}$ is the set of all rational numbers.
- $\mathbb{R}$ is the set of all real numbers.
- $\mathbb{C}$ is the set of all complex numbers.

Warning 2.6. Some authors use $\mathbb{N}$ to refer to the set of positive integers $\{1, 2, 3, \dots\}$. We will use the notation $\mathbb{Z}^+$ to refer to this set.

Notice that

$$\mathbb{N} \subsetneq \mathbb{Z} \subsetneq \mathbb{Q} \subsetneq \mathbb{R} \subsetneq \mathbb{C}.$$

Also note that some mathematicians exclude 0 from $\mathbb{N}$. If it matters whether $0 \in \mathbb{N}$, we'll try to indicate whether it's true or not.

2.1.2 Set Operations

Now, we will define several useful sets and operations on them, many of which you will encounter in almost any upper division math class.

The simplest operations involve two sets, so we begin with those. We begin with operations which act on single sets.

Definition 2.7 (union, intersection, product)**.** Let X and Y be sets.

- The *union* of X and Y, written $X \cup Y$, is the set consisting of elements in X or Y:

$$X \cup Y \coloneqq \{z : z \in X \text{ or } z \in Y\}.$$

- The *intersection* of X and Y, written $X \cap Y$, is the set consisting of elements in X and Y:

$$X \cap Y \coloneqq \{z : z \in X \text{ and } z \in Y\}.$$

- The *product* of X and Y, written $X \times Y$, is the set of all ordered pairs of elements in X and in Y:

$$X \times Y \coloneqq \{(x, y) : x \in X \text{ and } y \in Y\}.$$

With that said, there are operations on single sets.

Definition 2.8 (power set, union, intersection)**.** Let X be a set.

- The *power set* of X, written $\mathcal{P}(X)$ or 2^X, is the set of all subsets of X:
$$\mathcal{P}(X) \coloneqq \{Y : Y \text{ is a set and } Y \subseteq X\}.$$
- If $\mathcal{F} = X$ is a "collection" or "family" of sets (i.e., a set containing sets), then the *union* of all the sets in $\mathcal{F}$, written $\bigcup \mathcal{F}$, is
$$\bigcup \mathcal{F} \coloneqq \{z : \text{there is a set } Z \in \mathcal{F} \text{ such that } z \in Z\}.$$
- If $\mathcal{F} = X$ is a collection of sets, then the *intersection* of all the sets in $\mathcal{F}$, written $\bigcap \mathcal{F}$, is
$$\bigcap \mathcal{F} \coloneqq \{z : z \in Z \text{ for all sets } Z \in \mathcal{F}\}.$$

Example 2.9. The intersection of the set P of presidents of the United States and the set R of royalty of the United Kingdom is empty: $P \cap R = \varnothing$. Their product $P \times R$ consists of all the different ways we could pair a president with a royal; for example, $(\text{Abraham Lincoln}, \text{Queen Elizabeth I}) \in P \times R$.

Remark 2.10. An ordered pair is not the same thing as a set with two elements: ordered pairs allows for repetition, and order matters.

Here are a few exercises for you to try.

Exercise 2.11. Explain why it's true that every element of $\varnothing$ is even. Is it also true that every element of $\varnothing$ is odd?

Exercise 2.12. Suppose A, B, C, and D are sets. Is it true that $(A \times B) \cup (C \times D) = (A \cup C) \times (B \cup D)$? If yes, prove this claim; if not, find a counterexample, and decide whether one of these sets contains the other.

Exercise 2.13. Let X be the set $\{1, 2, 3\}$. Determine the following.

(a) What is the power set $\mathcal{P}(X)$?

(b) Is 1 an element of $\mathcal{P}(X)$? What about $\{1\}$?

(c) Is $\{2, 3\}$ a subset of $\mathcal{P}(X)$? What about $\{\{2, 3\}\}$? What about $\{\{2\}, \{3\}\}$?

(d) Is $\varnothing$ an element of $\mathcal{P}(X)$? Is $\varnothing$ a subset of $\mathcal{P}(X)$?

(e) Is X an element of $\mathcal{P}(X)$? Is X a subset of $\mathcal{P}(X)$?

Let's give a few examples of theorems about sets.

Theorem 2.14. Let X, Y, and Z be sets. Then the following are true.

(a) $X \subseteq X$.

(b) If $X \subseteq Y$ and $Y \subseteq X$ then $X = Y$.

(c) If $X \subseteq Y$ and $Y \subseteq Z$ then $X \subseteq Z$.

Proof. We show these one at a time.

(a) For every $x \in X$, we see $x \in X$. So $X \subseteq X$.

(b) Assume $X \subseteq Y$ and $Y \subseteq X$. Then $x \in X$ if and only if $x \in Y$, so $X = Y$ follows.

(c) Assume $X \subseteq Y$ and $Y \subseteq Z$. Let $x \in X$; we must show $x \in Z$. Because $x \in X$, it follows that $x \in Y$. Because $x \in Y$, it follows that $x \in Z$. ■

Remark 2.15. When we study relations in section 2.2, we will see that Theorem 2.14 implies that the relation $\subseteq$ is a "partial order" of $\mathcal{P}(X)$. One also says that $\mathcal{P}(X)$ is a "poset," for "partially ordered set."

Here is the last operation of this section, requiring two or three sets depending on viewpoint.

Definition 2.16 (complement)**.** Suppose A and B are sets, both contained in a set X. The *complement* of A in X, denoted A^c, is the set

$$A^c = \{x \in X : x \notin A\}$$

Similarly, write $A - B$ or $A \setminus B$ for the *relative complement* (or *set difference*) of A with B,

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

Exercise 2.17. If A is a subset of a set X, express A^c as a relative complement.

Exercise 2.18. Let A and B be subsets of a set X Prove the following.

(a) $A \subseteq A \cup B$.

(b) $A \cap B \subseteq B$.

(c) $A \setminus B \subseteq A$.

Theorem 2.19 (de Morgan's laws)**.** Suppose that A, B, and C are subsets of X. Then the following are true.

(a) $(A^c)^c = A$.

(b) $(A \cap B)^c = A^c \cup B^c$.

(c) $(A \cup B)^c = A^c \cap B^c$.

Proof. Generally speaking, to show that two sets are equal, it suffices by Theorem 2.14 to show that they are subsets of each other. We will use this approach for (b) and (c).

(a) By definition, $x \in (A^c)^c$ if and only if x is not in A^c. But A^c consists of precisely those elements of X not in A, so $x \notin A^c$ if and only if $x \in A$. In total,

$$x \in (A^c)^c \text{ if and only if } x \in A,$$

so $A = (A^c)^c$ follows.

(b) In one direction, let $x \in (A \cap B)^c$, and we show $x \in A^c \cup B^c$. Then $x \notin A \cap B$. If $x \in A$ and $x \in B$, then this is a contradiction, so $x \notin A$ or $x \notin B$. So $x \in A^c \cup B^c$. Therefore, $(A \cap B)^c \subseteq A^c \cup B^c$.

For the other direction, let $x \in A^c \cup B^c$, and we show $x \in (A \cap B)^c$. Well, either $x \notin A$ or $x \notin B$, so $x \notin A \cap B$, so $x \in (A \cap B)^c$. Therefore, $(A \cap B)^c \supseteq A^c \cup B^c$.

(c) This proof is similar to (b). In one direction, let $x \in (A \cup B)^c$. Thus, $x \notin A \cup B$, so $x \notin A$ or $x \notin B$. It follows $x \in A^c \cap B^c$ and so $(A \cup B)^c \subseteq A^c \cap B^c$.

In the other direction, let $x \in A^c \cup B^c$. Then $x \notin A$ and $x \notin B$, so $x \notin A \cup B$. It follows $x \in (A \cup B)^c$ and so $(A^c \cup B^c) \subseteq (A \cup B)^c$. ■

2.1.3 Problems

Problem 2.1. Let X, Y, and Z be sets. Show the following.

(a) $X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$.

(b) $X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$.

(Hint: Try drawing a Venn diagram to visualize each set!)

Problem 2.2. Suppose A, B, C are subsets of X. Write $A \triangle B$ for the *symmetric difference* of A and B in X, which is

$$A \triangle B = (A \setminus B) \cup (B \setminus A).$$

In the Venn diagram of A and B, $A \triangle B$ consists of the parts of the diagram that are in exactly one of A and B, but not both.

Show the following.

(a) $x \in X$ has $x \in A \triangle B$ if and only if $x \in A$ or $x \in B$ but not both.

(b) $(A \triangle B) \triangle (B \triangle C) = A \triangle C$. (Hint: break each step into cases and apply part 1)

(c) $(A \triangle B) \triangle C = A \triangle (B \triangle C)$.

(d) $A \cap (B \triangle C) = (A \cap B) \triangle (A \cap C)$.

Problem 2.3. Suppose X is some set.

(a) Find a subset $E_1 \subseteq X$ such that for every $A \subseteq X$, we have $E_1 \cup A = A$. Is E_1 unique?

(b) Find a subset $E_2 \subseteq X$ such that for every $A \subseteq X$, we have $E_2 \cap A = A$. Is E_2 unique?

(c) Find a subset $E_3 \subseteq X$ such that for every $A \subseteq X$, we have $A \setminus E_3 = A$. Is E_3 unique?

Problem 2.4. Let X be a set. Show that $\bigcup \mathcal{P}(X) = X$ and $\bigcap \mathcal{P}(X) = \varnothing$.

2.2 Week 5: Functions and Relations

Functions appear throughout mathematics. In linear algebra, you will see functions called "linear transformations." In Math 104, you may deal with sequences, metrics, and homeomorphisms—all different types of functions. In Math 113, you will learn about homomorphisms, another special kind of function. We begin this section by introducing functions and some of the vocabulary associated with them.

2.2.1 Functions

We begin with a few definitions.

Definition 2.20 (function)**.** Let X and Y be sets. A *function*, *mapping*, *morphism*, or *transformation* $f\colon X \to Y$ is a "rule" by which each element of X is assigned exactly one element of Y. If f sends $x \in X$ to $y \in Y$, we write $f(x) = y$, or $f\colon x \mapsto y$.

Remark 2.21. The truly pedantic may prefer to think of a function as its "graph," which is the set

$$\{(x, f(x)) : x \in X\}.$$

From this perspective, a function f is a subset of $X \times Y$ satisfying the following condition: for each $x \in X$, there is exactly one $y \in Y$ such that $(x, y) \in f$. We will not use this perspective going forward.

Example 2.22. Here are some functions.

- Sending natural numbers $n \in \mathbb{N}$ to their square $n^2 \in \mathbb{N}$ defines a function $f\colon \mathbb{N} \to \mathbb{N}$. In other words, we may define a function $f\colon \mathbb{N} \to \mathbb{N}$ by $f(n) = n^2$ for each $n \in \mathbb{N}$.
- We can define a function $g\colon \mathbb{N} \to \mathbb{R}$ by $g(n) \coloneqq \sqrt{n}$ for each $n \in \mathbb{N}$.

Non-Example 2.23. A function $X \to Y$ assigns exactly one element of Y to each element of X. As such, the rule which sends $x \in \mathbb{R}$ to the $y \in \mathbb{R}$ such that (x, y) lies on the unit circle is not a function. Indeed, if $y \neq 0$, then $(x, -y)$ also lies on the unit circle, so we are sending the single x-value to multiple y-values.

A rule between sending elements of X to elements of Y is said to be "well-defined" if it defines a function. If you are asked to prove that a rule $f\colon X \to Y$ is well-defined, then you should show that for all $x, x' \in X$, then $f(x) \in Y$, and if $x = y$, then $f(x) = f(y)$.

Non-Example 2.24. Consider $f\colon \mathbb{Q} \to \mathbb{Q}$ defined by the rule

$$f\left(\frac{a}{b}\right) \coloneqq a + b.$$

This f is not well-defined. To see this, we must find $x, y \in \mathbb{Q}$ such that $x = y$ but $f(x) \neq f(y)$. Let $x = \frac{1}{2}$ and $y = \frac{2}{4}$. Then $x = y$, but $f(x) = 3$ and $f(y) = 6$, so f is not well-defined.

A function $f\colon X \to Y$ helps us understand the sets X and Y. Here are the corresponding nouns.

Definition 2.25 (domain, codomain, pre-image)**.** Let $f\colon X \to Y$ be a function. Then X is called the *domain* of f, and Y is called the *codomain* or *target* of f. More generally, given a subset $Y' \subseteq Y$, the set

$$f^{-1}(Y') \coloneqq \{x \in X : f(x) \in Y'\}$$

is called the *pre-image* of Y' under f.

Functions relate sets together, but we might also want to relate functions together. Composition is how this is done.

Definition 2.26 (composition)**.** Let X, Y, and Z be sets, and let $f\colon X \to Y$ and $g\colon Y \to Z$ be functions. Then we define the *composition* of f and g, denoted $(g \circ f)$, to be the function $(g \circ f)\colon X \to Z$ given by

$$(g \circ f)(x) \coloneqq g(f(x)).$$

Exercise 2.27. Define the functions $f, g, h\colon \mathbb{Z} \to \mathbb{Z}$ by $f(n) \coloneqq 2n$ and $g(n) \coloneqq n + 1$ and $h(n) \coloneqq 2n - 2$. Compute $f \circ (g \circ h)$ and $(f \circ g) \circ h$.

To understand functions, we will want to give them adjectives. Here are a few.

Definition 2.28 (injective, surjective)**.** Let $f\colon X \to Y$ be a function.

- We say f is *injective* or *one-to-one* if and only if $f(x_1) = f(x_2)$ implies $x_1 = x_2$.
- We say f is *surjective* or *onto* Y if and only if $f(X) = Y$. In other words, for each $y \in Y$, there exists $x \in X$ such that $f(x) = y$.
- If f is both injective and surjective, we say that f is a *bijection* or *one-to-one correspondence*.

Intuitively, injective functions are "efficient" in that they don't send two points to the same point. Surjective functions are "effective" in that they hit every point. As such, bijective functions are both "efficient and effective."

In some sense, a bijection $f\colon X \to Y$ tells us that X and Y are essentially the same. As such, we expect to be able to go backwards $Y \to X$ along f. Indeed, this is true.

Proposition 2.29. Let $f\colon X \to Y$ be a function. A function $f^{-1}\colon Y \to X$ is an *inverse* for f if and only if $f^{-1}(f(x)) = x$ for all $x \in X$ and $f\left(f^{-1}(y)\right) = y$ for all $y \in Y$. The function f is bijective if and only if its has an inverse.

Proof. This proof has two directions.

- Suppose f is bijection. We need to show that f has an inverse. To do this, we construct a function which looks like it could be an inverse, and then we prove that it actually is one.

  We begin by defining f^{-1}. Because f is surjective, for every $y \in Y$, we can find a $x \in X$ such that $f(x) = y$; in fact, this x is unique because if we found another x' with $f(x') = y$, then $f(x) = y = f(x')$ implies $x = x'$ because f is injective. Thus, we define $f^{-1}\colon Y \to X$ by sending $y \in Y$ to the unique $x \in X$ such that $f(x) = y$.

  We now check that f^{-1} defines our inverse function. By construction, $f\left(f^{-1}(y)\right) = y$. Further, for any $x \in X$, we set $y \coloneqq f(x)$ and see that $f(x) = y$ implies

  $$x = f^{-1}(y) = f^{-1}(f(x)),$$

  finishing.

- Suppose f has an inverse $f^{-1}\colon Y \to X$. We show that f is bijective. To show that f is injective, suppose $y \coloneqq f(x) = f(x')$ for some $x, x' \in X$. Then $f^{-1}(f(x)) = x$ and $f^{-1}(f(x')) = x'$, so $x = f^{-1}(y) = x'$.

  To show f is surjective, for each $y \in Y$, we note that $x \coloneqq f^{-1}(y)$ has $f(x) = y$ by definition of f^{-1}.

The above implications complete the proof. ∎

Exercise 2.30. Find sets X and Y and functions $f\colon X \to Y$ and $g\colon Y \to X$ such that $g(f(x)) = x$ for all $x \in X$, but there exists $y \in Y$ such that $f(g(y)) \neq y$. Is f injective?

Remark 2.31. Even though we defined a bijective function as being both injective and surjective, in practice is often easier to construct an inverse function instead. For the most part, this will be our strategy when exhibiting a bijection in the future.

On reason to care about the adjectives we have defined so far is that they can "build" up all functions, in the following sense.

Proposition 2.32. Let $f\colon X \to Z$ be a function. There exists a set Y and functions $\pi\colon X \to Y$ and $\iota\colon Y \to Z$ such that π is surjective, ι is injective, and

$$f = \iota \circ \pi.$$

Proof. As in Proposition 2.29, the strategy here will be to construct all of our objects first and then show that they satisfy the desired properties. To get some intuition of what is going on here, the idea is that π is supposed to do all the squishing that f does, and all that is left over is some injective part.

More explicitly, define $Y \coloneqq f(X)$. Then for each $x \in X$, we see that $f(x) \in f(X)$, so we define $\pi\colon X \to Y$ by $\pi(x) \coloneqq f(x)$. Continuing, we note that $f(X) \subseteq Z$, so we define $\iota\colon Y \to Z$ by $\iota(y) \coloneqq y$. We now check that this works.

- We check that π is surjective. Indeed, for any $y \in f(X)$, by definition of $f(X)$, there exists some $x \in X$ such that $f(x) = y$.
- We check that ι is injective. Indeed, given $y, y' \in f(X)$ such that $\iota(y) = \iota(y')$, we note that $\iota(y) = y$ and $\iota(y') = y'$, so $y = y'$ follows.
- We check that $f = \iota \circ \pi$. Well, for any $x \in X$, we see that

  $$\iota(\pi(x)) = \iota(f(x)) = f(x)$$

  where we have used the definitions of π and ι.

The above checks complete the proof. ∎

Remark 2.33. At the cost of an extra function $\widetilde{f}$, we can control Y a bit more. Namely, one can show the following: there are sets $X_0 \subseteq X$ and $Z_0 \subseteq Z$ and functions $\pi\colon X \to X_0$, $\widetilde{f}\colon X_0 \to Z_0$, and $\iota\colon Z_0 \to Z$ such that π is surjective, $\widetilde{f}$ is bijective, ι is injective, and

$$f = \iota \circ \widetilde{f} \circ \pi.$$

The reader is encouraged to try to prove this, but it is nontrivial.

2.2.2 Relations and Orders

Relations allow us to compare elements within a set.

Definition 2.34 (relation)**.** Let X be a set, and let n be a positive integer. An *n-ary* relation on X is a subset of X^n. If R is an n-ary relation on X and $(x_1, \ldots, x_n) \in X^n$, then we write $R(x_1, \ldots, x_n)$ to mean $(x_1, \ldots, x_n) \in R$. If $n = 2$, we might also write $x_1 R x_2$ to mean $(x_1, x_2) \in R$.

We might say "binary" instead of 2-ary.

Example 2.35. Notice that a 1-ary relation on a set X is just a subset of X. For example, consider the 1-ary relation E on $\mathbb{N}$ where $E(n)$ means "n is even." Then E is the set of even natural numbers.

Example 2.36. Let $R \subseteq \mathbb{Z} \times \mathbb{Z}$ be the set of all ordered pairs (x, y) such that $x + y$ is even. Then R is relation. For example, we have $1R21$.

Example 2.37. Define the function $f\colon \mathbb{R} \to \mathbb{R}$ by $f(x) \coloneqq x^2$. Then we define $R \subseteq \mathbb{R} \times \mathbb{R}$ by

$$\{(x_1, x_2) \in \mathbb{R} \times \mathbb{R} : f(x_1) = f(x_2)\}.$$

Then R is relation. For example, $0R0$ and $2R(-2)$.

Orderings provide special examples of relations.

Example 2.38. Define $L \subseteq \mathbb{N} \times \mathbb{N}$ by

$$L \coloneqq \{(a, b) \in \mathbb{N} \times \mathbb{N} : a \leq b\}.$$

Then L is a relation. For example, $1L2$, but $(2, 1) \notin L$.

Let's codify what we mean by an order.

Definition 2.39 (parital order)**.** A *partially ordered set*, or *poset*, is a set X with a binary relation $\preceq$ on X with the following properties.

(a) Reflexivity: for all $x \in X$, we have $x \preceq x$.

(b) Antisymmetry: for all $x, y \in X$, if $x \preceq y$ and $y \preceq x$, then $x = y$.

(c) Transitivity: for all $x, y, z \in X$, if $x \preceq y$ and $y \preceq z$, then $z \preceq z$.

In this case, we call $\preceq$ a *partial order* on X.

Exercise 2.40. For natural numbers a and b, we write $a \mid b$ to mean "b is divisible by a." Note $\mid$ defines a binary relation on $\mathbb{N}$.

(a) Write down three pairs of natural numbers which are elements of the relation and three pairs of natural numbers which are not in the relation.

(b) Prove that $(\mathbb{N}, \mid)$ is a partially ordered set.

(c) For natural numbers a and b, write $a \nmid b$ to mean "b is not divisible by a." It is also true that $\nmid$ is a binary relation on $\mathbb{N}$. Is $\nmid$ a partial order?

However, partial orders are a bit weak. Continuing Exercise 2.40, we note that 2 does not divide 3, and 3 does not divide 2. If we want our order to be a good notion of "size," then we would like this sort of thing to not happen. We want total orders.

Definition 2.41 (total order)**.** A *totally ordered set* is a set X with a partial order $\preceq$ satisfying the following fourth property.

(d) Totality: for all $x, y \in X$, we have $x \preceq y$ or $y \preceq x$.

In this case, we call $\preceq$ a *total order* on X.

Example 2.42. The relation L of Example 2.38 is a total order.

2.2.3 Equivalence Relations

A common way to compare two things is to say that they are similar to each other. For sets, the way to say that two elements of a set are similar to each other is with an equivalence relation.

Definition 2.43 (equivalence relation)**.** Let X be a set, and let $\sim$ be a binary relation on X. We will use the notation $x \sim y$ to mean that $\sim$ holds of the pair (x, y). Say that $\sim$ is an *equivalence relation* on X if all the following hold:

(a) Reflexivity: for all $x \in X$, we have $x \sim x$.

(b) Symmetry: for all $x, y \in X$, if $x \sim y$, then $y \sim x$.

(c) Transitivity: for all $x, y, z \in X$, if $x \sim y$ and $y \sim z$, then $x \sim z$.

Here is an example that will be relevant to us later in section 5.1.2.

Proposition 2.44. Let n be a positive integer, and let $\sim$ be the binary relation on $\mathbb{Z}$ where $x \sim y$ if and only if $x - y$ is divisible by n. Then $\sim$ is an equivalence relation.

Proof. To prove that $\sim$ is an equivalence relation, we must show that $\sim$ is reflexive, symmetric, and transitive.

(a) Reflexivity: let $x \in \mathbb{Z}$. Then $x - x = 0$ is divisible by n because $0 = 0 \cdot n$, so $x \sim x$.

(b) Symmetry: let $x, y \in \mathbb{Z}$, and suppose $x \sim y$. Then $x - y$ is divisible by n, so we can find an integer a such that $x - y = an$. But then

$$y - x = -1 \cdot (x - y) = -a \cdot n,$$

so $y - x$ is divisible by n, so $y \sim x$.

(c) Transitivity: let $x, y, z \in \mathbb{Z}$, and suppose that $x \sim y$ and $y \sim z$. Then $x - y$ and $y - z$ are divisible by n, so we can find integers a and b such that $x - y = an$ and $y - z = bn$. Summing, we see

$$x - z = (x - y) + (y - z) = an + bn = (a + b)n.$$

Thus $x - z$ is divisible by n, so $x \sim z$. ■

Exercise 2.45. Show that the relation R of Example 2.37 is an equivalence relation.

One can generalize Example 2.37 as follows.

Proposition 2.46. Let X be a set and $f \colon X \to Y$ be a function. Then define the relation $\sim_f$ on X by $x_1 \sim_f x_2$ if and only if $f(x_1) = f(x_2)$. Then $\sim_f$ is an equivalence relation.

Proof. As before, to show that $\sim_f$ is an equivalence relation, we must show that $\sim_f$ is reflexive, symmetry, and transitive.

(a) Reflexivity: for each $x \in X$, we note $f(x) = f(x)$, so $x \sim_f x$.

(b) Symmetry: given $x, y \in X$ such that $x \sim_f y$, we see $f(x) = f(y)$. But then $f(y) = f(x)$, so $y \sim_f x$ as well.

(c) Transitivity: suppose $x, y, z \in X$ have $x \sim_f y$ and $y \sim_f z$. Then $f(x) = f(y)$ and $f(y) = f(z)$, so $f(x) = f(z)$, meaning $x \sim_f z$. ■

We said that equivalence relations declare elements similar to each other, so it is useful to talk about all the elements similar to each other as one object.

Definition 2.47 (equivalence class)**.** Let X be a set and $\sim$ be an equivalence relation on X. An *equivalence class* is a subset $Y \subseteq X$ such that, for all $y_1, y_2 \in Y$, we have $y_1 \sim y_2$. If $x \in X$, then the set

$$[x] \coloneqq \{y \in X : x \sim y\}$$

is the equivalence class of x.

Remark 2.48. Technically, we must check that $[x]$ is in fact an equivalence class. For completeness, we do this: if $y_1, y_2 \in [x]$, we must show $y_1 \sim y_2$. Well, by definition of $[x]$, we see $x \sim y_1$ and $x \sim y_2$. But $\sim$ is an equivalence relation! It follows $y_1 \sim x$ and $x \sim y_2$, so $y_1 \sim y_2$.

Example 2.49. Let n be a positive integer. Using Proposition 2.44, consider the equivalence relation $\sim$ on $\mathbb{Z}$ where $x \sim y$ if and only if $x - y$ is divisible by n. Then $[0]$ is the set of multiples of n, which are the integers with last digit 0. One can check that the other equivalence classes are $[1], [2], \ldots, [n-1]$.

Example 2.50. Using the relation R of Example 2.37, we see that

$$[1] = \left\{x \in \mathbb{R} : x^2 = 1^2\right\} = \{\pm 1\}.$$

More generally, $[y] = \{\pm y\}$.

Having divided our set into equivalence classes, we now pick up the equivalence classes back again.

Definition 2.51. Let X be a set and $\sim$ an equivalence relation on X. Then $X/\!\sim$, usually pronounced "X mod $\sim$," is the set of equivalence classes of X under the equivalence relation $\sim$.

Example 2.52. Let n be a positive integer. Using Proposition 2.44, consider the equivalence relation $\sim$ on $\mathbb{Z}$ where $x \sim y$ if and only if $x - y$ is divisible by n. Then we saw $\mathbb{Z}/\!\sim$ is the set $\{[0], [1], [2], \ldots, [n]\}$.

Exercise 2.53. The following problem is similar to one you will likely see in Math 113. Fix a positive integer n, and consider again the equivalence relation $\sim$ on $\mathbb{Z}$ where $x \sim y$ if and only if $x - y$ is divisible by n. Recalling $\mathbb{Z}/\!\sim = \{[0], [1], [2], \ldots, [n-1]\}$, show that the function $f\colon (\mathbb{Z}/\!\sim) \times (\mathbb{Z}/\!\sim) \to \mathbb{Z}/\!\sim$, given by

$$f\colon ([a], [b]) \mapsto [a + b],$$

is well-defined.

The notion of $X/\!\sim$ allows us to build the following reversed version of Proposition 2.46. This tells us that functions and equivalence relations are inherently intertwined: any function gives an equivalence relation, and any equivalence relation gives back a function.

Exercise 2.54. Let $\sim$ be an equivalence relation on a set X. Define the function $f\colon X \to X/\!\sim$ by sending each $x \in X$ to the equivalence class $[x] \in X/\!\sim$. Show that $x \sim x'$ if and only if $x \sim_f x'$, where $\sim_f$ has been defined as in Proposition 2.46.

This next definition gives another way to think about equivalence relations.

Definition 2.55 (partition)**.** Let X be a set. A *partition* of X is a family $\mathcal{F}$ of subsets of X such that the following hold.

(a) Distinct elements of the partition are disjoint: if $A, B \in \mathcal{F}$ and $A \neq B$, then $A \cap B = \varnothing$.

(b) Every element of the partition is nonempty: if $A \in \mathcal{F}$, then $A \neq \varnothing$.

(c) The union of $\mathcal{F}$ is the entire set X: in other words, every element of X is found in some $A \in \mathcal{F}$.

Exercise 2.56. Let X be a set, and let $\mathcal{F}$ be a partition of X. Define $\sim$ to be the binary relation on X such that $x \sim y$ if and only if there is $A \in \mathcal{F}$ such that $x \in A$ and $y \in A$. Prove that $\sim$ is an equivalence relation.

Exercise 2.57. Let X be a set, and let $\sim$ be an equivalence relation on X. Prove that $X/\!\sim$ is a partition of X.

Again, the previous two exercises tell us that partitions are intertwined with equivalence relations: any partition gives an equivalence relation, and any equivalence relation gives back a partition.

2.2.4 Problems

Problem 2.5. Let $f\colon X \to Y$ and $g\colon Y \to Z$ be functions. Show the following.

(a) If f and g are injective, then $g \circ f$ is injective.

(b) If f and g are surjective, then $g \circ f$ is surjective.

(c) If f and g are bijective, then $g \circ f$ is bijective.

(d) If $g \circ f$ is injective, then f is injective.

(e) If $g \circ f$ is surjective, then g is surjective.

Problem 2.6. Let $f\colon X \to Y$ be a function, $A, B \subseteq X$ be subsets of X, and $C, D \subseteq Y$ be subsets of Y.

(a) Show that $f(A \cup B) = f(A) \cup f(B)$.

(b) Show that $f(A \cap B) \subseteq f(A) \cap f(B)$.

(c) If f is injective, show that $f(A \cap B) = f(A) \cap f(B)$.

(d) Construct an example where $f(A \cap B) \neq f(A) \cap f(B)$.

(e) Show that $f^{-1}(C \cup D) = f^{-1}(C) \cup f^{-1}(D)$.

(f) Show that $f^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D)$.

Problem 2.7. Let S be a set and $\mathcal{P}(S)$ denote the power set of S. For any $A \subseteq S$, define $1_A\colon S \to \{0, 1\}$ be defined by

$$1_A(s) = \begin{cases} 1 & \text{if } s \in A, \\ 0 & \text{if } s \notin A. \end{cases}$$

Prove that 1_A is well-defined. This function 1_A is called the "characteristic function of A in S."

Problem 2.8. Extend Proposition 2.32 as follows: let $T\colon V \to W$ be a linear transformation of vector spaces. Show there exists a vector space U and linear transformations $\pi\colon V \to U$ and $\iota\colon U \to W$ such that π is surjective, ι is injective, and

$$T = \iota \circ \pi.$$

Problem 2.9. For each of the following rules, either prove the rule defines a function or show it is not well-defined.

(a) $f\colon \mathbb{Q} \to \mathbb{Q}$ by $\frac{a}{b} \mapsto ab$

(b) $f\colon \mathbb{N} \to \mathbb{N}$ by $n \mapsto n+1$.

(c) $f\colon \mathbb{Z}^+ \times \mathbb{Z}^+ \to \mathbb{Z}$ by $(a,b) \mapsto \frac{a}{b}$, where $\mathbb{Z}^+$ is the set of positive integers.

(d) $f\colon \mathbb{Q}^+ \to \mathbb{Q}^+$ by $\frac{a}{b} \mapsto \frac{b}{a}$, where $\mathbb{Q}^+$ is the set of positive rational numbers.

(e) $f\colon X \to (X/\sim)$ by $x \mapsto [x]$, where $\sim$ is an equivalence relation on a set X.

(f) $f\colon (X/\sim) \to X$ by $[x] \mapsto x$, where $\sim$ is an equivalence relation on a set X.

Problem 2.10. This exercise will teach you how to construct the set of integers from equivalence classes on the set of pairs of natural numbers. For $(a,b),(c,d) \in \mathbb{N}\times\mathbb{N}$, say $(a,b) \sim (c,d)$ if and only if $a+d = b+c$. We claim that $(\mathbb{N}\times\mathbb{N})/\sim$ looks very much like $\mathbb{Z}$.

(a) Prove that $\sim$ is an equivalence relation.

(b) Write $(\mathbb{N}\times\mathbb{N})/\sim$ as $\mathcal{Z}$. Define $\alpha\colon \mathbb{Z} \mapsto \mathcal{Z}$ by

$$z \mapsto \begin{cases} [(z,0)] & \text{if } z \geq 0, \\ [(0,-z)] & \text{if } z < 0. \end{cases}$$

Prove that α is well-defined and a bijection. Where does the inverse function send $[(a,b)]$?

(c) Define $+_{\mathcal{Z}}\colon \mathcal{Z}\times\mathcal{Z} \to \mathcal{Z}$ by

$$[(a,b)] +_{\mathcal{Z}} [(c,d)] = [(a+c, b+d)]$$

Show that $+_{\mathcal{Z}}$ is well-defined. In fact, show $\alpha(a) +_{\mathcal{Z}} \alpha(b) = \alpha(a+b)$ for any $a,b \in \mathbb{Z}$.

2.3 Week 6: Cardinality

Cardinality is one way mathematicians formalize the concept of the "size" of a set. If we are given two sets X and Y, we might ask whether X and Y have the same size, or whether one is larger than the other. For finite sets, this is not complicated: we can compare the number of elements in X and in Y. What do we do for infinite sets? Do $\mathbb{N}$ and $\mathbb{Z}$ have the same size? How about $\mathbb{Q}$ and $\mathbb{R}$? In this section, we will be able to give formal answers to these questions using functions and cardinality.

2.3.1 Cardinalities

In an auditorium where each audience member is seated in exactly one seat, we can say confidently that the total number of seats is the number of audience members even if we do not know how many audience members or seats there are. In other words, by providing a bijection between audience members and seats, we know the number of each is the same.

With this motivation, we are ready to define cardinality.

Definition 2.58 (cardinality)**.** Two sets X and Y have the same *cardinality* if and only if there is a function $f\colon X \to Y$ which is a bijection.

Example 2.59. Let X be the set of students in MUSA 74, and let Y be the set of student IDs of the students in MUSA 74. Do X and Y have the same cardinality? Yes! There is a bijection $f\colon X \to Y$ which sends each student to their student ID. Each student in MUSA 74 has a unique student ID, so f is a bijection, so by definition of *cardinality*, X and Y have the same cardinality.

Cardinality has allowed to declare that two sets are the same size, but it is also interesting to compare sizes. For finite sets, we can again just count and compare the numbers, but for general sets we will want a more function-based approach as with cardinality.

Example 2.60. Every Berkeley student has a unique student ID. Thus, there is an injective function from the set S of all Berkeley students to the set $\mathbb{Z}$ of all integers, taking students to their IDs. This injective functions convinces us that there are at least as many integers as Berkeley students, even without knowing how many Berkeley students there are.

Example 2.61. Let $\sim$ be an equivalence relation on a nonempty set X. Then the function $\pi\colon X \to X/\!\sim$ sending an element $x \in X$ to its equivalence class $[x] \in X/\!\sim$ is surjective. Thus, the surjection π convinces us that the number of elements of X is at least the number of equivalence classes.

The above two examples give us two ways to think about comparing cardinalities, and it will turn out that they are equivalent in favorable circumstances.

Definition 2.62. Let X and Y be sets. Then we say that the cardinality of X is less than or equal to the cardinality of Y if and only if there is an injective function $\iota\colon X \to Y$.

Example 2.63. There is an injective function $\iota\colon \mathbb{Z} \to \mathbb{Q}$ given by $\iota(x) \coloneqq x$. Thus, the cardinality of $\mathbb{Z}$ is less than or equal to the cardinality of $\mathbb{Z}$.

We now explain Example 2.61.

Proposition 2.64. Let X and Y be sets. Suppose X is nonempty. Then the following are equivalent.

(a) There is an injective function $i\colon X \to Y$.

(b) There is a surjective function $p\colon Y \to X$.

In other words, (a) implies (b), and (b) implies (a).

Proof. We have two claims to show.

- We show that (a) implies (b). Because X is nonempty, we may find some element $a \in X$. We now construct our surjective map $p\colon Y \to X$. Note that $y \in i(X)$ is equivalent to having some $x \in X$ such that $y = i(x)$; because i is injective, this $x \in X$ is unique. Thus, we define

$$p(y) \coloneqq \begin{cases} x & \text{if } i(x) = y, \\ a & \text{if no } x \in X \text{ has } i(x) = y. \end{cases}$$

  For each $x \in X$, we see $p(i(x)) = x$, so p is surjective.

- We show that (b) implies (a). For each $x \in X$, we know that there is some $y \in Y$ such that $p(y) = x$. As such, for each $x \in X$, we define $i(x)$ to be some chosen $y \in Y$ such that $p(y) = x$. This defines a map $i\colon X \to Y$, and we see that
$$p(i(x)) = x$$
for each $x \in X$ by construction. Thus, i is injective: $i(x) = i(x')$ for $x, x' \in X$ implies $x = p(i(x)) = p(i(x')) = x'$. ■

Remark 2.65. One might expect that, if the cardinality of X is less than or equal to the cardinality of Y, and the cardinality of Y is less than or equal to the cardinality of X, then X and Y have the same cardinality. In other words, given injections $i\colon X \to Y$ and $j\colon Y \to X$, then there is a bijection $X \to Y$. This is in fact true, but it is quite nontrivial to prove. For the interested, this result is known as the Cantor–Schröder–Bernstein theorem.

2.3.2 Finite Sets

Cardinality was defined to work for arbitrary sets, but as an aside, we note that finite sets remain well-defined.

Definition 2.66 (finite)**.** A set X is *finite* if and only if there is some $n \in \mathbb{N}$ such that X has the same cardinality as $\{1, 2, \ldots, n\}$. In this case, we say that X has n elements and write $|X| = n$. If no such n exists, we say that X is *infinite*.

This definition might feel a little weird because the intuitive way to think of a set as having, say, 2 elements is not via some bijection. To explain this, saying that a set X has n elements intuitively means that we can enumerate the elements of X as
$$X = \{x_1, x_2, \ldots, x_n\}.$$
However, given such an enumeration, we can then define a bijection $f\colon \{1, 2, \ldots, n\} \to X$ by $f(k) \coloneqq x_k$. And conversely, given a bijection $f\colon \{1, 2, \ldots, n\} \to X$, we can enumerate the elements of X as
$$X = \{f(1), f(2), \ldots, f(n)\},$$
showing visually that X has n elements.

This idea gives a very useful proof technique known as *combinatorial proof*: to show that two natural numbers n and m are the same, just show that there is a bijection between a set with n elements and an element with m elements. Let's see an example of this.

Definition 2.67. Let X be a set and $k \in \mathbb{N}$. By $X^{(k)}$, we mean the set of all subsets of X of cardinality k. If X has n elements, we let $\binom{n}{k}$ denote the cardinality of $X^{(k)}$.

Remark 2.68. It is not terribly difficult to show that if two sets X and Y have the same cardinality, then $X^{(k)}$ and $Y^{(k)}$ have the same cardinality for any $k \in \mathbb{N}$. We outline the proof. Let $f\colon X \to Y$ be a bijection with inverse function $g\colon Y \to X$. Then the functions $f^{(k)}\colon X^{(k)} \to Y^{(k)}$ and $g^{(k)}\colon Y^{(k)} \to X^{(k)}$ given by
$$f^{(k)}\colon A \mapsto f(A) \qquad \text{and} \qquad g^{(k)}\colon B \mapsto g(B)$$
are inverse functions and thus give a bijection $X^{(k)} \to Y^{(k)}$. The reader is encouraged to check as many of these details as they would like.

Example 2.69. For any $n, k \in \mathbb{N}$ with $n \geq k$, we have
$$\binom{n}{k} = \binom{n}{n-k}.$$

Proof. Let X be a set of n elements. We need to show that $X^{(k)}$ and $X^{(n-k)}$ have the same cardinality. If $A \subseteq X$ has k elements then $X \setminus A$ has $n-k$ elements, so we define the map $f\colon X^{(k)} \to X^{(n-k)}$ by $f\colon A \mapsto (X \setminus A)$.

We claim that f is a bijection. By Proposition 2.29, it suffices to show that f has an inverse function, so we define $g\colon X^{(n-k)} \to X^{(k)}$ by $g\colon B \mapsto (X \setminus B)$. Then for any $A \in X^{(k)}$ and $B \in X^{(n-k)}$, we can compute

$$g(f(A)) = g(X \setminus A) = A \qquad \text{and} \qquad f(g(B)) = f(X \setminus B) = B.$$

So g is an inverse for f, so f is a bijection, completing the proof. ∎

For finite sets, checking that a function can be simplified if we at least know our sets have the same size already.

Proposition 2.70. Suppose that X and Y are finite sets of the same cardinality, and let $f\colon X \to Y$ be a function. Then the following are equivalent.

(a) f is injective.

(b) f is surjective.

(c) f is bijective.

In other words, if any one of (a), (b), or (c) is true, then all are true.

Proof. To prove that multiple properties are equivalent, we show that (a) implies (b), that (b) implies (c), and that (c) implies (a). Then, for example, if (b) is true, we know (c) is implied, and then (a) is implied from (c). Anyway, this lets us break down our proof into three parts.

- We show (a) implies (b). Suppose that f is injective. We need to show that f is surjective, which means we want to show $f(X) = Y$. Note f maps X surjectively onto its image $f(X)$, so X and $f(X)$ have the same cardinality.

  However, $f(X)$ is a subset of Y, and because $f(X)$ and Y are both finite sets with the same cardinality, we conclude that $f(X) = Y$. More explicitly, if $Y \setminus f(X)$ had any elements, then Y would have strictly larger cardinality than $f(X)$, which we know to be false.[1]

- Suppose that f is surjective. We need to show that f is bijective. By Proposition 2.29, we just need to find an inverse of f. Strap in—this proof is going to be a little wild.

  We now define a candidate inverse function $g\colon Y \to X$. Using the recipe of (b) implies (a) in Proposition 2.64, we get the surjective function $f\colon X \to Y$ defines an injective function $g\colon X \to X$ such that

  $$f(g(y)) = y \tag{2.1}$$

  for each $y \in Y$. Using the argument of the previous point, because g is injective, we see that g is actually bijective.

  We now finish checking that g is an inverse for f. Namely, given $x \in X$, we must check $g(f(x)) = x$. Well, g is surjective, so we know there is some $y \in Y$ such that $g(y) = x$. But then

  $$f(x) = f(g(y)) = y$$

  by (2.1), so we conclude $g(f(x)) = x$ by definition of y. This completes the proof.

- By definition, if f is bijective, then f is injective.

The above implications complete the proof. ∎

Here are a couple combinatorial proofs for you to try. Feel free to use Proposition 2.70.

[1] We are using the following fact: if X is a finite set, and $A \subseteq X$ is a subset such that A and X have the same cardinality, then $A = X$. We do not quite have the tools to prove this yet (we need induction), so we postpone a formal proof until Proposition 3.8.

Exercise 2.71. Let X be a set, and let $B(X)$ denote the set of functions $X \to \{0,1\}$. Show that $\mathcal{P}(X)$ has the same cardinality as $B(X)$ by sending subsets $A \subseteq X$ to the function $1_A \colon X \to \{0,1\}$ defined by

$$1_A(x) \coloneqq \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Exercise 2.72. Let X be a finite set with n elements. Use the previous exercise to show that $\mathcal{P}(X)$ has 2^n elements. Conclude that

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

2.3.3 Countable Sets

The "next largest" type of set after finite sets are countable sets.

Definition 2.73 (countable)**.** Let X be a set. Then X is *countable* if and only if there is an injective function $f \colon X \to \mathbb{N}$. If X is not countable, we say X is *uncountable*.

Remark 2.74. If X is nonempty, being countable is equivalent to having some surjective function $p \colon \mathbb{N} \to X$.

It's not obvious that there are any uncountable sets at all, and we'll need a powerful proof technique known as diagonalization to show that they exist. We'll have to come back to that later. For now, let's show that there are at a large quantity of countable sets.

Example 2.75. Define the function $i \colon \mathbb{Z} \to \mathbb{N}$ by

$$i(n) \coloneqq \begin{cases} -2n+1 & \text{if } n < 0, \\ 2n & \text{if } n \geq 0. \end{cases}$$

Then i is injective: if $i(n) = i(m)$, then $i(n)$ and $i(m)$ are both even or both odd, so n and m are both negative or both nonnegative. If n and m are both negative, then $-2n+1 = -2m+1$, so $n = m$. The case of n and m both being nonnegative is similar.

Lemma 2.76. Let X, Y be sets. If X is countable and there is an injection $i \colon Y \to X$, then Y is countable.

Proof. Because X is countable, there is an injection $f \colon X \to \mathbb{N}$. Thus, by Problem 2.5, the composition $(f \circ i)$ defines an injection $Y \to \mathbb{N}$, which makes Y countable. ■

Proposition 2.77. Let X be a set. If X is finite, then X is countable.

Proof. Being finite means that there is some $n \in \mathbb{N}$ with a bijection $f \colon X \to \{1, 2, \ldots, n\}$. But $\{1, 2, \ldots, n\} \subseteq \mathbb{N}$, so f actually extends to an injection $f \colon X \to \mathbb{N}$, making X countable. ■

Corollary 2.78. Let X be a set. If X is uncountable, then X is infinite.

Proof. This is the contraposition of Proposition 2.77. ■

It is a pretty powerful result that products of countable sets remain countable. Let's build towards this result.

Lemma 2.79. The set $\mathbb{N} \times \mathbb{N}$ is countable.

Proof. The idea is to write out the following grid.

$$\begin{array}{ccccccc}
(0,0) & \to & (0,1) & & (0,2) & \to & (0,3) & \cdots \\
& \swarrow & & \nearrow & & \swarrow & & \nearrow \\
(1,0) & & (1,1) & & (1,2) & & (1,3) & \cdots \\
\downarrow & \nearrow & & \swarrow & & \nearrow & & \swarrow \\
(2,0) & & (2,1) & & (2,2) & & (2,3) & \cdots \\
& \swarrow & & \nearrow & & \swarrow & & \nearrow \\
(3,0) & & (3,1) & & (3,2) & & (3,3) & \cdots \\
\downarrow & \nearrow & & \swarrow & & \nearrow & & \\
\vdots & & \vdots & & \vdots & & \vdots & \ddots
\end{array}$$

The layout of the grid suggests a surjective function $\mathbb{N} \to \mathbb{N} \times \mathbb{N}$: send 0 to $(0,0)$, then send 1 to $(0,1)$, then send 2 to $(1,0)$, then send 3 to $(2,0)$, and continue the process. This surjective function $\mathbb{N} \to \mathbb{N} \times \mathbb{N}$ shows that $\mathbb{N} \times \mathbb{N}$ is countable by Proposition 2.64. ■

Exercise 2.80. In fact, show that the function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ defined by

$$f(a,b) \coloneqq \frac{(a+b)(a+b+1)}{2} + b$$

is a bijection. This tells us directly that $\mathbb{N} \times \mathbb{N}$ and $\mathbb{N}$ have the same cardinality.

Theorem 2.81. Let X, Y be sets. If X and Y are countable, then $X \times Y$ is countable.

Proof. Because X and Y are countable, there are injections $f : X \to \mathbb{N}$ and $g : Y \to \mathbb{N}$. It follows that the function $i : (X \times Y) \to (\mathbb{N} \times \mathbb{N})$ defined by

$$i(x,y) \coloneqq (f(x), g(y))$$

is injective: if $i(x,y) = i(x',y')$, then $f(x) = f(x')$ and $g(y) = g(y')$, so $x = x'$ and $y = y'$. Because $\mathbb{N} \times \mathbb{N}$ is countable by Lemma 2.79, we conclude that $X \times Y$ is countable by Lemma 2.76. ■

Corollary 2.82. The set $\mathbb{Q}$ is countable.

Proof. The idea here is that any rational number can be written as $\frac{a}{b}$ for integers a and b. Note that the set $\mathbb{Z}^+$ of positive integers is a subset of the countable set $\mathbb{N}$ and thus countable. By Theorem 2.81, we see that $\mathbb{Z} \times \mathbb{Z}^+$ is countable. Thus, we note we have a function $f : \mathbb{Z} \times \mathbb{Z}^+ \to \mathbb{Q}$ defined by

$$f(a,b) \coloneqq \frac{a}{b}.$$

Note that there are no division-by-zero problems here because $b \in \mathbb{Z}^+$. Now, because all rational numbers can be written a/b for a positive integer b, we see that f is surjective. It follows from Proposition 2.64 that there is an injection $\mathbb{Q} \to \mathbb{Z} \times \mathbb{Z}^+$, so $\mathbb{Q}$ is countable by Lemma 2.76. ■

Exercise 2.83. Let $\mathcal{F}$ be a countable family of countable sets. Use Theorem 2.81 to show that the union of all the sets in $\mathcal{F}$ is countable.

Thus far we have built countable sets up from small countable sets. In fact, we can go in reverse and extract countable sets from larger ones.

Theorem 2.84. Every infinite set X has an infinite, countable subset.

Proof. The idea here is that any finite subset $X_n \subseteq X$ must have $X \setminus X_n$ still finite, so we can keep removing finitely many elements from X as long as we please. As such, we construct our countable subset in steps.

0. To begin, we note X is infinite and thus nonempty, so we can find some $x_0 \in X$.
1. Next, $\{x_0\}$ is finite, so $X \setminus \{x_0\}$ is infinite and thus nonempty, so we can find some $x_1 \in X \setminus \{x_0\}$.
2. Next, $\{x_0, x_1\}$ is finite, so $X \setminus \{x_0, x_1\}$ is infinite and thus nonempty, so we can find some $x_2 \in X \setminus \{x_0, x_1\}$.

Continuing the above process, we product a subset

$$Y \coloneqq \{x_0, x_1, x_2, \ldots\}$$

of X. Note that Y is not finite because it has more than n elements for any $n \in \mathbb{N}$. However, Y is countable, because the function $f : Y \to \mathbb{N}$ defined by $f : x_n \mapsto n$ is injective. ∎

2.3.4 Diagonalization and Uncountable Sets

Let's now give an especially powerful contradiction trick, invented in 1891 by Georg Cantor. The trick, called the "diagonal argument," shows that certain sets are uncountable. It is best seen by example.

Theorem 2.85 (Cantor's diagonal argument)**.** The set of real numbers is uncountable.

Proof. It is sufficient to show that the interval $(0, 1)$ in $\mathbb{R}$ is uncountable. By Proposition 2.64, it suffices to show that no function $p : \mathbb{N} \to (0, 1)$ is surjective. Well, pick up some function $p : \mathbb{N} \to (0, 1)$, which allows us to enumerate the image of p as follows, for some specific p.

$$\begin{aligned}
x_1 &= 0.\mathbf{1}23456\ldots \\
x_2 &= 0.1\mathbf{4}1592\ldots \\
x_3 &= 0.10\mathbf{1}010\ldots \\
x_4 &= 0.500\mathbf{0}00\ldots \\
x_5 &= 0.4142\mathbf{1}3\ldots \\
x_6 &= 0.23571\mathbf{1}\ldots \\
&\vdots
\end{aligned}$$

(Explicitly, we have set $x_{i+1} \coloneqq p(i)$ for each i.) We must show that p is not surjective, so we find a real number $x \in (0, 1)$ not in the image of p. Well, for each i, let the ith decimal place of x (after the decimal point) be a 1 if the ith decimal place of x_i is not a 1 and a 2 if the ith decimal place of x_i is a 2. For the above example, we have

$$x = 0.212122\ldots.$$

For all i, the ith decimal place of x differs from the ith decimal place of x_i, so $x \neq x_i$. Thus, x is not in the image of p, which is what we wanted to prove. ∎

Exercise 2.86. Modify the proof Theorem 2.85 to show that the set of functions $\mathbb{N} \to \{0, 1, 2, \ldots, 9\}$ is uncountable.

Uncountability is essentially a size result, so we have essentially proven that $\mathbb{R}$ is a "pretty big set." This has surprising applications.

Definition 2.87 (computable)**.** Let $x \in \mathbb{R}$. Then x is *computable* if and only if there is a computer program which takes a $n \in \mathbb{N}$ as input, and returns the nth digit of x as output.

Corollary 2.88. There is a real number which is not computable.

Proof. Let C denote the subset of all computable real numbers. We claim that C is countable. This will finish because $\mathbb{R}$ is uncountable by Theorem 2.85, so it will imply that $C \subsetneq \mathbb{R}$.

Note every computer program is stored as a finite sequence of zeroes and ones. For every $k \in \mathbb{N}$, the set X_k of all sequences of zeroes and ones of length k is finite, hence countable. Therefore the set

$$X = X_1 \cup X_2 \cup X_3 \cup \cdots$$

of all finite sequences of zeroes and ones is countable by Exercise 2.83 because it is the union of countably many countable sets. Therefore the set Z of all computer programs is countable by Lemma 2.76.

Now, define a function $f \colon Z \to \mathbb{R}$ as follows: if $P \in Z$ computes a real number x, then we set $f(P) \coloneqq x$. Otherwise the program P does not compute a real number, so we don't care about P and define $f(P) \coloneqq 0$. By definition of a countable real number, the function f surjects onto C, so C is countable. This finishes the proof. ■

Here is another application of the diagonal argument, extending Exercise 2.86.

Theorem 2.89. Let X be any set. Each function $f \colon X \to \mathcal{P}(X)$ is not surjective.

For fun, let's begin with a mysterious proof of this result, and then we'll give another proof to explain what's going on.

Proof 1. We claim that the subset

$$Y \coloneqq \{x \in X : x \notin f(x)\}$$

is not in the image of f. Indeed, suppose for the sake of contradiction that $Y = f(x)$ for some $x \in X$. Then $x \in Y$ is equivalent to $x \notin f(x)$, which is equivalent to $x \notin Y$. This is a contradiction. ■

Proof 2. Let's explain what's going on in the above proof. Recall from Exercise 2.71 that $\mathcal{P}(X)$ is the same size as the set of functions $X \to \{0, 1\}$; let F denote this set of functions. It suffices to show that any function $f \colon X \to F$ fails to be surjective. For notational ease, we let $f_x \colon X \to \{0, 1\}$ denote the function $X \to \{0, 1\}$ which f returns when evaluated at $x \in X$.

Well, imitating Exercise 2.86, we imagine that we could list all the elements of X linearly to make a grid as follows.

	x_1	x_2	x_3	$\cdots$
f_{x_1}	0	0	0	$\cdots$
f_{x_2}	1	1	1	$\cdots$
f_{x_3}	0	1	0	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

(For conreteness, we have labeled the elements of X by x_k for $k \in \mathbb{N}$, even though X need not be countable.)

Here, each row is describing a function f_{x_i}, and each column explains what happens when a function f_{x_i} is evaluated at some input. We would like to find a function $g \colon X \to \{0, 1\}$ which is not in the image of f, for

which we employ the diagonal argument: define g to disagree with each f_{x_i} along the diagonal! Namely, we simply define

$$g(x_i) \coloneqq \begin{cases} 1 & \text{if } f_{x_i}(x_i) = 0, \\ 0 & \text{if } f_{x_i}(x_i) = 1. \end{cases}$$

Removing the indices, we are defining

$$g(x) \coloneqq \begin{cases} 1 & \text{if } f_x(x) = 0, \\ 0 & \text{if } f_x(x) = 1. \end{cases}$$

Now, $g(x) \neq f_x(x)$ for each $x \in X$, so $g \neq f_x$ for each $x \in X$. Thus, g is not in the image of f, which is what we wanted. ■

Remark 2.90. To finish explaining the first proof, we note that we can translate everything in the second proof back into subsets of X so that g corresponds to the subset $Y \subseteq X$.

It is a consequence of Theorem 2.89 that $\mathcal{P}(X)$ has strictly larger cardinality than X, which we will see more explicitly in Problem 2.16. In fact, if X was a finite set, then this would be quite clear: if there are n elements in X then there would be 2^n elements in $\mathcal{P}(X)$, and $2^n > n$.

As an application of our more general diagonalization, we can build larger and larger sets.

Corollary 2.91. There is no set of largest cardinality.

Proof. For any set X was a set with the largest cardinality, we set $\mathcal{P}(X)$ has strictly larger cardinality by Problem 2.16, so X does not have the largest cardinality. ■

Corollary 2.92. There is no set U such that $X \in U$ for each set X.

Proof. If such a set U existed, then note that each element of $\mathcal{P}(U)$ is also a set, so $\mathcal{P}(U) \subseteq U$. Thus, there is an injection $\mathcal{P}(U) \to U$, so by Proposition 2.64, there is a surjection $U \to \mathcal{P}(U)$, which contradicts Theorem 2.89. ■

Remark 2.93. It is possible to prove Corollary 2.92 directly with a diagonization similar to Theorem 2.89. To see this, suppose some U exists, and define

$$Y \coloneqq \{x \in U : x \notin x\}.$$

Now, $Y \in U$ because Y is a set, but $Y \in Y$ is equivalent to $Y \notin Y$, which is a contradiction. In some sense, this proof is a rephrasing of Example 1.43, where Y is "the set of all sets which do not contain themselves."

Exercise 2.94. A polynomial with rational coefficients is a function $f : \mathbb{R} \to \mathbb{R}$ of the form

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

where $a_0, a_1, \ldots, a_n \in \mathbb{Q}$. Such a polynomial is "nonzero" if there exists $x \in \mathbb{R}$ such that $f(x) \neq 0$. A real number r is "algebraic" if there is a nonzero polynomial f with rational coefficients such that $f(r) = 0$. Show that the set of algebraic real numbers is countable. Conclude there are real numbers which are not algebraic.

2.3.5 Problems

Problem 2.11. Let X be a set. Given subsets $A, B \subseteq X$, write $X \sim Y$ to mean that A and B have the same cardinality. Prove that $\sim$ is an equivalence relation on $\mathcal{P}(X)$.

Problem 2.12 (inclusion-exclusion principle)**.** Let X and Y be finite sets. Show that

$$|X \cup Y| = |X| + |Y| - |X \cap Y|.$$

Problem 2.13 (Hilbert's grand hotel)**.** The result of Proposition 2.70 is not true for infinite sets. To see why, answer the following riddle.

After mathematicians die, they go to a grand hotel in the heavens with infinitely many rooms. Suppose that every room in the hotel is taken, but that a new mathematician has just arrived at the front door. The usher at the front desk tells her, "Just wait a minute, I need to move some people around." Five minutes later, the usher returns, and though no mathematician has vacated the hotel, there is a room for the new guest! What happened?

Problem 2.14 (Dedekind's definition of infinity)**.** Prove the following.

(a) Suppose X is a finite set. Then any injective function $f\colon X \to X$ is bijective.

(b) Show that there is an injective function $f\colon \mathbb{N} \to \mathbb{N}$ which is not surjective.

(c) Suppose X is any infinite set. Then there exists an injective function $f\colon X \to X$ which is not surjective.

We have described "Dedekind's definition of infinity." It shows that a set is infinite if and only if it could be the set of rooms in Hilbert's grand hotel.

Problem 2.15 (Don't deal with the Devil!)**.** Suppose that you have infinitely many $\$1$-bills, labeled 1, 3, 5, and so on. A rather hellish merchant makes you an offer: he will give you $\$2$ for each of your $\$1$ bills, as follows:

(a) After 30 minutes, he will take the bill labeled 1 and give you \$2 in bills labeled 2 and 4.

(b) After 15 more minutes, he will again take \$1, namely the bill labeled 2, and give you another \$2, in bills labeled 6 and 8.

(c) After another 7.5 minutes, he will take the bill labeled 3 and give you bills labeled 10 and 12.

(d) After another 3.75 minutes, he will take the bill labeled 4 and give you bills labeled 14 and 16.

(e) And so on, until 60 minutes have passed. (Note that $30 + 15 + 7.5 + 3.75 + 1.825 + \cdots = 60$.)

Would you take this offer? Why or why not?

Problem 2.16 (Cantor's paradox)**.** Let X be a set.

(a) Show there exists an injective function $f\colon X \to \mathcal{P}(X)$.

(b) Show there does not exist an injective function $f\colon \mathcal{P}(X) \to X$.

Problem 2.17 (Poincare recurrence)**.** Let X be a set, and let $T\colon X \to X$ be a bijection. Further, for each positive integer n, let $T^{\circ n}\colon X \to X$ denote the n-fold application of T as $T \circ T \circ \cdots \circ T$ (where T is repeated n times).

(a) Suppose X is finite. Show that for every $x \in X$ there is an $n > 0$ such that $T^{\circ n}(x) = x$.

(b) Suppose X is finite. Show that there are infinitely many $n > 0$ such that $T^{\circ n}(x) = x$.

(c) Show that there exists a bijection $T\colon \mathbb{Z} \to \mathbb{Z}$ such that $T^{\circ n}(0) \neq 0$ for each positive integer n.

This phenomenon is known as "Poincare recurrence." If you are very brave (and know the prerequisite physics), interpret Poincare recurrence as the following, highly paradoxical statement: "If an ideal gas is allowed to travel between two chambers and starts in one chambers, eventually all the gas molecules will collect in one of the chambers."

For the last exercises, we require the following definition.

Definition 2.95. Let $X_k = \{1, 2, \ldots, k\}$ and let $X_k^n = X_k \times X_k \times \cdots \times X_k$ (n copies of X_k). The *infinite tree* with k branches, denoted T_k, is the set

$$T_k = X_k \cup X_k^2 \cup X_k^3 \cup \cdots .$$

Let $A \subseteq T_k$. An "infinite path" through A is an infinite set of the form

$$\{(a_1), (a_1, a_2), (a_1, a_2, a_3), \ldots\}$$

where the $a_j \in A$.

So a typical element of T_2, for example, looks like

$$(2, 1, 2, 2, 1, 2, 1, 1, 1, 1, 2, 1, 2),$$

which for convenience we may just choose to write as 2122121111212. An infinite path would look like

$$\{2, 21, 212, 2122, 21221, \ldots\},$$

which for convenience we may just choose to write as $21221\cdots$.

Problem 2.18. Let T_k denote the infinite tree. Show that Ω_k, the set of infinite paths through T_k, is countable if and only if $k = 1$.

Problem 2.19. Let T_k denote the infinite tree. Say that a path

$$\{(a_1), (a_1, a_2), (a_1, a_2, a_3), \ldots\}$$

through T_k is "uncomputable path" if and only if there does not exist a computer program which takes a number $n \in \mathbb{N}$ as input and returns $a_n \in \mathbb{N}$ as output. Show that if $k \geq 2$, then there is an uncomputable path.

CHAPTER 3
INDUCTION AND RECURSION

One fish, two fish, red fish, blue fish.

—Dr. Suess, [Gei60]

3.1 Week 7: Mathematical Induction

In science, "inductive reasoning" is the act of using empirical evidence about the world we live in to come to some sort of conclusion. For example, the following is a valid inductive argument.

1. The sun rose in the east every day of my life so far.
2. Therefore, the sun will rise in the east tomorrow.

However, the above reasoning is not valid in mathematics! For example, consider the following reasoning.

Bad Theorem 3.1. For any nonnegative integer $k \in \mathbb{Z}$, the number $2^{2^k} + 1$ is prime.

Bad proof. The numbers $2^{2^0} + 1 = 3$, $2^{2^1} + 1 = 5$, $2^{2^2} + 1 = 17$, $2^{2^3} + 1 = 257$, and $2^{2^4} + 1$ are all prime, so by inductive reasoning every number of the form $2^{2^k} + 1$ is prime. ■

However $2^{2^5} + 1$ is not prime: it factors as $641 \cdot 6700417$. So, the above sort of inductive reasoning is invalid in mathematics. Instead, we will often make use of mathematical induction, a powerful proof technique which we will use to prove claims about certain types of infinite sets.

3.1.1 Induction

Here is mathematical induction.

Theorem 3.2 (Principle of induction)**.** Let P be a statement indexed by $\mathbb{N}$, the set of all natural numbers. To show $P(n)$ is true for all n, it suffices to show the following.

1. Base case: $P(0)$ is true.
2. Inductive step: Let $k \in \mathbb{N}$. If $P(k)$ is true, then $P(k+1)$ is true.

For philosophical reasons, we will not provide a formal proof of induction, but we will explain why you should believe it.[1]

Explanation. To begin, we observe that by the base case $P(0)$ is true. But $0 \in \mathbb{N}$, so by the inductive step $P(1)$ is also true. Again, since $1 \in \mathbb{N}$ and $P(1)$ is true, by the inductive step $P(2)$ is also true. Proceeding in this fashion, we conclude that $P(n)$ is true for any $n \in \mathbb{N}$ because

$$n = (n-1) + 1 = \cdots = 0 + \underbrace{1 + \cdots + 1}_{n}.$$

In other words, by starting at 0 and repeatedly applying the inductive step, we can reach any natural number n, making $P(n)$ true. ■

We note that we did not need to start at 0; in other words, our base case need not be proving that $P(0)$ is true. Here is an example with a different base case.

Example 3.3. Let's show that every positive integer greater than 4 is at least 5. We show this using induction.

1. Base case: we use 5, where we see $5 \geq 5$.
2. Inductive step: if $k \geq 5$ then $k + 1 \geq k \geq 5$, so by the transitivity of $\geq$ we conclude that $k + 1 \geq 5$ as well.

Exercise 3.4. Explain why we can freely change the base case of an inductive proof. Could we start with a negative number? Can you perform induction on $\mathbb{Z}$? If not, is it possible to adapt the principle of induction so that it can be used on $\mathbb{Z}$?

Let's try a harder example.

Example 3.5. For all $n \in \mathbb{N}$, the number $S_n \coloneqq 1 + 3 + \cdots + (2n + 1)$ is a perfect square; in other words, there is some integer j such that $j^2 = S_n$.

Proof Attempt. We proceed by induction.

1. For our base case, we see that $S_0 = 1$, but $1 = 1^2$ is a perfect square.
2. For the inductive step, assume that we have shown that S_k is a perfect square. For the induction step we have

   $$S_{k+1} = 1 + 3 + \ldots + 2k - 1 + 2k + 1 = S_k + 2k + 1$$

   By assumption S_k is a perfect square so $S_{k+1} = j^2 + 2k + 1$ for some $j \in \mathbb{N}$.

   At first glance, you might think that this is the perfect square $(j+1)^2 = j^2 + 2j + 1$. However, we don't know that $k = j$, so we're stuck. ■

[1] From some perspectives, the principle of induction is a defining property of $\mathbb{N}$ and therefore is not proven.

Because we're stuck, let's try computing a few simple cases. This is often a good way to get a feel for what you're actually trying to prove. Indeed,

$$\begin{aligned} S_0 &= 1 \\ &= 1 = 1^2 \\ S_1 &= 1 + 3 \\ &= 4 = 2^2 \\ S_2 &= 1 + 3 + 5 \\ &= 9 = 3^2 \\ S_3 &= 1 + 3 + 5 + 7 \\ &= 16 = 4^2. \end{aligned}$$

It seems that $S_n = (n+1)^2$, which is a stronger statement than what we were supposed to prove, but maybe we can show that instead.

Proof of Example 3.5. We claim that $S_n = (n+1)^2$ for each $n \in \mathbb{N}$. The base case is the same as before, so we focus on the inductive step.

Indeed, let us assume that $S_k = k^2$. Then

$$\begin{aligned} S_{k+1} &= (1 + \cdots + (2k-1)) + (2k+1) \\ &= S_k + (2k+1) \\ &= k^2 + 2k + 1 \\ &= (k+1)^2, \end{aligned}$$

so we're done. ■

Here's a similar example for you to try.

Exercise 3.6. Prove that for each positive integer n we have

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

Induction can do more than prove equalities.

Example 3.7. We show $2^n > n$ for each positive integer n.

Proof. We proceed by induction.

1. Base case: using $n = 1$ as our base case, we see $2^n = 2^1 = 2 > 1$.
2. Inductive step: suppose $2^k > k$ for some positive integer k. We aim to show that $2^{k+1} > k + 1$. To accomplish this, we first note that $2^{k+1} = 2 \cdot 2^k$. But $2^k > k$ by the inductive hypothesis, so

   $$2^{k+1} > 2k.$$

   Finally, since $k \geq 1$ we have $2k = k + k \geq k + 1$, so putting everything together, we have

   $$2^{n+1} = 2 \cdot 2^n > 2n \geq n + 1.$$

   This completes the induction. ■

Here is a more conceptual inductive result, which we used in the proof of Proposition 2.70.

Proposition 3.8. Let X be a finite set with n elements, where $n \in \mathbb{N}$. If A is a subset of X with n elements, then $A = X$.

Proof. Because of the way we have arranged our definitions, this statement has some content. Unsurprisingly, we will induct on n.

1. Base case: when $n = 0$, we see X has no elements, so $X = \varnothing$. Thus, $A \subseteq X$ forces $A = \varnothing = X$, which is what we wanted.

2. Inductive step: suppose that all k-element subsets A' of k-elements sets X' have $A' = X'$. Now, let X be a set with $k+1$ elements and A be a subset with $k+1$ elements. Because A has more than one elements, we can find $a \in A$. Then
$$A \setminus \{a\} \subseteq X \setminus \{a\},$$
but $A \setminus \{a\}$ and $X \setminus \{a\}$ both have $(k+1) - 1 = k$ elements! Thus, $A \setminus \{a\} = X \setminus \{a\}$ by the inductive hypothesis, so we conclude $A = X$ after adding the element a back in. ■

We close this section with a few more examples for you.

Exercise 3.9. Show that $2^{2n} - 1$ is divisible by 3 for each positive integer n.

Exercise 3.10. The "factorial," denoted $n!$, is the product of the first n positive integers:
$$n! \coloneqq 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n,$$
and by convention, $0! \coloneqq 1$. Additionally, we define the "gamma function" $\Gamma \colon \mathbb{N} \to \mathbb{R}$ by
$$\Gamma(n) \coloneqq \int_0^\infty x^{n-1} e^{-x} dx.$$
Prove that $\Gamma(n+1) = n!$ for each $n \in \mathbb{N}$.

Exercise 3.11. Define the Fibonacci sequence $F_0, F_1, F_2, \ldots$ by $F_n = F_{n-1} + F_{n-2}$ and $F_0 = 0$ and $F_1 = 1$. Show
$$F_0 + F_1 + \cdots + F_n = F_{n+2} - 1.$$
for each positive integer n.

Exercise 3.12. Define the Fibonacci sequence $F_0, F_1, F_2, \ldots$ by $F_n = F_{n-1} + F_{n-2}$ and $F_0 = 0$ and $F_1 = 1$. Show
$$F_0^2 + F_1^2 + \cdots + F_n^2 = F_n F_{n+1}.$$
for each positive integer n.

The moral to this section is as follows.

Idea 3.13. whenever you see which looks like

"For all natural numbers $n \in \mathbb{N}$, property $P(n)$ is true,"

you should try induction first.

3.1.2 Problems

Problem 3.1. Show that

$$1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

for each positive integer n.

Problem 3.2. Show that

$$1^3 + 2^3 + \cdots + n^3 = \left(\frac{n(n+1)}{2}\right)^2$$

for each positive integer n.

Problem 3.3. Show that

$$1 + \frac{1}{4} + \cdots + \frac{1}{n^2} \leq 2 - \frac{1}{n}$$

for each positive integer n.

Problem 3.4. Define the Fibonacci sequence $F_0, F_1, F_2, \ldots$ by $F_n = F_{n-1} + F_{n-2}$ and $F_0 = 0$ and $F_1 = 1$. Show

$$F_0 - F_1 + F_2 - \cdots - F_{2n-1} + F_{2n} = F_{2n-1} - 1.$$

for each positive integer n.

Problem 3.5. Using induction and Theorem 2.19, prove the following.

> **Theorem 3.14** (generalized de Morgan's laws)**.** Let n be a positive integer, and suppose $A_1, \ldots, A_n$ are sets such that each A_i is contained in a set X. Then
>
> (a) $\left(\bigcap_{i=1}^n A_i\right)^c = \bigcup_{i=1}^n A_i^c$
>
> (b) $\left(\bigcup_{i=1}^n A_i\right)^c = \bigcap_{i=1}^n A_i^c$

If you would like, prove Theorem 3.14 again without induction.

Problem 3.6. Let $a_1, a_2, a_3, \ldots$ be a sequence of real numbers such that $a_{m+n} = a_m + a_n$ for any positive integers m and n. Show that $a_n = n \cdot a_1$ for each positive integer n.

3.2 Week 8: Strong Induction and Well Ordering

We've already seen how induction can serve as a powerful proof-writing tool under the right conditions. We will now look at how it can be generalized to attack a broader class of problem, allowing us to perform induction with fewer constraints and over a wider variety of sets.

3.2.1 Strong Induction

First, we will turn our attention to "strong induction," which looks very similar to the induction we've already been introduced to. In fact, strong induction is equivalent to "regular" induction, as we will see in Theorem 3.36.

Theorem 3.15 (Principle of strong induction)**.** Let P be a statement indexed by $\mathbb{N}$. To show $P(n)$ is true for all $n \in \mathbb{N}$, it suffices to show the following.

1. Base case: $P(0), P(1) \ldots, P(m)$ are true for some nonnegative integer m.
2. Inductive step: Let $k \in \mathbb{N}$. If $P(\ell)$ is true for every $\ell < k$, then $P(k)$ is true.

The important difference between induction and strong induction is that the former has a single base case and only advances one step at a time. On the other hand, strong induction allows you to assume all previous cases are true, which is often necessary to prove certain results, several of which we will explore subsequently. We defer a formal proof of Theorem 3.36, but we will provide an explanation similar to what we did for Theorem 3.2.

Explanation. By the case, we are given that $P(0), P(1), \ldots, P(m)$ are true for some nonnegative integer m. To show $P(m+1)$, we see that $P(\ell)$ is true for $\ell < m+1$ already, so the inductive step finishes. Similarly, to show $P(m+2)$, we now know that $P(\ell)$ is true for all $\ell < m+2$ (including $m+1$), so the inductive step finishes. This process is able to show $P(n)$ is true for all $n \in \mathbb{N}$ eventually. ■

Observe that, like "regular" induction, our base case need not start at 0. The explanation that we could start with any consecutive $m+1$ integers rather than $0, \ldots, m$ is identical to what we did above.

Example 3.16. Define the Fibonacci sequence $F_0, F_1, F_2, \ldots$ by $F_n = F_{n-1} + F_{n-2}$ and $F_0 = 0$ and $F_1 =$. For each $n \in \mathbb{N}$, we have

$$F_n = \frac{\varphi^n - (1-\varphi)^n}{\sqrt{5}},$$

where $\varphi = \frac{1+\sqrt{5}}{2}$.

Proof. We proceed by strong induction on n, using $n = 0, 1$ as our base cases.

1. Base case: note $F_0 = 0$ and $\varphi^0 = (1-\varphi)^n = 1$, so the claim holds for $n = 0$. For $n = 1$, we compute

$$\frac{\varphi - (1-\varphi)}{\sqrt{5}} = \frac{\frac{1+\sqrt{5}}{2} - \frac{1-\sqrt{5}}{2}}{\sqrt{5}} = 1,$$

   which is indeed F_1.

2. Inductive step: let $k \in \mathbb{N}$ be greater than 1, and suppose we have already shown that

$$F_\ell = \frac{\varphi^\ell - (1-\varphi)^\ell}{\sqrt{5}}$$

   for each $\ell < k$. (In practice, we will only need $\ell = k-1$ and $\ell = k-2$.) By definition, we have

$$F_k = F_{k-1} + F_{k-2},$$

   so from our inductive hypothesis we obtain

$$F_k = \frac{\varphi^{k-1} - (1-\varphi)^{k-1}}{\sqrt{5}} + \frac{\varphi^{k-2} - (1-\varphi)^{k-2}}{\sqrt{5}}. \tag{3.1}$$

   This doesn't look great, but we promise that all that remains is some algebraic manipulation.

   We could try to expand the binomial in the expression above and then try to cancel terms, but that would be messy and tedious. Instead, it helps to make the observation that $\varphi = \frac{1+\sqrt{5}}{2}$ and $1-\varphi = \frac{1-\sqrt{5}}{2}$

are the conjugate roots of the polynomial $x^2 - x - 1$. (This an be verified by direct computation or the quadratic formula.) In other words,

$$\varphi^2 = \varphi + 1,$$
$$(1-\varphi)^2 = (1-\varphi) + 1.$$

Multiplying both sides of the first equation by φ^{k-2}, we obtain

$$\varphi^k = \varphi^{k-1} + \varphi^{k-2}. \tag{3.2}$$

Similarly, we have

$$(1-\varphi)^k = (1-\varphi)^{k-1} + (1-\varphi)^{k-2}. \tag{3.3}$$

Substituting (3.2) and (3.3) into (3.1), we have

$$F_k = \frac{\varphi^k - (1-\varphi)^k}{\sqrt{5}},$$

which completes the proof of the inductive step. ■

Note that regular induction could not have proven the statement in the previous example without modifications. This is because in order to prove a claim about F_n, we actually need to look at the previous two iterations of the Fibonacci sequence, namely F_{n-1} and F_{n-2}. Regular induction does not give us the ability to do this, because we proved only a single base case: we would run into trouble as soon as we looked at $F_2 = F_1 + F_0$.

However, with that said, it is possible to make some modifications in order for regular induction to go through. The idea here is to add base cases by hand to the assertion we're trying to prove. Here's what that looks like concretely.

Exercise 3.17. For example, for $n \in \mathbb{N}$, let $P(n)$ denote the assertion

$$F_n = \frac{\varphi^n - (1-\varphi)^n}{\sqrt{5}} \qquad \text{and} \qquad F_{n+1} = \frac{\varphi^{n+1} - (1-\varphi)^{n+1}}{\sqrt{5}}.$$

(Namely, $P(n)$ has two equalities.) Show $P(n)$ is true for all $n \in \mathbb{N}$ using regular induction.

As another application of strong induction, we prove part of the Fundamental theorem of arithmetic. To do so, we define prime factorizations.

Definition 3.18 (prime)**.** A positive integer $n \in \mathbb{N}$ is *prime* if and only if $n > 1$ and n cannot be written as $n = ab$ for positive integers a and b greater than 1.

Remark 3.19. Some authors prefer the word "irreducible" to "prime" in the above definition.

Definition 3.20 (prime factorization)**.** If $n \in \mathbb{N}$ is a natural number, a *prime factorization* of n is a product of the form

$$n = p_1 \cdot p_2 \cdot \ldots \cdot p_m,$$

where each p_i is a prime number.

Theorem 3.21 (Fundamental theorem of arithmetic, existence)**.** Every natural number $n \geq 2$ has a prime factorization.

Proof. Because this is a statement indexed by natural numbers, we will proceed by strong induction. Our base case is $n = 2$, which is prime, so its prime factorization is just "$2 = 2$."

For the inductive step, we may suppose $n > 2$ and that all positive integers k between 2 and $n - 1$ have prime factorizations. We now argue by cases: either n is prime, or n is not prime.

- If n is prime, then like the $n = 2$ case above, we see that "$n = n$" is our prime factorization.
- Otherwise, n is not prime. However, $n > 2$, so there are positive integers a and b such that $a, b > 1$ while $n = ab$. Because $b > 1$, we see $a < n$, and similarly $b < n$. Thus, by the inductive hypothesis, we see a and b both have prime factorizations, which we write as

  $$a = p_1 \cdot p_2 \cdot \ldots \cdot p_m \qquad \text{and} \qquad b = q_1 \cdot q_2 \cdot \ldots \cdot q_n.$$

  But $n = ab$, so we may write

  $$n = ab = (p_1 \cdot p_2 \cdot \ldots \cdot p_m) \cdot (q_1 \cdot q_2 \cdot \ldots \cdot q_n)$$

  to provide a prime factorization of n.

The above cases finish the inductive step. ■

Remark 3.22. This proof of the fundamental theorem of arithmetic actually gives instructions for explicitly writing down the prime factorization: just keep factoring until you eventually get prime factors.

Remark 3.23. Another induction is able to show that prime factorizations of positive integers are unique, up to permutation of the factors. This proof is a little more involved (what does "up to permutation of the factors" even mean?), so we have not assigned it as an exercise, but the interested reader should attempt a proof.

Exercise 3.24. Explain why the proof of Theorem 3.21 requires strong induction and could not have simply used regular induction. In other words, where did we use the strong induction hypothesis?

Exercise 3.25. Despite Exercise 3.24, prove Theorem 3.21 by regular induction as follows. Imitating Exercise 3.17, let $P(n)$ denote the assertion "all positive integers $k \geq 2$ such that $k \leq n$ have a prime factorization." Prove $P(n)$ is true for all $n \geq 2$ by regular induction.

When using strong induction, be vigilant! It's easy to make silly mistakes if you don't complete the whole process of an inductive argument. Here's an example.

Bad Theorem 3.26. For all n, we have $\frac{d}{dx}(x^n) = 0$.

Bad proof. Our base case is $n = 0$, where $\frac{d}{dx}(1) = 0$. For our inductive step, suppose that $\frac{d}{dx}(x^k) = 0$ for $k < n$. Then, by the product rule, we compute

$$\frac{d}{dx}\left(x^{n+1}\right) = \frac{d}{dx}\left(x^n \cdot x^1\right) = x^n \cdot \frac{d}{dx}\left(x^1\right) + x^1 \cdot \frac{d}{dx}\left(x^n\right) = x^n \cdot 0 + x^1 \cdot 0 = 0,$$

which finishes. ■

So what went wrong? It turns out that while the above manipulation is valid for all $n \geq 1$, it isn't for $n = 0$: because the inductive step breaks for x^1, this incorrect step allowed the rest to follow. In other words, the issue is that we did not prove the base case $n = 1$, but we assumed it was true when performing the inductive step.

3.2.2 Well-Ordering

We'll now turn our attention to the notion of "well-ordering," which formalizes both forms of induction we've seen so far, and will allow us to easily deduce their equivalence. In addition, well-ordering will let us use induction on much more exotic sets than $\mathbb{N}$.

Intuitively, a "well-ordering" is a total order with minimums. Let's make this precise.

Definition 3.27. Let $\leq$ be a total order on a set X, and let $S \subseteq X$ be a subset. An element $x \in S$ is *minimal* (in S) if and only if $x \leq y$ for each $y \in S$. An element $x \in S$ is *maximal* (in S) if and only if $y \leq x$ for each $y \in S$.

Exercise 3.28. Let $\leq$ be a total order on a set X, and let $S \subseteq X$ be a subset. Further, let $x, y \in S$.

- If x and y are minimal in S, then $x = y$.
- If x and y are maximal in S, then $x = y$.

We are now ready to define well-orders.

Definition 3.29. A total order $\leq$ on a set X is a *well-ordering* if any nonempty subset $Y \subseteq X$ has a minimal element. In this case, we say that X is *well-ordered* under $\leq$.

Example 3.30. The set $\mathbb{N}$ is well-ordered under its usual ordering. Intuitively, we can see this as follows: for any nonempty set S, find some element $s \in S$. Then the set

$$S' \coloneqq S \cap \{0, 1, 2, \ldots, s\}$$

is finite (S' has at most $s+1$ elements) and nonempty (S' contains s), so S' has a minimal element. But the minimal element of $S \cap \{0, 1, 2, \ldots, s\}$ will also be minimal in S, so S has a minimal element.

Exercise 3.31. Is $\mathbb{Z}$ well-ordered under its usual ordering? If so, prove that it is. If not, can you come up with a different total order on $\mathbb{Z}$ under which it is well-ordered?

There turns out to be a connection between well-ordering and induction. This is best seen by example. Let's redo the proof of Example 3.5.

Example 3.32. For each $n \in \mathbb{N}$, we use the well-ordering of $\mathbb{N}$ in order to show that the number

$$S_n \coloneqq 1 + 3 + \cdots + (2n+1)$$

is always a perfect square.

Proof. In fact, we claim that $S_n = (n+1)^2$ for each n. To see this, we proceed by contradiction: suppose for the sake of contradiction that the set

$$S \coloneqq \left\{n \in \mathbb{N} : S_n \neq (n+1)^2\right\}$$

is nonempty. By the well-ordering principle, we may find a minimal element $n_0 \in S$. Then there are two cases, which correspond to the base case and the inductive step of our induction.

1. Base case: note that $n_0 > 0$ because $n_0 = 0$ has $S_n = 1 = (0+1)^2$, so $n_0 \notin S$.

2. Inductive step: because $n_0 > 0$, we see that $n_0 - 1 > 0$ and so $n_0 - 1 \in \mathbb{N}$. However, because n_0 is minimal, we see $n_0 - 1 \in S$, so

$$1 + 3 + \cdots + (2(n_0 - 1) + 1) = S_{n_0 - 1} = (n_0 - 1 + 1)^2 = n_0^2.$$

Adding $2n_0 + 1$ to both sides, we conclude $S_{n_0} = (n_0 + 1)^2$, so $n_0 \notin S$. This is our contradiction. ■

Let's explain this connection more abstractly. In the following theorem, we abbreviate the statement that $x \leq y$ and $x \neq y$ by $x < y$ (which should hopefully agree with your intuition for strict inequalities).

Theorem 3.33. Let X be a set, and let $\leq$ be a total order on X. The following are equivalent.

(a) The total order $\leq$ is actually a well-ordering on X.

(b) Let P be any statement indexed by X. Suppose that for every $x \in X$, if $P(y)$ is true for each $y \in X$ such that $y < x$, then $P(x)$ is true. Then $P(x)$ is true for all $x \in X$.

It is helpful to read the below proof imagining that we set $X = \mathbb{N}$ the entire time.

Proof. We have to show that (a) implies (b) and that (b) implies (a). Roughly speaking, the idea in this proof is to figure out how to translate between "properties" and "sets."

- We show (a) implies (b). The main idea is to construct a subset of X that we can use the well-ordering on. Imitating the construction of Example 3.32, we set

$$S \coloneqq \{x \in X : P(x) \text{ is false}\}.$$

If S is nonempty, then $P(x)$ is true for all $x \in X$, so we are done.

Thus, we suppose for the sake of contradiction that S is nonempty. But X is well-ordered! As such, we may let s_0 be the minimal element of S. However, if $y < y_0$, then y cannot be in S because y_0 is minimal in S, so $P(y)$ must be true. It follows by hypothesis on P that $P(y_0)$ is true, so $y_0 \notin S$. But $y_0 \in S$ by construction, so this is our contradiction.

- We show (b) implies (a). We would like to show that all nonempty sets have a minimal element. This proof will use contraposition a few times, so pay attention. Arguing by contraposition, we show that any set S without a minimal element must be empty.

Reversing the previous proof, the main idea is to construct a property P that we can use the hypothesis on. Thus, for $x \in X$, we let $P(x)$ be the property that "$x \notin S$." We would like to show that $P(x)$ holds for all $x \in X$.

Well, given any $x \in X$, we claim if $P(y)$ is true for each $y \in X$ with $y < x$, then $P(x)$ is true. This will finish by hypothesis. To see this, we again argue by contraposition: given that $P(x)$ is false, we need to show that there is some $y \in X$ with $y < x$ and $P(y)$ false.

Translating, we would like to show that, if $x \in S$, then there is some $y \in X$ with $y < x$ and $y \in S$. However, this is exactly the statement that $x \in S$ is not a minimal element, which is true because S has no minimal elements!

The above implications complete the proof. ■

3.2.3 Well-Ordering for $\mathbb{N}$

In this section, we explain the somewhat cryptic comments on "explanations" of regular induction and strong induction. The issue here is that, when one wants to define the natural numbers $\mathbb{N}$, one often just assumes that regular induction or strong induction or something similar holds. Thus, any "proof" of these will likely end up being rather circular.

With that said, one common way to define the natural numbers $\mathbb{N}$ is by assuming that they are well-ordered, as we remarked in Example 3.30. We will take this approach.

Axiom 3.34 (well-ordering principle)**.** The usual total ordering $\leq$ on $\mathbb{N}$ is a well-order.

We call the Well-ordering principle an "axiom" to remind ourselves that it is not a theorem: it's part of the definition of $\mathbb{N}$! It is not so different from the axioms "if p is a sentence which is not false, then p is true" or "if X is a nonempty set, then we can pick an arbitrary element of X" that we have taken for granted since day one.

Exercise 3.35. Let m and n be positive integers. Show, using Axiom 3.34, that there exist integers q and r with $0 \leq r < m$ such that

$$n = qm + r$$

Theorem 3.36. The following are equivalent.

(a) The well-ordering principle: any subset $S \subseteq \mathbb{N}$ has a minimal element.

(b) Regular induction: let P be any statement indexed by $\mathbb{N}$ such that the following hold.

1. Base case: the statement $P(0)$ is true.
2. Inductive step: for any $n \in \mathbb{N}$, if $P(n)$ is true, then $P(n+1)$ is true.

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

(c) Strong induction: let P be any statement indexed by $\mathbb{N}$ such that the following hold for some $m \in \mathbb{N}$.

1. Base case: the statements $P(0), P(1), \ldots, P(m)$ are all true.
2. Inductive step: for any $n > m$, if $P(k)$ is true for all $k < n$, then $P(n)$ is true.

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

Proof. By Theorem 3.33, statement (a') is equivalent to the following more inductive statement.

(a') Let P be any statement indexed by $\mathbb{N}$. Suppose that for every $x \in \mathbb{N}$, if $P(y)$ is true for each $y \in \mathbb{N}$ such that $y < x$, then $P(x)$ is true. Then $P(x)$ is true for all $x \in \mathbb{N}$.

Now, to prove a theorem claiming more than two statements are equivalent, the most efficient approach is to typically show that each statement implies the next. In other words, we show (a') implies (b), that (b) implies (c), and that (c) implies (a'). Intuitively, we can see somewhat visually that (a') looks a lot like (c), so the portions of this proof to really pay attention to are (a') implies (b) and (b) implies (c).

- We show (a') implies (b). We expect strong induction (which (a') is similar to) to be "stronger" than regular induction, so this proof will be direct. Let P be a statement satisfying the hypotheses of induction. We show that $P(n)$ is true for all $n \in \mathbb{N}$ by using (a').

  Indeed, select some $x \in \mathbb{N}$ such that $P(y)$ is true for each $y \in \mathbb{N}$ with $y < x$. We need to show $P(x)$ is true. There are two cases.

  - If $x = 0$, then $P(0)$ is true by the base case of the induction.
  - If $x > 0$, then note $x - 1 \in \mathbb{N}$ and $x - 1 < x$, so $P(x-1)$ is true by hypothesis on x. Thus, $P(x)$ is true by the inductive step.

  The above checks show that the hypotheses of (a') holds for P, so $P(n)$ is true for all $n \in \mathbb{N}$.

- We show (b) implies (c). This is the hardest part of the proof. Given our P and $m \in \mathbb{N}$ satisfying the hypotheses of (c), the main idea is to again to adjust P to some P' to apply regular induction. Imitating Exercise 3.25, we let $P'(n)$ denote the assertion "$P(k)$ holds for all $k \leq n + m$."

  We now show $P'(n)$ is true for all $n \in \mathbb{N}$ by regular induction.

1. Base case: note $P'(0)$ is just the base case of (c).
2. Inductive step: given $n \in \mathbb{N}$ such that $P'(0)$ is true, we note the statements $P(0), P(1), \ldots, P(n+m)$ are all true. It follows by the inductive step of (c) that $P(n+m+1)$ is also true. So we see all statements $P(0), P(1), \ldots, P(n+m+1)$ are true, makaing $P'(n+1)$ true.

It follows that $P'(n)$ is true for all $n \in \mathbb{N}$ by regular induction.

- We show (c) implies (a'). Suppose P satisfies the condition of (a'). Then we set $m \coloneqq 0$ and apply strong induction on P to show $P(n)$ for all $n \in \mathbb{N}$.

  1. Base case: we show $P(0)$ is true. Well, there are no $y \in \mathbb{N}$ such that $y < 0$, so $P(y)$ vacuously holds for all of them. We conclude that $P(0)$ is true.
  2. Inductive step: suppose $n > 0$ and that $P(k)$ is true for all $k < n$. Then $P(n)$ is true directly from (a').

  We conclude that $P(n)$ is true for all $n \in \mathbb{N}$ by strong induction.

The above implications complete the proof. ■

3.2.4 Problems

Problem 3.7. Prove the following.

(a) For any positive integer n, there exists an integer $k \geq 0$ such that $2^k \leq n < 2^{k+1}$.

(b) Every positive integer n can be written as the sum of distinct powers of 2.

Problem 3.8. Deduce the well-ordering principle, Axiom 3.34, from the more following plausible-sounding statement: for every $n \in \mathbb{N}$, there are only finitely many $m \in \mathbb{N}$ such that $m < n$.

Problem 3.9. Two positive integers a and b are said to be "coprime" if and only if they share no positive common divisors besides 1. In other words, if a and b are both divisible by a positive integer d, then $d > 0$. We will prove the following theorem.

> **Theorem 3.37** (Bezout)**.** Let x and y be coprime nonzero integers. Then there exist integers a and b such that
> $$ax + by = 1$$

Fill in the sketch below.

(a) Let S denote the set of integers of the form $ax + by$ where a and b are any integers. Show that S contains a positive integer and thus a least positive integer g.

(b) Use the division algorithm to show that g divides x. Similarly, show that g divides y.

(c) Show that there exist integers a and b such that $ax + by = 1$.

Problem 3.10. Use regular induction to show the following: for any natural number $n \in \mathbb{N}$, any nonempty subset $S \subseteq \mathbb{N}$ containing a natural number less than or equal to n has a minimal element.

Part II

Concepts

Chapter 4
Introduction to Topology

Rarely is a picture a proof, but I hope a good picture will cement your understanding of why something is true. Seeing is believing.

—Charles C. Pugh, [Pug15]

4.1 Week 9: Metric Spaces

In calculus, you learned that a function is continuous if "you can draw its graph without lifting your pencil from the paper." However, this seems hopelessly difficult to prove properties about in $\mathbb{R}$ using the machinery we've developed so far: we have learned how to write down symbols, not draw pictures![1] In order for us to develop a solid intuition of the concepts we were introduced to in calculus, we must start from the ground and work our way up.

Sequences and their characteristics are the foundation of limits, continuity, differentiation, and integration. We will first generalize these definitions and concepts by stepping away from the comfort of working in $\mathbb{R}$ and by doing so, we will begin to realize why these concepts need such rigorous definitions and explanations.

4.1.1 Metric Spaces

Metric spaces are an extension of the notion of a set, which allows us to talk about the distance between points in the set. You have had plenty of metric space experience because you have been working with them in most of your mathematics courses.

[1] However, pictures are a useful tool for developing proofs!

Definition 4.1 (metric space)**.** Let X be a set. A *metric* d on X is a function $d\colon X \times X \to \mathbb{R}$ that satisfies the following properties for any $x, y, z \in X$.

- Positive: $d(x, y) \geq 0$.
- Zero: $d(x, y) = 0$ if and only if $x = y$.
- Symmetric: $d(x, y) = d(y, x)$.
- Triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$.

If d is a metric on a set X, the ordered pair (X, d) is called a *metric space*.

Remark 4.2. When it is understood, we may say "X is a metric space with metric d," but we emphasize here that the metric d is essential data to the metric space. This is the same idea behind defining a partially ordered set as a set X equipped with a partial order $\preceq$; in fact, many texts define a partially ordered set as an ordered pair $(X, \preceq)$!

Definition 4.1 is pretty abstract. Let's see some examples.

Example 4.3. Set $X \coloneqq \mathbb{R}$, and define $d\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by $d(x, y) \coloneqq |x - y|$. We show d is a metric.

Proof. Fix any real numbers $x, y, z \in \mathbb{R}$.

- Positive: note $|x - y| \geq 0$.
- Zero: if $x = y$, then $|x - y| = |0| = 0$. Conversely, if $|x - y| = 0$, then $x - y = 0$, so $x = y$.
- Symmetric: note $x - y = -(y - x)$, so $|x - y| = |y - x|$. Thus, $d(x, y) = d(y, x)$.
- Triangle inequality: this is a bit tricky. The trick is to set $a \coloneqq x - y$ and $b \coloneqq y - z$ so that we want to show
$$|a + b| \leq |a| + |b|.$$
Now, $|a| \geq a$ and $|b| \geq b$, so $|a| + |b| \geq a + b$. Additionally, $|a| \geq -a$ and $|b| \geq -b$, so $|a| + |b| \geq -a - b$. However, $|a + b|$ equals either $a + b$ or $-a - b$, so the inequality follows. ∎

More generally, we have the following.

Proposition 4.4. Let n be a positive integer, and set $X \coloneqq \mathbb{R}^n$, which is the set of tuples $(x_1, \ldots, x_n)$ of real numbers. Define $d\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by
$$d((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}.$$
We show d is a metric.

We will want the following lemma for this proof.

Lemma 4.5 (Cauchy–Schwarz)**.** Let n be a positive integer and let $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$ be real numbers. Then
$$(a_1 b_1 + \cdots + a_n b_n)^2 \leq \left(a_1^2 + \cdots + a_n^2\right)\left(b_1^2 + \cdots + b_n^2\right).$$

Proof. This proof is technically by direct expansion, but it is somewhat annoying to prove directly. Summation notation helps with the book-keeping, but we will use induction on n. At $n = 0$, there are no real numbers anywhere, so the inequality is $0 \leq 0$, which is true.

We now proceed with the inductive step. Assume the inequality for n, and we show it for $n+1$. As such, let $a_1, \ldots, a_{n+1}$ and $b_1, \ldots, b_{n+1}$ be real numbers. On one hand, we can expand

$$\begin{aligned}(a_1b_1 + \cdots + a_nb_n + a_{n+1}b_{n+1})^2 &= ((a_1b_1 + \cdots + a_nb_n) + a_{n+1}b_{n+1})^2 \\ &= (a_1b_1 + \cdots + a_nb_n)^2 + a_{n+1}^2b_{n+1}^2 + 2(a_1b_1 + \cdots + a_nb_n)a_{n+1}b_{n+1}.\end{aligned}$$

On the other hand, we expand

$$\begin{aligned}\left(a_1^2 + \cdots + a_n^2 + a_{n+1}^2\right)\left(b_1^2 + \cdots + b_n^2 + b_{n+1}^2\right) &= \left(a_1^2 + \cdots + a_n^2\right)\left(b_1 + \cdots + b_n^2\right) \\ &\quad + a_{n+1}^2\left(b_1^2 + \cdots + b_n^2\right) + \left(a_1^2 + \cdots + a_n^2\right)b_{n+1}^2 \\ &\quad + a_{n+1}^2b_{n+1}^2.\end{aligned}$$

Now, to show

$$(a_1b_1 + \cdots + a_nb_n + a_{n+1}b_{n+1})^2 \stackrel{?}{\leq} \left(a_1^2 + \cdots + a_n^2 + a_{n+1}^2\right)\left(b_1^2 + \cdots + b_n^2 + b_{n+1}^2\right),$$

we note that we already have

$$(a_1b_1 + \cdots + a_nb_n)^2 \leq \left(a_1^2 + \cdots + a_n^2\right)\left(b_1^2 + \cdots + b_n^2\right)$$

by the inductive hypothesis, so combining our above inequalities means that we want to show

$$2(a_1b_1 + \cdots + a_nb_n)a_{n+1}b_{n+1} \stackrel{?}{\leq} a_{n+1}^2\left(b_1^2 + \cdots + b_n^2\right) + \left(a_1^2 + \cdots + a_n^2\right)b_{n+1}^2.$$

This rearranges into

$$0 \stackrel{?}{\leq} \left(a_{n+1}^2b_1^2 - 2a_{n+1}b_1a_1b_{n+1} + a_1^2b_{n+1}^2\right) + \cdots + \left(a_{n+1}^2b_n^2 - 2a_{n+1}b_na_nb_{n+1} + a_n^2b_{n+1}^2\right).$$

However, for each k, we see $\left(a_{n+1}^2b_1^2 - 2a_{n+1}b_1a_1b_{n+1} + a_n^2b_{n+1}^2\right) = (a_{n+1}b_1 - a_1b_{n+1})^2$, so each term is nonnegative, so the total sum is nonnegative. ■

Exercise 4.6. Rewrite the proof of Lemma 4.5 using summation notation but no induction.

Remark 4.7. Lemma 4.5 is usually stated as follows: let a and b be vectors in $\mathbb{R}^n$, and let $\langle \cdot, \cdot \rangle$ be the usual inner product. Then

$$\langle a, b\rangle^2 \leq \langle a, a\rangle \cdot \langle b, b\rangle.$$

We are now ready to prove Proposition 4.4.

Proof of Proposition 4.4. Find $x, y, z \in \mathbb{R}^n$, where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ and $z = (z_1, \ldots, z_n)$.

- Positive: note $d(x, y)$ is the square root of nonnegative real number, so $d(x, y) \geq 0$.
- Zero: if $x = y$, then $x_k = y_k$ for each k, so

  $$d(x, y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2} = \sqrt{0 + \cdots + 0} = 0.$$

  Conversely, if $d(x, y) = 0$, then for each index k, we see

  $$(x_k - y_k)^2 \leq (x_1 - y_1)^2 + \cdots + (x_n - y_n)^2 = d(x, y)^2 = 0.$$

  However, $(x_k - y_k)^2 \geq 0$ as well, so $(x_k - y_k)^2 = 0$, so $x_k = y_k$ for each k. Thus, $x = y$.
- Symmetric: for any real numbers $a, b \in \mathbb{R}$, we see $(a - b) = -(b - a)$, so $(a - b)^2 = (b - a)^2$. Applying this to each term,

  $$d(x, y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2} = \sqrt{(y_1 - x_1)^2 + \cdots + (y_n - x_n)^2} = d(y, x).$$

- Triangle inequality: this is again tricky. For each k, we set $a_k \coloneqq x_k - y_k$ and $b_k \coloneqq y_k - z_k$ so that $a_k + b_k = x_k - z_k$. Thus, the triangle inequality is equivalent to the inequality

$$\sqrt{(a_1+b_1)^2 + \cdots + (a_n+b_n)^2} \overset{?}{\leq} \sqrt{a_1^2 + \cdots + a_n^2} + \sqrt{b_1^2 + \cdots + b_n^2}.$$

  Squaring both sides, we want to show

$$(a_1+b_1)^2 + \cdots + (a_n+b_n)^2 \overset{?}{\leq} \left(a_1^2 + \cdots + a_n^2\right) + \left(b_1^2 + \cdots + b_n^2\right) + 2\sqrt{\left(a_1^2 + \cdots + a_n^2\right)\left(b_1^2 + \cdots + b_n^2\right)}.$$

  Now, for each k, we see $(a_k + b_k)^2 = a_k^2 + b_k^2 + 2a_kb_k$. Thus, after expanding these and cancelling all the a_k^2 and b_k^2 terms, we want to show

$$2(a_1b_1 + \cdots + a_nb_n) \overset{?}{\leq} 2\sqrt{\left(a_1^2 + \cdots + a_n^2\right)\left(b_1^2 + \cdots + b_n^2\right)}.$$

  Dividing by two and squaring, we would like to show

$$(a_1b_1 + \cdots + a_nb_n)^2 \overset{?}{\leq} \left(a_1^2 + \cdots + a_n^2\right)\left(b_1^2 + \cdots + b_n^2\right).$$

  This is exactly Lemma 4.5. ■

Here are few more metrics.

Exercise 4.8. Let $X \coloneqq (0, \infty)$ be the set of positive real numbers. Show that the function $d\colon X \times X \to \mathbb{R}$ defined by $d(x, y) = |\log(y/x)|$ is a metric space.

Example 4.9. Suppose that d is a metric on a set X. Show that $d_1((x, y), (x', y')) \coloneqq d(x, x') + d(y, y')$ is a metric on the set $X \times X$.

Proof. We check the conditions one at a time. Fix $(x, y), (x', y'), (x'', y'') \in X \times X$.

- Positive: note $d_1((x, y), (x', y')) = d(x, x') + d(y, y') \geq 0 + 0 = 0$.
- Zero: if $d_1((x, y), (x', y')) = 0$, then $d(x, x') + d(y, y') = 0$. However, $d(x, x') \geq 0$ and $d(y, y') \geq 0$, so the only way to get $d(x, x') + d(y, y') = 0$ is for both $d(x, x') = 0$ and $d(y, y') = 0$. Thus, $x = x'$ and $y = y'$ because d is a metric, so $(x, y) = (x', y')$ follows.
- Symmetric: note $d_1((x, y), (x', y')) = d(x, x') + d(y, y') = d(x', x) + d(y', y) = d_1((x', y'), (x, y))$. Note that we have used the fact that d is symmetric.
- Triangle inequality: using the triangle inequality for d, we see

$$\begin{aligned} d_1((x, y), (x', y')) + d_1((x', y'), (x'', y'')) &= \big(d(x, x') + d(y, y')\big) + \big(d(x', x'') + d(y', y'')\big) \\ &= \big(d(x, x') + d(x', x'')\big) + \big(d(y, y') + d(y', y'')\big) \\ &\geq d(x, x'') + d(y, y'') \\ &= d_1((x, y), (x'', y'')), \end{aligned}$$

  which is what we wanted. ■

Exercise 4.10. Suppose d_1 and d_2 are both metrics on the set X. Show that $d_1 + d_2$ is also a metric on X.

Exercise 4.11. Suppose d is a metric on the set X. If a is a positive real number, show that the function $ad\colon X \times X \to \mathbb{R}$ defined by $(ad)(x, x') \coloneqq a \cdot d(x, x')$ is also a metric on X.

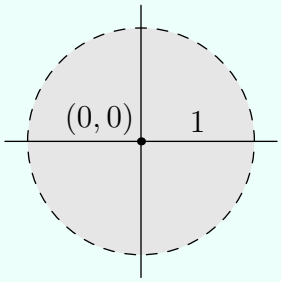
4.1.2 Open Sets

Roughly speaking, we will try to understand metric spaces by paying attention to certain special subsets. Here, we will define open subsets, which are built from open sets.

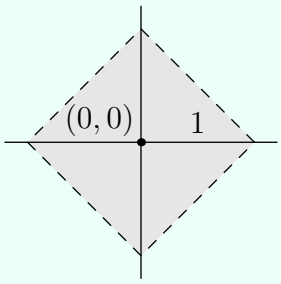
Definition 4.12 (open ball)**.** Let (X, d) be a metric space. Fix $x_0 \in X$ and $\varepsilon > 0$. The *open ball* centered at x_0 with radius ε is the set

$$B(x_0, \varepsilon) = \{x \in X : d(x, x_0) < \varepsilon\}.$$

Example 4.13. Give $\mathbb{R}^2$ the metric d by $d((x_1, y_1), (x_2, y_2)) \coloneqq \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. Then the open ball $B((0, 0), 1)$ is as follows.



Example 4.14. Give $\mathbb{R}^2$ the metric d by $d((x_1, y_1), (x_2, y_2)) \coloneqq |x_1 - x_2| + |y_1 - y_2|$. Then the open ball $B((0, 0), 1)$ is as follows.



Definition 4.15 (interior point)**.** Let (X, d) be a metric space, and fix a subset $A \subseteq X$. We say that $a \in A$ is an *interior point* of A if there is exists a $\varepsilon > 0$ such that $B(a, \varepsilon) \subseteq A$.

Definition 4.16 (interior)**.** Let (X, d) be a metric space, and fix a subset $A \subseteq X$. The *interior* of A, denoted A°, is the set of all interior points of A.

Example 4.17. Give $\mathbb{R}$ the metric $d(x, y) \coloneqq |x - y|$. Set $A \coloneqq [0, \infty)$, and we compute A°.

Proof. We fix some $a \in \mathbb{R}$ and test if $a \in A^\circ$ in various cases.

- If $a \leq 0$ no $\varepsilon > 0$ can possibly yield $B(a, \varepsilon) \subseteq A$ because $a - \varepsilon/2 \in B(a, \varepsilon)$ while $-\varepsilon/2 \notin A$ because $a - \varepsilon/2 < 0$. Thus, $a \notin A^\circ$.
- If $a > 0$, then we claim $a \in A^\circ$. Indeed, set $\varepsilon \coloneqq a$, and we see $\varepsilon > 0$ because $0 < a < 1$. We must show $B(a, \varepsilon) \subseteq [0, \infty)$. Well, if $x \in B(a, \varepsilon)$, then

$$|x - a| = d(x, a) < \varepsilon = a,$$

so $-a < x - a < a$, so $x > 0$, so $x \in [0, \infty)$.

Thus, the interior of $[0, \infty)$ is $(0, \infty)$. In fact, we note that the above argument also shows that the interior of $(0, \infty)$ is $(0, \infty)$. ■

Exercise 4.18. Give $\mathbb{R}$ the metric $d(x, y) \coloneqq |x - y|$. Compute the interior of the sets $(-\infty, 1]$ and $[0, 1]$.

The above examples have the feature $A^\circ \subseteq A$, and this is true in general: if $a \in A^\circ$, then a is an interior point of A, so there is $\varepsilon > 0$ such that $B(a, \varepsilon) \subseteq A$. But $a \in B(a, \varepsilon)$, so $a \in A$. Thus, $A^\circ \subseteq A$ is also true. The equality case will be of special interest to us.

Definition 4.19 (open set)**.** Let (X, d) be a metric space. A subset $A \subseteq X$ is *open* if and only if $A^\circ = A$. Equivalently, A is *open* if and only if each point $a \in A$ has some $\varepsilon > 0$ such that $B(a, \varepsilon) \subseteq A$.

Take a moment to convince yourself that the conditions stated in the above definition are in fact equivalent.

Remark 4.20. In some sense, the interior A° of a set A is the largest open subset A. We will make this precise later, but this explains how to think about the interior.

Example 4.21. Example 4.17 shows that $[0, \infty)$ is not an open subset of $\mathbb{R}$, but $(0, \infty)$ is.

Example 4.22. Let (X, d) is a metric space. Then $A \coloneqq \varnothing$ is an open set, though not for interesting reasons. Indeed, $A^\circ \subseteq A = \varnothing$, so $A^\circ = \varnothing = A$.

Example 4.23. Let (X, d) is a metric space. Then $A \coloneqq X$ is also open, though again not for interesting reasons. Indeed, for any $a \in X$, we see that

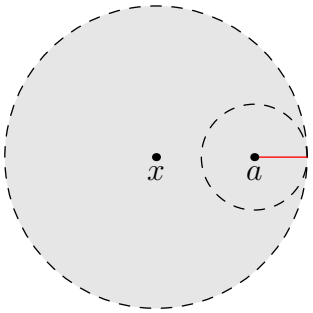
$$B(a, 1) = \{x \in X : d(a, x) < 1\} \subseteq X.$$

Thus, X is open.

It should be somewhat uncomfortable that we have defined "open balls" and "open sets," but it is not yet obvious that open balls are open sets. Let's see this.

Proposition 4.24. Let (X, d) be a metric space. Given any $x \in X$ and $r > 0$, the open ball $B(x, r)$ is an open set.

Proof. This is somewhat technical. Given $a \in B(x, r)$, we must produce $\varepsilon > 0$ such that $B(a, \varepsilon) \subseteq B(x, r)$. Our guide will be the following image, where we have shown such an open ball $B(a, \varepsilon)$ around $a \in B(x, r)$.



Solving for the length of the red segment, we see that it is $\varepsilon := r - d(x,a)$. Note $a \in B(x,r)$ implies $d(x,a) < r$, so $r - d(x,a) > 0$, so $\varepsilon > 0$.

It remains to show $B(a,\varepsilon) \subseteq B(x,r)$. Hopefully this is visually clear from the image, so we will try to prove it. Given $y \in B(a,\varepsilon)$, we want to show $y \in B(x,r)$. In other words, we are given $d(a,y) < \varepsilon$, and we want to show $d(x,y) < r$. The key is to use the triangle inequality, which implies

$$d(x,y) \le d(x,a) + d(a,y) < d(x,a) + \varepsilon = d(x,a) + r - d(x,a) = r$$

by plugging into the definition of ε. This completes the proof. ■

We now explain the sentence "open sets are built from open balls," as follows.

Proposition 4.25. Let (X,d) be a metric space. Given an open subset $U \subseteq X$, let $\mathcal{B}$ denote the collection of open balls contained in U. Then

$$U = \bigcup_{B \in \mathcal{B}} B.$$

Proof. We have two inclusions to show.

- We show $\bigcup_{B\in\mathcal{B}} B \subseteq U$. Well, we pick up any $x \in \bigcup_{B\in\mathcal{B}} B$. Then there exists some $B \in \mathcal{B}$ such that $x \in B$. However, $B \in \mathcal{B}$ forces $B \subseteq U$ by definition of $\mathcal{B}$, so $x \in U$ follows. This finishes.
- We show $U \subseteq \bigcup_{B\in\mathcal{B}} B$. Well, we pick up any $x \in U$. Because U is open, there exists some $\varepsilon > 0$ such that $B(x,\varepsilon) \subseteq U$. However, $B(x,\varepsilon)$ is an open ball contained in U, so it follows $B(x,\varepsilon) \in \mathcal{B}$. Thus,

  $$x \in B(x,\varepsilon) \subseteq \bigcup_{B\in\mathcal{B}} B,$$

  from which the desired inclusion follows. ■

4.1.3 Building Open Sets

Proposition 4.24 provides us a wealth of open sets, but we are going to want access to many more. We begin by explaining Remark 4.20, which gives an open subset from any subset.

Lemma 4.26. Let (X,d) be a metric space. Given any subset $A \subseteq X$, and an open subset $U \subseteq X$ such that $U \subseteq A$, we actually have $U \subseteq A^\circ$.

Proof. This is a matter of unwinding all the definitions. Given $x \in U$, we want to show $x \in A^\circ$. Well, U is open, so $x \in U$ promises some $\varepsilon > 0$ such that $B(x,\varepsilon) \subseteq U$. However, $U \subseteq A$, so we see $B(x,\varepsilon) \subseteq A$. It follows $a \in A^\circ$. ■

Proposition 4.27. Let (X,d) be a metric space. Given any subset $A \subseteq X$, the interior A° is open. In other words, $(A^\circ)^\circ = A^\circ$.

Proof. This follows by combining Proposition 4.24 and lemma 4.26, but it requires care, so pay attention. Given $a \in A^\circ$, we must find $\varepsilon > 0$ such that $B(a,\varepsilon) \subseteq A^\circ$. However, $a \in A^\circ$ promises some $\varepsilon > 0$ such that

$$B(a,\varepsilon) \subseteq A.$$

To finish the proof, we note $B(a,\varepsilon)$ is open by Proposition 4.24 and so $B(a,\varepsilon) \subseteq A^\circ$ by Lemma 4.26. ■

Exercise 4.28. Use Lemma 4.26 and proposition 4.24 to show the following: for a subset A of a metric space (X, d), we have

$$A^\circ = \bigcup_{U \subseteq A} U,$$

where the union is over all open sets U contained in A.

Another way to build sets is with set operations, so we now investigate how open sets behave with unions and intersections. Let's first discuss intersections.

Proposition 4.29. Let (X, d) be a metric space. Given open subsets $A, B \subseteq X$, the set $A \cap B$ is also open.

Proof. Given $x \in A \cap B$, we must find ε such that $B(x, \varepsilon) \subseteq A \cap B$. Well, $x \in A$, and A is open, so there is some $\varepsilon_A > 0$ such that $B(x, \varepsilon_A) \subseteq A$. Similarly, $x \in B$, and B is open, so there is some $\varepsilon_B > 0$ such that $B(x, \varepsilon_B) \subseteq B$. To finish the proof, we set

$$\varepsilon \coloneqq \min\{\varepsilon_A, \varepsilon_B\}.$$

Note $\varepsilon > 0$, so we want to show $B(x, \varepsilon) \subseteq A \cap B$. Well, for $y \in B(x, \varepsilon)$, we see $d(x, y) < \varepsilon \leq \varepsilon_A$, so $y \in B(x, \varepsilon_A)$, so $y \in A$ because $B(x, \varepsilon_A) \subseteq A$. A similar argument shows $y \in B$, so we conclude $y \in A \cap B$. ■

Exercise 4.30. Show the following: given finitely many open subsets $A_1, A_2, \ldots, A_n$ of a metric space (X, d), the intersection

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n$$

is a closed set. One can show this directly, arguing similarly as in Proposition 4.29; alternatively, one can use induction with Proposition 4.29.

Non-Example 4.31. In spite of Exercise 4.30, a countable intersection of open sets need not be open. A similar argument to Example 4.17 shows that the sets $A_n \coloneqq (-1/n, \infty)$ are all open for any positive integer n, but their intersection is

$$\bigcap_{n=1}^{\infty} A_n = [0, \infty),$$

and $[0, \infty)$ is not open.

We quickly note that we can translate Proposition 4.29 into a statement about the interior.

Corollary 4.32. Let (X, d) be a metric space. Given subsets $A, B \subseteq X$, we have $(A \cap B)^\circ = A^\circ \cap B^\circ$.

Proof. We show this in two inclusions. As usual, this argument is somewhat confusing.

- We show $(A \cap B)^\circ \subseteq A^\circ \cap B^\circ$. Note $(A \cap B)^\circ$ is open by Proposition 4.27, and $(A \cap B)^\circ \subseteq A \cap B \subseteq A$. Thus, $(A \cap B)^\circ \subseteq A^\circ$ follows from Lemma 4.26. A similar argument shows $(A \cap B)^\circ \subseteq B^\circ$, so we are done.
- We show $A^\circ \cap B^\circ \subseteq (A \cap B)^\circ$. Note $A^\circ \subseteq A$ and $B^\circ \subseteq B$, so $A^\circ \cap B^\circ \subseteq A \cap B$. However, A° and B° are both open by Proposition 4.27, so $A^\circ \cap B^\circ$ is open by Proposition 4.29, so $A^\circ \cap B^\circ \subseteq A \cap B$ follows from Lemma 4.26. ■

And now we discuss unions.

Exercise 4.33. Let (X, d) be a metric space. Given open sets $A, B \subseteq X$, show directly that $A \cup B$ is open.

In fact, we don't have to settle for merely taking the union of two open sets.

Proposition 4.34. Let (X, d) be a metric space. Let $\mathcal{U}$ be a collection of open subsets of X. Then the union

$$A := \bigcup_{U \in \mathcal{U}} U$$

is also open.

Proof. Given $a \in A$, we must find $\varepsilon > 0$ such that $B(a, \varepsilon) \subseteq A$. Well, A is the union of the open sets in $\mathcal{U}$, so there is some $U \in \mathcal{U}$ such that $x \in U$. However, U is open, so there is some $\varepsilon > 0$ such that $B(x, \varepsilon) \subseteq U$. Because $U \subseteq A$, it follows $B(x, \varepsilon) \subseteq A$ as well, which completes the proof. ■

Exercise 4.35. It is not true that the interior of the union is the union of the interiors. For example, given $\mathbb{R}$ the usual metric $d(x, y) := |x - y|$. Then we set $A = [0, \infty)$ and $B = (-\infty, 0]$. Arguing similarly as in Example 4.17, we see

$$A^\circ \cup B^\circ = (0, \infty) \cup (-\infty, 0) = \mathbb{R} \setminus \{0\},$$

but

$$(A \cup B)^\circ = \mathbb{R}^\circ = \mathbb{R}.$$

Lastly, we might be interested in taking the complement of open sets, but this operation does not produce open sets. Instead, the complement of an open set is a "closed set," which will be discussed in the next section.

4.1.4 Problems

Problem 4.1. The *Chebyshev metric* on $\mathbb{R}^n$ is the function $d \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by

$$d((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) := \max\{|x_k - y_k| : k \in \{1, 2, \ldots, n\}\}.$$

Verify that d is a metric on $\mathbb{R}^n$.

Problem 4.2. The *discrete metric* on $\mathbb{R}$ is the function $d \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined for all $x, y \in \mathbb{R}^n$ by

$$d(x, y) := \begin{cases} 0 & \text{if } x = y, \\ 1 & \text{if } x \neq y. \end{cases}$$

Verify that d is a metric.

Problem 4.3. Define the function $d \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by $d(x, y) = \frac{|x-y|}{1+|x-y|}$. Determine if d is a metric on $\mathbb{R}$.

Problem 4.4. Let $d \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be the metric on $\mathbb{R}$ defined by $d(x, y) := |x - y|$. The *standard bounded metric* on $\mathbb{R}$ is the function $\overline{d} \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined by

$$\overline{d}(x, y) := \min\{1, d(x, y)\}.$$

Verify that $\overline{d}$ is a metric on $\mathbb{R}$.

Problem 4.5. Let X be a set. Define the function $d\colon X \times X \to \mathbb{R}$ by

$$d(x,y) \coloneqq \begin{cases} 0 & x = y, \\ 1 & x \neq y. \end{cases}$$

Verify that d is a metric on X.

Problem 4.6. Consider the metric space $(\mathbb{R}, d)$, where $d(x,y) \coloneqq |x-y|$. Find the the interior (without proof) of the following subsets of $\mathbb{R}$.

(a) $A = (0,1)$

(b) $A = [0,1]$

(c) $A = \mathbb{Z}$

(d) $A = \mathbb{Q}$

(e) $A = \mathbb{R}$

Problem 4.7. Consider the metric space $(\mathbb{R}, d)$ where $d(x,y) \coloneqq |x-y|$. Given real numbers $a, b \in \mathbb{R}$ such that $a < b$, show that $[a,b]^\circ = (a,b)$.

Problem 4.8. Let (X,d) be a metric space. Given a subset $A \subseteq X$, let $\mathcal{B}$ denote the collection of open balls contained in A. Show that

$$A^\circ = \bigcup_{B \in \mathcal{B}} B.$$

4.2 Week 10: Closed Sets

In this lecture, we discuss everything you ever wanted to know about closed sets. We will give two definitions of a closed subset of a metric space (X,d) and explain why they are equivalent. Along the way, we will also give some properties of closed sets.

4.2.1 Convergence

Intuitively, a closed set is one which is closed under convergence of sequences. We will make this precise, but only very slowly. We must first define what it means for a sequence to convergence.

Definition 4.36 (sequence)**.** Fix a set X. Then an infinite *sequence* of elements in X is a function $a\colon \mathbb{N} \to X$. We will often write this as $\{a_n\}_{n\in\mathbb{N}}$, where a_n refers to $a(n)$.

Definition 4.37 (limit)**.** Fix a metric space (X,d) and a sequence $\{a_n\}_{n\in\mathbb{N}}$ of elements in X. The sequence $\{a_n\}_{n\in\mathbb{N}}$ is *converges* to the *limit* $a \in X$ if and only if, for each $\varepsilon > 0$, there exists some N such that

$$n > N \implies d(a_n, a) < \varepsilon.$$

We might write this situation as $\lim a_n = a$.

Intuitively, we are saying that the elements $\{a_n\}_{n\in\mathbb{N}}$ get closer and closer to a as n gets larger. Let's see some examples.

Example 4.38. Let (X, d) be a metric space and $a \in X$ an element, and define the sequence $\{a_n\}_{n\in\mathbb{N}}$ by $a_n := a$ for each $n \in \mathbb{N}$. Then we claim $\lim a_n = a$. Indeed, for any $\varepsilon > 0$, we set $N := 0$ so that $n > N$ implies

$$d(a_n, a) = d(a, a) = 0 < \varepsilon.$$

Here is a harder example.

Example 4.39. Give $\mathbb{R}$ the metric d by $d(x, y) := |x - y|$. Define $a_n := 1/(n + 1)$ for each $n \in \mathbb{N}$. Then $\lim a_n = 0$.

Proof. We plug directly into the definition. Fix some $\varepsilon > 0$, and we want to find some N such that $n > N$ implies

$$\frac{1}{n+1} = d(a_n, a) \overset{?}{<} \varepsilon.$$

Rearranging this, we are asking for $n + 1 > 1/\varepsilon$, or $n > \varepsilon - 1$. As such, we just set $N := -1 + 1/\varepsilon$. Then $n > N$ implies

$$\frac{1}{n+1} < \frac{1}{N+1} = \varepsilon,$$

as desired. ∎

Importantly, in the above proof, we had N be a function of ε. This is a common feature of such proofs of convergence. Here is are a few examples for you to try.

Exercise 4.40. Give $\mathbb{R}$ the metric d by $d(x, y) := |x - y|$. Define $a_n := 2/(n + 1)$ for each $n \in \mathbb{N}$. Then show $\lim a_n = 0$.

Exercise 4.41. Give $\mathbb{R}$ the metric d by $d(x, y) := |x - y|$. Define $a_n := \frac{2n+\sin(n)}{3n+\cos(n)}$ for each $n \in \mathbb{N}$. Then show $\lim a_n = \frac{2}{3}$.

Let's give an example where convergence fails.

Example 4.42. Give $\mathbb{R}$ the metric d by $d(x, y) := |x - y|$. Define $a_n := (-1)^n$ for each $n \in \mathbb{N}$. Then there does not exist $a \in \mathbb{R}$ such that $\lim a_n = a$.

Proof. We proceed by contradiction. Suppose for the sake of contradiction that we have found $a \in \mathbb{R}$ such that $\lim a_n = a$. Thus, for all $\varepsilon > 0$, we are promised some N such that $n > N$ implies

$$d(a_n, a) < \varepsilon.$$

Intuitively, we expect to arrive at a contradiction because the sequence $\{a_n\}_{n\in\mathbb{N}}$ does not get infinitely close to any particularly real number—it oscillates between $+1$ and -1!

Let's make this intuition rigorous: actually, we can see that the terms $\{a_n\}_{n\in\mathbb{N}}$ never stay within $\varepsilon := 0.1$ of any given real number due to this oscillation. Thus, we will use $\varepsilon := 0.1$ for our contradiction: we are given some N such that

$$n > N \implies d(a_n, a) < \varepsilon = 0.1.$$

However, choosing some $n > N$ which is even, we see $d(1, a) < 0.1$, and choosing some $n > N$ which is odd, we see $d(-1, a) < 0.1$. To get our contradiction, we cleverly apply the triangle inequality, which asserts

$$2 = d(-1, 1) \leq d(-1, a) + d(a, 1) = d(-1, a) + d(1, a) = 0.1 + 0.1 = 0.2.$$

The above inequality is obviously false, so we have arrived at our contradiction. ∎

Exercise 4.43. Give $\mathbb{R}$ the metric d by $d(x,y) \coloneqq |x-y|$. Define $a_n \coloneqq n^2$ for each $n \in \mathbb{N}$. Then show that there does not exist $a \in \mathbb{R}$ such that $\lim a_n = a$.

4.2.2 Closures

We are now almost ready to give our first definition of a closed set. Similar to how we defined open sets by defining interiors, we will define closed sets by defining closures.

Definition 4.44 (closure)**.** Let (X,d) be a metric space and $A \subseteq X$ be a subset. Then the *closure* is the set $\overline{A} \subseteq X$ such that each $a \in \overline{A}$ satisfies the following property: for each $\varepsilon > 0$, the set $A \cap B(a,\varepsilon)$ is nonempty.

Roughly speaking, the closure $\overline{A}$ consists of all the points which are "arbitrarily close" to points in A.

Example 4.45. Give $\mathbb{R}$ the metric d by $d(x,y) \coloneqq |x-y|$. Let $A \coloneqq (0,\infty)$. We show $\overline{A} = [0,\infty)$.

Proof. For each $a \in \mathbb{R}$, we must decide if $a \in \overline{A}$. We have the following cases.

- Suppose $a > 0$. We show $a \in \overline{A}$. Well, for each $\varepsilon > 0$, we note that $a \in A$ and $a \in B(a,\varepsilon)$, so $A \cap B(a,\varepsilon)$ and is thus nonempty.
- Suppose $a = 0$. We show $a \in \overline{A}$. Well, for each $\varepsilon > 0$, we must show that $A \cap B(0,\varepsilon)$ is nonempty. For this, we note $\varepsilon/2 > 0$ has $\varepsilon/2 \in A$ and $\varepsilon/2 \in B(0,\varepsilon)$, so $A \cap B(0,\varepsilon)$ is nonempty. This is what we wanted.
- Suppose $a < 0$. We show $a \notin \overline{A}$. Intuitively, we expect to have a contradiction because a has some positive distance from A, so elements of A cannot get arbitrarily close to a.

  To make this rigorous, we expect that elements of A are a distance of at least $|a|$ away from a. So we set $\varepsilon \coloneqq |a|$ and claim that

  $$A \cap B(a,\varepsilon) \stackrel{?}{=} \varnothing,$$

  from which $a \notin \overline{A}$ will follow. Indeed, if $x \in A$, then $x > 0$. But then we note $x > 0 > a$, so

  $$d(x,a) = |x-a| = |x| + |a| > |a| = \varepsilon,$$

  so it follows $x \notin B(a,\varepsilon)$. Thus, $A \cap B(a,\varepsilon) = \varnothing$.

Synthesizing the above cases, we see $\overline{A} = [0,\infty)$. In fact, we note that the above argument also shows $\overline{[0,\infty)} = [0,\infty)$. ∎

Exercise 4.46. Give $\mathbb{R}$ the metric d by $d(x,y) \coloneqq |x-y|$. Show that $\overline{(0,1)} = [0,1]$.

Before continuing, we explain why we introduced convergence at the beginning of this section.

Proposition 4.47. Let (X,d) be a metric space. Given a subset $A \subseteq X$, we have $a \in \overline{A}$ if and only if there exists a sequence $\{a_n\}_{n\in\mathbb{N}}$ of elements in A such that $\lim a_n = a$.

Proof. We have two directions to show.

- Suppose $a \in \overline{A}$. Then we must construct a sequence of elements of A which gets arbitrarily close to A. For this, we use the definition of the closure: for any $n \in \mathbb{N}$, we note that $A \cap B(a, 1/(n+1))$ is nonempty, so we can find a point
$$a_n \in A \cap B(a, 1/(n+1)).$$
The idea here is that the term a_n is a distance of at most $1/(n+1)$ away from 0, but this distance goes to 0 as n gets large, so the sequence $\{a_n\}_{n\in\mathbb{N}}$ should converge to a.

  As such, we claim that $\{a_n\}_{n\in\mathbb{N}}$ is the desired sequence. Note $a_n \in A$ for each $n \in \mathbb{N}$ by construction, so it remains to show $\lim a_n = a$. For this, we fix any $\varepsilon > 0$ and see that we want N such that
$$n > N \implies d(a, a_n) < \varepsilon.$$
However, by construction, we see $d(a, a_n) < 1/(n+1)$, so it will be enough for $1/(n+1) < \varepsilon$. Rearranging, we see that we may set $N := 1/\varepsilon$: indeed, for any $n > N$, we see
$$d(a, a_n) < \frac{1}{n+1} < \frac{1}{n} < \frac{1}{N} = \varepsilon.$$

- Suppose there is a sequence $\{a_n\}_{n\in\mathbb{N}}$ of elements in A such that $\lim a_n = a$. Then we claim $a \in \overline{A}$. Indeed, for any $\varepsilon > 0$, we must show $A \cap B(a, \varepsilon)$ is nonempty. Intuitively, we are asking for an element of A which is very close to a, so we will use the given sequence $\{a_n\}_{n\in\mathbb{N}}$.

  We are given some N such that $n > N$ implies $d(a, a_n) < \varepsilon$. Thus, choosing any $n > N$ provides some $a_n \in A$ such that $d(a, a_n) < \varepsilon$ and so $a_n \in A \cap B(a, \varepsilon)$. It follows $A \cap B(a, \varepsilon)$ is nonempty. ■

Exercise 4.48. Give $\mathbb{R}$ the metric d by $d(x, y) := |x - y|$. Find a sequence $\{a_n\}_{n\in\mathbb{N}}$ of elements in $(0, \infty)$ such that $\lim a_n = 0$.

4.2.3 Closed Sets

At this point, we should remark that the argument in Example 4.45 generalizes to show $A \subseteq \overline{A}$ in general. Indeed, for any $a \in A$, we claim $a \in \overline{A}$: for any $\varepsilon > 0$, we see $a \in B(a, \varepsilon)$ and $a \in A$, so $A \cap B(a, \varepsilon)$ is nonempty.

As with interiors and open sets, the equality case is what we are interested in.

Definition 4.49 (closed sets)**.** Let (X, d) be a metric space and $A \subseteq X$ be a subset. Then A is *closed* if and only if A satisfies one of the following equivalent conditions.

- $\overline{A} = A$.
- $\overline{A} \subseteq A$. In other words, suppose $a \in X$ makes the set $A \cap B(a, \varepsilon)$ nonempty for all $\varepsilon > 0$. Then $a \in A$.
- If a sequence $\{a_n\}_{n\in\mathbb{N}}$ of elements in A which converges to an element $a \in X$, then actually $a \in A$.

Again, take a moment to convince yourself that the two conditions we gave above are equivalent. The equivalence of the last two uses Proposition 4.47.

Remark 4.50. Similar to before, the closure $\overline{A}$ will turn out to be the smallest closed set containing A. We will make this precise later.

Here are some examples.

Example 4.51. Give $\mathbb{R}$ the metric d by $d(x, y) := |x - y|$. The argument of Example 4.45 shows that $(0, \infty)$ is not closed, but $[0, \infty)$ is.

Example 4.52. Let (X, d) be a metric space. Then we claim the subset X of X is closed. Indeed, suppose $a \in X$ makes the set $X \cap B(a, \varepsilon)$ nonempty for all $\varepsilon > 0$. Then of course $a \in X$.

Example 4.53. Let (X, d) be a metric space. Then we claim $\varnothing \subseteq X$ is closed. Indeed, there is no $a \in X$ such that the sets $\varnothing \cap B(a, \varepsilon)$ are nonempty for all $\varepsilon > 0$. Thus, $\overline{\varnothing} = \varnothing$.

Warning 4.54. Sets are not doors! They can be both open and closed! (They can also be neither open nor closed!) Most notably, given a metric space (X, d), the sets X and $\varnothing$ are both open and closed.

Example 4.55. Let (X, d) be a metric space. Given an element $a \in X$, the set $\{a\}$ is closed. Indeed, we use the last equivalent condition in Definition 4.49: for any sequence $\{a_n\}_{n \in \mathbb{N}}$ of elements in $\{a\}$, we see we must have $a_n = a$ for each $n \in \mathbb{N}$. Thus, $\lim a_n = a$ follows, so the limit lives in $\{a\}$, which is what we wanted.

Here is a last, more interesting example which provides us with a wealth of closed sets.

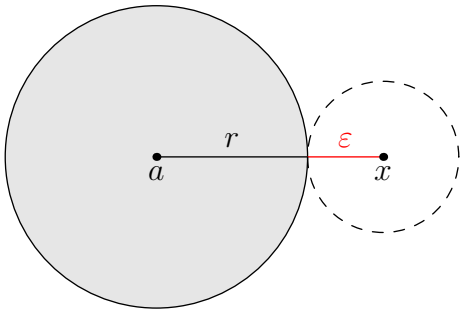
Proposition 4.56. Let (X, d) be a metric space. Given $a \in X$ and $r > 0$, the "closed ball"

$$\overline{B}(a, r) \coloneqq \{x \in X : d(a, x) \leq r\}$$

is closed.

Proof. This is somewhat technical. We would like to show the following: given $x \in X$, if the sets $\overline{B}(a, r) \cap B(x, \varepsilon)$ are nonempty for all $\varepsilon > 0$, then $x \in \overline{B}(a, r)$. Instead, we will show the contrapositive, which is the following: given $x \in X$, if $x \notin \overline{B}(a, r)$, then there is some $\varepsilon > 0$ such that $\overline{B}(a, r) \cap B(x, \varepsilon)$ is empty.

Unpacking our definitions, we are given some $x \in X$ such that $d(a, x) > r$, and we want $\varepsilon > 0$ such that $\overline{B}(a, r) \cap B(x, \varepsilon)$ is empty. For this, we draw the following picture.



Solving for ε, we hope that $\varepsilon \coloneqq d(a, x) - r$ will work. Note $\varepsilon > 0$ by construction.

It remains to show that $\overline{B}(a, r) \cap B(x, \varepsilon)$ is empty. Indeed, suppose $x' \in \overline{B}(a, r)$, and we will show $x' \notin B(x, \varepsilon)$, which will finish the proof because it implies that no element lives in both $\overline{B}(a, r)$ and $B(x, \varepsilon)$ at once. Well, we see $d(a, x') \leq r$. To compute $d(x, x')$, we use the triangle inequality, writing

$$d(x, x') + r \geq d(x, x') + d(x', a) \geq d(x, a) = \varepsilon + r.$$

Rearranging, we see $d(x, x') \geq \varepsilon$, so $x' \notin B(x, \varepsilon)$ follows. This completes the proof. ∎

4.2.4 Building Closed Sets

We now mirror the discussion of section 4.1.3 to build lots of closed sets. To begin, we really should check that the closure of a set is closed, which produces a closed set from any subset.

Proposition 4.57. Let (X, d) be a metric space. Given a subset $A \subseteq X$, the closure $\overline{A}$ is closed. In other words, $\overline{\overline{A}} = \overline{A}$.

Proof. Suppose $a \in X$ makes the sets $\overline{A} \cap B(a, \varepsilon)$ nonempty for all $\varepsilon > 0$. Then we must show $a \in \overline{A}$. However, to show $a \in \overline{A}$, we must show that the sets $A \cap B(a, \varepsilon)$ are all nonempty.

For this proof, we employ a "$\varepsilon/2$" trick. Let's first attempt the proof without this trick. Fix any $\varepsilon > 0$, and we want to show $A \cap B(a, \varepsilon)$ is nonempty. We construct two elements.

- By definition of a, we know $\overline{A} \cap B(a, \varepsilon)$ is nonempty, so we find $a' \in \overline{A} \cap B(a, \varepsilon)$.
- By definition of $\overline{A}$, we know $A \cap B(a', \varepsilon)$ is nonempty, so we find $a'' \in A \cap B(a', \varepsilon)$.

We might hope that a'' is the desired element in $A \cap B(a, \varepsilon)$ because it is in A after all, but we cannot quite show this: doing our best, we use the triangle inequality to bound

$$d(a, a'') \leq d(a, a') + d(a', a'') < \varepsilon + \varepsilon = 2\varepsilon,$$

so actually $a'' \in A \cap B(a, 2\varepsilon)$. But we wanted to show $A \cap B(a, \varepsilon)$ is nonempty!

To fix this problem, we divide all the relevant ε terms in the above paragraph by 2, which is the aforementioned "$\varepsilon/2$" trick. We rewrite the above paragraph for clarity. Note $\varepsilon/2 > 0$.

- By definition of a, we know $\overline{A} \cap B(a, \varepsilon/2)$ is nonempty, so we find $a' \in \overline{A} \cap B(a, \varepsilon/2)$.
- By definition of $\overline{A}$, we know $A \cap B(a', \varepsilon/2)$ is nonempty, so we find $a'' \in A \cap B(a', \varepsilon/2)$.

Now, we see $a'' \in A$, and by the triangle inequality, we have

$$d(a, a'') \leq d(a, a') + d(a', a'') < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

so $a'' \in B(a, \varepsilon)$ as well. Thus, $A \cap B(a, \varepsilon)$ is indeed nonempty. ∎

We can now explain Remark 4.50: we see $\overline{A}$ is a closed set, and it is the smallest closed set which contains A in the following sense.

Proposition 4.58. Let (X, d) be a metric space. Given a set A and a closed set C such that $A \subseteq C$, then actually $\overline{A} \subseteq C$.

Proof. Suppose $a \in \overline{A}$, and we must show $a \in C$. Then for any $\varepsilon > 0$, the set $A \cap B(a, \varepsilon)$ is nonempty. However, $A \subseteq C$, so the set $C \cap B(a, \varepsilon)$ is also nonempty for all $\varepsilon > 0$. Thus, because C is closed, it follows that $a \in C$. ∎

Exercise 4.59. Use Proposition 4.58 to prove the following alternate characterization of the closure: given a subset A of a metric space (X, d), we have

$$\overline{A} = \bigcap_{A \subseteq C} C,$$

where the intersection is over all closed subsets $C \subseteq X$ which contain A. This statement is in some sense the closed-set analogue to Proposition 4.25.

We now discuss some set operations. We begin with intersections. The following is analogous to Proposition 4.34.

Proposition 4.60. Let (X, d) be a metric space. Given a collection $\mathcal{C}$ of closed subsets of X, the intersection

$$A \coloneqq \bigcap_{C \in \mathcal{C}} C$$

is closed.

Proof. Suppose $x \in X$ makes the sets $A \cap B(x, \varepsilon)$ nonempty for all $\varepsilon > 0$. We show $x \in A$. By definition of A, it's enough to show $x \in C$ for any given $C \in \mathcal{C}$. However, we note $A \subseteq C$, so for any $\varepsilon > 0$, the set $C \cup B(x, \varepsilon)$ is nonempty because $A \cap B(x, \varepsilon)$ is nonempty. Because C is closed, it follows that $x \in C$, which is what we wanted. ∎

Example 4.61. Similar to Exercise 4.35, it is not true that the closure of an intersection is the intersection of the closures. For example, give $\mathbb{R}$ the usual metric $d(x, y) \coloneqq |x - y|$. Then set $A \coloneqq (-1, 0)$ and $B \coloneqq (0, 1)$, and we see

$$\overline{A \cap B} = \overline{\varnothing} = \varnothing,$$

but

$$\overline{A} \cap \overline{B} = [-1, 0] \cap [0, 1] = \{0\}.$$

The computation of $\overline{A}$ and $\overline{B}$ follow from computations similar to Exercise 4.46.

And now we discuss unions. The following result is analogous to Proposition 4.29.

Proposition 4.62. Let (X, d) be a metric space. Given closed sets $A, B \subseteq X$, the union $A \cup B$ is also closed.

Proof. Suppose $x \in X$ has a sequence $\{x_n\}_{n \in \mathbb{N}}$ of elements in $A \cup B$ which converges to x. We show that $x \in A \cup B$.

Note that there are infinitely many elements in the sequence $\{x_n\}_{n \in \mathbb{N}}$, so infinitely many elements must live in A, or infinitely many elements must live in B. Indeed, if only finitely many elements of the sequence live in either A or B, then only finitely many elements of the sequence live in $A \cup B$, which is does not make any sense.

Without loss of generality, we will say that infinitely many elements of our sequence $\{x_n\}_{n \in \mathbb{N}}$ live in A. Then, for each $k \in \mathbb{N}$, we may let x_{n_k} denote the kth element of the sequence in A. We claim that $\lim x_{n_k} = x$, which will imply $x \in A$ because A is closed, so this will complete the proof.

Well, we know $\lim x_n = x$, and $\{x_{n_k}\}_{k \in \mathbb{N}}$ is just a subsequence, so it had better have the same limit. Indeed, for any $\varepsilon > 0$, we are promised some N such that $n > N$ implies $d(x, x_n) < \varepsilon$. Now, for each k, we must have $n_k \geq k$, so

$$k > N \implies n_k > N \implies d(x, x_{n_k}) < \varepsilon,$$

which is what we wanted. ∎

Exercise 4.63. Let (X, d) be a metric space. Given finitely many closed sets $A_1, A_2, \ldots, A_n$, show that the union

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n$$

is also closed.

Non-Example 4.64. The finite union of closed sets is closed by Exercise 4.63, but the countable union of closed sets need not be closed. For example, give $\mathbb{R}$ the metric d by $d(x,y) \coloneqq |x-y|$. Arguing similar to Example 4.45, the sets $[1/(n+1), \infty)$ are closed for all $n \in \mathbb{N}$. However, their union

$$\bigcup_{n\in\mathbb{N}} [1/(n+1), \infty) = (0, \infty)$$

is not closed by Example 4.45.

Unsurprisingly, analogous to Corollary 4.32, we can also translate Proposition 4.62 into closures.

Exercise 4.65. Let (X,d) be a metric space. Given subsets $A, B \subseteq X$, show that $\overline{A \cup B} = \overline{A} \cup \overline{B}$.

4.2.5 Complements of Open Sets

The purpose of this subsection is to establish the following result.

Theorem 4.66. Let (X,d) be a metric space. Given a subset $A \subseteq X$, the following are equivalent.

(a) A is closed.

(b) $X \setminus A$ is open.

Proof. We work very slowly done equivalent characterizations of these statements. At each step, be convinced that the statement is equivalent to the one provided before.

1. By definition, "$X \setminus A$ is open" is equivalent to "if $x \in X \setminus A$, there exists $\varepsilon > 0$ such that $B(x,\varepsilon) \subseteq X \setminus A$."
2. By contraposition, this is equivalent to "if $x \in A$, then any $\varepsilon > 0$ has $B(x,\varepsilon) \not\subseteq X \setminus A$."
3. Unpacking the statement "$B(x,\varepsilon) \not\subseteq X \setminus A$," we see that it promises some $y \in B(x,\varepsilon)$ such that $y \notin X \setminus A$. However, $y \notin X \setminus A$ is equivalent to $y \in A$, so "$B(x,\varepsilon) \not\subseteq X \setminus A$," is equivalent to asserting "$A \cap B(x,\varepsilon)$ is nonempty."

   Thus, "$X \setminus A$ is open" is now equivalent to "if $x \in A$, then any $\varepsilon > 0$ makes the sets $A \cap B(x,\varepsilon)$ nonempty."
4. Finishing, the last condition stated is equivalent to "A is closed," completing the proof. ∎

Exercise 4.67. Use Theorem 4.66 to show the following: given a subset A of a metric space (X,d), the following are equivalent.

(a) A is open.

(b) $X \setminus A$ is closed.

Roughly speaking, Theorem 4.66 provides another characterization of closed sets by relating them to how we understand open sets. Indeed, many of the results we proved in the previous subsection can be given new proofs by appealing to results of open sets. Here are a few for you to try.

Exercise 4.68. Use Theorem 4.66 and results of section 4.1.3 to give new proofs of Propositions 4.60 and 4.62.

We can translate Theorem 4.66 into a statement about closures and interiors, as follows.

Corollary 4.69. Let (X, d) be a metric space. For any subset $A \subseteq X$, we have $\overline{X \setminus A} = X \setminus A^\circ$.

Proof. We show this in two inclusions.

- We show $\overline{X \setminus A} \subseteq X \setminus A^\circ$. Note A° is open by Proposition 4.27, so $X \setminus A^\circ$ is closed by Exercise 4.68. However, $A^\circ \subseteq A$ implies $X \setminus A \subseteq X \setminus A^\circ$, so
$$\overline{X \setminus A} \subseteq X \setminus A^\circ$$
follows from Proposition 4.58.

- We show $X \setminus A^\circ \subseteq \overline{X \setminus A}$. Taking complements, we may instead show $X \setminus \overline{X \setminus A} \subseteq A^\circ$. Now, $\overline{X \setminus A}$ is closed by Proposition 4.57, so $X \setminus \overline{X \setminus A}$ is open by Theorem 4.66. Further, $X \setminus A \subseteq \overline{X \setminus A}$ implies $X \setminus \overline{X \setminus A} \subseteq A$, so Lemma 4.26 implies
$$X \setminus \overline{X \setminus A} \subseteq A^\circ,$$
which is what we wanted. ∎

Exercise 4.70. Use Corollaries 4.32 and 4.69 to give a new proof of Exercise 4.65. It might be helpful to show $\overline{A} = X \setminus (X \setminus A)^\circ$ for subsets $A \subseteq X$.

4.2.6 Problems

Problem 4.9. Give a set X the discrete metric d defined by
$$d(x, y) \coloneqq \begin{cases} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y. \end{cases}$$
Suppose that a sequence $\{a_n\}_{n \in \mathbb{N}}$ of elements in X converges to the point $a \in X$. Show that there exists some N such that $a_n = a$ for all $n > N$.

Problem 4.10. Repeat exercise Problem 4.6 for the closure instead of the interior.

Problem 4.11. Consider the metric space $(\mathbb{R}, d)$ where $d(x, y) \coloneqq |x - y|$. Given real numbers $a, b \in \mathbb{R}$ such that $a < b$, show that $\overline{(a, b)} = [a, b]$.

Problem 4.12. Consider the metric space $\left(\mathbb{R}^2, d\right)$, where $d((x_1, y_1), (x_2, y_2)) \coloneqq \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. Show that the set
$$\{(x, x) : x \in \mathbb{R}\}$$
is a closed subset of $\mathbb{R}^2$.

Problem 4.13. We investigate the closure of open balls.

(a) Let (X, d) be a metric space. For any $a \in X$ and $r > 0$, show that $\overline{B(a,r)} \subseteq \overline{B}(a,r)$.

(b) Give $\mathbb{R}$ the discrete metric defined by

$$d(x,y) \coloneqq \begin{cases} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y. \end{cases}$$

Find $r > 0$ such that $\overline{B(0,r)} \neq \overline{B}(0,r)$.

(c) Give $\mathbb{R}$ the usual metric defined by $d(x,y) \coloneqq |x-y|$. Show that $\overline{B(a,r)} = \overline{B}(a,r)$ for each $a \in \mathbb{R}$ and $r > 0$.

4.3 Week 11: Continuity

In this section, we use our understanding of open and closed sets to discuss continuity. At the end, we give a small taste of topology.

4.3.1 Continuous Functions

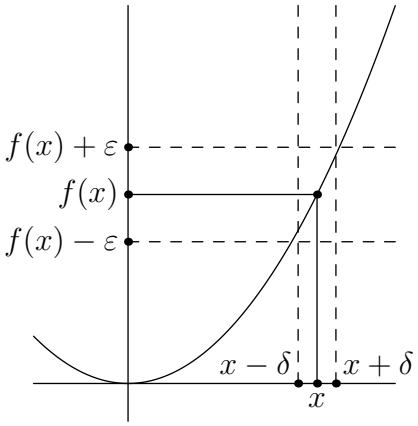
Roughly speaking, the idea here is that mathematical objects are not understood in isolation but in how they relate to one another. Metric spaces relate to each other by continuous maps.

Definition 4.71 (continuous)**.** Let (X, d) and (X', d') be metric spaces. A function $f\colon X \to X'$ is *continuous* at an element $x_0 \in X$ if and only if, for all $\varepsilon > 0$, there exists some $\delta > 0$ such that

$$d(x_0, x) < \delta \implies d'(f(x_0), f(x)) < \varepsilon.$$

The function f is *continuous* if and only if it is continuous at all elements $x_0 \in X$.

The intuition here is that points x sufficiently close to x_0 should have $f(x)$ sufficiently close to $f(x_0)$. The definition makes this statement rigorous. Here is the image for continuous functions $\mathbb{R} \to \mathbb{R}$, where $\mathbb{R}$ has been given the usual metric $d(x,y) \coloneqq |x-y|$.



Let's see a few examples.

Example 4.72. Give $\mathbb{R}$ the metric $d(x,y) \coloneqq |x-y|$. Then the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) \coloneqq 3x+5$ is continuous.

Proof 1. We would like to show that f is continuous at any given $x_0 \in \mathbb{R}$. This means that, for any $\varepsilon > 0$, we must find some $\delta > 0$ such that $|x_0 - x| < \delta$ implies $|f(x_0) - f(x)| < \varepsilon$.

The conclusion $|f(x_0) - f(x)| < \varepsilon$ is currently mysterious to us, so we will try to simplify that. Expanding, we see

$$|f(x_0) - f(x)| = |(3x_0 + 5) - (3x + 5)| = 3|x_0 - x|.$$

Because we want $|f(x_0) - f(x)| < \varepsilon$, we see that it will be enough for $|x_0 - x| < \varepsilon/3$. As such, we set $\delta \coloneqq \varepsilon/3$ and can compute that $|x_0 - x| < \delta$ implies

$$|f(x_0) - f(x)| = 3|x_0 - x| < 3\delta = \varepsilon,$$

which is what we wanted. ■

In the above proof, note that δ depended on ε. This is a common feature in such proofs.

Now, the above proof was written to explain why we chose δ from that ε. For clarity reasons, it is more common for a proof to simply construct ε without this intermediate explanation. Here is what that looks like.

Proof 2. We would like to show that f is continuous at any given $x_0 \in \mathbb{R}$. Well, for any $\varepsilon > 0$, set $\delta \coloneqq \varepsilon/3$. Then $d(x_0, x) < \delta$ implies

$$d(f(x_0), f(x)) = |(3x_0 + 5) - (3x + 5)| = 3|x_0 - x| < 3\delta = \varepsilon,$$

which is what we wanted. ■

For the remainder of this subsection, we will write two proofs of each our continuity results (one explaining how δ is chosen, and one where δ is merely constructed), but we will opt for the second style of proof afterwards. On homework, it is encouraged to write proofs closer to the second proofs.

Example 4.73. Give $\mathbb{R}$ the metric $d(x, y) \coloneqq |x - y|$. Then the function $f \colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) \coloneqq x^2$ is continuous.

Proof 1. This argument is harder. We would like to show that f is continuous at any given $x_0 \in \mathbb{R}$. Namely, for any $\varepsilon > 0$, we would like some $\delta > 0$ such that $|x_0 - x| < \delta$ implies $|f(x_0) - f(x)| < \varepsilon$.

Again, we begin by simplifying $|f(x_0) - f(x)|$ as

$$|f(x_0) - f(x)| = \left|x_0^2 - x^2\right| = |(x_0 - x)(x_0 + x)| = |x_0 - x| \cdot |x_0 + x|.$$

Now, we can force $|x_0 - x|$ to be small (because we enforce $|x_0 - x| < \delta$), but it's not obvious how to make $|x_0 + x|$ to be small. It will actually be enough to just make $|x_0 + x|$ have some bounded size. Explicitly, note that $|x_0 - x| \leq 1$ implies that $-1 \leq x - x_0 \leq 1$ and so

$$-(2|x_0| + 1) \leq 2x_0 - 1 \leq x + x_0 \leq 2x_0 + 1 \leq 2|x_0| + 1,$$

so $|x + x_0| \leq 2|x_0| + 1$ follows. Thus, we see that

$$|f(x_0) - f(x)| = |x_0 - x| \cdot |x_0 + x| \leq |x_0 - x| \cdot (2|x_0| + 1),$$

where we have also plugged in $|x_0 - x| < \delta$. To make the right-hand side less than ε, we see that we must also have $|x_0 - x| < \varepsilon/(2|x_0| + 1)$.

In total, the above argument requires $|x_0 - x| \leq 1$ and $|x_0 - x| < \varepsilon/(2|x_0| + 1)$. Thus, we set $\delta \coloneqq \frac{1}{2}\min\{1, \varepsilon/(2|x_0| + 1)\}$ so that both of these conditions are satisfied; note $\delta > 0$. Tracking through the argument of the previous paragraph, we see that $|x_0 - x| < \delta$ implies

$$|f(x_0) - f(x)| \leq |x_0 - x| \cdot (2|x_0| + 1) < \delta \cdot (2|x_0| + 1) = \varepsilon,$$

which is what we wanted. ■

Proof 2. We would like to show that f is continuous at a given $x_0 \in \mathbb{R}$. Well, for any $\varepsilon > 0$, set $\delta \coloneqq \frac{1}{2}\min\{1, \varepsilon/(2|x_0|+1)\}$. We show that $d(x_0, x) < \delta$ implies $d(f(x_0), f(x)) < \varepsilon$.

To begin, we note $d(x_0, x) < \delta$ implies $|x_0 - x| < 1$. Then we see $-1 < x - x_0 < 1$, so $2x_0 - 1 < x + x_0 < 2x_0 + 1$, so

$$|x + x_0| < 2|x_0| + 1.$$

But now we can bound

$$d(f(x_0), f(x)) = \left|x_0^2 - x^2\right| = |x_0 - x| \cdot |x_0 + x| < \delta \cdot (2|x_0| + 1).$$

By definition of δ, the right-hand side above is less than ε, which completes the proof. ■

In the above proofs, we see that δ depended on both ε and x_0. This is also common.

Exercise 4.74. Give $\mathbb{R}$ the metric $d(x, y) \coloneqq |x - y|$. Show that the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) \coloneqq \frac{1}{x^2+1}$ is continuous.

Let's start working with different metric spaces now.

Example 4.75. Give $\mathbb{R}$ the metric $d(x, y) \coloneqq |x - y|$, and give $\mathbb{R}^2$ the Euclidean metric $d_2((x, y), (x', y')) \coloneqq \sqrt{(x - x')^2 + (y - y')^2}$. Then the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) \coloneqq x + y$ is continuous.

Proof 1. Fix some point $(x_0, y_0) \in \mathbb{R}^2$ so that we would like to show that f is continuous at (x_0, y_0). Namely, for any $\varepsilon > 0$, we would like $\delta > 0$ such that

$$d_2((x_0, y_0), (x, y)) < \delta \overset{?}{\implies} d(f(x_0, y_0), f(x, y)) < \varepsilon.$$

Unwinding our definitions, we we want

$$\sqrt{(x_0 - x)^2 + (y_0 - y)^2} < \delta \overset{?}{\implies} |(x_0 - x) + (y_0 - y)| < \varepsilon.$$

To compress our information somewhat, we set $a \coloneqq (x_0 - x)$ and $b \coloneqq (y_0 - y)$ to be some real numbers, and we want

$$\sqrt{a^2 + b^2} < \delta \overset{?}{\implies} |a + b| < \varepsilon.$$

Now, the idea is that $\sqrt{a^2 + b^2} < \delta$ should actually force both a and b to be small, which makes $|a + b|$ also small. Indeed, $|a| < \sqrt{a^2 + b^2}$, so $|a| < \delta$ is forced. Similarly, we see $|b| < \delta$, and then it follows $|a + b| \leq |a| + |b| < 2\delta$. Thus,

$$\sqrt{a^2 + b^2} < \delta \implies |a + b| < 2\delta.$$

Setting $\delta \coloneqq \varepsilon/2$ completes the proof. ■

Proof 2. We would like to show that f is continuous at a given $(x_0, y_0) \in \mathbb{R}^2$. Well, for any $\varepsilon > 0$, we set $\delta \coloneqq \varepsilon/2$. Then $d_2((x_0, y_0), (x, y)) < \delta$ implies

$$|x_0 - x| < \sqrt{(x_0 - x)^2 + (y_0 - y)^2} = d_2((x_0, y_0), (x, y)) < \delta.$$

Similarly, we see $|y_0 - y| < \delta$. It follows

$$d(f(x_0, y_0), f(x, y)) = |(x_0 + y_0) - (x + y)| \leq |x_0 - x| + |y_0 - y| < 2\delta = \varepsilon,$$

which finishes the proof. ■

4.3.2 Working with Abstract Metric Spaces

As an intermission, here are a few examples of continuous functions between abstract metric spaces.

Example 4.76. Let (X, d) and (X', d') be metric spaces. Fixing some point $a' \in X'$, define the function $f\colon X \to X'$ by $f(x) \coloneqq a'$ for all $x \in X$. Then the function f is continuous: at any $x_0 \in X$, for each $\varepsilon > 0$, we may set $\delta \coloneqq 1$ so that

$$d(x_0, x) < \delta \implies d'(f(x_0), f(x)) = d'(a, a) = 0 < \varepsilon.$$

Example 4.77. Let (X', d') be a metric space. Give a nonempty set X the discrete metric

$$d(x, y) \coloneqq \begin{cases} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y. \end{cases}$$

Then any function $f\colon X \to X'$ is continuous.

Proof. This is a bit silly. To show that f is continuous at $x_0 \in X$, for any $\varepsilon > 0$, we set $\delta \coloneqq 1$. Then $d(x_0, x) < \delta$, meaning $d(x_0, x) < 1$. However, checking the definition of d, we see $d(x_0, x) < 1$ forces $d(x_0, x) = 0$ and so $x_0 = x$. But then $f(x_0) = f(x)$, so $d'(f(x_0), f(x)) = 0 < \varepsilon$, as needed. ∎

Example 4.78. Let (X, d) be a metric space. Further, give $X \times X$ the metric $d_1((x, y), (x', y')) \coloneqq d(x, x') + d(y, y')$ of Example 4.9, and give $\mathbb{R}$ the usual metric $d_{\mathbb{R}}(x, y) \coloneqq |x - y|$. Then the metric function $d\colon X \times X \to \mathbb{R}$ is continuous.

This statement is scary, so we will give our two proofs, as in the previous subsection.

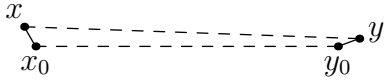
Proof 1. Fix some $(x_0, y_0) \in X \times X$ so that we will show d is continuous at (x_0, y_0). Indeed, fix any $\varepsilon > 0$, and we need $\delta > 0$ such that

$$d_1((x_0, y_0), (x, y)) < \delta \overset{?}{\implies} d_{\mathbb{R}}(d(x_0, y_0), d(x, y)) < \varepsilon.$$

Unwinding our definitions, we are asking for

$$d(x_0, x) + d(y_0, y) < \delta \overset{?}{\implies} |d(x_0, y_0) - d(x, y)| < \varepsilon.$$

Now, the idea is that $d(x_0, x) + d(y_0, y) < \delta$ forces x_0 and x to be close together and forces y_0 and y to be close together. Thus, the distance between x_0 and y_0 should be approximately equal to the distance between x and y. Here is the image, where the solid lines are small distances.



With this in mind, we use $d(x_0, x) + d(y_0, y) < \delta$ and the triangle inequality to bound

$$d(x_0, y_0) \leq d(x_0, x) + d(x, y_0) \leq d(x_0, x) + d(x, y) + d(y, y_0) < d(x, y) + \delta.$$

Reversing the roles of (x_0, y_0) and (x, y), we see

$$d(x, y) < d(x_0, y_0) + \delta.$$

Thus, we see

$$d(x_0, x) + d(y_0, y) < \delta \implies |d(x_0, y_0) - d(x, y)| < \delta.$$

As such, we set $\delta \coloneqq \varepsilon$ to complete the proof. ∎

Proof 2. We show that d is continuous at a given $(x_0, y_0) \in X \times X$. Indeed, for any $\varepsilon > 0$, we set $\delta := \varepsilon$. Then $d_1((x_0, y_0), (x, y)) < \delta$ means

$$d(x_0, x) + d(y_0, y) < \varepsilon.$$

Now, by the triangle inequality, we see

$$d(x_0, y_0) \leq d(x_0, x) + d(x, y) + d(y, y_0) \leq d(x, y) + \varepsilon.$$

A similar argument shows $d(x, y) \leq d(x_0, y_0) + \varepsilon$, so

$$d_{\mathbb{R}}(d(x_0, y_0), d(x, y)) = |d(x_0, y_0) - d(x, y)| < \varepsilon,$$

which is what we wanted. ■

Exercise 4.79. Let (X, d) be a metric space. Further, give $X \times X$ the metric $d_1((x, y), (x', y')) := d(x, x') + d(y, y')$ of Example 4.9. Show that the projection function $p_1 \colon X \times X \to X$ given by $p_1(x, y) := x$ is continuous.

Exercise 4.80. Let (X, d) be a metric space. Show that the function $i \colon X \to X$ defined by $i(x) := x$ is continuous.

Exercise 4.81. Let (X, d) and (X', d') and (X'', d'') be metric spaces. If the functions $f \colon X \to X'$ and $f' \colon X' \to X''$ are continuous, show that the function $(f' \circ f) \colon X \to X''$ is also continuous.

4.3.3 Continuity by Open Sets

Now that we are familiar with continuous functions, we explain how they relate to open sets. The idea is that the statement

$$d(x_0, x) < \delta \implies d'(f(x_0), f(x)) < \varepsilon$$

can be restated as

$$f(B(x_0, \delta)) \subseteq B(f(x_0), \varepsilon),$$

or

$$B(x_0, \delta) \subseteq f^{-1}(B(f(x_0), \varepsilon)).$$

(Be careful: the two open balls live in different metric spaces!) Indeed, the condition $x \in B(x_0, \delta)$ is the same as $d(x_0, x) < \delta$. Further, the condition $x \in f^{-1}(B(f(x_0), \varepsilon)$ is equivalent to $f(x) \in B(f(x_0), \varepsilon)$, which is equivalent to $d'(f(x_0), f(x)) < \varepsilon$.

The above paragraph explains how pre-images are going to enter our discussion.

Theorem 4.82. Let (X, d) and (X', d') be metric spaces, and let $f \colon X \to X'$. The following are equivalent.

(a) f is continuous.

(b) For all open subsets $U' \subseteq X'$, the set $f^{-1}(U') \subseteq X$ is also open.

Proof. We have two implications to show.

- We show that (a) implies (b). Suppose that $U' \subseteq X'$ is open, and we want to show that $f^{-1}(U')$ is open. Well, fix some $x_0 \in f^{-1}(U')$, and we would like some $\delta > 0$ such that $B(x_0, \delta) \subseteq f^{-1}(U')$.

However, we note that $f(x_0) \in U'$, and U' is open, so there exists some ε such that $B(f(x_0), \varepsilon) \subseteq U'$. Arguing as above, we note that continuity of f promises some $\delta > 0$ such that

$$d(x_0, x) < \delta \implies d'(f(x_0), f(x)) < \varepsilon.$$

But now, $d'(f(x_0), f(x)) < \varepsilon$ implies $f(x) \in B(f(x_0), \varepsilon)$, implying $f(x) \in U'$, implying $x \in f^{-1}(U')$. Thus, $B(x_0, \delta) \subseteq f^{-1}(U')$, which completes the proof.

- We show (b) implies (a). We would like to show that f is continuous at a given $x_0 \in X$. Indeed, for any $\varepsilon > 0$, we would like some $\delta > 0$ such that $d(x_0, x) < \delta$ implies $d'(f(x_0), f(x)) < \varepsilon$.

  Well, we know $B(f(x_0), \varepsilon)$ is open by Proposition 4.24, so $f^{-1}(B(f(x_0), \varepsilon))$ is also open by hypothesis on f. However, $x_0 \in f^{-1}(B(f(x_0), \varepsilon))$, so by definition of an open set, there exists $\delta > 0$ such that

  $$B(x_0, \delta) \subseteq f^{-1}(B(f(x_0), \varepsilon)).$$

  As explained previously, this is equivalent to the assertion $d(x_0, x) < \delta$ implies $d'(f(x_0), f(x)) < \varepsilon$. ■

Theorem 4.82 is a very powerful result because we understand how open sets behave reasonably well, so we will be able to show properties of continuous functions fairly cleanly by appealing to facts about open sets. As an example, we will give a new proof of Exercise 4.81, which is fairly annoying with Theorem 4.82.

Proof of Exercise 4.81. Using Theorem 4.82, for any open subset $U'' \subseteq X''$, we would like to show that $(f' \circ f)^{-1}(U'')$ is open in X. Well, we compute

$$(f' \circ f)^{-1}(U'') = \{x \in X : f'(f(x)) \in U''\} = \left\{x \in X : f(x) \in (f')^{-1}(U'')\right\} = f^{-1}\left((f')^{-1}(U'')\right).$$

Now, because f' is continuous, we see $(f')^{-1}(U'')$ is open by Theorem 4.82, and $f^{-1}\left((f')^{-1}(U'')\right)$ is also open for the same reason. This completes the proof. ■

Exercise 4.83. Give new proofs of Example 4.76, Example 4.77, and Exercise 4.80 using Theorem 4.82.

4.3.4 Primer on Point-Set Topology

To close out this section, we make a few remarks on point-set topology. The idea here is that Theorems 4.66 and 4.82 explain that open sets are central to the way we understand metric spaces, arguably more central than the metric itself in some respects. In point-set topology, this idea becomes key, and we totally forget about the metric and focus on the open sets only.

So how should we define an open set of some "space" X? Well, the trick is that we're not going to! We are simply going to declare that some of the sets in X are open. To ensure that these open sets behave as we would expect, we are going to requite that they satisfy Examples 4.22 and 4.23 and propositions 4.29 and 4.34. Here is our definition.

Definition 4.84 (topology)**.** Let X be a set. A collection $\mathcal{T}$ of subsets of X is a *topology* on X if and only if it satisfies the following conditions.

- $\varnothing, X \in \mathcal{T}$.
- Arbitrary union: for a subcollection $\mathcal{U} \subseteq \mathcal{T}$, the union $\bigcup_{U \in \mathcal{U}} U$ lives in $\mathcal{T}$.
- Finite intersection: for $U, V \in \mathcal{T}$, we have $U \cap V \in \mathcal{T}$.

In this case, we call the ordered pair $(X, \mathcal{T})$ a *topological space*, and the sets in $\mathcal{T}$ are called *open* sets.

Example 4.85. Let X be a set. Then the collection $\mathcal{T} \coloneqq \{\varnothing, X\}$ is a topology on X, called the "indiscrete topology." Here are the checks.

- Note $\varnothing, X \in \mathcal{T}$ by assumption.
- Arbitrary union: fix a subcollection $\mathcal{U} \subseteq \mathcal{T}$. If $\mathcal{U}$ does not contain X, then we must have $\mathcal{U} \subseteq \{\varnothing\}$, so $\bigcup_{U \in \mathcal{U}} U = \varnothing$, which is in $\mathcal{T}$. Otherwise, $\mathcal{U}$ does contain X, so $\bigcup_{U \in \mathcal{U}} U = X$.
- Finite intersection: fix $U, V \in \mathcal{T}$. If $U = \varnothing$ or $V = \varnothing$, then $U \cap V = \varnothing$, so $U \cap V \in \mathcal{T}$. Otherwise, $U = V = X$, so $U \cap V = X$, which is still in $\mathcal{T}$.

Exercise 4.86. Let X be a set. Show that the collection $\mathcal{T} \coloneqq \mathcal{P}(X)$ of all subsets of X is a topology on X, called the "discrete topology."

Exercise 4.87. Let (X, d) be a metric space. Show that the collection $\mathcal{T}$ of open sets forms a topology on X.

Remark 4.88. In fact, Exercise 4.87 is a generalization of Exercise 4.86. Indeed, any subset U of X is open when X is given the discrete metric.

However, Example 4.85 does not come from a metric in general. To see this, suppose that X is a set with at least two distinct elements $x, y \in X$. Then we show that the discrete topology $\mathcal{T}$ does not come from any metric d on X. Indeed, suppose for the sake of contradiction that $\mathcal{T}$ is the set of open sets of some metric $d \colon X \times X \to \mathbb{R}$. But then the set

$$U \coloneqq B(x, d(x, y))$$

is an open set in X by Proposition 4.24, so $U \in \mathcal{T}$. However, $x \in U$, but $d(x, y)$ is not less than $d(x, y)$, so $y \notin U$. This is a contradiction because the only nonempty set in $\mathcal{T}$ is X, so $x \in U$ forces $U = X$, but then this would imply $y \in U$.

Thus, we see that considering topological spaces is in fact more general than just thinking about metric spaces. Let's begin trying to push the rest of the theory we developed into our new framework. We will be very fast in our exposition, leaving most of this to exercises, because many arguments are similar to ones we've already given. For interiors, we note that Exercise 4.28 provides a construction of the interior which only discusses open sets, so we will use this as our definition.

Definition 4.89 (interior)**.** Let $(X, \mathcal{T})$ be a topological space. For a subset $A \subseteq X$, we define the *interior* to be

$$A^\circ \coloneqq \bigcup_{\substack{U \in \mathcal{T} \\ U \subseteq A}} U.$$

Here, the union is over all open sets contained in A.

Example 4.90. We generalize Lemma 4.26. Let $(X, \mathcal{T})$ be a topological space. If $A \subseteq X$ has $U \in \mathcal{T}$ and $U \subseteq A$, then the definition of A° gives $U \subseteq A^\circ$.

Exercise 4.91. Let $(X, \mathcal{T})$ be a topological space. Show the following properties of the interior for subsets $A, B \subseteq X$, generalizing the case of metric spaces.

- $\varnothing^\circ = \varnothing$ and $X^\circ = X$.
- $A^\circ \subseteq A$.
- $A \in \mathcal{T}$ if and only if $A^\circ = A$.
- $(A^\circ)^\circ = A^\circ$.
- $(A \cap B)^\circ = A^\circ \cap B^\circ$.

Now, we discuss closed sets. Theorem 4.66 tells us how to think about closed sets purely in terms of open sets, so we will take this as our definition.

Definition 4.92 (closed)**.** Let $(X, \mathcal{T})$ be a topological space. A subset $A \subseteq X$ is *closed* if and only if $X \setminus A \in \mathcal{T}$.

Exercise 4.93. Let $(X, \mathcal{T})$ be a topological space. By taking complements in the definition of a topology, show the following.

- $\varnothing$ and X are closed.
- Arbitrary intersection: given a collection $\mathcal{C}$ of closed sets, the intersection $\bigcap_{C \in \mathcal{C}} C$ is closed.
- Finite union: given closed sets A and B, the set $A \cup B$ is also closed.

Remark 4.94. By taking complements, note also that $A \subseteq X$ is open if and only if $X \setminus A$ is closed. Thus, we could also have defined open sets in terms of closed ones.

We note that we can also build an analogue to our original definition Definition 4.49 of a closed set by replacing open balls $B(a, \varepsilon)$ with general open sets U.

Proposition 4.95. Let $(X, \mathcal{T})$ be a topological space. For subsets $A \subseteq X$, the following are equivalent.

(a) A is closed.

(b) If $a \in X$ makes the sets $A \cap U$ nonempty for all $U \in \mathcal{T}$ containing a, then $a \in A$.

Proof. We have two implications to show.

- We show (a) implies (b). We will show the contrapositive: if $a \in X \setminus A$, then there exists $U \in \mathcal{T}$ containing a such that $A \cap U$ is empty. Indeed, we see that $X \setminus A$ is open because A is closed, so $U := (X \setminus A)$ will both contain a and make $A \cap U$ empty.
- We show (b) implies (a). We would like to show that $X \setminus A$ is open. Well, by taking the contrapsitive of (b), we know that any $a \in X \setminus A$ has some $U_a \in \mathcal{T}$ containing a such that $A \cap U_a$ is empty. Now, $A \cap U_a$ being empty means that all elements of U_a live in $X \setminus A$, so $U_a \subseteq X \setminus A$. Looping through all $a \in X \setminus A$, we see
$$X \setminus A = \bigcup_{a \in X \setminus A} \{a\} \subseteq \bigcup_{a \in X \setminus A} U_a \subseteq \bigcup_{a \in X \setminus A} X \setminus A.$$
Thus, equality of all the above sets holds, so
$$X \setminus A = \bigcup_{a \in X \setminus A} U_a.$$

It follows that $X \setminus A$ is the union of some sets $U_a \in \mathcal{T}$, so $X \setminus A \in \mathcal{T}$ because $\mathcal{T}$ is a topology on X. ■

Next up, we discuss closures. Similar to our discussion of interiors, we have the statement Exercise 4.59 which defines the closure just in terms of closed sets, so we will use this as our definition.

Definition 4.96 (closure)**.** Let $(X, \mathcal{T})$ be a topological space. For a subset $A \subseteq X$, we define the *closure* to be

$$\overline{A} \coloneqq \bigcap_{\substack{X \setminus C \in \mathcal{T} \\ A \subseteq C}} C.$$

Here, the intersection is over all closed sets C containing A.

Here are the appropriate generalizations of our facts about closures.

Proposition 4.97. Let $(X, \mathcal{T})$ be a topological space. For any subset $A \subseteq X$, we have $\overline{X \setminus A} = X \setminus A^\circ$.

Proof. We directly compute

$$X \setminus A^\circ = X \setminus \bigcup_{\substack{U \in \mathcal{T} \\ U \subseteq A}} U = \bigcap_{\substack{U \in \mathcal{T} \\ U \subseteq A}} (X \setminus U).$$

The main claim, now, is that the last intersection is precisely $\overline{X \setminus A}$. We now define the collection $\mathcal{C} \coloneqq \{X \setminus U : U \in \mathcal{T}, U \subseteq A\}$. Note that each $U \in \mathcal{T}$ with $U \subseteq A$ produces a closed set $X \setminus U$ such that $X \setminus A \subseteq X \setminus U$. Thus, each set in $\mathcal{C}$ is a closed set containing $X \setminus A$.

Conversely, if C is a closed set containing $X \setminus A$, then $U \coloneqq X \setminus C$ is an open set contained in A, so $C = X \setminus U$ is in $\mathcal{C}$. Thus, $\mathcal{C}$ is exactly the collection of closed sets containing $X \setminus A$, so

$$X \setminus A^\circ = \bigcap_{C \in \mathcal{C}} C = \overline{X \setminus A}.$$

This completes the proof. ■

Remark 4.98. Proposition 4.97 tells us that we could also have defined the closure as $\overline{A} \coloneqq X \setminus (X \setminus A)^\circ$.

Exercise 4.99. Let $(X, \mathcal{T})$ be a topological space. Show the following properties of the closure for subsets $A, B \subseteq X$.

- $\overline{\varnothing} = \varnothing$ and $\overline{X} = X$.
- $A \subseteq \overline{A}$.
- A is closed if and only if $\overline{A} = A$.
- $\overline{\overline{A}} = \overline{A}$.
- $\overline{A \cup B} = \overline{A} \cup \overline{B}$.

Exercise 4.100. Show the following generalization of Definition 4.44: let $(X, \mathcal{T})$ be a topological space, and let $A \subseteq X$ be a subset. Then $a \in \overline{A}$ if and only if $A \cap U$ is nonempty for all open subsets $U \in \mathcal{T}$ containing a.

Lastly, we discuss continuous functions. Theorem 4.82 tells us what our definition should be.

Definition 4.101 (continuous)**.** Let $(X, \mathcal{T})$ and $(X', \mathcal{T}')$ be topological spaces. A function $f\colon X \to X'$ is *continuous* if and only if, for all $U' \in \mathcal{T}'$, we have $f^{-1}(U') \in \mathcal{T}$.

Here are some immediate consequences of this definition.

Example 4.102. Let $(X, \mathcal{T})$ and $(X', \mathcal{T}')$ be topological spaces, where $\mathcal{T}' = \{\varnothing, X'\}$. (In other words, $\mathcal{T}'$ is the indiscrete topology on X'.) Then any function $f\colon X \to X'$ is continuous. Indeed, we just have to compute $f^{-1}(\varnothing) = \varnothing$ and $f^{-1}(X') = X$ are both in $\mathcal{T}$.

Exercise 4.103. Let $(X, \mathcal{T})$ and $(X', \mathcal{T}')$ be topological spaces, where $\mathcal{T} = \mathcal{P}(X)$. (In other words, $\mathcal{T}$ is the discrete topology on X.) Show that any function $f\colon X \to X'$ is continuous.

Exercise 4.104. Let $(X, \mathcal{T})$ be a topological space. Show that the function $i\colon X \to X$ defined by $i(x) \coloneqq x$ is continuous.

Exercise 4.105. Let $(X, \mathcal{T})$ and $(X', \mathcal{T}')$ and $(X'', \mathcal{T}'')$ be topological spaces. Given continuous functions $f\colon X \to X'$ and $f'\colon X' \to X''$, show that $f' \circ f$ is continuous.

4.3.5 Problems

Problem 4.14. Give $\mathbb{R}$ the metric $d(x, y) \coloneqq |x - y|$. Show the following.

(a) The function $f(x) = |x|$ is continuous.

(b) The function $f(x) \coloneqq \sqrt{|x|}$ is continuous at each $x_0 \neq 0$.

(c) The function $f(x) \coloneqq \sqrt{|x|}$ is continuous at $x_0 = 0$.

(d) The function

$$f(x) \coloneqq \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0, \end{cases}$$

is not continuous at $x_0 = 0$.

Problem 4.15. Give $\mathbb{R}$ the metric $d(x, y) \coloneqq |x - y|$. Show the following.

(a) If $f, g\colon \mathbb{R} \to \mathbb{R}$ are continuous functions, then $f + g$ is continuous. Here, $(f + g)\colon \mathbb{R} \to \mathbb{R}$ is defined by $(f + g)(x) \coloneqq f(x) + g(x)$.

(b) If $f\colon \mathbb{R} \to \mathbb{R}$ is continuous, and $a \in \mathbb{R}$, then the function af is continuous. Here, $(af)\colon \mathbb{R} \to \mathbb{R}$ is defined by $(af)(x) \coloneqq a \cdot f(x)$.

Problem 4.16. Let (X, d) and (X', d') be metric spaces. Suppose that the function $f\colon X \to X'$ has some real number $c \in \mathbb{R}$ such that

$$d'(f(x_1), f(x_2)) \leq c \cdot d(x_1, x_2)$$

for any $x_1, x_2 \in X$. Show that f is continuous.

Problem 4.17. We solve Problem 4.12 another way. Let $(\mathbb{R}^2, d)$ be a metric space, where $d((x_1, y_1), (x_2, y_2)) \coloneqq \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. Further, let $(\mathbb{R}, d)$ be a metric space, where $d(x, y) \coloneqq |x - y|$.

(a) Define $p \colon \mathbb{R}^2 \to \mathbb{R}$ by $p(x, y) \coloneqq y - x$. Show that p is continuous.

(b) Show that $\{(x, y) \in \mathbb{R}^2 : y \neq x\}$ is an open subset of $\mathbb{R}^2$.

(c) Show that $\{(x, x) : x \in \mathbb{R}\}$ is a closed subset of $\mathbb{R}^2$.

Problem 4.18. Let $(X, \mathcal{T})$ and $(X', \mathcal{T}')$ be topological spaces. Given a function $f \colon X \to X'$, show that the following are equivalent.

(a) f is continuous.

(b) For all closed subsets $A' \subseteq X'$, the set $f^{-1}(A') \subseteq X$ is also closed.

Problem 4.19. Let $(X, \mathcal{T})$ be a topological space. Given some $A \subseteq X$, we define the *boundary* of A as $\partial A \coloneqq \overline{A} \setminus A^\circ$.

(a) Give $\mathbb{R}$ the usual metric $d(x, y) \coloneqq |x - y|$. Show that $\partial\mathbb{Q} = \mathbb{R}$ and $\partial\mathbb{R} = \varnothing$.

(b) Given $A \subseteq X$, show that ∂A is closed.

(c) If $A \subseteq X$ is closed, show that $\partial A \subseteq A$. In particular, show $\partial(\partial A) \subseteq \partial A$.

(d) If $A \subseteq X$ is closed, show that $(\partial A)^\circ = \varnothing$. Deduce $\partial(\partial A) = A$ in this case.

(e) Given $A \subseteq X$, show that $\partial(\partial(\partial A)) = \partial(\partial A)$.

Chapter 5
Introduction to Group Theory

The philosophy is that any time the reader sees a definition or a theorem about such an object, they should test it against the prototypical example.

—Evan Chen, [Che22]

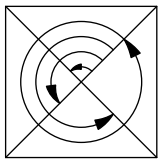
5.1 Week 12: Groups

Groups, like partially ordered sets and metric spaces, are sets endowed with some structure. The goal of this section is to define groups and then give many examples.

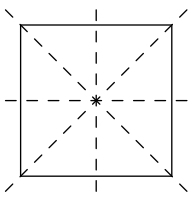
5.1.1 Symmetries of the Square

Intuitively, a group is the set of symmetries on an object. For example, let D_4 denote the set of symmetries of a square. There are eight elements in D_4, as follows.

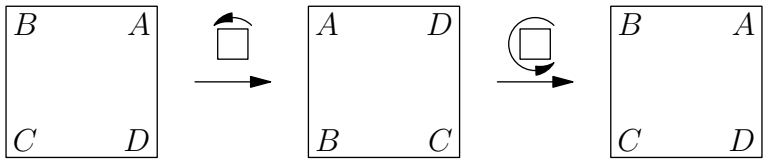
- One can counterclockwise rotate by 0°, but 90°, by 180°, or by 270°.



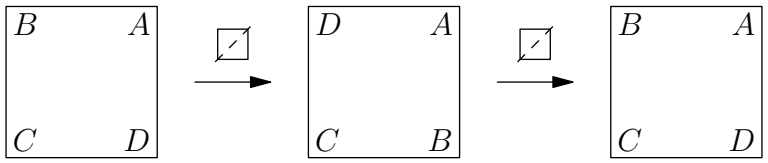
- One can also reflect along one of the following lines.



There are two central points to make about our eight symmetries: we can invert them, and we can compose them. Indeed, we expect a symmetry to be some action on the square which we can undo, and that undoing action is precisely inversion. For example, we can undo a $90°$ rotation by rotating $270°$, as follows.

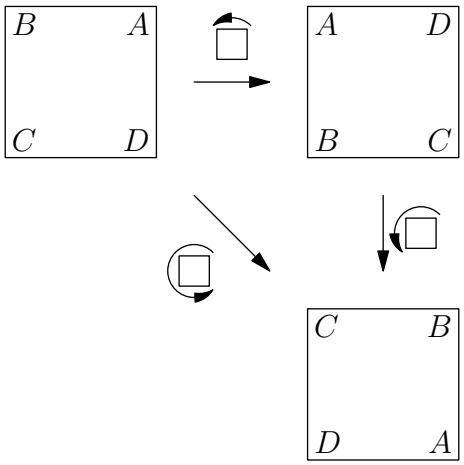


(We have labeled the vertices of the square for clarity.) Similarly, we can undo any reflection by just doing the reflection again.

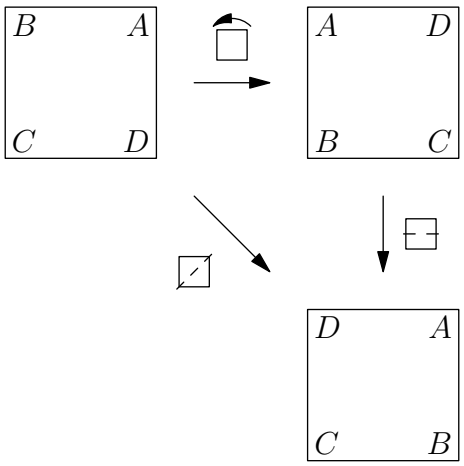


It will be a general observation that we want each operation to have an inverse operation.

The second central point is that we can compose two symmetries to get a third symmetry. Here, composition means that if we apply one symmetry, and then we apply a second symmetry, the total operation applied makes a symmetry. For example, if we rotate twice, we get out another rotation.



More interestingly, if we rotate and then reflect, we will get out another reflection.



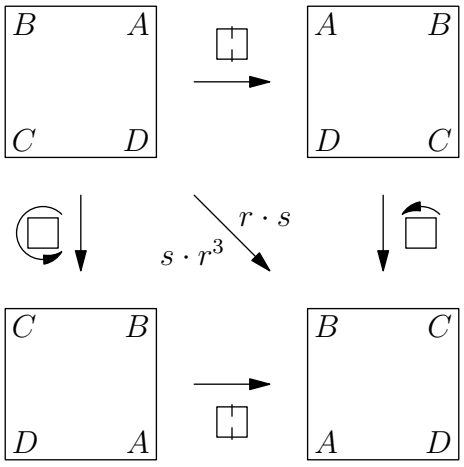
Exercise 5.1. What happens if we reflect and then rotate? What happens if we reflect and then reflect again?

Let's start to label what's going on. Given two symmetries of the square g and h, we will let $g \cdot h$ denote the symmetry obtained by applying h and then applying g. The reason that we go right-to-left is to mimic function composition.

Let r denote the $90°$ counterclockwise rotation. Then we can compute, as we did above, that $r^2 = r \cdot r$ is the $180°$ rotation and that $r^3 = r \cdot r \cdot r$ is the $270°$ rotation. Further, $r^4 = r \cdot r \cdot r \cdot r$ is the $360°$ rotation, but this is a special operation: rotating by $360°$ does nothing, so we will call this operation e.[1] Note $s \cdot e = s$ and $e \cdot s = s$ for any symmetry s because applying the symmetry e does nothing.

Let's discuss inversion. In general, a symmetry $g \in D_4$ will have an inverse symmetry $g^{-1} \in D_4$ such that $g \cdot g^{-1}$ and $g^{-1} \cdot g$ are both the do-nothing symmetry e. For example, we see that $r \cdot r^3 = r^4 = e$, so r^3 is the operation "undoing" r. As such, we think of r^3 as the inverse symmetry to r, so we might write $r^3 = r^{-1}$. Similarly, we can see that $r^2 = \left(r^2\right)^{-1}$ or even that $\left(r^{-1}\right)^{-1} = r$.

To add in reflections, we will just let s denote the reflection of the square across the vertical axis. Note $s^2 = s \cdot s = e$ because reflecting over an axis twice sends the square back to where it started. Now, r and s actually relate to each other: we claim $r \cdot s = s \cdot r^3$, which we can see directly by drawing our squares.



Having access to a relation like $r \cdot s = s \cdot r^3$ allows us to manipulate our symmetries algebraically without ever having to draw squares. For example, we can compute

$$\begin{aligned} r \cdot s \cdot r \cdot s &= r \cdot (s \cdot r) \cdot s \\ &= r \cdot \left(r^3 \cdot s\right) \cdot s \\ &= r^4 \cdot s^2 \\ &= e \cdot e \\ &= e. \end{aligned}$$

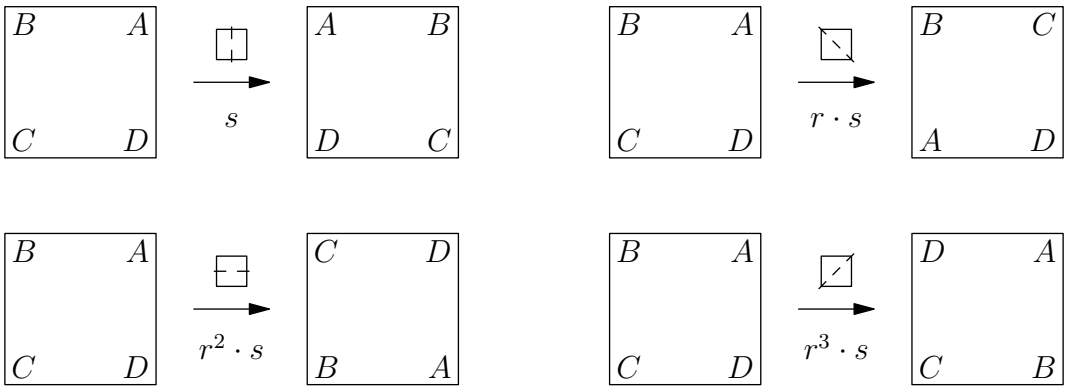
Thus, reflecting along s, rotating by r, reflecting along by s, and then rotating by r one more time in total does the same symmetry as nothing at all! This is not at all obvious by just stating it out loud, but it was not difficult to show with our algebraic manipulation.

Exercise 5.2. Verify by drawing squares that $r \cdot s \cdot r \cdot s = e$.

Remark 5.3. In the above algebraic manipulation, we have used the fact that $(g \cdot h) \cdot k = g \cdot (h \cdot k)$ for symmetries g, h, and k. However, because symmetries are functions that we apply to a square, and function composition associates, the operation $\cdot$ that we defined will also associate.

[1] The letter e stands for "eidentity."

We close our discussion of D_4 by enumerating the remaining the reflections in terms of r and s. Feel free to verify these as exercises.



Notably, we see that $D_4 = \{e, r, r^2, r^3, r \cdot s, r^2 \cdot s, r^3 \cdot s\}$.

5.1.2 Modular Arithmetic

In this subsection, we give another central example of a group, but it will not obvious how the group behaves as symmetries.

Definition 5.4. Let n be a positive integer. Let C_n denote the set of equivalence classes of the equivalence relation $\sim$ on $\mathbb{Z}$ given by $a \sim b$ if and only if $n \mid (a - b)$. We will write $a \equiv b \pmod{n}$ instead of $a \sim b$. We will denote an equivalence class by $[a]_n$, where the equivalence class is represented by $a \in \mathbb{Z}$.

Remark 5.5. Fix a positive integer n. For concreteness, we note that an integer a has equivalence class

$$[a]_n = \{b \in \mathbb{Z} : n \mid (b - a)\} = \{a + nk : k \in \mathbb{Z}\}.$$

As such, we might write $[a]_n = a + n\mathbb{Z}$.

Later on, we will use the notation $\mathbb{Z}/n\mathbb{Z}$ instead of C_n, but we will not do so until we can explain this notation.

Example 5.6. For every integer $k \in \mathbb{Z}$, there exists exactly one element in $a \in \{0, 1, 2, 3, 4\}$ such that $k - a$ is divisible by 5: indeed, divide k by 5 and take the remainder to retrieve a. Thus, we see $C_5 = \{[0]_5, [1]_5, [2]_5, [3]_5, [4]_5\}$. For concreteness, we note that

$$[1]_5 = \{b \in \mathbb{Z} : 5 \mid (b - 1)\} = \{1 + 5k : k \in \mathbb{Z}\}.$$

As such, we might write $[1]_5 = 1 + 5\mathbb{Z}$.

To generalize Example 5.6, we will need to be precise about what we mean by "division." For our purposes, we will want the division algorithm.

Theorem 5.7. Let a be an integer, and let b be a positive integer. Then there exists integers q and r such that

$$a = bq + r,$$

where $0 \leq r < b$.

Proof. The idea is to keep subtracting bs away from a until we get a remainder which is less than b. Intuitively, we know that this process should terminate eventually, though we don't necessarily know how long

it will take. Because we don't know how long it will take, we will use the well-ordering principle to non-constructively tell us how long it should take. Indeed, we claim that the set of our possible remainders

$$R \coloneqq \{a - bq : q \in \mathbb{Z}\}$$

contains a nonnegative integer. Indeed, if $a \geq 0$, then we can take $q \coloneqq 0$ so that $a = a + bq \in R$ is the needed nonnegative integer. Otherwise, $a < 0$, so we set $q \coloneqq a$ so that $a - bq = -a(b-1)$. But if $a < 0$, then $-a > 0$, and $b - 1 \geq 0$ because b is a positive integer, so $a - bq = -a(b-1)$ is a nonnegative integer which is in R.

Because R contains a nonnegative integer, the well-ordering principle implies that R contains a least nonnegative integer, which we denote r. We expect r to be the desired remainder. By definition of S, we know that there exists an integer $q \in \mathbb{Z}$ such that $a - bq = r$, or

$$a = bq + r.$$

It remains to show that $0 \leq r < b$. Because r is a nonnegative integer, we know that $r \geq 0$ automatically, so we have left to show $r < b$.

Suppose for the sake of contradiction that $r \geq b$. Continuing our intuition, having a remainder which is greater than or equal to b means that we can actually subtract out an additional b: set $r' \coloneqq r - b$ and $q' \coloneqq q + 1$, and we see

$$a - bq' = a - bq - b = r - b = r',$$

so $r' \in R$. However, $0 \leq r' < r$, so r' is a strictly smaller nonnegative integer in R, which violates the construction of r. This completes the proof. ■

Corollary 5.8. For any positive integer n, we have $C_n = \{[0]_n, [1]_n, \ldots, [n-1]_n\}$. In particular, C_n has n elements.

Proof. Note that each $k \in \{0, 1, \ldots, n-1\}$ does indeed produce an equivalence class $[k]_n \in C_n$. Furthermore, we see that these are all the needed equivalence classes by Theorem 5.7: for any integer a, there exist integers q and $r \in \{0, 1, \ldots, n-1\}$ such that

$$a = nq + r,$$

so $a \equiv r \pmod{n}$ follows, meaning $[a]_n = [r]_n$.

To finish up, we show that all the equivalence classes listed are in fact distinct. In other words, if k and ℓ are distinct elements of $\{0, 1, \ldots, n-1\}$, then $[k]_n$ and $[\ell]_n$ are distinct equivalence classes. To show this, we argue by contraposition: we show that if $k, \ell \in \{0, 1, \ldots, n-1\}$ have $[k]_n = [\ell]_n$, then $k = \ell$. Indeed, $[k]_n = [\ell]_n$ implies

$$n \mid (k - \ell).$$

Now, without loss of generality, suppose $k \geq \ell$. Then $k, \ell \in \{0, 1, 2 \ldots, n-1\}$, so $0 \leq k - \ell \leq (n-1) < n$. But for $k - \ell$ to be divisible by n, we see that the only option here is for $k - \ell = 0$, so $k = \ell$. This completes the proof. ■

For now, our focus will be on the fact that we can add elements of C_n together. Observe that there is some ambiguity here. To see this, suppose we wanted to add elements of C_5 together to get an element of $\mathbb{Z}$. We might hope that we can just do

$$[a]_5 + [b]_5 \coloneqq a + b.$$

However, this addition operation isn't well-defined! For example, we would have

$$[0]_5 + [0]_5 = 0 + 0 = 0,$$

but surely $[0]_5 = [5]_5$ because $5 \equiv 0 \pmod{5}$, so we would also have

$$[5]_5 + [5]_5 = 5 + 5 = 10.$$

Thus, our addition has suddenly required that $0 = 10$, which is false!

To fix this issue, we will add two elements in C_n together to produce a third element of C_n. Nonetheless, it still requires a bit of work to show that this addition operation is well-defined.

Lemma 5.9. The function $+\colon C_n \times C_n \to C_n$ given by $[a]_n + [b]_n \coloneqq [a+b]_n$ for any $[a]_n, [b]_n \in C_n$ is a well-defined function.

Proof. We check that ambiguities of the type described above do not arise. Namely, if $[a]_n = [a']_n$ and $[b]_n = [b']_n$, we must show that $[a]_n + [b]_n = [a']_n + [b']_n$. Unwinding how we defined $+$, we want to show

$$[a+b]_n = [a'+b']_n.$$

Because $[a]_n = [a']_n$, know that $n \mid (a-a')$, so there exists $k \in \mathbb{Z}$ such that $a - a' = kn$. Similarly, $[b]_n = [b']_n$ implies that there exists $\ell \in \mathbb{Z}$ such that $b - b' = \ell n$, so we see

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

Thus, $n \mid (a+b) - (a'+b')$, meaning $[a+b]_n = [a'+b']_n$. ∎

Let's try to draw a few parallels between D_4 with its operation $\cdot$ and C_n with its operation $+$.

- Both D_4 and C_n have constructed a way to construct a third element via the operation if given two elements.
- Note D_4 has a special "do-nothing" element $e \in D_4$ such that $g \cdot e = e \cdot g = g$ for all $g \in D_4$. Similarly, C_n has a special "zero" element $[0]_n \in C_4$ such that

  $$[k]_n + [0]_n = [k+0]_n = [k]_n \qquad \text{and} \qquad [0]_n + [k]_n = [0+k]_n = [k]_n$$

  for all $[k]_n \in C_n$.
- Lastly, for D_4, we saw that each symmetry $g \in D_4$ had an inverse symmetry g^{-1} such that $g \cdot g^{-1} = g^{-1} \cdot g = e$. Similarly, each $[k]_n \in C_n$ has the element $[-k]_n \in C_n$ such that

  $$[k]_n + [-k]_n = [k + -k]_n = [0]_n \qquad \text{and} \qquad [-k]_n + [k]_n = [-k+k]_n = [0]_n.$$

The goal of group theory is to give one generalized theory that is able to talk about both of the above examples in a clean way.

5.1.3 Defining Groups

Having done two extended examples, we will now give the abstract definition of a group. Similar to metric spaces, a group will be a set endowed with a special function satisfying some properties. The function of interest has a special name.

Definition 5.10 (binary operation)**.** A *binary operation* on a set S is a function $S \times S \to S$.

Intuitively a binary operation on S is rule to combine two elements of S into another elements of S. Here are some examples.

Example 5.11. Let D_4 denote the set of symmetries of square. Then we defined the operation $\cdot\colon D_4 \to D_4$ by composition: given $g, h \in D_4$, we defined $g \cdot h$ as the symmetry obtained by applying h and then applying g to the square.

Example 5.12. The function $f\colon \mathbb{N} \times \mathbb{N} \to \mathbb{Z}$ by $f(m,n) \coloneqq m+n$ is a binary operation.

Example 5.13. The function $f\colon \mathbb{N} \times \mathbb{N} \to \mathbb{Q}$ by $f(m,n) \coloneqq m/n$ is a binary operation.

Example 5.14. Let X be a set, and let $S \coloneqq \mathrm{Mor}(X, X)$ denote the set of functions $X \to X$. Then composition is a binary operation $\circ \colon S \times S \to S$: given two functions $f, g \colon X \to X$, we produce a third function $(f \circ g) \colon X \to X$.

Notation for binary operations differs from usual functions though, as we usually write the operation symbol in between the inputs, as in Examples 5.11, 5.12 and 5.14. In fact, when the operation is clear from context the symbol is often omitted all together and "a times b" can be written as just ab.

We are now ready to define groups.

Definition 5.15. A *group* is an ordered pair $(G, \cdot)$ consisting of a set G along with a binary operation $\cdot \colon G \times G \to G$ satisfying the following axioms.

- Associativity: for all $a, b, c \in G$, we have that $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
- Identity: there exists an element $e \in G$ such that, for all $a \in G$, we have that $a \cdot e = e \cdot a = a$. We call e the "identity element."
- Inverse: for each $a \in G$, there exists $b \in G$ such that $a \cdot b = b \cdot a = e$. We call b the "inverse" of a.

We notably do not require the operation $\cdot$ on the group G to satisfy $g \cdot h = h \cdot g$. Such groups are called *commutative* or *abelian*.

In practice, we will write the group as just its underlying set G, with the binary operation left implied. With this convention, most group operators are written "multiplicatively" where the multiplication is denoted $a \cdot b$ or simply ab. We'll adopt this convention when proving general theorems.

Here are some examples of groups. For each, be sure that you are convinced that axioms (a)–(c) of Definition 5.15 are satisfied, though do not feel compelled to write them all out on paper.

Example 5.16. From section 5.1.1, the set D_4 forms a group under the operation $\cdot$.

Example 5.17. From section 5.1.2, the set C_n forms a group under the operation $+$, for any positive integer n.

Example 5.18. The set of integers form a group with operation given by addition $(\mathbb{Z}, +)$. The same holds with the set of rationals $\mathbb{Q}$, the set of reals $\mathbb{R}$, and the set of complex numbers $\mathbb{C}$.

Example 5.19. Let $\mathbb{R}^\times$ denote the nonzero real numbers. Then $\mathbb{R}^\times$ forms a group with operation given by multiplication. The same holds for $\mathbb{C}^\times$.

Example 5.20. The set $\mathrm{GL}_n(\mathbb{C})$ of invertible $n \times n$ matrices with complex coefficients forms a group under matrix multiplication. Similarly, the set $\mathrm{SL}_n(\mathbb{C})$ of invertible $n \times n$ matrices with complex coefficients and determinant 1 forms a group under matrix multiplication.

Exercise 5.21. Which of the above groups are commutative?

Here are a few non-examples.

Non-Example 5.22. The set $\mathbb{Z}$ of integers does not form a group under the operation subtraction $- \colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$. Indeed, $-$ is not even associative: note that

$$(1 - 2) - 3 = -4 \neq 2 = 1 - (2 - 3).$$

Non-Example 5.23. Let $\mathbb{Z}^\times$ denote the set of nonzero integers. Then $\mathbb{Z}^\times$ does not form a group under multiplication. Indeed, the only possible identity element $e \in \mathbb{Z}^\times$ such that $e \cdot a = a \cdot e = a$ is $e = 1$: taking $a = 1$, we see

$$e = e \cdot 1 = 1.$$

However, with identity 1, we don't have inverses: there is no integer $b \in \mathbb{Z}^\times$ such that $2 \cdot b = b \cdot 2 = 1$.

Non-Example 5.24. Let S denote the set of functions $\mathbb{Z} \to \mathbb{Z}$. Then S does not form a group under the operation of composition. Again, the problem is that we do not have inverses. Indeed, suppose for the sake of contradiction that S does form a group. Let $e \in S$ denote the identity. Note that the identity function $i \colon \mathbb{Z} \to \mathbb{Z}$ defined by $i(x) \coloneqq x$ must then have

$$e(x) = e(i(x)) = (e \circ i)(x) = i(x) = x$$

because $e \circ i = i$. Thus, $e = i$. But this implies that each $f \in S$ has some $g \in S$ such that $f \circ g = i$. For example, taking $f \colon \mathbb{Z} \to \mathbb{Z}$ defined by $f(x) \coloneqq 0$ for all $x \in \mathbb{Z}$. Then such a function g implies

$$1 = i(1) = (f \circ g)(1) = f(g(1)) = 0,$$

which is a contradiction.

The previous two non-examples are both a bit technical because we must know what the identity is before we can actually talk intelligently about inverses. The trick we used to extract the identity element from the operation will be used again shortly in Lemma 5.25.

5.1.4 Basic Group Theory

Let's collect a few lemmas about groups.

Lemma 5.25. Let $(G, \cdot)$ be a group. Then the identity element of G is unique.

The above lemma justifies us saying "the" identity of the group.

Proof. For a uniqueness statement like this, we suppose that we have two identities e and e', and we show that $e = e'$. For this, we must use the definition of the identity, which tells us that $e \cdot g = g \cdot e = g$ and $e' \cdot g = g \cdot e' = g$ for all $g \in G$. Well, plugging these into each other, we see

$$e = e \cdot e' = e',$$

which is what we wanted. ∎

Lemma 5.26. Let $(G, \cdot)$ be a group. Given $g \in G$, the inverse of g is unique.

Again, the above lemma justifies us saying "the" inverse of g.

Proof. Suppose that both h and h' are inverses of g, and we show $h = h'$. Letting e denote the identity of G, we thus see $g \cdot h = h \cdot g = e$ and $g \cdot h' = h' \cdot g = e$. Now, the key trick to make these interact is to write $h = h \cdot e$. Then

$$h = h \cdot e = h \cdot (g \cdot h') = (h \cdot g) \cdot h' = e \cdot h' = h'.$$

Notably, we have applied the associativity of $\cdot$ above. ∎

Notation 5.27. Let $(G, \cdot)$ be a group. We will let e_G or sometimes just e denote the identity of G. For each $g \in G$, we will let g^{-1} denote the inverse of G. Extending the negative exponents, we will write $g^{-k} \coloneqq \left(g^{-1}\right)^k$ for any positive integer k.

To explain the exponents, we will say out loud that $g^{a+b} = g^a \cdot g^b$ and $(g^a)^b = g^{ab}$ for any integers $a, b \in \mathbb{Z}$. Checking this rigorously is somewhat annoying, so we will leave it only for the particularly determined.

Here are a few short properties of inverses.

Lemma 5.28. Let $(G, \cdot)$ be a group. For each $g \in G$, we have $\left(g^{-1}\right)^{-1} = g$.

Proof. By definition of g^{-1}, we see $g \cdot g^{-1} = g^{-1} \cdot g = e$. However, these equations also imply that g is the inverse of g^{-1}! In other words, $\left(g^{-1}\right)^{-1} = g$, which is what we wanted. ∎

Here are a few exercises for you to try.

Exercise 5.29. Let $(G, \cdot)$ be a group. For $g, h \in G$, we have $(g \cdot h)^{-1} = h^{-1} \cdot g^{-1}$. Note that the order of the terms has switched!

Exercise 5.30. Let $(G, \cdot)$ be a group, and fix $a, b, c \in G$. Show the following.

(a) If $ab = ac$, then $b = c$.

(b) If $ba = ca$, then $b = c$.

This is called the "cancellation law."

Thinking about groups as symmetries means that we expect the elements to be bijections of some kind. We can rigorize this intuition into the following lemma.

Proposition 5.31. Let $(G, \cdot)$ be a group, and fix $g \in G$. Then the function $\mu_g \colon G \cdot G$ given by $\mu_g(h) \coloneqq g \cdot h$ is a bijection with inverse given by $\mu_{g^{-1}}$.

Proof. To show μ_g is a bijection it suffices to define the inverse function, and we should expect the inverse element to give the inverse function. As such, let g^{-1} denote the inverse of g, and define the function $\mu_{g^{-1}} \colon G \to G$ by $\mu_{g^{-1}}(h) \coloneqq g^{-1} \cdot h$. To check that the functions μ_g and $\mu_{g^{-1}}$ are inverse, for each $h \in H$ we compute

$$\mu_g\left(\mu_{g^{-1}}(h)\right) = \mu_g\left(g^{-1} \cdot h\right) = g \cdot g^{-1} \cdot h = e \cdot h = h,$$

and

$$\mu_{g^{-1}}(\mu_g(h)) = \mu_{g^{-1}}(g \cdot h) = g^{-1} \cdot g \cdot h = e \cdot h = h,$$

which is what we wanted. ∎

Exercise 5.32. Show the left-side analogue of Proposition 5.31: for $g \in G$, show that the function $\mu_g \colon G \cdot G$ given by $\mu_g(h) \coloneqq h \cdot g$ is a bijection.

5.1.5 Subgroups

Sometimes, it feels like there is a group "inside" of another group. For example, the integers $\mathbb{Z}$ forms a group under addition, but $\mathbb{Q}$ also forms a group under addition, and $\mathbb{Z}$ is contained in $\mathbb{Q}$. It will be useful for us to have language to describe this relationship.

Definition 5.33 (subgroup)**.** Let $(G, \cdot)$ be a group. A subset $H \subseteq G$ is a *subgroup* if and only if $(H, \cdot)$ forms a group, where we have restricted $\cdot$ to H appropriately. More precisely, $H \subseteq G$ is a subgroup if and only if the following conditions hold.

- Closure: if $h, h' \in H$, then $h \cdot h' \in H$.
- Identity: the identity e of G has $e \in H$.
- Inverse: for each $h \in H$, the inverse h^{-1} is in H.

Remark 5.34. Note that the identity element of G remains the identity element of H, and the inverses element from G remain the inverse elements of H. To see the first claim, we note that

$$h \cdot e = e \cdot h = h$$

for all $h \in H$ because in fact $h \in G$, and e is the identity of G. A similar argument shows that the inverse of $h \in H$ in the subgroup H is also the inverse element in G.

Let's see some examples.

Example 5.35. Let $(\mathbb{Q}, +)$ denote group of rationals under addition. Then $\mathbb{Z} \subseteq \mathbb{Q}$ is a subgroup. Here are our checks.

- Identity: the identity of $(\mathbb{Q}, +)$ is 0, which is an integer.
- Closure: if $a, b \in \mathbb{Z}$, then $a + b$ is also an integer.
- Inverse: for each $a \in \mathbb{Z}$, the inverse for addition is $-a$, which is an integer.

Exercise 5.36. Show that $\mathbb{Q}$ is a subgroup of $(\mathbb{R}, +)$.

Example 5.37. Consider the group D_4 of symmetries of the square, with operation given by $\cdot$. Then the set $R \coloneqq \{e, r, r^2, r^3\}$ is a subgroup. Here are our checks.

- Identity: the identity $e \in D_4$ is in R by construction.
- Closure: given two $r^a, r^b \in R$ where $a, b \in \{0, 1, 2, 3\}$, we note $r^a \cdot r^b = r^{a+b}$ is in R after taking $a + b \pmod 4$. For example,
$$r^2 \cdot r^3 = r^5 = r \cdot r^4 = r \cdot e = r.$$
Convince yourself that this works in general.
- Inverse: for each $r^a \in R$ for $a \in \{0, 1, 2, 3\}$, we note that $r^{4-a} \in R$ still, and
$$r^a \cdot r^{4-a} = r^4 = e.$$

Exercise 5.38. Show that $\{e, s\}$ is a subgroup of $(D_4, \cdot)$.

Exercise 5.39. Show that $\{e, s, r^2, sr^2\}$ is a subgroup of $(D_4, \cdot)$.

Here are some non-examples.

Non-Example 5.40. The set of positive integers $\mathbb{Z}^+$ is not a subgroup of $(\mathbb{Z}, +)$. Indeed, the identity element $0 \in \mathbb{Z}$ is not a positive integer.

Non-Example 5.41. The set of nonnegative integers $\mathbb{Z}_{\geq 0}$ is not a subgroup of $(\mathbb{Z}, +)$. Indeed, even though $3 \in \mathbb{Z}_{\geq 0}$, the inverse -3 is not in $\mathbb{Z}_{\geq 0}$.

Non-Example 5.42. Let $(D_4, \cdot)$ denote the group of symmetries of the square. Then the subset $S \coloneqq \{e, s, rs, r^2s, r^3s\}$ is not a subgroup. Indeed, $s \in S$ and $rs \in S$, but

$$rs \cdot s = r \cdot s^2 = r \cdot e = r$$

is not in S.

Here are a few more abstract examples.

Proposition 5.43. Let $(G, \cdot)$ be a group, and let $g \in G$ be an element. Then the subset

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

is a subgroup of G. Here, $g^{-k} \coloneqq \left(g^{-1}\right)^k$ for any positive integer k.

Proof. We check our conditions by hand. The main content here is that we need to prove our exponent rules, but we will not be very formal about it. Feel free to ignore the footnotes.

- Identity: note that $e = g^0$ is in $\langle g \rangle$ by definition.
- Closure: pick up elements $g^k, g^\ell \in \langle g \rangle$. Then $g^k \cdot g^\ell = g^{k+\ell}$ is in $\langle g \rangle$.[2]
- Inverse: suppose $g^k \in \langle g \rangle$. Then $\left(g^k\right)^{-1} = g^{-k} \in \langle g \rangle$.[3]

■

Proposition 5.44. Let $(G, \cdot)$ be a group. Then the subset

$$Z(G) \coloneqq \{g \in G : g \cdot h = h \cdot g \text{ for all } h \in G\}$$

is a subgroup of G.

Proof. We check our conditions by hand.

- Identity: note that $e \cdot h = h = h \cdot e$ for all $h \in G$. Thus, $e \in Z(G)$.
- Closure: suppose that $g, g' \in Z(G)$. We would like to show that $g \cdot g' \in Z(G)$. Well, for each $h \in G$, we want to show
$$(g \cdot g') \cdot h \stackrel{?}{=} h \cdot (g \cdot g').$$
For this, we associate and use the fact that $g, g' \in Z(G)$, rearranging as
$$\begin{aligned} g \cdot g' \cdot h &= g \cdot (g' \cdot h) \\ &= g \cdot (h \cdot g') \\ &= (g \cdot h) \cdot g' \\ &= (h \cdot g) \cdot g' \\ &= h \cdot (g \cdot g'), \end{aligned}$$

[2] We have not technically proven that $g^k \cdot g^\ell = g^{k+\ell}$ for any $k, \ell \in \mathbb{Z}$. This is just a lot of casework. For $k, \ell \geq 0$, there is nothing to say. If $k, \ell < 0$, then $g^k \cdot g^\ell = \left(g^{-1}\right)^{-k} \cdot \left(g^{-1}\right)^{-\ell} = \left(g^{-1}\right)^{k+\ell} = g^{-k-\ell}$. Lastly, if one is nonnegative and the other negative, say $k \geq 0$ and $\ell < 0$. If $k \geq -\ell$, then $g^k \cdot g^\ell = g^{k-(-\ell)} \cdot g^{-\ell} \cdot \left(g^{-1}\right)^{-\ell} = g^k$. A similar argument works in the case where $k \leq -\ell$.

[3] Again, this equality requires an argument. If $k \geq 0$, then $\left(g^k\right)^{-1} = (g \cdot \ldots \cdot g)^{-1} = \left(g^{-1} \cdot \ldots \cdot {g^{-1}}_k\right) = \left(g^{-1}\right)^k = g^{-k}$, where each of the iterated multiplications happens k times. If $k < 0$, then note $g^k = \left(g^{-1}\right)^{-k}$ by definition, so $-k \geq 0$ implies $\left(g^k\right)^{-1} = \left(g^{-1}\right)^k$ by prior work. Then $\left(g^{-1}\right)^k = \left(\left(g^{-1}\right)^{-1}\right)^{-k}$ by definition, which is g^k by Lemma 5.28.

which is what we wanted.

- Inverse: suppose that $g \in Z(G)$. We would like to show that $g^{-1} \in Z(G)$. Well, for any $h \in H$, we want to show
$$g^{-1} \cdot h \stackrel{?}{=} h \cdot g^{-1}. \tag{5.1}$$
We should use the fact that $h \cdot g = g \cdot h$, so we take this equation and multiply both sides by g^{-1}, giving
$$g^{-1} \cdot (h \cdot g) \cdot g^{-1} = g^{-1} \cdot (g \cdot h) \cdot g^{-1}.$$
Simplifying both sides of the above equation yields (5.1). ■

Exercise 5.45. Let $(D_4, \cdot)$ be the symmetries of the square. Show that $Z(D_4) = \{e, r^2\}$.

Exercise 5.46. Let $(G, \cdot)$ be a group. Show that the subsets $\{e\}$ and G are a subgroup of G.

In general, it can be an interesting question to classify the subgroups of a particular group. As an example, let's classify the subgroups of $(\mathbb{Z}, +)$.

Exercise 5.47. Let H be a subgroup of $(\mathbb{Z}, +)$. If H contains 5 and 3, then show that H contains 2. In fact, show that H contains 1.

Proposition 5.48. Let $(\mathbb{Z}, +)$ denote the group of integers.

(a) For each $d \in \mathbb{Z}$, the subset $d\mathbb{Z} \coloneqq \{dk : k \in \mathbb{Z}\}$ is a subgroup of $\mathbb{Z}$. Here, $d\mathbb{Z}$ is the set of multiples of d.

(b) If $H \subseteq \mathbb{Z}$ is a subgroup, then there exists a nonnegative integer $d \in \mathbb{Z}$ such that $H = d\mathbb{Z}$.

Proof. We show the parts separately.

(a) We run the checks directly.

- Identity: we see $0 = d \cdot 0$ lives in $d\mathbb{Z}$.
- Closure: given $da, db \in d\mathbb{Z}$ where $a, b \in \mathbb{Z}$, we see that $da + db = d(a + b)$ is again an element of $d\mathbb{Z}$.
- Inverse: for any $dk \in \mathbb{Z}$, we see that its inverse is $-(dk) = d \cdot (-k)$, which is again in $d\mathbb{Z}$.

(b) This proof requires us to be a little careful because we must account for the subgroup $\{0\}$ of $\mathbb{Z}$. Indeed, if H contains no nonzero integers, then we note that H must certainly contain the identity 0, so $H = \{0\}$. Thus, $H = 0\mathbb{Z} = \{0k : k \in \mathbb{Z}\}$.

Otherwise, H contains a nonzero integer h. Now, the main difficulty in this proof is finding the element d. Note that either h or $-h$ is positive, so H also contains a positive integer. By the well-ordering principle, H thus contains a least positive integer d.

Thus, we claim that $H = d\mathbb{Z}$. We have two inclusions to show. We begin by showing $d\mathbb{Z} \subseteq H$. For each positive integer k, we see that
$$dk = \underbrace{d + \cdots + d}_{k}$$
will live in H as well because H is closed under $+$.[4] Additionally, $d \cdot 0 = 0$ lives in H. Lastly, for each negative integer k, we note that $-dk = d \cdot -k$. But then $-k$ is a positive integer, so $d \cdot -k$ lives in H. However, H is also closed under taking inverses, so $-dk \in H$ forces $dk \in H$.

[4] More formally, one can show this claim by induction, but we won't bother.

Lastly, we show that $H \subseteq d\mathbb{Z}$. This requires using Theorem 5.7. Pick up some $h \in H$. By Theorem 5.7, there is are integers $q, r \in \mathbb{Z}$ such that

$$h = dq + r,$$

where $0 \leq r < d$. We would like to show that $r = 0$, for this would imply $h = dq \in d\mathbb{Z}$.

The fact that $r = 0$ will follow from the minimality of d—note we have used this minimality yet! Indeed, note $r \in H$: indeed, $dq \in d\mathbb{Z}$ is in H, so $-dq \in H$, so $r = h + -dq$ is also in H. Further, $r < d$ by definition of d, but d is the least positive integer in H, so r cannot be a positive integer! Because $r \geq 0$, we conclude that $r = 0$ is forced. This completes the proof. ■

5.1.6 Problems

Problem 5.1. Let X be a set, and let $\mathrm{Sym}(X)$ denote the set of bijections $X \to X$. Show that this forms a group under function composition $\circ$. The group $(\mathrm{Sym}(X), \circ)$ is called the "symmetric group" of X. In the case where $X = \{1, 2, \ldots, n\}$, we might write $S_n \coloneqq \mathrm{Sym}(X)$.

Problem 5.2. Let X be a set. For two subsets $A, B \subseteq X$, recall the definition of the symmetric differences

$$A \triangle B \coloneqq (A \setminus B) \cup (B \setminus A).$$

Show that the operation $\triangle$ on the set of all subsets $\mathcal{P}(X)$ is a group.

Problem 5.3. Determine if the following are groups. No justification is required.

(a) The integers $\mathbb{Z}$ where the operation is subtraction.

(b) The integers $\mathbb{Z}$ where the operation is multiplication.

(c) The nonzero integers $\mathbb{Z} \setminus \{0\}$ where the operation is multiplication.

(d) The positive rational numbers $\mathbb{Q}^+$ where the operation is addition.

(e) The positive rational numbers $\mathbb{Q}^+$ where the operation is multiplication.

(f) The set $\mathbb{Z} \times \mathbb{Z}$ of ordered pairs of integers, where the operation is given by $(a, b) \cdot (c, d) \coloneqq (a + b, c + d)$.

Problem 5.4. Let $(G, \cdot)$ be a finite group, where $G = \{g_1, g_2, \ldots, g_n\}$. Further, suppose that the group is abelian. Define $p \coloneqq g_1 \cdot g_2 \cdot \ldots \cdot g_n$. Show that $p^2 = e$, where e is the identity element of G.

Problem 5.5. Let n be a positive integer.

(a) If $[a]_n = [a']_n$ and $[b]_n = [b']_n$ for integers $a, a', b, b' \in \mathbb{Z}$, then $[a \cdot a']_n = [b \cdot b']_n$. Conclude that the binary operation $\cdot : C_n \times C_n \to C_n$ given by

$$[a]_n \cdot [b]_n \coloneqq [a \cdot b]_n$$

is well-defined.

(b) Set $n = 5$. Does C_n form a group under the operation $\cdot : C_n \times C_n \to C_n$?

Problem 5.6. Let n be a positive integer.

(a) Suppose that $d \in \mathbb{Z}$. Show that $\{[dk]_n : k \in \mathbb{Z}\}$ is a subgroup of C_n.

(b) Suppose that $H \subseteq C_n$ is a subgroup. Show that

$$\{k \in \mathbb{Z} : [k]_n \in H\}$$

is a subgroup of $\mathbb{Z}$. Conclude that there exists an integer $d \in \mathbb{Z}$ such that $H = \{[dk]_n : k \in \mathbb{Z}\}$.

Problem 5.7. Let $(G, \cdot)$ be a group. Given a subset $S \subseteq G$, define the *centralizer* by

$$C_G(S) \coloneqq \{g \in G : g \cdot s = s \cdot g \text{ for all } s \in S\}.$$

For example, $C_G(\{e\}) = G$, where $e \in G$ is the identity element.

(a) Show that $C_G(S)$ is a subgroup of G.

(b) Given subsets $S, T \subseteq G$, show that $S \subseteq C_G(T)$ implies $T \subseteq C_G(S)$.

(c) Given subsets $S, T \subseteq G$, show that $S \subseteq T$ implies $C_G(T) \subseteq C_G(S)$.

(d) Show that $S \subseteq C_G(C_G(S))$.

(e) Use the above parts to show that $C_G(C_G(C_G(S))) \subseteq C_G(S)$ and $C_G(S) \subseteq C_G(C_G(C_G(S)))$. Conclude that $C_G(C_G(C_G(S))) = C_G(S)$.

5.2 Week 13: Cosets

In this section, we will more closely examine the way that subgroups relate to the larger group. Along the way, we will show Lagrange's theorem (Theorem 5.72) and discuss quotient groups.

5.2.1 Cosets

Given a positive integer n, the subgroup $n\mathbb{Z}$ sits inside the group $(\mathbb{Z}, +)$. This viewpoint allows us to understand the construction of C_n better: as in Remark 5.5, we may think about the element $[a]_n \in C_n$ as

$$[a]_n = \{a + nk : k \in \mathbb{Z}\} = a + n\mathbb{Z}.$$

This motivates the following definition.

Definition 5.49 (coset)**.** Let $(G, \cdot)$ be a group and $H \subseteq G$ a subgroup. Fix $g \in G$.

- The *left* coset of g is $g \cdot H \coloneqq \{g \cdot h : h \in H\}$. The set of all left cosets is denoted G/H.
- The *right coset* of g is $H \cdot g \coloneqq \{h \cdot g : h \in H\}$. The set of all right cosets is denoted $H\backslash G$.

The "left" and "right" refers to where the g is with respect to H.

It can be difficult to read the difference between $H\backslash G$ (right cosets) and $H \setminus G$ (set difference), and it is essentially for this reason that we will try to reason with left cosets instead of right cosets when we can. However, confusion will not usually arise because there is no reason to subtract the full group G from a subgroup H because this is just $H \setminus G = \varnothing$.

Notation 5.50. As explained at the start of this subsection, we may now write C_n as $\mathbb{Z}/n\mathbb{Z}$. We might write the equivalence classes $[k]_n$ as $k + n\mathbb{Z}$.

Warning 5.51. At this point in the notes, G/H and $H\backslash G$ have been constructed as sets, not groups. Later on we will see that G/H can sometimes be turned into a group, like $\mathbb{Z}/n\mathbb{Z}$.

Exercise 5.52. Let n be a positive integer and k an integer. Verify that $k + n\mathbb{Z} = n\mathbb{Z} + k$ as sets.

Let's see some more examples.

Example 5.53. Let $(D_4, \cdot)$ be the symmetries of the square. The set $S \coloneqq \{e, s\}$ is a subgroup of D_4.

- The left cosets are

$$e \cdot S = \{e, s\}, \quad r \cdot S = \{r, r \cdot s\}, \quad r^2 \cdot S = \left\{r^2, r^2 \cdot s\right\}, \quad r^3 \cdot S = \left\{r^3, r^3 \cdot s\right\}.$$

  There are no more cosets because we have alrady represented each element of D_4 above.

- The right cosets are

$$S \cdot e = \{e, s\}, \quad S \cdot r = \left\{r, r^3 \cdot s\right\}, \quad S \cdot r^2 = \left\{r^2, r^2 \cdot s\right\}, \quad S \cdot r^3 = \left\{r^3, r \cdot s\right\}.$$

  (Why there are no more cosets?)

Notably, $r \cdot S \neq S \cdot r$, so the distinction between left and right cosets is necessary.

Exercise 5.54. Let $(D_4, \cdot)$ be the symmetries of the square. Compute the left and right cosets of the subgroup $R \coloneqq \left\{e, r, r,^2, r^3\right\}$.

Example 5.55. The group $(\mathbb{Z}/6\mathbb{Z}, +)$ has a subgroup $3\mathbb{Z}/6\mathbb{Z} = \{[0]_6, [3]_6\}$.

- The left cosets are

$$[0]_6 + 3\mathbb{Z}/6\mathbb{Z} = \{[0]_6, [3]_6\}, \quad [1]_6 + 3\mathbb{Z}/6\mathbb{Z} = \{[1]_6, [4]_6\}, \quad [2]_6 + 3\mathbb{Z}/6\mathbb{Z} = \{[2]_6, [5]_6\}.$$

- The right cosets are

$$3\mathbb{Z}/6\mathbb{Z} + [0]_6 = \{[0]_6, [3]_6\}, \quad 3\mathbb{Z}/6\mathbb{Z} + [1]_6 = \{[1]_6, [4]_6\}, \quad 3\mathbb{Z}/6\mathbb{Z} + [2]_6 = \{[2]_6, [5]_6\}.$$

Note that the cosets are equal this time. Also note that $(\mathbb{Z}/6\mathbb{Z})/(3\mathbb{Z}/6\mathbb{Z})$ resembles $\mathbb{Z}/3\mathbb{Z}$.

Example 5.56. Consider the subgroup $\mathbb{Z}$ of the group $(\mathbb{R}, +)$. Then we can consider the left cosets $\mathbb{R}/\mathbb{Z}$, which are the sets $a + \mathbb{Z}$, where $a + \mathbb{Z} = a' + \mathbb{Z}$ if and only if $a - a' \in \mathbb{Z}$. Convince yourself that each coset in $\mathbb{R}/\mathbb{Z}$ has a unique representative of the form $a + \mathbb{Z}$ such that $a \in [0, 1)$.

Here are a few more abstract examples.

Example 5.57. Let $(G, \cdot)$ be a group. Then $H = \{e\}$ is a subgroup by Exercise 5.46. The left coset of some $g \in G$ is

$$g \cdot H = \{g \cdot h : h \in H\} = \{g \cdot e\} = \{g\}.$$

Similarly, $H \cdot g = \{g\}$.

Example 5.58. Let $(G, \cdot)$ be a group. Then $H = G$ is a subgroup by Exercise 5.46. We claim that the left coset of some $g \in G$ is $g \cdot G = G$. For one, note that

$$e \cdot G = \{e \cdot g : g \in G\} = \{g : g \in G\} = G.$$

However, this implies that $g \in e \cdot G$, so $[g] = [e]$, where we are using the equivalence relation of Exercise 5.60. It follows $g \cdot H = e \cdot H$.

5.2.2 Cosets by Equivalence Relation

Recall that we technically defined $C_n = \mathbb{Z}/n\mathbb{Z}$ be equivalence relation. This proves to be a fruitful way to think about cosets, so we generalize the notion here. Again, the point is to focus on the subgroup $n\mathbb{Z} \subseteq \mathbb{Z}$. The equivalence relation $a \equiv b \pmod{n}$ is equivalent to $n \mid a-b$, which we now see is equivalent to $a-b \in n\mathbb{Z}$. In total, we have

$$a \equiv b \pmod{n} \iff a - b \in n\mathbb{Z}.$$

Thus, trying to generalize the equivalence relation of $\mathbb{Z}/n\mathbb{Z}$, we have the following lemma.

Lemma 5.59. Let $(G, \cdot)$ be a group and $H \subseteq G$ a subgroup. Define the relation $\sim$ on G by $g \sim h$ if and only if $g \cdot h^{-1} \in H$. Then $\sim$ is an equivalence relation.

Proof. We have the following checks. Fix $g, h, k \in G$.

- Reflexive: we would like to show $g \sim g$, or $g \cdot g^{-1} \in H$. However, $g \cdot g^{-1} = e$, and $e \in H$, so we are done.
- Symmetric: if $g \sim h$, we would like to show $h \sim g$. Unwinding the definition of $\sim$, we are given that $g \cdot h^{-1} \in H$, and we want to show $h \cdot g^{-1} \in H$. However, we see

  $$\left(g \cdot h^{-1}\right)^{-1} = \left(h^{-1}\right)^{-1} \cdot g^{-1} = h \cdot g^{-1}.$$

  Here, we have used Exercise 5.29 in the first equality and Lemma 5.28 in the second one. Thus, $h \cdot g^{-1} \in H$ because H contains inverses.
- Transitive: given $g \sim h$ and $h \sim k$, we want to show $g \sim k$. Unwinding $\sim$, we are given $g \cdot h^{-1} \in H$ and $h \cdot k^{-1} \in H$, and we want to show $g \cdot k^{-1}$. Well, we the product

  $$\left(g \cdot h^{-1}\right) \cdot \left(h \cdot k^{-1}\right) = g \cdot \left(h^{-1} \cdot h\right) \cdot k^{-1} = g \cdot e \cdot k^{-1} = g \cdot k^{-1}$$

  must be in H because H is a subgroup, so we are done. ∎

Exercise 5.60. Let $(G, \cdot)$ be a group and $H \subseteq G$ a subgroup. Define the relation $\sim$ on G by $g \sim h$ if and only if $g^{-1} \cdot h \in H$. Show that $\sim$ is an equivalence relation. This does not follow immediately from Lemma 5.59.

As with C_n, the equivalence classes of the above relations are what interest us.

Lemma 5.61. Let $(G, \cdot)$ be a group and $H \subseteq G$ a subgroup. Define the equivalence relation $\sim$ from Lemma 5.59. Then the equivalence class represented by $g \in G$ is the set

$$H \cdot g \coloneqq \{h \cdot g : h \in H\}.$$

Proof. For now, let $[g]$ denote the equivalence class represented by g. Very quickly, we unwind the definition of $[g]$: note that $g' \in [g]$ if and only if $g' \sim g$, which is in turn equivalent to $g' \cdot g^{-1} \in H$. Now, we have two inclusions to show.

- We show $[g] \subseteq H \cdot g$. Note that $g' \in [g]$ if and only if $g' \cdot g^{-1} \in H$, as discussed above. Thus, we define $h \coloneqq g' \cdot g^{-1}$ and see that
$$h \cdot g = g' \cdot g^{-1} \cdot g = g' \cdot e = g',$$
so $g' \in H \cdot g$ follows.
- We show $H \cdot g \subseteq [g]$. Indeed, suppose we have some $g' = h \cdot g$ in $H \cdot g$. To show $g' \cdot g^{-1} \in H$, we compute
$$g' \cdot g^{-1} = h \cdot g \cdot g^{-1} = h \cdot e = h,$$
finishing. ■

Exercise 5.62. Let $(G, \cdot)$ be a group and $H \subseteq G$ a subgroup. Define the equivalence relation $\sim$ from Exercise 5.60. Show that equivalence class represented by $g \in G$ is the set
$$g \cdot H \coloneqq \{g \cdot h : h \in H\}.$$

5.2.3 How to Think About Cosets

The following proposition explains how to think about cosets. Being able to interface with all the equivalent conditions is important.

Proposition 5.63. Let $(G, \cdot)$ be a group and $H \subseteq G$ be a subgroup. For $g_1, g_2 \in G$, the following are equivalent.

(a) $g_1^{-1} \cdot g_2 \in H$.

(b) $g_2^{-1} \cdot g_1 \in H$.

(c) $g_1 \in g_2 \cdot H$.

(d) $g_2 \in g_1 \cdot H$.

(e) $g_1 \cdot H \subseteq g_2 \cdot H$.

(f) $g_2 \cdot H \subseteq g_1 \cdot H$.

(g) $g_1 H = g_2 H$.

We will give two proofs of this result: first, we will show the equivalences by direct computation of cosets. Second, we will provide a proof using Exercise 5.62 in order to showcase its power.

Proof by computation. We proceed in steps.

1. We show that (a) and (b) are equivalent. Note $g_1^{-1} \cdot g_2 \in H$ implies that $g_2^{-1} \cdot g_1 = \left(g_1^{-1} \cdot g_2\right)^{-1} \in H$ because H is a subgroup. Thus, (a) implies (b), and switching the roles of g_1 and g_2 shows that (b) implies (a).
2. We show that (a) implies (c); switching the roles of g_1 and g_2 will show that (b) implies (d). Well, $h \coloneqq g_1^{-1} \cdot g_2 \in H$ implies
$$g_1 = g_2 \cdot h^{-1} \in g_2 \cdot H.$$
3. We show that (c) implies (e); switching the roles of g_1 and g_2 will show that (d) implies (f). Well, $g_1 \in g_2 \cdot H$ implies that we may write $g_1 = g_2 \cdot h_0$ for some $h_0 \in H$. Thus,
$$g_1 \cdot H = \{g_2 \cdot (h_0 \cdot h) : h \in H\} \subseteq \{g_2 \cdot h : h \in H\} = g_2 \cdot H$$
because $h_0 \cdot h \in H$ for any $h \in H$.

4. We show that (e) implies (a); switching the roles of g_1 and g_2 will show (f) implies (b). Well, $g_1 \cdot H \subseteq g_2 \cdot H$ implies that $g_1 \cdot e \in g_1 \cdot H$ lives in $g_2 \cdot H$, so we may write $g_1 = g_2 \cdot h$ for some $h \in H$. Thus, $g_1^{-1} \cdot g_2 = h^{-1} \in H$.
5. The above work shows that (a)–(f) are all equivalent. It remains to show that (a)–(f) are equivalent to (g). In one direction, note that (g) implies (e). In the other direction, note that (e) implies (f) by the above work, so (e) implies (e) and (f), which is (g). ■

Proof by equivalence relation. We will be a little terser in this proof. Indeed, this proof contains no "hard work." Instead, we will essentially reduce everything to the equivalence relation $\sim$ of Exercise 5.60 so that the sets $g \cdot H$ are the equivalence classes by Exercise 5.62.

For example, because $g_1 \sim g_2$ is equivalent to $g_2 \sim g_1$, we see that (a) and (b) are equivalent. Furthermore, because cosets are the equivalence classes, we see that $g_1 \in g_2 \cdot H$ is equivalent to $g_1^{-1} \cdot g_2 \in H$, establishing (a) and (c) are equivalent. Similarly, (b) and (d) are equivalent. Thus, all of (a)–(d) are equivalent.

Next, we show that (a)–(d) are equivalent to (e). On one hand, given (e), we note $g_1 \in g_1 \cdot H$, so $g_1 \cdot H \subseteq g_2 \cdot H$. Conversely, suppose $g_1 \in g_2 \cdot H$, and we show $g_1 \cdot H \subseteq g_2 \cdot H$. Using the equivalence relation, we note $g \in g_1 \cdot H$ is equivalent to $g \sim g_1$. However, we know $g_1 \sim g_2$, so it follows $g \sim g_2$ as well. Thus, $g \in g_1 \cdot H$ implies $g \in g_2 \cdot H$, which is what we wanted.

A similar argument shows that (a)–(d) are equivalent to (f). It remains to show that the conditions (a)–(f) are equivalent to (g). Well, (g) is equivalent to (e) and (f) combined. So in one direction, (g) certainly implies (e). In the other direction, we know (e) implies (f), and then (e) and (f) implies (g). This completes the proof. ■

Remark 5.64. The way to remember Proposition 5.63 is to imagine trying to manipulate the expression $g_1 \cdot H = g_2 \cdot H$ symbolically. For example, from $g_1 \cdot H = g_2 \cdot H$, we expect to have $\left(g_2^{-1} \cdot g_1\right) \cdot H = e \cdot H$ by multiplying on the left by g_2^{-1}. This then should imply $g_2^{-1} \cdot g_1 \in H$, as we expect. We will partially rigorize this kind of thinking in Lemma 5.71.

Exercise 5.65. Let $(G, \cdot)$ be a group and $H \subseteq G$ a subgroup. Show directly from the definition $g \cdot H \coloneqq \{g \cdot h : g \in G\}$ that $g_1 \in g_2 \cdot H$ implies $g_1 \cdot H \subseteq g_2 \cdot H$.

Exercise 5.66. State and prove the following right-coset version of Proposition 5.63: let $(G, \cdot)$ be a group and $H \subseteq G$ a subgroup. For $g_1, g_2 \in G$, the following are equivalent.

(a) $g_1 \cdot g_2^{-1} \in H$.

(b) $g_1 \in H \cdot g_2$.

(c) $H \cdot g_1 = H \cdot g_2$.

Feel free to add more equivalent conditions.

Let's see an example of how to use these conditions.

Corollary 5.67. Let $(G, \cdot)$ be a group and $H \subseteq G$ a subgroup. For $g_1, g_2 \in G$, the following are equivalent.

- $(g_1 \cdot H) \cap (g_2 \cdot H)$ is nonempty.
- $g_1 \cdot H = g_2 \cdot H$.

Proof. In the easier direction, if $g_1 \cdot H = g_2 \cdot H$, then $(g_1 \cdot H) \cap (g_2 \cdot H)$ is nonempty; for example, $g_1 \in g_1 \cdot H$, so $g_1 \in g_1 \cdot H \cap g_2 \cdot H$.

The converse requires some attention. Suppose $(g_1 \cdot H) \cap (g_2 \cdot H)$ is nonempty. Then there is some $g \in G$ such that $g \in g_1 \cdot H$ and $g \in g_2 \cdot H$. But by Proposition 5.63, this implies $g \cdot H = g_1 \cdot H$ and $g \cdot H = g_2 \cdot H$, so $g_1 \cdot H = g_2 \cdot H$ follows. ■

Corollary 5.68. Let $(G, \cdot)$ be a group and $H \subseteq G$ a subgroup. For $g_1, g_2 \in G$, if $g_1 \cdot H = g_2 \cdot H$, then $H \cdot g_1^{-1} = H \cdot g_2^{-1}$.

Proof. Note $g_1 \cdot H = g_2 \cdot H$ implies $g_1^{-1} \cdot g_2 \in H$ by Proposition 5.63. However, $g_2 = \left(g_2^{-1}\right)^{-1}$, so $g_1^{-1} \cdot \left(g_2^{-1}\right)^{-1} \in H$. By Exercise 5.66, this implies $H \cdot g_1^{-1} = H \cdot g_2^{-1}$, which is what we wanted. ■

Exercise 5.69. Show Corollary 5.68 without using any results of this subsection.

Exercise 5.70. Find an example of a group $(G, \cdot)$ and subgroup $H \subseteq G$ such that there are elements $g_1, g_2 \in G$ with $g_1 \cdot H = g_2 \cdot H$ and $H \cdot g_1 \neq H \cdot g_2$.

5.2.4 Lagrange's Theorem

Examples 5.53, 5.55 and 5.57 all have the common feature that the cosets seem to all have the same size. (Even Example 5.56 has this property if one considers cardinality.) This is not a coincidence, as we now explain.

Lemma 5.71. Let $(G, \cdot)$ be a group and $H \subseteq G$ a subgroup. For each $g \in G$, define the function $\mu_g \colon G \to G$ by $\mu_g(g') \coloneqq g \cdot g'$. Then $\mu_g(g' \cdot H) = (g \cdot g') \cdot H$.

Proof. This is essentially the associative law: intuitively, we should read this result as $g \cdot (g' \cdot H) = (g \cdot g') \cdot H$, but we have not defined how to multiply $g \in G$ by the coset $g' \cdot H$. Anyway, we compute

$$\begin{aligned} \mu_g(g' \cdot H) &= \{\mu_g(x) : x \in g' \cdot H\} \\ &= \{\mu_g(g' \cdot h) : h \in H\} \\ &= \{g \cdot g' \cdot h : h \in H\} \\ &= (g \cdot g') \cdot H, \end{aligned}$$

which is what we wanted. ■

Theorem 5.72 (Lagrange)**.** Let $(G, \cdot)$ be a group and $H \subseteq G$ a subgroup. Then all cosets in G/H have the same cardinality.

Proof. For psychological reasons, we note that it suffices to show that each coset $g \cdot H$ has the same cardinality as $e \cdot H$. Indeed, this will imply that any two cosets $g \cdot H$ and $g' \cdot H$ have the same cardinality as $e \cdot H$ and thus have the same cardinality.

Well, define the function $\mu_g \colon G \to G$ as in Proposition 5.31 and note that μ_g actually restricts to a function $\mu_g \colon (e \cdot H) \to (g \cdot H)$ by Lemma 5.71. (Indeed, $g \cdot e = g$.) We claim that this restriction is bijective, which will complete the proof.

- We show $\mu_g \colon (e \cdot H) \to (g \cdot H)$ is surjective. Indeed, this is exactly Lemma 5.71.
- We show $\mu_g \colon (e \cdot H) \to (g \cdot H)$ is injective. Well, the full function $\mu_g \colon G \to G$ is already injective, so $\mu_g(g_1) = \mu_g(g_2)$ implies $g_1 = g_2$ for any $g_1, g_2 \in G$. Thus, $\mu_g(h_1) = \mu_g(h_2)$ implies $h_1 = h_2$ for any $h_1, h_2 \in e \cdot H$, which is what we wanted. ■

Theorem 5.72 has the following surprising corollary.

Corollary 5.73. Let $(G, \cdot)$ be a finite group and $H \subseteq G$ a subgroup. Then

$$|G| = |G/H| \cdot |H|.$$

In particular, $|H|$ divides $|G|$.

Proof. The idea here is that the cosets form a partition of G, which has $|G|$ elements. But there are $|G/H|$ total cosets, each of size $|H|$, which will give the result.

Let's be a bit more explicit. Enumerate the cosets in G/H as $\{g_1 \cdot H, g_2 \cdot H, \dots, g_n \cdot H\}$, where $n \coloneqq |G/H|$. Because cosets are equivalence classes (by Exercise 5.62), we see that each element of G lives in exactly one of these cosets. Taking cardinalities, it follows that

$$|G| = \sum_{i=1}^{n} |g_i \cdot H|.$$

However, by Theorem 5.72, we see $|g_i \cdot H| = |H|$, so actually

$$|G| = \sum_{i=1}^{n} |H| = |G/H| \cdot |H|,$$

which is what we wanted. ∎

Corollary 5.73 is amazing. It is essentially the first time in group theory that we really see the structure of a group impact what a group can possibly be. Of course, we have been seeing this all along in our examples of cosets at the start of this section.

Example 5.74. Let $(G, \cdot)$ be a finite group. We saw in Proposition 5.44 that

$$Z(G) \coloneqq \{g \in G : g \cdot h = g \cdot g \text{ for all } h \in G\}$$

is a subgroup of G. Thus, $|Z(G)|$ divides $|G|$. This is not at all obvious a priori!

Here is a more involved consequence.

Proposition 5.75. Let $(G, \cdot)$ be a finite group. For any $g \in G$, we have $g^{|G|} = e$, where e is the identity of G. In fact, the smallest positive integer k such that $g^k = e$ divides $|G|$.

Before jumping into the proof, we note that Proposition 5.75 is "sharp" in the following sense: there do exist groups G such that $|G|$ is the smallest positive integer n such that $g^n = e$.

Example 5.76. Fix a positive integer n and consider the group $(\mathbb{Z}/n\mathbb{Z}, +)$. Suppose that n_0 is a positive integer such that $n_0 \cdot [k]_n = [0]_n$ for any $[k]_n \in \mathbb{Z}/n\mathbb{Z}$; we claim that $n_0 \geq n$. Well, we see that

$$[n_0]_n = n_0 \cdot [1]_n = [0]_n,$$

so n divides n_0. To finish, we may write $n_0 = nq$ for some integer q, so because $n_0, n > 0$, we see that $q \geq 1$, so $n_0 \geq n$ follows.

Anyway, let's move on with the proof.

Proof of Proposition 5.75. The main character of this proof is the subgroup

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

of G defined in Proposition 5.43. For clarity, we proceed in steps.

1. We show that there exists some positive integer n such that $g^n = e$. Indeed, $\langle g \rangle$ is a finite set because it is a subset of the finite set G, so because $\mathbb{Z}$ is infinite, there must integers m and n such that $g^m = g^n$. Switching n and m if necessary, we may assume that $m > n$, so we see
$$g^{m-n} = g^m \cdot (g^n)^{-1} = e.$$
Thus, $m - n > 0$ is the desired positive integer.

2. Let n be any positive integer such that $g^n = e$. We claim that
$$\langle g \rangle \stackrel{?}{=} \left\{e, g, g^2, \ldots, g^{k-1}\right\}$$
and that all elements on the right-hand side are distinct. To begin, note $\left\{e, g, g^2, \ldots, g^{k-1}\right\} \subseteq \langle g \rangle$ by definition of $\langle g \rangle$. In the other direction, for any $g^m \in \langle g \rangle$ where $m \in \mathbb{Z}$ is an integer, use Theorem 5.7 to write $m = nq + r$ where $0 \leq r < n$. Then
$$g^m = g^{nq} \cdot g^r = (g^n)^q \cdot g^r = e^q \cdot g^r = e \cdot g^r = g^r \in \left\{e, g, g^2, \ldots, g^{n-1}\right\}.$$

3. Now, let k be the least positive integer such that $g^k = e$. We claim that $|\langle g \rangle| = k$. By the previous step, we know that
$$\langle g \rangle = \left\{e, g, g^2, \ldots, g^{k-1}\right\},$$
so it is enough to show that the elements in the set on the right-hand side are distinct. Well, suppose that $g^m = g^n$ for some $0 \leq m, n < k$. By swapping n and m if necessary, we may assume that $m \geq n$. As before, we note $g^{m-n} = g^m \cdot (g^n)^{-1} = e$, but $0 \leq m - n < k$, so $m - n = 0$ by the minimality of k. Thus, $m = n$.

4. We complete the proof. Set $n \coloneqq |G|$ for brevity. By Corollary 5.73, we see that the size of $\langle g \rangle$ divides n. By the previous step, we see that $|\langle g \rangle| = k$, so we see $k \mid n$. Finishing up, write $n = qk$ for some integer q. Then
$$g^n = g^{qk} = \left(g^k\right)^q = e^q = e,$$
which is what we wanted. ■

Exercise 5.77. Verify by hand that $g^8 = e$ for any element $g \in D_4$.

Remark 5.78. Proposition 5.75 is not "sharp" in the following sense: there exist groups G and positive integers n less than $|G|$ such that g^n is the identity for any $g \in G$. For example, one can verify by hand that g^4 is the identity for any $g \in D_4$.

5.2.5 Quotient Groups

We continue trying to generalize our construction of $\mathbb{Z}/n\mathbb{Z}$. Given a group $(G, \cdot)$ with subgroup $H \subseteq G$, we might hope that we can make G/H into a group with the operation
$$(g_1 \cdot H) \cdot (g_2 \cdot H) \coloneqq (g_1 \cdot g_2) \cdot H.$$
However, this operation is not well-defined in general.

Example 5.79. We work in the context of Example 5.53. We would like
$$(e \cdot S) \cdot (r \cdot S) = r \cdot S,$$
but $e \cdot S = s \cdot S$, so we would also like
$$(s \cdot S) \cdot (r \cdot S) = (s \cdot r \cdot r) \cdot S = \left(r^3 \cdot s\right) \cdot S = r^3 \cdot S,$$
and $r \cdot S \neq r^3 \cdot S$.

Thus, one cannot in general make G/H into a group they way that we would like.

Let's investigate this further. Thinking symbolically about our cosets (for example, see Remark 5.64), we might hope we can write

$$(g_1 \cdot H) \cdot (g_2 \cdot H) = g_1 \cdot (H \cdot g_2) \cdot H \overset{*}{=} g_1 \cdot (g_2 \cdot H) \cdot H = (g_1 \cdot g_2) \cdot H, \tag{5.2}$$

where maybe $H \cdot H = H$. However, there is an issue at the marked equality $\overset{*}{=}$: we won't always have $g_2 \cdot H = H \cdot g_2$! For example, in Example 5.53, we saw $r \cdot S \neq S \cdot r$, which was more or less the problem in Example 5.79.

However, in our construction of $\mathbb{Z}/n\mathbb{Z}$, we saw in Exercise 5.52 that we do have $k + n\mathbb{Z} = n\mathbb{Z} + k$ for any $k \in \mathbb{Z}$, so sometimes we will have the property that $g_2 \cdot H = H \cdot g_2$. As such, we define a new adjective.

Definition 5.80 (normal)**.** Let $(G, \cdot)$ be a group. A subgroup $H \subseteq G$ is *normal* if and only if $g \cdot H = H \cdot g$ (as sets) for any $g \in G$.

Here are the examples we have easy access to.

Example 5.81. We showed in Exercise 5.52 that the subgroup $n\mathbb{Z}$ of $\mathbb{Z}$ is normal.

Example 5.82. The subgroup $R \coloneqq \{e, r, r^2, r^3\}$ of $(D_4, \cdot)$ is normal. For $g \in D_4$, there are two cases.

- If $g \in R$, then $g \cdot R = e \cdot R = R = R \cdot e = R \cdot g$.
- If $g \notin R$, then $g \cdot R$ cannot have intersection with $e \cdot R$ by Corollary 5.67. However, D_4 has eight elements, and $e \cdot R = R$ and $g \cdot R$ must both have four elements by Theorem 5.72, so $g \cdot R$ must be $D_4 \setminus R$. The same argument shows $R \cdot g = D_4 \setminus R$, so $g \cdot R = R \cdot g$ follows.

Exercise 5.83. Show that the subgroup $\{e, r^2\}$ of $(D_4, \cdot)$ is normal.

Example 5.84. Let $(G, \cdot)$ be a group, and set $Z(G) \coloneqq \{g \in G : g \cdot h = h \cdot g \text{ for all } h \in G\}$. Then $Z(G)$ is a normal subgroup of G. Indeed, for any $g \in G$, we compute

$$\begin{aligned} g \cdot Z(G) &= \{g \cdot h : h \in Z(G)\} \\ &= \{h \cdot g : h \in Z(G)\} \\ &= Z(G) \cdot g. \end{aligned}$$

Non-Example 5.85. The subgroup $S \coloneqq \{e, s\}$ of $(D_4, \cdot)$ is not normal. Indeed, we saw in Example 5.53 that $r \cdot S \neq S \cdot r$.

Continuing our story, as we might hope from (5.2), our prayers about G/H are answered for normal subgroups.

Proposition 5.86. Let $(G, \cdot)$ be a group and $H \subseteq G$ a normal subgroup. Then G/H is a group with the operation

$$(g_1 \cdot H) \cdot (g_2 \cdot H) \coloneqq (g_1 \cdot g_2) \cdot H.$$

Proof. The main difficulty is showing that the operation is a well-defined function, which we do first. This proof is a lot of force. If $g_1 \cdot H = g_1' \cdot H$ and $g_2 \cdot H = g_2' \cdot H$, then we must show that

$$(g_1 \cdot H) \cdot (g_2 \cdot H) \overset{?}{=} (g_1' \cdot H) \cdot (g_2' \cdot H),$$

or

$$(g_1 \cdot g_2) \cdot H \stackrel{?}{=} (g_1' \cdot g_2') \cdot H.$$

The conclusion is the most complicated piece of the puzzle right now, so we will manipulate it first. By Proposition 5.63, it's enough to show that $(g_1' \cdot g_2') \cdot (g_1 \cdot g_2)^{-1} \in H$. Equivalently, it's enough to show $g_1' \cdot g_2' \cdot g_2^{-1} \cdot g_1^{-1} \in H$.

We're now in a position to use our hypotheses. Because H is normal, we see $g_2 \cdot H = g_2' \cdot H$ implies $H \cdot g_2 = H \cdot g_2'$, so Exercise 5.66 implies that $g_2' \cdot g_2^{-1} \in H$. Thus, we set $h \coloneqq g_2' \cdot g_2^{-1}$, and we want to show $g_1' \cdot h \cdot g_1^{-1} \in H$. To finish, we note Exercise 5.66 tells us that it's enough to show $g_1' \cdot h \in H \cdot g_1$, but $g_1' \cdot h \in g_1' \cdot H$ by definition of $g_1' \cdot H$, and $g_1' \cdot H = g_1 \cdot H = H \cdot g_1$ because H is normal.

So we have a well-defined binary operation. It remains to check our group properties. These all follow more or less directly from G. Fix any $g, g', g'' \in G$.

- Associative: note

  $$(g \cdot H) \cdot ((g' \cdot H) \cdot (g'' \cdot H)) = (g \cdot H) \cdot ((g' \cdot g'') \cdot H) = (g \cdot g' \cdot g'') \cdot H.$$

  A similar argument shows that $((g \cdot H) \cdot (g' \cdot H)) \cdot (g'' \cdot H) = (g \cdot g' \cdot g'') \cdot H$, so we are done.

- Identity: we claim that $e \cdot H$ is our identity element. Indeed, for any coset $g \cdot H$, we write

  $$(e \cdot H) \cdot (g \cdot H) = (e \cdot g) \cdot H = g \cdot H.$$

  Similarly, $(g \cdot H) \cdot (e \cdot H) = (g \cdot e) \cdot H = g \cdot H$.

- Inverse: we claim that the inverse of the coset $g \cdot H$ is $g^{-1} \cdot H$. Indeed,

  $$(g \cdot H) \cdot \left(g^{-1} \cdot H\right) = \left(g \cdot g^{-1}\right) \cdot H = e \cdot H.$$

  Similarly, $\left(g^{-1} \cdot H\right) \cdot (g \cdot H) = \left(g^{-1} \cdot g\right) \cdot H = e \cdot H$. ■

Example 5.87. We work in the context of Example 5.55. We can visually see that $3\mathbb{Z}/6\mathbb{Z}$ is a normal subgroup of $\mathbb{Z}/6\mathbb{Z}$. In the group $(\mathbb{Z}/6\mathbb{Z})/(3\mathbb{Z}/6\mathbb{Z})$, we can also see that the addition law is given by

$$([k]_6 + 3\mathbb{Z}/6\mathbb{Z}) + ([\ell]_6 + 3\mathbb{Z}/6\mathbb{Z}) = ([k]_6 + [\ell]_6) + 3\mathbb{Z}/6\mathbb{Z} = [k + \ell]_6 + 3\mathbb{Z}/6\mathbb{Z}.$$

Thus, this group really looks identical to $\mathbb{Z}/3\mathbb{Z}$. Next lecture we will be able to put into words what "identical" means.

5.2.6 Problems

Problem 5.8. Let $(D_4, \cdot)$ denote the symmetries of the square, and let $R \coloneqq \{e, r, r^2, r^3\}$.

(a) For each $g \in R$, show that $\{e, g \cdot s\}$ is a subgroup of D_4.

(b) For which $g \in R$ is $\{e, g \cdot s\}$ a normal subgroup of D_4?

Problem 5.9. Let $(G, \cdot)$ be a group and $H \subseteq G$ a subgroup. Show that the function $f \colon G/H \to H\backslash G$ defined by $f(g \cdot H) \coloneqq H \cdot g^{-1}$ is well-defined and a bijection. What is the inverse function?

Problem 5.10. Let $(G, \cdot)$ be an abelian group. Show that all subgroups $H \subseteq G$ are normal.

Problem 5.11. Let $(G, \cdot)$ be a group and $H \subseteq G$ a subgroup. For each $g \in G$, define the set

$$g \cdot H \cdot g^{-1} \coloneqq \left\{ g \cdot h \cdot g^{-1} : h \in H \right\}.$$

Show the following.

(a) Show that $g \cdot H \cdot g^{-1}$ is a subgroup of G.

(b) If $g \cdot H \cdot g^{-1} = H$, then $g \cdot H = H \cdot g$.

(c) If H is a normal subgroup, then $g \cdot H \cdot g^{-1} = H$ for all $g \in G$.

(d) If $g \cdot H \cdot g^{-1} = H$ for all $g \in G$, then H is normal.

Problem 5.12. Let $(G, \cdot)$ be a group.

(a) Let $H_1, H_2 \subseteq G$ be normal subgroups. Show that $H_1 \cap H_2$ is a normal subgroup.

(b) Let $\mathcal{S}_n$ be the set of all subgroups $H \subseteq G$ with n elements. Show that the intersection

$$\bigcap_{H \in \mathcal{S}_n} H$$

is normal.

Problem 5.13. Let $(G, \cdot)$. Suppose that $H \subseteq G$ is a subgroup such that G/H has two elements. Show that H is a normal subgroup of G.

Problem 5.14. Let $(G, \cdot)$ be a group and $H \subseteq G$ a subgroup. This exercise explains that the definition of "normal subgroup" is in some sense chosen exactly so that G/H is a group.

(a) Suppose $h \cdot g \in g \cdot H$ for all $h \in H$. Show that $H \cdot g = g \cdot H$.

(b) Suppose that G/H is a group with group law given by $(g_1 \cdot H) \cdot (g_2 \cdot H) = (g_1 \cdot g_2) \cdot H$. In particular, suppose that this operation is well-defined. Show that H is a normal subgroup of G.

5.3 Week 14: Homomorphisms

In mathematics, objects are not understood in isolation but in how they relate to each other by functions. With metric spaces, the special functions were continuous functions. With groups, the special functions are homomorphisms.

5.3.1 Isomorphisms

We begin with an extended example. In Example 5.87, we saw that the group $(\mathbb{Z}/6\mathbb{Z})/(3\mathbb{Z}/6\mathbb{Z})$ looks a lot like the group $\mathbb{Z}/3\mathbb{Z}$. Note that our association between these two groups was stronger than merely a bijection: we also noted that the group laws looked very similar: for example,

$$([1]_6 + 3\mathbb{Z}/6\mathbb{Z}) + ([2]_6 + 3\mathbb{Z}/6\mathbb{Z}) = [3]_6 + 3\mathbb{Z}/6\mathbb{Z} = [0]_6 + 3\mathbb{Z}/6\mathbb{Z}$$

in $(\mathbb{Z}/6\mathbb{Z})/(3\mathbb{Z}/6\mathbb{Z})$, and

$$[1]_3 + [2]_3 = [0]_3$$

in $\mathbb{Z}/3\mathbb{Z}$. If we write out all possible additions $k + \ell$ in both groups, we can make the following tables.

$+$	$[0]_6+3\mathbb{Z}/6\mathbb{Z}$	$[1]_6+3\mathbb{Z}/6\mathbb{Z}$	$[2]_6+3\mathbb{Z}/6\mathbb{Z}$
$[0]_6+3\mathbb{Z}/6\mathbb{Z}$	$[0]_6+3\mathbb{Z}/6\mathbb{Z}$	$[1]_6+3\mathbb{Z}/6\mathbb{Z}$	$[2]_6+3\mathbb{Z}/6\mathbb{Z}$
$[1]_6+3\mathbb{Z}/6\mathbb{Z}$	$[1]_6+3\mathbb{Z}/6\mathbb{Z}$	$[2]_6+3\mathbb{Z}/6\mathbb{Z}$	$[0]_6+3\mathbb{Z}/6\mathbb{Z}$
$[2]_6+3\mathbb{Z}/6\mathbb{Z}$	$[2]_6+3\mathbb{Z}/6\mathbb{Z}$	$[0]_6+3\mathbb{Z}/6\mathbb{Z}$	$[1]_6+3\mathbb{Z}/6\mathbb{Z}$

$+$	$[0]_3$	$[1]_3$	$[2]_3$
$[0]_3$	$[0]_3$	$[1]_3$	$[2]_3$
$[1]_3$	$[1]_3$	$[2]_3$	$[0]_3$
$[2]_3$	$[2]_3$	$[0]_3$	$[1]_3$

Checking visually, it really looks like these groups are the same, up to some relabeling. What is this relabeling? Well, we define the function $f\colon (\mathbb{Z}/6\mathbb{Z})/(3\mathbb{Z}/6\mathbb{Z})$ by $f([k]_6+3\mathbb{Z}/6\mathbb{Z}) \coloneqq [k]_3$. From what we've already worked out about these two groups, we see that f is a well-defined function and a bijection.[5] (This is what we mean by f being a "relabeling.")

Additionally, we can summarize the fact that the relabeling f also "preserves the table" by the equation

$$f(([k]_6+3\mathbb{Z}/6\mathbb{Z})+([\ell]_6+3\mathbb{Z}/6\mathbb{Z})) = f([k]_6+3\mathbb{Z}/6\mathbb{Z})+f([\ell]_6+3\mathbb{Z}/6\mathbb{Z}).$$

Indeed, we are saying that we can add elements in $(\mathbb{Z}/6\mathbb{Z})/(3\mathbb{Z}/6\mathbb{Z})$ first and then check where f relabels the sum, or we can relabel the elements to put them in $\mathbb{Z}/3\mathbb{Z}$ and then add the elements in $\mathbb{Z}/3\mathbb{Z}$.

This discussion motivates the following definition.

Definition 5.88 (isomorphism)**.** Let $(G,\cdot)$ and $(G',\cdot')$ be groups. Then a function $f\colon G\to G'$ is an *isomorphism* if and only if f is a bijection and

$$f(g\cdot h) = f(g)\cdot' f(h)$$

for any $g,h\in G$. Note that $g\cdot h$ is an operation which happens in G. If there is an isomorphism between G and G', we will say that G and G' are *isomorphic* and write $G\cong G'$.

Here are some examples.

Example 5.89. Let $(D_4,\cdot)$ be the group of symmetries of the square. Note that $R\coloneqq\{e,r,r^2,r^3\}$ is a subgroup of D_4. There is a function $f\colon \mathbb{Z}/4\mathbb{Z}\to S$ given by

$$f([0]_4)\coloneqq e,\qquad f([1]_4)\coloneqq r,\quad f([2]_4)\coloneqq r^2,\quad f([3]_4)\coloneqq r^3.$$

We claim that f is an isomorphism. We can see from the definition that f is a bijection. For the last check, we first note that $f([4]_4)=f([0]_4)=e=r^4$ and $f([5]_4)=f([1]_4)=r=r^5$ and $f([6]_4)=f([2]_4)=r^2=r^6$, so in fact $f([k]_4)=r^k$ for $k\in\{0,1,2,3,4,5,6\}$.

Thus, for any $[a]_4,[b]_4\in\mathbb{Z}/2\mathbb{Z}$ with $a,b\in\{0,1,2,3\}$, we see $a+b\in\{0,1,2,3,4,5,6\}$, so

$$f([a]_4+[b]_4)=f([a+b]_4)=r^{a+b}=r^a\cdot r^b=f([a]_4)\cdot f([b]_4).$$

Exercise 5.90. Let $(D_4,\cdot)$ be the group of symmetries of the square. Note that $S\coloneqq\{e,s\}$ is a subgroup of D_4. Show $S\cong\mathbb{Z}/2\mathbb{Z}$.

Exercise 5.91. Convince yourself that the function $f\colon(\mathbb{Z}/6\mathbb{Z})/(3\mathbb{Z}/6\mathbb{Z})\to\mathbb{Z}/3\mathbb{Z}$ defined above as $f([k]_6+3\mathbb{Z}/6\mathbb{Z})\coloneqq[k]_3$ is a well-defined function and in fact an isomorphism.

Exercise 5.92. Show that the function $f\colon\mathbb{Z}/3\mathbb{Z}\to(\mathbb{Z}/6\mathbb{Z})/(3\mathbb{Z}/6\mathbb{Z})$ defined by $f([k]_3)\coloneqq[k]_6+3\mathbb{Z}/6\mathbb{Z}$ is a well-defined function and in fact an isomorphism.

[5] If this sentence worries you, feel free to check it by hand.

Example 5.93. Consider the group $(\mathbb{Z}, +)$ and define the function $f\colon \mathbb{Z} \to \mathbb{Z}$ by $f(k) \coloneqq -k$. Then f is an isomorphism. To see that f is bijective, we note that f is its own inverse: for any $k \in \mathbb{Z}$, we have

$$f(f(k)) = f(-k) = -(-k) = k.$$

To see that f is an isomorphism, we also note that

$$f(k + \ell) = -(k + \ell) = -k + -\ell = f(k) + f(\ell)$$

for any $k, \ell \in \mathbb{Z}$.

Here are some more abstract examples.

Example 5.94. Let $(G, \cdot)$ and $(G', \cdot')$ be group such that $|G| = |G'| = 1$. Then we see $G = \{e\}$ and $G' = \{e'\}$, where e and e' are the identities. We claim that $G \cong G'$. Indeed, define $f\colon G \to G'$ by

$$f(e) \coloneqq e'.$$

We can see that f is a bijection. To finish, because G has just one element, it suffices to calculate

$$f(e \cdot e) = f(e) = e' = e' \cdot' e' = f(e) \cdot' f(e).$$

Exercise 5.95. Let $(G, \cdot)$ be a group such that $|G| = 2$. Then we can write $G = \{e, g\}$, where e is the identities, and g is the non-identity elements of G respectively. Show that the function $f\colon \mathbb{Z}/2\mathbb{Z} \to G$ defined by

$$f([0]_2) \coloneqq e \qquad \text{and} \qquad f([1]_2) \coloneqq g$$

is an isomorphism.

The above two examples say that there is exactly one group of order 1 and 2 "up to isomorphism," respectively.

Typically we think of isomorphic groups as being essentially the same. As such, we should expect isomorphisms to form an equivalence relation. This is roughly the case but requires some attention.

Lemma 5.96. Suppose $(G, \cdot)$ is a group. Define the function $i\colon G \to G$ given by $i(g) \coloneqq g$ for each $g \in G$. Then i is an isomorphism.

Proof. We can see directly that i is a bijection. For example, to show that i is injective, note $i(g) = i(g')$ implies $g = i(g) = i(g') = g'$ for any $g, g' \in G$.

To finish, we must show that

$$i(g \cdot g') = i(g) \cdot i(g')$$

for any $g, g' \in G$. Well, both sides of the above equation are $g \cdot g'$, so we are done. ∎

Lemma 5.97. Suppose $(G, \cdot)$ and $(G', \cdot')$ are groups. If $f\colon G \to G'$ is an isomorphism, then there exists an isomorphism $f'\colon G' \to G$ such that $f'(f(g)) = g$ and $f(f'(g')) = g'$ for all $g \in G$ and $g' \in G'$.

Proof. By Proposition 2.29, the fact that f is a bijection promises us an inverse function $f'\colon G' \to G$ such that $f'(f(g)) = g$ and $f(f'(g')) = g'$ for all $g \in G$ and $g' \in G'$. It remains to show that f' is an isomorphism. Namely, given any $g', h' \in G'$, we must show

$$f'(g' \cdot' h') \stackrel{?}{=} f'(g') \cdot f'(h').$$

The trick here is to relate everything about f' back to f because we know that f is an isomorphism. Indeed, f is injective, so it suffices to show

$$f(f'(g' \cdot' h')) \stackrel{?}{=} f(f'(g') \cdot f'(h')).$$

However, we can show this directly by computing

$$\begin{aligned} f(f'(g') \cdot f'(h')) &= f(f'(g')) \cdot' f(f'(h')) \\ &= g' \cdot' h' \\ &= f(f'(g' \cdot' h')), \end{aligned}$$

which is what we wanted. Note we have used the fact that f is an isomorphism in the first equality. ■

Lemma 5.98. Suppose $(G, \cdot)$ and $(G', \cdot')$ and $(G'', \cdot'')$ are groups. If $f\colon G \to G'$ and $f'\colon G' \to G''$ are isomorphisms, then $(f' \circ f)\colon G \to G'$ is an isomorphism.

Proof. By Problem 2.5, we already know that $(f' \circ f)$ is a bijection. To finish the proof, we must show that

$$(f' \circ f)(g \cdot h) = (f' \circ f)(g) \cdot'' (f' \circ f)(h)$$

for any $g, h \in G$. Well, using the fact that f and f' are already isomorphisms, we compute

$$\begin{aligned} (f' \circ f)(g \cdot h) &= f'(f(g \cdot h)) \\ &= f'(f(g) \cdot' f(h)) \\ &= f'(f(g)) \cdot'' f'(f(h)) \\ &= (f' \circ f)(g) \cdot'' (f' \circ f)(h), \end{aligned}$$

which is what we wanted. ■

Proposition 5.99. Suppose $(G, \cdot)$ and $(G', \cdot')$ and $(G'', \cdot'')$. The following are true.

(a) $G \cong G$.

(b) If $G \cong G'$, then $G' \cong G$.

(c) If $G \cong G'$ and $G' \cong G''$, then $G \cong G''$.

Proof. Here, (a) follows from Lemma 5.96. To show (b), if $G \cong G'$, then there is an isomorphism $f\colon G \to G'$, so Lemma 5.97 grants an inverse isomorphism $f'\colon G' \to G$, so $G' \cong G$. Lastly, to show (c), if $G \cong G'$ and $G' \cong G''$, then there are isomorphisms $f\colon G \to G'$ and $f'\colon G' \to G''$, so the isomorphism $(f' \circ f)\colon G \to G''$ show $G \cong G''$. ■

5.3.2 Homomorphisms

Isomorphisms dictate when two groups are basically the same. However, we do want to know how groups relate to each other even if they are not literally the same. This is the goal of homomorphisms. To avoid forcing groups with a homomorphism being literally the same, we will remove the bijective condition. Here is our definition.

Definition 5.100 (homomorphism)**.** Let $(G, \cdot)$ and $(G', \cdot')$ be groups. A function $f\colon G \to G'$ is a *homomorphism* if and only if

$$f(g \cdot h) = f(g) \cdot' f(h)$$

for all $g, h \in G$.

Notably, any isomorphism is automatically a homomorphism, so all the examples from the previous subsection apply here. Furthermore, by definition, an isomorphism is bijective homomorphism.

Let's give a few more examples.

Example 5.101. Consider the groups $(\mathbb{C}, +)$ and $(\mathbb{R}, +)$. Then the function $r\colon \mathbb{C} \to \mathbb{R}$ defined by $r(a + bi) \coloneqq a$ is a homomorphism. Indeed, for any $a + bi, a' + b'i \in \mathbb{C}$, we compute

$$r((a + bi) + (a' + b'i)) = r((a + a') + (b + b')i) = a + a' = r(a + bi) + r(a' + b'i).$$

Non-Example 5.102. Consider the groups $(\mathbb{C}^\times, \cdot)$ and $(\mathbb{R}^\times, \cdot)$. Then the function $r\colon \mathbb{C} \to \mathbb{R}$ defined by $r(a + bi) \coloneqq a$ is not a homomorphism. For example, we can compute

$$r(i \cdot i) = r(-1) = -1 \neq 0 = 0 \cdot 0 = r(i) \cdot r(i).$$

Example 5.103. Let n be a positive integer. Let $(\mathrm{GL}_n(\mathbb{C}), \cdot)$ be the group of invertible $n \times n$ matrices with complex coefficients. Then the function $\det\colon \mathrm{GL}_n(\mathbb{C}) \to \mathbb{C}^\times$ defines a group homomorphism. (Note that this function makes sense because the determinant of an invertible matrix is nonzero.) Indeed, it is a property of the determinant that

$$\det(A \cdot B) = (\det A) \cdot (\det B)$$

for any $A, B \in \mathrm{GL}_n(\mathbb{C})$.

Example 5.104. Consider the group $(\mathbb{Z}, +)$ and define the function $f\colon \mathbb{Z} \to \mathbb{Z}$ by $f(k) \coloneqq 2k$. Then f is a homomorphism: for any $k, \ell \in \mathbb{Z}$,

$$f(k + \ell) = 2(k + \ell) = 2k + 2\ell = f(k) + f(\ell)$$

Example 5.105. Consider the groups $(\mathbb{Z}, +)$ and $(D_4, \cdot)$. Define the function $f\colon \mathbb{Z} \to D_4$ by $f(k) \coloneqq r^k$. Then f is a homomorphism: for any $k, \ell \in \mathbb{Z}$,

$$f(k + \ell) = r^{k+\ell} = r^k \cdot r^\ell = f(k) \cdot f(\ell).$$

Exercise 5.106. Consider the groups $(\mathbb{Z}/4\mathbb{Z}, +)$ and $(D_4, \cdot)$. Show that the function $f\colon \mathbb{Z}/4\mathbb{Z} \to D_4$ given by $f([k]_4) \coloneqq r^k$ is well-defined and a homomorphism.

Example 5.107. More generally, let $(G, \cdot)$ be a group, and fix some $g \in G$. Then the function $f\colon \mathbb{Z} \to G$ given by $f(k) \coloneqq g^k$ is a homomorphism: for any $k, \ell \in \mathbb{Z}$, we see $f(k + \ell) = g^{k+\ell} = g^k \cdot g^\ell = f(k) \cdot f(\ell)$.

Example 5.108. Consider the groups $(\mathbb{Z}/6\mathbb{Z}, +)$ and $(\mathbb{Z}/3\mathbb{Z}, +)$ and define the function $f\colon \mathbb{Z}/6\mathbb{Z} \to \mathbb{Z}/3\mathbb{Z}$ by $f([k]_6) \coloneqq [k]_3$. Note that f is well-defined: if $[k]_6 = [\ell]_6$, then $k - \ell$ is divisible by 6. But then we see that $k - \ell$ is also divisible by 3, so $f([k]_6) = [k]_3$ is equal to $f([\ell]_6) = [\ell]_3$.

In fact, f is a homomorphism: for any $k, \ell \in \mathbb{Z}$, we see

$$f([k]_6 + [\ell]_6) = f([k + \ell]_6) = [k + \ell]_3 = [k]_3 + [\ell]_3 = f([k]_6) + f([\ell_6]).$$

Exercise 5.109. Generalize Example 5.108 as follows: let a be a positive integer divisible by the positive integer b. Show that the function $f\colon \mathbb{Z}/a\mathbb{Z} \to \mathbb{Z}/b\mathbb{Z}$ defined by $f([k]_a) \coloneqq [k]_b$ is well-defined and a homomorphism.

Exercise 5.110. Let $(G, \cdot)$ and $(G', \cdot')$ and $(G'', \cdot'')$ be groups. Given homomorphisms $f \colon G \to G'$ and $f' \colon G' \to G''$, show that the function $(f' \circ f) \colon G \to G''$ is a homomorphism. One way to do this is to adapt the proof of Lemma 5.98.

Exercise 5.111. Let $(G, \cdot)$ be a group and $H \subseteq G$ a subgroup. Show that the function $i \colon H \to G$ defined by $i(h) \coloneqq h$ is an injective homomorphism.

Here are a few quick facts about homomorphisms seen in the above examples.

Lemma 5.112. Let $f \colon G \to G'$ be a homomorphism between the groups $(G, \cdot)$ and $(G', \cdot')$.

(a) We have $f(e) = e'$, where e and e' are the identities of G and G', respectively.

(b) For any $g \in G$, we have $f\left(g^{-1}\right) = f(g)^{-1}$.

Proof. Here we go.

(a) The point here is that $f(e)$ behaves a lot like the identity element of G'; for example,

$$f(e) \cdot' f(e) = f(e \cdot e) = f(e).$$

Already this is enough to show $f(e) = e'$: indeed, multiplying both sides by $f(e)^{-1}$, we see

$$f(e) = f(e) \cdot e' = f(e) \cdot' f(e) \cdot' f(e)^{-1} = f(e) \cdot' f(e)^{-1} = e',$$

which is what we wanted.

(b) The point here is that inverses are unique in G'. Thus, we check that

$$f(g) \cdot' f\left(g^{-1}\right) = f\left(g \cdot g^{-1}\right) = f(e) = e',$$

and

$$f\left(g^{-1}\right) \cdot' f(g) = f\left(g^{-1} \cdot g\right) = f(e) = e',$$

so Lemma 5.26 promises $f\left(g^{-1}\right) = f(g)^{-1}$. ∎

We now take a moment to appreciate how simple one-element groups are.

Proposition 5.113. Let $(G, \cdot)$ be a one-element group with identity e. Given any other group $(G', \cdot')$, there exists exactly one homomorphsim $f \colon G \to G'$ and exactly one homomorphism $f \colon G' \to G$.

Proof. We show the two directions of homomorphism independently.

- We show there is a unique homomorphism $f \colon G \to G'$. Well, we merely have to decide where f goes, but Lemma 5.112 tells us that the identity G goes to the identity of G', so $f(e) \coloneqq e'$ is forced.

  Thus, there is at most one homomorphism $G \to G'$ because they must all send $e \mapsto e'$. It remains to show that $f(e) \coloneqq e'$ defines a homomorphism, for which we just have to check

$$f(e \cdot e) = f(e) = e' = e' \cdot' e' = f(e) \cdot' f(e)$$

because e is the only element of G. Thus, we have shown there is at least one homomorphism $G \to G'$.

- We show there is a unique homomorphism $f\colon G' \to G$. Well, for each $g' \in G$, we must have $f(g') \coloneqq e$ because $e \in G$ is the only possible output.

  Thus, there is at most one homomorphism $G \to G'$ because they must all send $g' \mapsto e$ for each $g' \in G$. It remains to show that $f(g') \coloneqq e$ defines a homomorphism, for which we compute

  $$f(g' \cdot' h') = e = e \cdot e = f(g') \cdot f(h')$$

  for any $g', h' \in G$. ■

In general, it is a hard problem to determine all the homomorphisms in and out of a given group. Roughly speaking, this requires perfect knowledge of the group.

We close this subsection with a special homomorphism.

Theorem 5.114 (Cayley)**.** Let $(G, \cdot)$ be a group, and let $(\mathrm{Sym}(G), \circ)$ be the group of bijections $G \to G$ under composition defined in Problem 5.1. For each $g \in G$, define the function $\mu_\bullet\colon G \to G$ by $\mu_g\colon g' \mapsto (g \cdot g')$. Then the function $\mu_\bullet\colon G \to \mathrm{Sym}(G)$ is an injective homomorphism.

Proof. Note that $\mu_\bullet\colon G \to G$ is a bijection for each $g \in G$ by Proposition 5.31, so the function $\mu\colon G \to \mathrm{Sym}(G)$ at least makes sense. It remains to check that $\mu_\bullet$ is an injective homomorphism.

- We show that $\mu_\bullet$ is injective. Indeed, suppose that $\mu_{g_1} = \mu_{g_2}$ as functions $G \to G$. The idea here is that we can "read" of g from the function $\mu(g)$. The quickest way to see this is that $\mu_g(e) = g \cdot e = g$ for any $g \in G$, so it follows

  $$g_1 = \mu_{g_1}(e) = \mu_{g_2}(e) = g_2.$$

- We show that μ is a homomorphism. Namely, for any $g_1, g_2 \in G$, we must show

  $$\mu_{g_1 \cdot g_2} \stackrel{?}{=} \mu_{g_1} \circ \mu_{g_2}.$$

  Well, two functions are equal if and only if they are equal on all inputs, so we pick up any $g \in G$ and compute

  $$\begin{aligned} \mu_{g_1 \cdot g_2}(g) &= (g_1 \cdot g_2) \cdot g \\ &= g_1 \cdot (g_2 \cdot g) \\ &= \mu_{g_1}(g_2 \cdot e) \\ &= \mu_{g_1}(\mu_{g_2}(g)) \\ &= (\mu_{g_1} \circ \mu_{g_2})(g), \end{aligned}$$

  which is what we wanted. ■

Remark 5.115. The reason why Theorem 5.114 has a name is that it says that all groups are (isomorphic to) some subgroup of a symmetric group. As such, we can think any group $(G, \cdot)$ as the permutations (i.e., bijections) of some object (here, the set G).

5.3.3 Kernels and Images

We claimed that homomorphisms tell us how groups relate to one another, but it is undeniable that some homomorphisms are more informative than others. For example, an isomorphism tell us that two groups are basically the same, but a homomorphism $\{e\} \to G$ from the one-element group doesn't really tell us anything at all about G because (by Proposition 5.113) there is only one such.

Kernels and images provide us with a way to measure what is lost in a group homomorphism $G \to G'$. Roughly speaking, if the kernel is small, then the homomorphism does a good job mapping G to G'. On the other hand, if the image is large, then the homomorphism does a good job covering G'. Here are our definitions.

Definition 5.116 (kernel, image)**.** Let $f\colon G \to G'$ be a homomorphism of groups $(G, \cdot)$ and $(G', \cdot')$. Let e' be the identity of G'.

- The *kernel* of f is $\ker f \coloneqq \{g \in G : f(g) = e\}$. Note $\ker f \subseteq G$.
- The *image* of f is $\operatorname{im} f \coloneqq f(G)$. Note $\operatorname{im} f \subseteq G'$.

Let's see some examples.

Example 5.117. Consider the homomorphism r of Example 5.101.

- We see $\ker r$ consists of the complex numbers $a + bi$ where $r(a + bi) = a$ is equal to 0. Thus, $\ker r = \{bi : b \in \mathbb{R}\}$.
- We see $\operatorname{im} r = \mathbb{R}$. For example, for any real number $a \in \mathbb{R}$, we see $r(a) = a$, so $a \in \operatorname{im} r$.

Example 5.118. Consider the homomorphism f of Example 5.104.

- To compute $\ker f$, we see $f(k) = 2k$ is zero if and only if $2k = 0$, which is equivalent to $k = 0$. Thus, $\ker f = \{0\}$.
- For $\operatorname{im} f$, we compute $\operatorname{im} f = \{f(k) : k \in \mathbb{Z}\} = \{2k : k \in \mathbb{Z}\} = 2\mathbb{Z}$.

Example 5.119. Consider the homomorphism $f\colon \mathbb{Z} \to D_4$ of Example 5.105. We compute $\ker f$ and $\operatorname{im} f$.

Proof. We run our computations separately.

- We claim $\ker f = 4\mathbb{Z}$. We have two inclusions to show. In one direction, if $n \in 4\mathbb{Z}$, then $r = 4q$ for some $q \in \mathbb{Z}$, so $f(n) = r^{4q} = \left(r^4\right)^q = e^q = e$. Thus, $4\mathbb{Z} \subseteq \ker f$.

  In the other direction, for any integer n, note Theorem 5.7 promises an integers $q \in \mathbb{Z}$ and $x \in \{0, 1, 2, 3\}$ such that $n = 4q + x$. Now, $f(n) = e$ if and only if $r^n = e$, which we expand as

  $$r^n = r^{4q+x} = \left(r^4\right)^q \cdot r^x = e^q \cdot r^x = e \cdot r^x = r^x.$$

  However, this means $r^n = e$ if and only if $r^x = e$, but because $x \in \{0, 1, 2, 3\}$, we can say $r^x = e$ if and only if $x = 0$. Thus, $n = 4q \in 4\mathbb{Z}$. It follows $\ker f \subseteq 4\mathbb{Z}$.

- We claim $\operatorname{im} f = \left\{e, r, r^2, r^3\right\}$. We have two inclusions to show. In one direction, note $f(n) = r^n$ for each $n \in \mathbb{Z}$, so the outputs $\{f(0), f(1), f(2), f(3)\}$ show $\left\{e, r, r^2, r^3\right\} \subseteq \operatorname{im} f$.

  In the other direction, the previous point showed that, for any $n \in \mathbb{Z}$, we have $f(n) = r^x$ for some $x \in \{0, 1, 2, 3\}$. Thus, $f(n) \in \left\{e, r, r^2, r^3\right\}$ for any $n \in \mathbb{Z}$, so $\operatorname{im} f \subseteq \left\{e, r, r^2, r^3\right\}$. ∎

Example 5.120. Consider the homomorphism $\det\colon \mathrm{GL}_n(\mathbb{C}) \to \mathbb{C}^\times$ of Example 5.103. We compute $\ker \det$ and $\ker \det$.

Proof. We run our computations separately.

- We note that $\ker \det$ consists of the matrices $M \in \mathrm{GL}_n(\mathbb{C})$ such that $\det M = 1$, which is equivalent to $M \in \mathrm{SL}_n(\mathbb{C})$. Thus, $\ker \det = \mathrm{SL}_n(\mathbb{C})$.

- We claim $\operatorname{im}\det = \mathbb{C}^\times$. Certainly $\operatorname{im}\det \subseteq \mathbb{C}^\times$. In the other direction, for any $z \in \mathbb{C}^\times$, we see that

$$\det\left(\begin{bmatrix} z & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}\right) = z,$$

  so $z \in \operatorname{im}\det$. Thus, $\mathbb{C}^\times \subseteq \operatorname{im}\det$. ■

Example 5.121. Consider the homomorphism $f\colon \mathbb{Z}/6\mathbb{Z} \to \mathbb{Z}/3\mathbb{Z}$ of Example 5.108. We compute $\ker f$ and $\operatorname{im} f$.

Proof. We run our computations separately.

- To compute $\ker f$, we note that $[k]_6 \in \ker f$ if and only if $f([k]_6) = [k]_3$ equals $[0]_3$, which is equivalent to $3 \mid k$. Checking the elements of $\mathbb{Z}/6\mathbb{Z}$, we see that $\ker f = \{[0]_6, [3]_6\} = 3\mathbb{Z}/6\mathbb{Z}$.
- We claim $\operatorname{im} f = \mathbb{Z}/3\mathbb{Z}$. Certainly $\operatorname{im} f \subseteq \mathbb{Z}/3\mathbb{Z}$. Conversely, for any $[k]_3 \in \mathbb{Z}/3\mathbb{Z}$, we see $f([k]_6) = [k]_3$, so $[k]_3 \in \operatorname{im} f$. It follows $\mathbb{Z}/3\mathbb{Z} \subseteq \operatorname{im} f$. ■

Example 5.122. Let $(G, \cdot)$ be a group and $H \subseteq G$ a subgroup. Consider the homomorphism $i\colon H \to G$ of Exercise 5.111.

- We see $g \in \ker i$ if and only if $i(g) = e$, but $i(g) = g$, so $g \in \ker i$ if and only if $g = e$. Thus, $\ker i = \{e\}$.
- We claim $\operatorname{im} i = H$. In one direction, note $h \in H$ implies $i(h) = h$, so we see $h \in \operatorname{im} i$. Thus, $H \subseteq \operatorname{im} i$.

  In the other direction, if $g \in \operatorname{im} i$, then there exists $h \in H$ such that $g = i(h)$. But $i(h) = h$, so $g = h \in H$. It follows $\operatorname{im} i \subseteq H$.

Example 5.123. Let $(G, \cdot)$ be a one-element group with identity e. For any group $(G', \cdot')$ consider the unique homomorphism $f\colon G' \to G$ defined in Proposition 5.113.

- We see $\ker f = G'$ because $f(g') = e$ for all $g' \in G'$.
- We see $\operatorname{im} f = G$ because $G = \{e\}$. Explicitly, we know $\operatorname{im} f \subseteq G = \{e\}$ immediately; conversely, identity element $e' \in G'$ has $f(e') = e$, so $\{e\} \subseteq \operatorname{im} f$.

Exercise 5.124. Let $(G, \cdot)$ be a one-element group with identity e. For any group $(G', \cdot')$ consider the unique homomorphism $f\colon G \to G'$ defined in Proposition 5.113. (Note the direction change!) Compute $\ker f$ and $\operatorname{im} f$.

Now let's prove a few facts which are hopefully not too surprising from the above discussion. We begin with the image.

Lemma 5.125. Let $f\colon G \to G'$ be a homomorphism of the groups $(G, \cdot)$ and $(G', \cdot')$.

(a) $\operatorname{im} f$ is a subgroup of G'.

(b) f is surjective if and only if $\operatorname{im} f = G'$.

Proof. We show these separately.

(a) We simply run our subgroups checks directly. Let e and e' be the identities of G and G', respectively.

- Identity: note $f(e) = e'$ by Lemma 5.112, so $e' \in \operatorname{im} f$.
- Closed: given $f(g), f(h) \in \operatorname{im} f$, we see that $f(g) \cdot' f(h) = f(g \cdot h)$ also lives in $\operatorname{im} f$.
- Inverse: given $f(g) \in \operatorname{im} f$, we see that $f(g)^{-1} = f\left(g^{-1}\right)$ by Lemma 5.112, so $f(g)^{-1}$ is also in $\operatorname{im} f$.

(b) Note $\operatorname{im} f = G'$ is equivalent to the following statement: for each $g' \in G'$, there exists $g \in G$ such that $f(g) = g'$. But this is equivalent to saying f is surjective, so we are done. ■

Exercise 5.126. Adapt the proof of Lemma 5.125 to show the following: let $f\colon G \to G'$ be a homomorphism of the groups $(G, \cdot)$ and $(G', \cdot')$. If $H \subseteq G$ is a subgroup, then $f(H)$ is also a subgroup.

Now we discuss the kernel.

Lemma 5.127. Let $f\colon G \to G'$ be a homomorphism of the groups $(G, \cdot)$ and $(G', \cdot')$.

(a) $\ker f$ is a subgroup of G.

(b) f is injective if and only if $\ker f = \{e\}$.

Proof. We show these separately.

(a) We show the subgroup properties by hand. Let e and e' denote the identities of G and G', respectively.

- Identity: by Lemma 5.112, we see that $f(e) = e'$, so $e \in \ker f$.
- Closed: if $g, h \in \ker f$, then to show $g \cdot h \in \ker f$ we compute

$$f(g \cdot h) = f(g) \cdot' f(h) = e' \cdot e' = e'.$$

- Inverse: if $g \in \ker f$, we would like to show $g^{-1} \in \ker f$. Well, by Lemma 5.112, we see $f\left(g^{-1}\right) = f(g)^{-1}$, but $f(g) = e'$, so $f\left(g^{-1}\right) = (e')^{-1} = e'$.

(b) This requires some care. In one direction, suppose f is injective. Then we know $f(e) = e'$ by Lemma 5.112. But now $g \in \ker f$ is equivalent to $f(g) = e' = f(e)$. Because f is injective, we thus see $g \in \ker f$ is equivalent to $g = e$, so $\ker f = \{e\}$.

In the other direction, suppose $\ker f = \{e\}$. Suppose $g, h \in G$ have $f(g) = f(h)$ so that we would like to show $g = h$. The key claim is to show that $g \cdot h^{-1} \in \ker f$. This will be enough because $\ker f = \{e\}$, so $g \cdot h^{-1} \in \ker f$ will imply $g \cdot h^{-1} = e$ and so $g = h$.

We now show $g \cdot h^{-1} \in \ker f$ by direct computation: we write

$$f\left(g \cdot h^{-1}\right) = f(g) \cdot' f\left(h^{-1}\right) = f(g) \cdot' f(h)^{-1}.$$

However, $f(g) = f(h)$, so this is $f(g) \cdot' f(g)^{-1} = e'$. This completes the proof. ■

In fact, we can generalize the above proof to show that kernels have a special relationship to normal subgroups.

Proposition 5.128. Let $f\colon G \to G'$ be a homomorphism of the groups $(G,\cdot)$ and $(G',\cdot')$. The following are equivalent for $g, h \in G$.

(a) $f(g) = f(h)$.

(b) $g \cdot h^{-1} \in \ker f$.

(c) $g^{-1} \cdot h \in \ker f$.

It follows that $\ker f$ is a normal subgroup of G.

Proof. We begin by showing that (a) and (b) are equivalent. We are concerned if the element $g \cdot h^{-1}$ lives in $\ker f$, so the main point here is the computation

$$f\left(g \cdot h^{-1}\right) = f(g) \cdot' f\left(h^{-1}\right) = f(g) \cdot' f(h)^{-1}.$$

Thus, if given (a), then we see $f(g) \cdot' f(h)^{-1} = e'$, so $g \cdot h^{-1} \in \ker f$ follows. Conversely, if given (b), then $f\left(g \cdot g^{-1}\right) = e'$, so $f(g) \cdot' f(h)^{-1} = e'$, which rearranges into $f(g) = f(h)$.

The proof that (a) and (c) are equivalent is essentially the same. For completeness, we will note that the main point is again the computation

$$f\left(g^{-1} \cdot h\right) = f\left(g^{-1}\right) \cdot' f(h) = f(g)^{-1} \cdot' f(h).$$

We leave the rest of the proof to the following exercise.

Exercise 5.129. Complete the proof that (a) and (c) are equivalent.

We now turn to showing that $\ker f$ is a normal subgroup of G. Indeed, for any $g \in G$, we want to show $g \cdot (\ker f) = (\ker f) \cdot g$. By Proposition 5.63, we see that $h \in g \cdot (\ker f)$ is equivalent to $g^{-1} \cdot h \in \ker f$; similarly, $h \in (\ker f) \cdot g$ is equivalent to $g \cdot h^{-1} \in \ker f$. Thus, using the above work,

$$\begin{aligned} g \cdot (\ker f) &= \left\{h \in G : g^{-1} \cdot h \in \ker f\right\} \\ &= \left\{h \in G : g \cdot h^{-1} \in \ker f\right\} \\ &= (\ker f) \cdot g, \end{aligned}$$

which is what we wanted. ∎

Example 5.130. Let $(G,\cdot)$ be a group and $H \subseteq G$ be a normal subgroup. Then the function $f\colon G \to G/H$ given by $f(g) \coloneqq g \cdot H$ defines a homomorphism: for any $g_1, g_2 \in G$, we see

$$f(g_1 \cdot g_2) = (g_1 \cdot g_2) \cdot H = (g_1 \cdot H) \cdot (g_2 \cdot H) = f(g_1) \cdot f(g_2).$$

Now, we note $\ker f = H$. Indeed, $f(g) = g \cdot H$ is equal to $e \cdot H$ if and only if $g \in e \cdot H = H$.

Now, Proposition 5.128 tells us that all kernels are normal subgroups, and Example 5.130 tells us that all normal subgroups appear as the kernel of some map. Thus, we can (and should!) think about normal subgroups as the normal subgroups which arise as kernels.

5.3.4 Groups of Prime Order

As an application of the theory we have built, we will show the following result.

Theorem 5.131. Fix a prime p. All groups $(G,\cdot)$ with $|G| = p$ are isomorphic to each other. In fact, $G \cong \mathbb{Z}/p\mathbb{Z}$.

Theorem 5.131 is really amazing: given only knowledge about the size of G, we are immediately able to make deductions about the group structure of G. This result is one example of what we mean when we say "groups have structure": a small amount of information leads to a larger amount of information because of "structural" constraints.

Quickly, note that we have dealt with one-element groups in Example 5.94, so Theorem 5.131 is the next simplest "classification" result for groups. Further, observe that groups can in general be somewhat complicated. For example, there are non-isomorphic groups of size 4.

Example 5.132. Let V denote the subgroup $\{e, s, r^2, sr^2\}$ of $(D_4, \cdot)$. (See Exercise 5.39.) We claim that V and $\mathbb{Z}/4\mathbb{Z}$ are not isomorphic. To begin, note that $g^2 = e$ for all $g \in V$, which we can check directly: note $s^2 = r^4 = e$. Now, let $\varphi\colon \mathbb{Z}/4\mathbb{Z} \to V$ be any homomorphism, and we show that φ is not an isomorphism. Indeed, note

$$\varphi([2]_4) = \varphi([1]_4 + [1]_4) = \varphi([1]_4)^2 = e = \varphi([0]).$$

The key equality is $\varphi([1])^2 = e$, which holds because all elements of V square to e. Anyway, we see that φ thus fails to be injective.

Going further, we know that there are non-isomorphic groups G with $|G| = 8$: indeed, we have seen the two groups $(D_4, \cdot)$ and $(\mathbb{Z}/8\mathbb{Z}, +)$; see Problem 5.16 for details. One can also show that there are non-isomorphic groups of order 6, but such examples are not immediate given what we have developed so far.

Remark 5.133. In fact, there is a set of five groups of size 8 such that any group of order 8 is isomorphic to one of those five groups.

Anyway, let's go ahead and prove Theorem 5.131. The key is the following lemma.

Lemma 5.134. Fix a prime p, and let $(G, \cdot)$ be a group with $|G| = p$. Further, let e denote the identity of G. For any subgroup $H \subseteq G$, either $H = \{e\}$ or $H = G$.

Proof. The key input here is Theorem 5.72. Indeed, by Theorem 5.72, either $|H| = 1$ or $|H| = p$ because 1 and p are the only positive divisors of p. We thus have two cases.

- If $|H| = 1$, then we note that $e \in H$ because H is a subgroup, so the one-element set $\{e\}$ is a subset of H. Because $|\{e\}| = |H|$, we conclude $H = \{e\}$.
- If $|H| = p$, then we see that $H \subseteq G$ while $|H| = |G|$, so $H = G$ follows.

The above casework completes the proof. ∎

And now, here is our theorem.

Theorem 5.131. Fix a prime p. All groups $(G, \cdot)$ with $|G| = p$ are isomorphic to each other. In fact, $G \cong \mathbb{Z}/p\mathbb{Z}$.

Proof. Because $|G| > 1$, we may select an element $g \in G$ not equal to the identity $e \in G$. By Proposition 5.75, note that $g^p = e$, where $e \in G$ is the identity. For intuition, note that the proof of Proposition 5.75 shows that

$$G = \left\{e, g, g^2, \ldots, g^{p-1}\right\},$$

and the right-hand side looks like $\mathbb{Z}/p\mathbb{Z}$ if we replace g with $[1]_p$. With this in mind, we will show that the function $\varphi\colon \mathbb{Z}/p\mathbb{Z} \to G$ by

$$\varphi([k]_p) \coloneqq g^k$$

is an isomorphism. The main difficulty is showing that φ is well-defined. Anyway, here are our checks.

- Well-defined: suppose $[k]_p = [\ell]_p$, and we want to show that $g^k = g^\ell$. Well, $[k]_p = [\ell]_p$ implies that $p \mid k - \ell$, so we may use Theorem 5.7 to write $k = pq + \ell$ for some integer q. Then
$$g^k = g^{pq+\ell} = (g^p)^q \cdot g^\ell = e^q \cdot g^\ell = g^\ell.$$
- Injective: by Lemma 5.127, it is enough to show that $\ker \varphi = \{[0]_p\}$. Well, we know that $\ker \varphi$ is a subgroup of $\mathbb{Z}/p\mathbb{Z}$ already, so Lemma 5.134 implies that $\ker \varphi = \{[0]\}_p$ or $\ker \varphi = \mathbb{Z}/p\mathbb{Z}$. However, $\varphi([1]_p) = g \neq e$ by definition, so $[1]_p \notin \ker \varphi$, so $\ker \varphi \neq \mathbb{Z}/p\mathbb{Z}$, so $\ker \varphi = \{[0]_p\}$ follows.
- Surjective: by Lemma 5.125, it is enough to show that $\operatorname{im} \varphi = G$. Well, we know that $\operatorname{im} \varphi$ is a subgroup of G already, so Lemma 5.134 implies that $\operatorname{im} \varphi = \{e\}$ or $\operatorname{im} \varphi = G$. Well, $g = \varphi([1]_p) \in \operatorname{im} \varphi$, so $\operatorname{im} \varphi \neq \{e\}$, so $\operatorname{im} \varphi = G$ follows. ■

5.3.5 Problems

Problem 5.15. Let $2\mathbb{Z}$ denote the subgroup of even integers in $(\mathbb{Z}, +)$. Exhibit two distinct isomorphisms $\varphi\colon 2\mathbb{Z} \to \mathbb{Z}$.

Problem 5.16. The groups D_4 and $\mathbb{Z}/8\mathbb{Z}$ both have order 8.

(a) Show that $g^4 = e$ for all $g \in D_4$.

(b) Find an element $g \in \mathbb{Z}/8\mathbb{Z}$ such that $g + g + g + g \neq [0]$.

(c) Show that the groups D_4 and $\mathbb{Z}/8\mathbb{Z}$ are not isomorphic.

Problem 5.17. Let $(G, \cdot)$ be a group. Given $g \in G$, suppose that there is a least positive integer n such that $g^n = e$. Consider the homomorphism $f\colon \mathbb{Z} \to G$ defined by $f(k) \coloneqq g^k$ in Example 5.107.

(a) Show that $\ker f = n\mathbb{Z}$.

(b) Show that $\operatorname{im} f = \{e, g, g^2, \ldots, g^{n-1}\}$.

Problem 5.18. One can check that $\mathbb{Z}$ is a normal subgroup of $(\mathbb{Q}, +)$, so we may define the group $\mathbb{Q}/\mathbb{Z}$. Define $\varphi\colon \mathbb{Q}/\mathbb{Z} \to \mathbb{Q}/\mathbb{Z}$ by $\varphi(x) \coloneqq 5x$.

(a) Show that φ is a group homomorphism.

(b) Exhibit an isomorphism $\ker \varphi \cong \mathbb{Z}/5\mathbb{Z}$.

Problem 5.19. Consider the groups $(\mathbb{Z}, +)$ and $(\mathbb{Z}/10\mathbb{Z}, +)$.

(a) Compute the number of group homomorphisms $\varphi\colon \mathbb{Z} \to \mathbb{Z}/10\mathbb{Z}$.

(b) Compute the number of group homomorphisms $\varphi\colon \mathbb{Z}/10\mathbb{Z} \to \mathbb{Z}$.

(c) Compute the number of injective group homomorphisms $\varphi\colon \mathbb{Z} \to \mathbb{Z}/10\mathbb{Z}$.

(d) Compute the number of surjective group homomorphisms $\varphi\colon \mathbb{Z} \to \mathbb{Z}/10\mathbb{Z}$.

Problem 5.20. Let $(G, \cdot)$ be a group. Define the map $\varphi\colon G \to G$ by $\varphi(g) \coloneqq g^2$.

(a) Suppose φ is a group homomorphism. Show that G is commutative.

(b) Suppose G is commutative. Show that φ is a group homomorphism.

Problem 5.21. Let $(G, \cdot)$ be a group with identity element e. Suppose $g^2 = e$ for each $g \in G$. Show that G is commutative.

Problem 5.22. Let $(G, \cdot)$ be a group.

(a) Let $\mathrm{Aut}(G)$ denote the set of isomorphisms $G \to G$. Show that $\mathrm{Aut}(G)$ is a group where the operation is composition.

(b) Define the function $\varphi\colon \mathrm{Aut}(\mathbb{Q}) \to \mathbb{Q}^\times$ by $\varphi(f) \coloneqq f(1)$. Show that φ is an isomorphism.

(c) Let $g \in G$ be some element. Define $\varphi_g\colon G \to G$ by $\varphi_g(h) \coloneqq ghg^{-1}$. Show that φ_g is an automorphism.

(d) Define $\varphi\colon G \to \mathrm{Aut}(G)$ by $\varphi(g) \coloneqq \varphi_g$. Show that φ is a homomorphism.

(e) Show that $\ker \varphi = \{g \in G : gh = hg \text{ for all } h \in G\}$.

BIBLIOGRAPHY

[Jac90] Joseph Jacobs. *English Fairy Tales*. English Fairy Tales. David Nutt, 1890.

[Gei60] Theodor Geisel. *One Fish, Two Fish, Red Fish, Blue Fish*. Beginner Books. Beginner Books, 1960.

[Pug15] Charles C. Pugh. *Real Mathematical Analysis*. Undergraduate Texts in Mathematics. Springer International Publishing, 2015.

[Shu16] Neal Shusterman. *Scythe*. Arc of a Scythe. Simon & Schuster, 2016.

[Row17] Winston Rowntree. *People Watching: Why Nostalgia Is Total Bull*. 2017. URL: https://youtu.be/s9mfi0L6PC4?t=336.

[Ros19] Kenneth H. Rosen. *Discrete Mathematics and its Applications*. 8th ed. New York, NY: McGraw-Hill, 2019.

[Che22] Evan Chen. *An Infinitely Large Napkin*. 2022. URL: https://venhance.github.io/napkin/Napkin.pdf.

Index

abelian, 99

bijection, 34
binary operation, 98

cardinality, 41
closed, 76, 89
closure, 75, 90
codomain, 33
combinatorial proof, 42
commutative, 99
composition, 33
compound proposition, 12
computable, 47
conditional statement, 13
conjunction, 13
continuous, 82, 91
contradiction, 15
contraposition, 17
coset, 106
countable, 44

de Morgan's laws, 31
disjunction, 13
domain, 33

element, 28
empty set, 28
equivalence relation, 37
exclusive disjunction, 13
existential quantification, 18

finite, 42
function, 33
Fundamental theorem of arithmetic, 57

group, 99

Hilbert's grand hotel, 49
homomorphism, 119

image, 123
infinite, 42
infinite tree, 50
injective, 34
interior, 88
interior point, 68
intersection, 29
inverse, 34
isomorphism, 117

kernel, 123

limit, 73
logical operators, 12
logically equivalent, 15

mapping, 33
maximal, 59
metric, 65
metric space, 65
minimal, 59
morphism, 33

n-ary propositional function, 17
negation, 12
normal, 114

one-to-one, 34
one-to-one correspondence, 34
onto, 34
open, 69, 87
open ball, 68

partial order, 36
partially ordered set, 36

partition, 39
power set, 30
pre-image, 33
predicate, 17
prime, 57, 57
prime factorization, 57
primitive proposition, 12
Principle of induction, 51
principle of induction, 56
product, 29
proper subset, 28
proposition, 11
propositional variable, 12
Pythagorean theorem, 22

relation, 35

sentential variable, 12
sequence, 73
set, 28
subgroup, 102
subset, 28
surjective, 34
symmetric difference, 32

target, 33
tautology, 15
topological space, 87
topology, 87
total order, 36
totally ordered set, 36
transformation, 33
truth value, 12

uncountable, 44
union, 29
universal quantification, 18

well-ordering, 59
well-ordering principle, 61