MUSA 74: Proof-Writing Skills DeCal

Mathematics Undergraduate Student Association

Spring 2019

Contents

On these course notes

The following notes are collaborative notes meant for use by students at UC Berkeley. However, they may be read or edited by anyone interested in the material.

These notes were originally written by Cailan Li and the first class of students taking MUSA 74, in Spring 2018. They were edited by Aidan Backus, Andrew DeLapo, and Java Villano in preparation for Spring 2019.

Syllabus

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 Math 104 and Math 113.

MUSA 74 is intended for students who have some, but little, familiarity with writing proofs, but aren't sure if they're experienced enough to be ready for Math 113 and Math 104. Beyond this, we officially assume no prerequisites other than a little calculus (at the level of Math 1A) and linear algebra (at the level of Math 54), used as examples (though we shall also appeal to Math 55 for a few examples as we need). By the time you complete this course, you will be comfortable with writing proofs at the level required by the core upper-division sequence of Math 110, Math 113, Math 104, and Math 185.

Though MUSA 74 is designed around the material of Math 104 and Math 113, this is largely incidental, as the material is meant to serve as examples for learning how to write proofs. As such, the class is open to students taking neither class – for example, students of Math 55, Math 110, Math 128A, and CS 170 are all welcome to join.

The chapters on Math 113 and Math 104 are essentially independent of each other; the seminar will meet twice per week, covering each class once per week. Both are dependent on the introductory material, and both are prerequisites for the conclusion, a fun wrap-up about the Fourier transform. We encourage all students to come to both meetings per week, to ensure they get as much practice as possible.

Lectures will be formatted as follows: a facilitator will discuss the basic definitions and theorems and work through a few examples, and then the students will discuss the homework problems among each other and with facilitators. Incomplete homework problems should be completed at home before the next lecture, and turned in to a facilitator, who will give feedback on the writing and proof style.

We want to encourage a welcoming and inclusive learning environment. Questions, curiosity, and collaboration are all highly encouraged, and dismissive attitudes are strongly discouraged. Math is a difficult subject, and confusion is not a sign of weakness. If students would like help outside of class, they are highly encouraged to ask the course facilitators to meet one-on-one.

Grading policy, homework, and units

You will most likely have the opportunity to take MUSA 74 for a unit (though we cannot guarantee this with certainty at this time). If so, you will be graded on a "pass/no pass" basis, based on the completion of homework exercises. To pass the course, a student must complete and turn in homework assignments from at least half of lectures given (but note that the lectures on Fourier transforms will not have homework assigned). Homework will be assigned in each lecture as it is reached in class, and will be a due a week later. You're always welcome to ask facilitators for feedback.

Office hours

TBD, determined by MUSA office hours schedule. We encourage all students to come to 938 Evans during designated hours to discuss the material further.

Course outline

Here's a list of topics to be covered in the course.

1. 6 lectures on proofs: proofs, contradiction, induction, and existence and uniqueness

2. 8-10 Friday lectures on groups: groups, homomorphisms, subgroups, generators and relations, and quotients

3. 8-10 Monday lectures on metric spaces: metric spaces, limit and interior points, open and closed sets, and continuity and convergence

4. 2 lectures on special topics.

Chapter 1

Introduction

In science, we often discover facts by the scientific method. We make a hypothesis about how the world works, test our hypothesis, and make a conclusion based on the results. If there is sufficient evidence that a hypothesis is true, then we take the hypothesis as the truth, until some evidence comes to contradict the hypothesis.

Mathematics works differently, however. For mathematicians, the only such evidence that exists comes in the form of proofs. Proofs are infinitely more powerful than scientific evidence, as a proper proof can absolutely guarantee that a mathematical statement is true. For example, if you are told that $\sqrt{2}$ is irrational, how do you *know* this is true? How can you guarantee that no matter which integers a and b are picked, $\left(\frac{a}{b}\right)^2$ is never exactly 2? The answer is via mathematical proof.

You may have seen some proofs in your previous classes, especially in Math 54 and Math 55. Indeed, upper-division mathematics courses are almost entirely proof-based. In this section, we will see some of the methods of proof available to you as you encounter problems.

1.1 Basic proofs

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

Example 1.1. If $p(x) = x^2 + bx + c$ and $r_1 \neq r_2$ are 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 $p(r_1) = p(r_2) = 0$. Since $r_1 \neq r_2$, there is a polynomial f such that $p(x) = f(x)(x - r_1)(x - r_2)$, but p is a quadratic so f is a constant, which must be 1 since the coefficient on the x^2 term is 1. Therefore $p(x) = (x - r_1)(x - r_2)$. Expanding both sides,

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

So $-r_1 - r_2 = b$ and $r_1 r_2 = c$. $\square$

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 when someone else explains it to us things become clearer.

Example 1.2. Let n be a natural number. Prove that 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

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

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

Sometimes in order to prove that a statement is true, it is 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.

Example 1.3. There exist irrational numbers x and y such that x^y is rational.

Proof. We will prove in a later section (1.4) 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.

Case 1. 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.

Case 2. 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. Conclude that since either case 1 or case 2 must hold, and in both cases such x and y exist, the statement must be true. □

This proof is *non-constructive.* Notice that 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. (It turns out that $\sqrt{2}^{\sqrt{2}}$ is irrational, as in case 2, but proving this is non-trivial.) You will encounter plenty of non-constructive proofs in your upper-division math classes.

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

Example 1.4 (Euler's formula)**.** For all $x \in \mathbf{R}$,

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

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 $\exp' = \exp$, while $\sin' = \cos$ and $\cos' = -\sin$. 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 that the Taylor series

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

Moreover,

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots$$

and

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} - \ldots.$$

Plugging in ix and using the definition of i, namely $i^2 = -1$ we have

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

So $e^{ix} = \cos x + i \sin x$. □

It can be helpful to break up a big proof into lots of smaller parts, called *lemmata*, and work on each one separately.

Notice that in all of the above proofs, we needed to use every assumption we made. If you finished a proof and didn't use some assumption that the theorem made, then something went wrong, and either:

1. You assumed too much. In this case, the statement of the theorem you were trying to prove should be rephrased without the unnecessary assumptions.

2. 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.

3. You made an error elsewhere in the proof. In this case, fix your proof!

If you've taken linear algebra, you should try to follow this proof.

Example 1.5. If $T : V \to W$ is a linear transformation between vector spaces V and W, then the kernel $\ker(T)$ of the transformation is a vector subspace of V.

Proof. Recall the definition of the *kernel* of a linear transformation:

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

where 0_W is the zero vector of W. Also, 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 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. First, it is clear that $\ker(T)$ is a subset of V. Now, we check that $\ker(T)$ is closed under addition.

To check closure under addition, we must show that given any two arbitrary vectors in $\ker(T)$, their sum is also in $\ker(T)$. Let $x, y \in \ker(T)$. We consider $x + y$. Since T is a linear transformation,

$$T(x + y) = T(x) + T(y)$$

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

$$T(x + y) = 0_W + 0_W$$

Then, by definition of the zero vector,

$$T(x + y) = 0_W$$

By definition of $\ker(T)$, it follows that $x + y \in \ker(T)$. We can deduce that $\ker(T)$ is closed under addition.

The remaining step is to check that $\ker(T)$ is closed under scalar multiplication.

This completes the proof. □

Discussion topic 1.6 (triangle inequality). Show that for every triple of real numbers x, y, z,

$$|x - z| \leq |x - y| + |y - z|.$$

If you are brave, prove this for vectors as well.

Discussion topic 1.7 (Pythagorean theorem). Show that if a right triangle has leg lengths a and b and hypotenuse length c, then $a^2 + b^2 = c^2$. (Hint: Let ABC be a right triangle with right angle at C. Let the leg opposite A be a and the leg opposite B be b. Draw a line segment ℓ from the hypotenuse c to C. Let H be the point where c touches ℓ, and show that

$$\frac{a}{c} = \frac{BH}{a}$$

and a similar result for AH using similar triangles.)

Discussion topic 1.8. Prove that if n is an integer, then $3n^2 + n + 10$ is odd.

Discussion topic 1.9. Prove that:

1. If x is even and y is odd then $x + y$ is odd.

2. If x and y are both even then $x + y$ is even.

3. If x and y are both odd then $x + y$ is even.

Discussion topic 1.10 (Euclid's lemma)*.* Show that if p is prime and $p|ab$ (i.e. p is divisible by ab) then $p|a$ or $p|b$.

Homework 1.11*.* Show that if $a|x$ (that is, x is divisible by a) and $b|y$ then $ab|xy$.

Here's a problem for students of Math 54.

Homework 1.12*.* Verify that, in fact, $\ker(T)$ is closed under scalar multiplication.

1.2 Sets

Before we go on to learn more about proofs, we should discuss the language of mathematics. Mathematicians often talk about "the set of all x". A set is just a collection of objects, which we call elements.

Given a set X, we write $x \in X$ to mean that x is an element of X. We say that two sets X and Y are *equal*, and write $X = Y$, if *and only if* the elements of X are exactly the same as the elements of Y.

If there are finitely many elements in a set, we can just list them using commas, and begin and end the list using curly brackets. For example, $\{1, 2, 3\}$ is a set. Its elements are the numbers 1, 2, and 3.

But this might be tedious, and is hard if the set is infinite. For this, we use a special notation. If $P(x)$ is some property that elements x might have, then by

$$\{x : P(x)\}$$

we mean the set of all x such that $P(x)$ is true. In other words, $y \in \{x : P(x)\}$ if and only if $P(y)$ is true. For example,

$$\{n : n \text{ is an even integer}\}$$

is a set, whose elements are 0, 2, -2, 4, -4, and so on.

Definition 1.13. If X and Y are sets, and Y has the property that if $y \in Y$, then $y \in X$as well, then we say that Y is a *subset* of X, and write

$$Y \subseteq X.$$

If also $Y \neq X$, we write

$$Y \subset X,$$

and say that Y is a *proper subset* of X.

(Note that some books will use $Y \subset X$ just to mean that Y is a subset, not necessarily proper, of X!)

For example, Barack Obama is an element of the set P of all presidents of the United States; and if Q denotes the set of all world leaders, then $P \subset Q$. So, in particular, Barack Obama is an element of Q.

Note carefully 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.

Let's define some more sets.

Definition 1.14. Let X and Y be sets, and let $\mathcal{F}$ be a "family" of sets: a set whose elements are sets. Then:

1. The *empty set*, written $\emptyset$, is the set with no elements whatsoever,
$$\emptyset = \{\}.$$
2. The *power set* of X, written $\mathcal{P}(X)$ or 2^X, is the set of all subsets of X,
$$\mathcal{P}(X) = \{Y : Y \text{ is a set and } Y \subseteq X\}.$$
3. The *union* of X and Y, written $X \cup Y$, is the set consisting of elements in X or Y,
$$X \cup Y = \{z : z \in X \text{ or } z \in Y\}.$$
4. The *intersection* of X and Y, written $X \cap Y$, is the set consisting of elements in X and Y,
$$X \cap Y = \{z : z \in X \text{ and } z \in Y\}.$$
5. The *union* of all the sets in $\mathcal{F}$, written $\bigcup \mathcal{F}$, is
$$\bigcup \mathcal{F} = \{z : \text{ There is a set } Z \in \mathcal{F} \text{ such that } z \in \mathcal{F}\}.$$
6. The *intersection* of all the sets in $\mathcal{F}$, written $\bigcap \mathcal{F}$, is
$$\bigcap \mathcal{F} = \{z : \text{ If } Z \in \mathcal{F} \text{ then } z \in Z\}.$$
7. The *product* of X and Y, written $X \times Y$, is the set of all pairs,
$$X \times Y = \{(x, y) : x \in X \text{ and } y \in Y\}.$$

For example, the intersection of the set P of presidents of the United States and the set R of royalty of the United Kingdom, $P \cap R = \emptyset$. These definitions will be used every day in your math classes, and should be committed to memory.

Let's define some special sets, written in bold to emphasize their importance.

Definition 1.15. The following sets will be used throughout your mathematical career:

1. $\mathbf{N}$ is the set of all natural numbers, $\mathbf{N} = \{1, 2, \dots\}$.
2. $\mathbf{Z}$ is the set of all integers, $\mathbf{Z} = \mathbf{N} \cup \{0, -1, -2, \dots\}$.
3. $\mathbf{Q}$ is the set of all rational numbers.
4. $\mathbf{R}$ is the set of all real numbers.
5. $\mathbf{C}$ is the set of all complex numbers.

Notice that

$$\mathbf{N} \subset \mathbf{Z} \subset \mathbf{Q} \subset \mathbf{R} \subset \mathbf{C}.$$

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

Discussion topic 1.16. Let X, Y, and Z be sets. Show that:

1. $X \subseteq X$.
2. If $X \subseteq Y$ and $Y \subseteq X$ then $X = Y$.
3. If $X \subseteq Y$ and $Y \subseteq Z$ then $X \subseteq Z$.

Any "binary relation" (symbol you can write between a pair of elements pf a set) that has these properties is called a *partial ordering*. So you have just proved that for any set X, $\subseteq$ is a partial ordering of $\mathcal{P}(X)$. (One also says that $\mathcal{P}(X)$ is a *poset*, for "partially ordered set.")

Discussion topic 1.17 (de Morgan's laws). Suppose that A, B, C are subsets of X. Write A^c for the *complement* of A in X,

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

Show that:

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

Discussion topic 1.18. Suppose A, B, C are subsets of X. Write $A \Delta B$ for the *symmetric difference* of A and B in X,

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

Show that:

1. $x \in X$ has $x \in A \Delta B$ iff $x \in A$ or $x \in B$ but not both.

2. $(A \Delta B) \Delta (B \Delta C) = A \Delta C$.

3. $A \cap (B \Delta C) = (A \cap B) \Delta (A \cap C)$.

If you've taken some computer science classes, you might think about this in terms of logic gates. In fact, the ordered triple $(\mathcal{P}(X), \Delta, \cap)$ is often called the *Boolean ring* of X.

Discussion topic 1.19. Show that if X, Y, Z are sets, then:

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

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

Discussion topic 1.20. Find subsets e of X such that for every $A \subseteq X$:

1. $e \cup A = A$.

2. $e \cap A = A$.

3. $A \setminus e = A$.

Show that there is not a set e such that $\mathcal{P}(e) = e$. (This last one might take some ideas from later sections but it's not too hard to work out.)

Homework 1.21. Let X be a set. Show that $\bigcup \mathcal{P}(X) = X$ and $\bigcap \mathcal{P}(X) = \emptyset$.

1.3 Functions, cardinality, and infinity

One of the main purposes of the notion of set is to allow us to define what a function is.

Definition 1.22. Let X and Y be sets.

1. A *function*, *mapping*, *morphism*, or *transformation* $f : X \to Y$ 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 $x \mapsto y$.

2. X is called the *domain* of f, and Y is called the *codomain* of f.

3. The set
$$f(X) = \{y \in Y : \text{ There is a } x \in X \text{ such that } f(x) = y\}$$
is called the *image* of f.

4. If $g : Y \to Z$ is also a function, then the function $g \circ f : X \to Z$ is defined by
$$(g \circ f)(x) = g(f(x)).$$

For example, we can define a function $f : \mathbf{N} \to \mathbf{N}$ by requiring that $f(n) = n^2$. Then the imgae of f is the set of all square numbers, $f(\mathbf{N}) = \{1, 4, 9, 16, \dots\}$. We could also define a function $g : \mathbf{R} \to \mathbf{C}$ by requiring that $g(x) = \sqrt{x}$.

But a function $X \to Y$ assigns exactly one element of Y to each element of X. So the rule which sends $x \in \mathbf{R}$ to the $y \in \mathbf{R}$ such that (x, y) lies on the unit circle is not a function, since if $y \neq 0$, $(x, -y)$ also lies on the unit circle.

Definition 1.23. Let $f : X \to Y$ be a function. Then:

1. We say f is *injective*, or maps *one to one*, if, whenever $f(x_1) = f(x_2)$, we already know $x_1 = x_2$.
2. We say f is *surjective*, or maps *onto* Y, if its image $f(X) = Y$.
3. If f is both injective and surjective, we say that f is a *bijection*, or *correspondence.*
4. If f is a bijection, we say that X and Y have the same *cardinality.*

Proposition 1.24 (first isomorphism theorem of sets)**.** *For every function $f : X \to Y$, there are sets $X_0 \subseteq X$ and $Y_0 \subseteq Y$ and functions $\pi : X \to X_0$, $\tilde{f} : X_0 \to Y_0$, and $\iota : Y_0 \to Y$ such that π is surjective, $\tilde{f}$ is bijective, and ι is surjective, and such that*

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

In the context of set theory, "isomorphism" is just a synonym for bijection. (In other branches of math, such as linear algebra, an isomorphism is a special type of bijection!) There are other isomorphism theorems that prove that certain functions are bijections. We often express the conclusion of the first isomorphism theorem by saying that "the diagram

$$\begin{tikzcd}
X \arrow[r, "f"] \arrow[d, "\pi"] & Y \\
X_0 \arrow[r, "\tilde{f}"] & Y_0 \arrow[u, "\iota"']
\end{tikzcd}$$

commutes"; i.e. if you were to start with an $x \in X$ and apply the functions ("arrows") to x in the order that they appear, you'd end up with the same result no matter whether you took the path f or the path $\pi, \tilde{f}, \iota$.

Proof. Start with $X_0 = X$, and go through its elements one by one. If at any point we find a $x_1 \in X_0$ and a $x_2 \in X_0$ such that $x_2 \neq x_1$ and $f(x_1) = f(x_2)$, remove x_2 from X_0. We repeat this process until we have exhausted all elements of X. This defines X_0.

Let $\tilde{f}$ denote the restriction of f to X_0; that is, $\tilde{f}$ is the function $X_0 \to Y$ such that $\tilde{f}(x) = f(x)$ for all $x \in X_0$. If $\tilde{f}(x_1) = \tilde{f}(x_2)$, then $x_1 = x_2$ (or else x_2 was removed from X_0, so $\tilde{f}(x_2)$ is not defined), so $\tilde{f}$ is injective.

We define $\pi : X \to X_0$ to be function which sends $x_1 \in X$ to the $x_2 \in X_0$ such that $f(x_1) = \tilde{f}(x_2)$. Since $\tilde{f}$ is injective, π is actually a function. If $x \in X_0$, then $x \in X$ and $\pi(x) = x$. So π is surjective.

We now define $Y_0 = \tilde{f}(X_0)$. Then if we think of $\tilde{f}$ as a function $X_0 \to Y_0$, $\tilde{f}$ is surjective. Therefore $\tilde{f}$ is a bijection. Moreover, we let $\iota(y) = y$, so ι is defined on all of Y_0. If $\iota(y_1) = \iota(y_2)$ then $y_1 = y_2$ by definition of ι, so ι is injective.

Finally, notice that since ι is just the identity on Y_0,

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

By definition of π, $\tilde{f} \circ \pi = f$. Therefore

$$f = \iota \circ \tilde{f} \circ \pi,$$

as promised. □

Proposition 1.25. *Let $n \in \mathbf{N}$, let X be a set, and let $Z_n = \{1, 2, \ldots, n\}$. Then X has exactly n elements if and only if Z_n and X have the same cardinality.*

Proof. First, if X has n elements, then we can define a function $f : Z_n \to X$ as follows. Choose an arbitrary element $x_1 \in X$ and let $f(1) = x_1$. Then choose an arbitrary element $x_2 \in X$ such that $x_2 \neq x_1$ and let $f(2) = x_2$. Then choose an arbitrary element $x_3 \in X$ such that $x_3 \neq x_2$ and $x_3 \neq x_1$ and let $f(3) = x_3$. Repeat this process until we have defined $f(n)$ for every n. Since we chose a different element of X for each n, f is injective. Moreover, f is surjective, since there were only n elements of X to work with and we used all of them. So Z_n and X have the same cardinality.

On the other hand, if Z_n and X have the same cardinality, there is a bijection $f : Z_n \to X$. So the first element of X is $f(1)$, the second element of X is $f(2)$, and so on. Repeating, we count every single element of X in n steps since f is surjective and we don't double-count since f is injective. Therefore X has n elements. □

Definition 1.26. If X has n elements for some $n \in \mathbf{N}$ then we say that X is *finite*, and write $|X| = n$. Otherwise, we say that X is *infinite*.

Proposition 1.27. *Suppose that X and Y have the same cardinality and are finite, and $f : X \to Y$ is a function. Then* the following are equivalent *(i.e., if one of the following are true, then all of them automatically are):*

1. *f is injective.*

2. *f is surjective.*

3. *f is bijective.*

Proof. To prove that multiple properties are equivalent, we must show that if 1 is true, then 2 is true; that if 2 is true, then 3 is true; and that if 3 is true, then 1 is true. So the proof naturally breaks down into three parts:

1. Suppose that f is injective. We need to show that f is surjective, but f maps X onto its image $f(X)$, so X and $f(X)$ have the same cardinality. Therefore if X has n elements, so does $f(X)$. But we know that Y also has the same cardinality as the X, so Y has n elements. Therefore the complement $f(X)^c$ in Y, $f(X)^c = \{y \in Y : y \notin f(X)\}$, has $n - n$ elements, hence $f(X)^c = \emptyset$. Therefore $f(X) = Y$, so f is surjective.

2. Suppose that f is surjective. We need to show that f is bijective, and to do this we just need to find an inverse of f. Let $y \in Y$. Since f is surjective there is an $x \in X$ such that $f(x) = y$. Choose *any such* x; then g is injective by definition. So g is surjective, and hence a bijection, by the above argument (since X and Y have the same cardinality). Since $g = f^{-1}$ it follows that f is a bijection.

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

□

Let's wrap up with a definition that we'll use in the next section.

Definition 1.28. Let X be a set. If there is a surjective function $f : \mathbf{N} \to X$, then we say that X is *countable*. Otherwise, we say that X is *uncountable*.

Proposition 1.29. $\mathbf{Q}$ *is countable.*

Proof. Let us consider an infinite grid extending forever in both directions, where the horizontal part is indexed by the integers and the vertical part is indexed by the positive integers. So the first few entries of the grid look something like this:

$(-3,1)$	$(-2,1)$	$(-1,1)$	$(0,1)$	$(1,1)$	$(2,1)$	$(3,1)$
$(-3,2)$	$(-2,2)$	$(-1,2)$	$(0,2)$	$(1,2)$	$(2,2)$	$(3,2)$
$(-3,3)$	$(-2,3)$	$(-1,3)$	$(0,3)$	$(1,3)$	$(2,3)$	$(3,3)$

Let's now replace the grid with fractions, with the first coordinate corresponding to the numerator and the second coordinate corresponding to the denominator. Reducing the fractions, this grid looks like

-3	-2	-1	0	1	2	3
$-3/2$	-1	$-1/2$	0	$1/2$	1	$3/2$
-1	$-2/3$	$-1/3$	0	$1/3$	$2/3$	1

This grid contains every rational number, since if a/b is a fraction and $b > 0$, then a/b can be found at the coordinate (a, b) (and if $b < 0$, then it can be found at $(-a, -b)$. We leave it as an exercise to consider the case $b = 0$). Some are counted twice, but that's fine; the cardinality of $\mathbf{Q}$ will be bounded above by the cardinality of the grid, so we just have to show that the grid is countable.

Now draw a snake starting at $(0, 1)$, going up to $(1, 1)$, down to $(1, 2)$ and then left through $(0, 2)$ to $(-1, 2)$ and then up to $(-1, 1)$. Go left again to $(-2, 1)$, down through $(-2, 2)$ to $(-2, 3)$, and then right all the way to $(2, 3)$, and up to $(2, 1)$. Go to the right to $(3, 1)$ and repeat this snaking process forever, which will cover the whole grid eventually. This gives an enumeration

1. 0
2. 1
3. 1/2
4. 0

5. $-1/2$
6. -1
7. -2
8. -1
9. $-2/3$
10. $-1/3$
11. 0
12. $1/3$
13. $2/3$
14. 1
15. 2
16. 3
17. $\cdots$

of $\mathbf{Q}$. □

Discussion topic 1.30. Let us write $X \sim Y$ to mean that X and Y have the same cardinality. Prove that, for any sets X, Y, and Z:

1. $X \sim X$.
2. If $X \sim Y$, then $Y \sim X$.
3. If $X \sim Y$ and $Y \sim Z$, then $X \sim Z$.

Any binary relation with these three properties is called an *equivalence relation*. So you have just proven that $\sim$ is an equivalence relation.

Discussion topic 1.31 (inclusion-exclusion principle). Let X and Y be finite sets. Show that

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

Discussion topic 1.32. Let $f : X \to Y$ and $g : Y \to Z$ be functions. Show that:

1. If f and g are injective, then $g \circ f$ is injective.
2. If f and g are surjective, then $g \circ f$ is surjective.
3. If f and g are bijective, then $g \circ f$ is bijective.

Discussion topic 1.33 (Hilbert's grand hotel). The result of Proposition 1.27 is not true in infinite sets. To see why, answer the following riddle, and prove that your proposed solution is actually possible.

Suppose that 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?

Prove that the above phenomenon is possible in any infinite set. In other words, the conclusion of Proposition 1.27 holds if and only if a set is infinite. (This was *Dedekind's definition of infinity*.)

Discussion topic 1.34. Prove the following basic facts about countability:

1. If X is finite, then X is countable.

2. If X is countable and $Y \subseteq X$, then Y is countable.

3. If X is countable, then $X \times X$ is countable.

4. If X is countable and infinite, then X has the same cardinality as $\mathbf{N}$.

5. If X is uncountable, then X is infinite.

Homework 1.35 (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:

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

2. 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.

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

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

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

Would you take this offer?

Homework 1.36. Let $f : X \to Y$ be any function. Let $A, B \subseteq X$. Show that $f(A \cap B) \subseteq f(A) \cap f(B)$ and $f(A \cup B) = f(A) \cup f(B)$.

Why didn't I ask you to prove that $f(A \cap B) = f(A) \cap f(B)$?

Homework 1.37. Let X be a set, and let $B(X)$ denote the set of functions $X \to \{0,1\}$. Show that $B(X)$ has the same cardinality as $\mathcal{P}(X)$.

Why is $\mathcal{P}(X)$ often called 2^X?

Homework 1.38. For those who have taken Math 54, show that there is a "first isomorphism theorem of vector spaces". That is, if V and W are (say, finite-dimensional) vector spaces and $T : V \to W$ is a linear transformation, there are subspaces (not just subsets!) $V_0 \subseteq V$ and $W_0 \subseteq W$, and linear transformations (not just functions!) $\pi : V \to V_0$, $\iota : W_0 \to W$, and $\tilde{T} : V_0 \to W_0$ such that π is surjective, ι is injective, and $\tilde{T}$ is invertible, and such that the diagram

$$\begin{tikzcd}
V \arrow[r, "T"] \arrow[d, "\pi"] & W \\
V_0 \arrow[r, "\tilde{T}"] & W_0 \arrow[u, "\iota"']
\end{tikzcd}$$

commutes; in other words,

$$T = \iota \circ \tilde{T} \circ \pi.$$

Moreover, prove that $\dim V_0 + \dim \ker T = \dim V$ and $\dim W_0 = \operatorname{rank} T$. To accomplish this, you'll probably want to use rank theorem and the invertible matrix theorem.

1.4 Proof by contradiction

Proofs by contradiction follow this outline: if the statement were false, then we can prove contradictory facts from its falsehood, so the statement must be true. Perhaps the most classic example of proof by contradiction is the proof that $\sqrt{2}$ is irrational.

Example 1.39. $\sqrt{2}$ is irrational.

Proof. Suppose $\sqrt{2}$ is rational. We will prove contradictory facts from this assumption. If $\sqrt{2}$ is rational, then there exist non-zero integers a and b such that

$$\sqrt{2} = \frac{a}{b}$$

Furthermore, we may assert that such a and b exist where they are coprime to each other (they share no common factors besides 1) so that the fraction $\frac{a}{b}$ is in its simplest form. Square each side of the equation to get

$$2 = \frac{a^2}{b^2}$$

It follows that

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

so a^2 is even and therefore a is even. Let $a = 2c$, for some integer c. We now have

$$\begin{aligned} 2b^2 &= (2c)^2 \\ 2b^2 &= 4c^2 \\ b^2 &= 2c^2 \end{aligned}$$

Then b^2 is even and therefore b is even. We have shown that both a and b are even, which means they share 2 as a common factor. From our assumption that $\sqrt{2}$ is rational, we managed to show that its fractional representation $\frac{a}{b}$ both exists in reduced form and *does not* exist in reduced form. This is a contradiction. It must be the case, then, that $\sqrt{2}$ is irrational. □

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.

Theorem 1.40 (Cantor's diagonal argument)**.** *The set of real numbers is uncountable.*

Proof. It is sufficient to show that the interval $(0, 1)$ in $\mathbf{R}$ is uncountable. Suppose the interval $(0, 1)$ is countable. Then we can enumerate the reals in $(0, 1)$ in a numbered list.

$$\begin{aligned}
x_1 &= 0.\mathbf{0}12345...\\
x_2 &= 0.1\mathbf{4}1592...\\
x_3 &= 0.10\mathbf{1}010...\\
x_4 &= 0.500\mathbf{0}00...\\
x_5 &= 0.1020\mathbf{0}3...\\
x_6 &= 0.05159\mathbf{8}...\\
&\vdots
\end{aligned}$$

We will construct another real number $x \in (0, 1)$ that is not on this list. The i-th digit (after the decimal place) of x will be 1 if the i-th digit of x_i is 0. Otherwise, the i-th digit of x will be 0. Thus in our example,

$$x = 0.100110...$$

For all i, the i-th decimal place of x differs from the i-th decimal place of x_i, so it cannot be that x is in the list. This contradicts our assumption that $(0, 1)$ could be enumerated. Conclude that the reals are uncountable. □

Between 1890 and 1920, there was a flurry of interest in the properties of sets due to Cantor's discovery. Many mathematicians thought that if P was *any* property whatsoever, they could define

$$\{x : P(x)\}$$

without any trouble. Unfortunately, Bertrand Russell and Ernst Zermelo showed that this was nonsense.

Theorem 1.41 (Russell's paradox)**.** *There does not exist a set R such that*

$$R = \{X : X \textit{ is a set, and } X \notin X\}.$$

Proof. Assume that R exists. Then either $R \in R$ or $R \notin R$, so let's do a proof by cases.

1. If $R \in R$, then it is not true that $R \notin R$, so $R \notin R$, by definition of R. Therefore $R \notin R$, which is a contradiction.

2. Otherwise, $R \notin R$. So since R is a set, $R \in R$ by definition of R. So this is also a contradiction.

Therefore contradiction is unavoidable, and R does not exist. □

To avoid this paradox, we usually will assume that sets do not contain themselves. That is, if X is a set, then $X \notin X$, regardless of how we defined X.

Here's a famous application of proof by contradiction to computer science.

Example 1.42 (The undecidability of the halting problem)**.** It is impossible to design a computer program that can determine whether any computer program will finish running.

Proof. Suppose that such a program P exists: given any program Q, P will say "halts" if Q will eventually stop running, but "doesn't halt" otherwise.

Now consider a program R which obeys the following rules:

1. First, R runs P on itself.

2. If P says "halts", R keeps running forever.

3. Otherwise, if P says "doesn't halt", then R stops running.

If you know a programming language such as Python, you can easily write code for R, so R definitely exists.

Either P says "halts" when applied to R, or not, so again we do a proof by cases, similar to in the proof of Russell's paradox:

1. If P says "halts," then R does not halt. So P actually must say "doesn't halt," which is a contradiction.

2. If P says "doesn't halt," then R halts, so P actually must say "halts."

Either way, we cannot avoid a contradiction. □

Discussion topic 1.43 (pigeonhole principle). Suppose that $n > m$ are natural numbers, and that we have put n elements in m disjoint sets (i.e., if two of the sets are X and Y, and $X \neq Y$, then $X \cap Y = \emptyset$). Prove that at least one of the sets has at least two elements. (That is, if n pigeons came to roost in m nests ["pigeonholes"], then one of the nests must contain at least one pigeon.)

Discussion topic 1.44 (Cantor's paradox). Prove that if X is a set, then there is an injective function $X \to \mathcal{P}(X)$, but not an injective function $\mathcal{P}(X) \to X$. (Because of this, we say that $\mathcal{P}(X)$ has a higher cardinality, or simply that $\mathcal{P}(X)$ is bigger, than X, and write $|\mathcal{P}(X)| > |X|$.) Why does this imply that there are infinitely many infinite cardinalities?

Discussion topic 1.45 (Euclid's theorem on primes). Prove that there are infinitely many prime numbers. (Hint: If there are only n prime numbers $p_1, \dots, p_n$, prove that $p_1 \dots p_n + 1$ is composite.)

Discussion topic 1.46. Let a, b, c be odd integers. Show that the quadratic equation $ax^2 + bx + c = 0$ has no solutions in $\mathbf{Q}$, without appealing to the quadratic formula.

Discussion topic 1.47. Prove that if $k \in \mathbf{N}$ is not a perfect square, then $\sqrt{k}$ is irrational. If you are brave, do the same for if k is not a pth power, for p some prime number. (This proof technique is called *proof by infinite descent*. Do you see why?)

Homework 1.48. Let n be a *perfect cube* (i.e. a number $n \in \mathbf{N}$ such that there exists $m \in \mathbf{N}$ with $m^3 = n$.) Show that:

1. If n is even, then n is divisible by 8.

2. If n is odd, then n is not divisible by $64 = 8^3$.

Homework 1.49. Let $n \in \mathbf{N}$. Show that for every $p \in \mathbf{N}$ there exist $k, \ell \in \mathbf{N}$ such that $n^k - n^\ell$ is not divisible by p.

Homework 1.50. Let $A \subset \mathbf{N}$ be a set of $n \in \mathbf{N}$ numbers. Show that two of them have the same remainder when divided by p, as long as $p < n$.

1.5 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:

"Theorem" 1.51. *For any $k \in \mathbf{N}$, 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. □

But $2^{2^5} + 1$ is not prime; it factors as $641 \cdot 6700417$. So, *inductive reasoning is invalid in mathematics.*

But we have something even more powerful.

Theorem 1.52 (principle of induction)**.** *Let $P(n)$ be a statement indexed by $n \in \mathbf{N}$, the set of all natural numbers.*

To show $P(n)$ is true for all n, it suffices to show that

1. *(Zero stage) $P(1)$ is true.*

2. *(Successor stages) Assume $P(k)$ is true, then show $P(k+1)$ is true.*

We're assuming that the first natural number is 1, but some writers will take $0 \in \mathbf{N}$. Be aware!

Let us see an example of this in practice.

Example 1.53. For all n, $S_n = 1 + 3 + \ldots + 2n - 1$ is a *perfect square*; i.e. there is some j such that $j^2 = S_n$.

Proof. First we see that $S_1 = 1$. This is a perfect square, so the base case is done.

Now we 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 $j \in \mathbf{N}$.

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. □

Since 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,

$$\begin{aligned} S_1 &= 1 & &= 1 = 1^2; \\ S_2 &= 1 + 3 & &= 4 = 2^2; \\ S_3 &= 1 + 3 + 5 & &= 9 = 3^2; \\ S_4 &= 1 + 3 + 5 + 7 & &= 16 = 4^2. \end{aligned}$$

It seems that $S_k = k^2$! This is a stronger statement than what we were supposed to prove, and if it's true, then it's easier to prove, since we know what the j in the perfect square is – it's just k.

Proof of Example 1.53, encore. Again, $S_1 = 1$ and we're done with the base case.

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

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

and we're done. □

Let us rephrase the principle of induction a bit.

Theorem 1.54 (principle of induction)**.** *Let $S \subseteq \mathbf{N}$. Suppose that*

1. *$1 \in S$.*

2. *If $k \in S$, then $k + 1 \in S$.*

Then $S = \mathbf{N}$.

Proof. Let $P(n)$ be the statement that $n \in S$. Then $P(1)$ is true, and if $P(k)$ is true, then $P(k + 1)$ is true. So $P(n)$ is true for each n, by the principle of induction. □

Definition 1.55 (extreme elements)**.** Let $X \subseteq \mathbf{R}$ and $x \in X$. Then:

1. x is the *least element* or *minimum* of X, written $x = \min X$, if for each $y < x$, $y \notin X$.

2. x is the *greatest element* of *maximum* of X, written $x = \max X$, if for each $y > x$, $y \notin X$.

Theorem 1.56 (well-ordering principle)**.** *Any non-empty subset of* $\mathbf{N}$ *has a least element.*

Theorem 1.57. *The well-ordering principle is enough to prove the principle of induction, and the principle of induction is enough to prove the well-ordering principle.*

Proof. First we'll assume the well-ordering principle.

Given a set S satisfying (i) and (ii) of 1.54 we want to show $S = \mathbf{N}$. Certainly S is non-empty, as $1 \in S$, so applying WOP to it gives us ... nothing, as 1 is the smallest element in $\mathbf{N}$, so we already knew that S has a smallest element. Now it seems we are stuck. What can we do? □

We'll have to use proof by contradiction! But what is the negation of the principle of induction? The principle of induction is an implication: *if* there is some set S satisfying (i) and (ii), *then* $S = \mathbf{N}$. So how does this go, that is what is the negation of $P \implies Q$? Well, it turns out that $P \implies Q$ is the same as $(\neg P) \vee Q$. Besides this, de Morgan's law says that

$$\neg(A \vee B) = (\neg A) \wedge (\neg B)$$

so

$$\begin{aligned}\neg(P \implies Q) &= \neg(\neg P \vee Q)\\ &= P \wedge \neg Q\end{aligned}$$

so our job is to assume that there is a set S satisfying (i) and (ii) but $S \neq \mathbf{N}$, and contradict the well-ordering principle.

Proof of Theorem 1.57, encore. Returning to our problem, suppose that the principle of induction isn't true. Then there is a set S such that (i) and (ii) are true and $S \neq \mathbf{N}$. Because $S \neq \mathbf{N}$ this means that $\mathbf{N} \setminus S$ is nonempty so applying the well-ordering principle we have a minimal element of $\mathbf{N} \setminus S$, say k. What does this mean? Well, $k - 1 < k$, so $k \notin \mathbf{N} \setminus S$, so $k - 1 \in S$. But by (ii) we see that $(k - 1) + 1 = k \in S$, which is a contradiction. So the principle of induction must be true, and this direction is done.

Now suppose the principle of induction. We want to show the well-ordering principle, so given a nonempty subset S of $\mathbf{N}$, we need to show that it has a least element.

I have no information about S; in particular S may not satisfy conditions (i) and (ii), so we can't apply induction. We're stuck again.

So we argue by contradiction. Suppose S does not have a least element. Then what elements could possibly be in S?

If 1 is in S, then 1 is the least element of S, since 1 is the least element of $\mathbf{N}$. So $1 \notin S$. By the same reasoning, 2 is not in S, 3 is not in S, and so on...

We'll formalize this using induction. We have that $1 \notin S$, and in particular $1 \in \mathbf{N} \setminus S$. Similarly, if $1, \ldots, k \notin S$, then $1, \ldots, k \in \mathbf{N} \setminus S$, which is enough to imply that $k+1 \in \mathbf{N} \setminus S$

(since otherwise $k+1$ would be the least element of S). Therefore by induction, $\mathbf{N} \setminus S = \mathbf{N}$, which is to say that $S = \emptyset$, the empty set.

But we already have assumed that S is not empty. So this is a contradiction, and thus the well-ordering principle must be true. □

Notice that the above theorem implies that the only property of $\mathbf{N}$ that we used was that it was well-ordered. So if X is *any* well-ordered set and we want to prove some statement about every $x \in X$, we can do induction on X. On the other hand, if X is finite, we could just check every element manually. So $\mathbf{N}$ is the "happy medium": there are infinitely many cases to check, one for each $n \in \mathbf{N}$, but $\mathbf{N}$ is well-ordered, so we can use induction.

The moral is that whenever you see a statement of the form

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

you should try induction first.

But be wary. It's easy to make silly mistakes if you don't complete the whole process of an inductive argument.

"Theorem" 1.58. *For all n, $\frac{d}{dx}(x^n) = 0$.*

Bad proof. Clearly $\frac{d}{dx}(1) = 0$. Our inductive hypothesis will be $(x^k)' = 0 \ \forall k \leq n$. For the inductive step,

$$(x^{n+1})' = (x^n \cdot x)' = x^n(x)' + x(x^n)' = x^n 0 + x0 = 0.$$

□

So what went wrong? It turns out that while the above manipulation is valid for all $n \geq 1$ it isn't for $n = 0$ – since it's false for x^1, this incorrect step allowed the rest to follow.

Discussion topic 1.59. Let

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x}\, dx$$

and prove that

$$\Gamma(n+1) = n!$$

where $n! = (n-1)!n$ if $n \geq 1$, and $0! = 1$.

Discussion topic 1.60. Let F_n be the *Fibonacci sequence* defined by $F_n = F_{n-1} + F_{n-2}$ and $F_0 = 0, F_1 = 1$, and show

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

Discussion topic 1.61. Let $f : \mathbf{Q} \to \mathbf{R}$ be a function such that for any x, y,

$$f(x+y) = f(x) + f(y).$$

(Such a function is called a *morphism of groups*.) Show that for every x, one has $f(x) = xf(1)$.

Discussion topic 1.62. Assume that for each two cities x, y in the country of Topologia, there is a one-way road connecting x and y (but possibly not a two-way road!) Show that there is a path that passes through every city in Topologia. (Hint: Assume that there is a path that passes through the first k cities $C_1, \dots, C_k$. If there is a path from C_k to C_{k+1}, we're done. Otherwise, find $i < k$ such that there are roads from C_i to C_{k+1} and C_{k+1} to C_{i+1}.)

Discussion topic 1.63 (Pick's theorem). Let P be a polygon in $\mathbf{R}^2$ with area A_P. Say that a point $p \in P$ is a *lattice point* if both coordinates of p are integers. Let i be the number of lattice points inside P and b the number of lattice points on the boundary of P. Show that

$$A_P = i + \frac{b}{2} - 1.$$

The "hard part" is that Pick's theorem is true for *triangles*, which you can take as a given (but try to prove it for fun!) (Hint: Write $P = P' \cup T$ where P' is a polygon, T is a triangle, and $P' \cap T$ is a line segment. Assume that Pick's theorem is true for P' and T and use $A_P = A_{P'} + A_T$.)

Homework 1.64. Show that

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

Homework 1.65 (Generalized Euclid's lemma). Prove that if p is prime and $p | a_1 \dots a_n$ then there is some $i \in \{1, \dots, n\}$ such that $p | a_i$.

Homework 1.66. Show that if $X_1, \dots, X_n$ are countable sets and

$$X = X_1 \times X_2 \times \cdots \times X_n,$$

then X is countable.

Homework 1.67. Show that

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

Homework 1.68. Show that

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

1.6 Existence and uniqueness

Let's say, in the course of everyday conversation, I were to say, "Aidan has a younger sister." From the way I phrased that sentence, it sounds like Aidan has one younger sister, and no others.

This use of language disagrees with how mathematicians speak. If I were to say "The number 6 has a prime factor," that would be technically true, since 2 is a prime factor of 6, but of course 6 has another prime factor, namely 3. For this reason, mathematicians (and so you!) are very careful about distinguishing between *existence* and *uniqueness*.

To say that some object x exists is to say that there is at least one example. On the other hand, to say that x is unique is to say that there is at most one example: *if* x exists at all (which is not guaranteed by the claim that x is unique!), there is only one.

Sometimes the most natural way to show existence is to give an explicit example of x, and in a subproof show that x is the kind of object we want to show it is.

Definition 1.69. If $n \in \mathbf{N}$ is a natural number, a *prime factorization* of n is a way of writing it as

$$n = p_1 p_2 \dots p_m$$

where each of the p_i is a prime number.

Theorem 1.70 (fundamental theorem of arithmetic, existence)**.** *Every natural number $n \geq 2$ has a prime factorization.*

Proof. Since this is a statement about every natural number (except 1 – so we have to start at 2), we proceed by induction.

Since 2 is prime, it has a prime factorization, namely itself. This is our base case (notice that since we don't care about 1, the induction starts at 2.)

Otherwise, suppose $n > 2$ and that all numbers from 2 to $n-1$ have prime factorizations. We now argue by cases: either n is prime, or n is composite.

If n is prime, then it has a prime factorization, namely itself, so we're done.

Otherwise, n is composite, so there are numbers a and b such that $1 < a \leq b < n$ and $n = ab$. But a and b both have prime factorizations, say

$$a = p_1 p_2 \dots p_n$$
$$b = q_1 q_2 \dots q_m.$$

So n has a prime factorization, namely

$$n = p_1 p_2 \dots p_n q_1 q_2 \dots q_m.$$

□

This proof of the fundamental theorem of arithmetic actually gives instructions for explicitly writing down the prime factorization: just keep dividing by factors until you eventually get prime factors. Cantor's diagonal argument was similar: given an enumeration, we got explicit instructions for finding a real number that wasn't in the enumeration.

Sometimes we aren't so lucky. The proof that there was a pair of irrational numbers x and y such that $x^y \in \mathbf{Q}$ was nonconstructive, because we proved that either $\sqrt{2}^{\sqrt{2}}$ was rational, or $\sqrt{2}^{\sqrt{2}^{\sqrt{2}}}$ was rational without checking which one. (It turns out that $\sqrt{2}^{\sqrt{2}}$ is irrational, but the proof of this is much harder.) Let's see an example where it is *impossible* to give constructive instructions.

Theorem 1.71 (Kőnig's lemma)**.** *Let G be a tree, such that every vertex is has only finitely many children. If G has infinitely many vertices, then G has an infinite path.*

For those who don't know what a tree is, imagine a family tree, except that there could be an infinite number of generations, and a parent could have infinitely many children. Each individual is called a "vertex," and the "root" is the oldest member of the family, who everyone is descended from.

Proof. Let v_1 be the root of the tree.

We want to prove the following lemma:

Lemma 1.72. *For each natural number n, if $v_1 \dots v_n$ is a path to a node v_n with infinitely many descendents, then there is a child of v_n, say v_{n+1}, such that v_{n+1} also has infinitely many descendents.*

Since we want to prove a statement about every natural number, we proceed by induction. If $n = 1$, then v_1 has infinitely many descendents, since it's the root of the tree.

Otherwise, assume that $v_1 \dots v_n$ is a path to a vertex v_n with infinitely many descendents. We need to prove that there is some child of v_n with infinitely many descendents.

Now suppose *not*, and argue by contradiction.

Indeed, if not, then every child of v_n has finitely many descendents. The set of descendents of v_n is the set of all children of v_n and the union of their descendents. But all of those sets are finite, and there's finitely many of them. So the set of descendents of v_n is finite. But we have already established that this set is infinite, so this is a contradiction. □

We don't know anything about this child, only that if it didn't exist, we'd get a contradiction. It would take an infinite amount of time to check every descendent of a child with an infinitely descending branch, so we couldn't just sit down and brute force a computation. Nevertheless, one exists!

Proof of Example 1.71, continued. Therefore, v_n has a child with infinitely many descendents, which completes the induction.

Choose one such descendent to get a child v_{n+1}, which extends the path. The path is $v_1 \dots v_n \dots$. □

Notice how we used induction to prove existence. In general if you want to prove existence of infinitely many things, ordered like $\mathbf{N}$, this isn't a bad way to do it.

What does it mean for a mathematical object to be *unique*? Again, uniqueness means that there is *at most one* of that object (but there possibly could be none at all). So existence and uniqueness together imply that there is *exactly* one.

A common method of showing that an object is unique is to show that any two instances of the object must be identical. This is best demonstrated through examples.

Example 1.73. Say that a function $f : \mathbf{R} \to \mathbf{R}$ is *strictly convex* if its second derivative $f''(t) > 0$ for all $t \in \mathbf{R}$. If f is strictly convex and f has a minimum, then its minimum is unique.

Proof. Assume that x and y are minima of f. *Without loss of generality*, we can assume that $x \leq y$. This means that we are making a harmless assumption (we don't know anything else about x and y, and either $x \leq y$ or $x \leq y$; the proofs will be exactly the same in either case but with the variables swapped, so we might as well assume $x \leq y$.)

Then, by some calculus, the first derivative $f'(x) = 0$ and $f'(y) = 0$. On the other hand, since $f''(t) > 0$ for all t, f' is strictly increasing: since $x \le y$, $f'(x) \le f'(y)$, and if $x < y$ then $f'(x) < f'(y)$. Clearly if $x < y$ this leads to the contradiction $0 < 0$. So $x = y$. □

Here's an example for students who have taken Math 54.

Example 1.74. The additive identity of a vector space is unique.

Proof. Let V be a vector space. To show that the additive identity of V is unique, we will suppose there are two instances of the additive identity, 0 and $0'$, in V, and show that they are in fact equal. Let $v \in V$. Then

$$v + 0 = v + 0'$$

Since V is a vector space, v has an additive inverse $-v$. Adding $-v$ to each side yields

$$\begin{gathered} v + 0 + (-v) = v + 0' + (-v) \\ 0 + 0 = 0 + 0' \\ 0 = 0' \end{gathered}$$

Conclude that the additive identity of a vector space is unique. □

Example 1.75 (Kőnig's lemma, again)**.** Let G be the *infinite binary tree*: the tree such that every node has exactly two children. Then the infinite path guaranteed by Kőnig's lemma (Theorem 1.71) is NOT unique.

Proof. We just need to show that there are two infinite paths through G which are distinct. By Kőnig's lemma, there is a path $v_1 \ldots v_n \ldots$ through G. Since G is the infinite binary tree, v_1 has a child $w_2 \neq v_2$. Since w_2 does not appear in the path $v_1 \ldots v_n \ldots$, any infinite path containing w_2 is not equal to $v_1 \ldots v_n \ldots$, disproving uniqueness.

We apply Kőnig's lemma again, to the tree consisting of w_2 and all its descendents, to get a path $w_2 \ldots w_n \ldots$. Then $v_1 w_2 w_3 \ldots w_n \ldots$ is an infinite path through G which is certainly not equal to $v_1 v_2 v_3 \ldots v_n \ldots$. So infinite paths through G are not unique. □

Here's an example from Math 54.

Example 1.76. Let V and W be vector spaces, and let $T : V \to W$ be a linear transformation from V to W. T is one-to-one if and only if $\ker(T) = \{0_V\}$.

Proof. ($\Longrightarrow$) Suppose T is one-to-one (injective). The proof in this direction is a uniqueness proof in disguise. We know that $0_V \in \ker(T)$ from the fact that T is a linear transformation. Then, our goal is to show that 0_V is the *only* element in $\ker(T)$.

Let $x \in \ker(T)$. Then $T(x) = 0_V$. Since T is a linear transformation, we also know $T(0_V) = 0_V$. So

$$T(x) = T(0_V)$$

Recall that T is one-to-one if for all $a, b \in V$, $T(a) = T(b)$ implies that $a = b$. Applying the definition of one-to-one to $T(x) = T(0_V)$ allows us to deduce that

$$x = 0_V$$

Thus we have shown that any arbitrary element in the kernel must be the 0_V, so $\ker(T)$ is exactly $\{0_V\}$.

($\Longleftarrow$) Suppose $\ker(T) = \{0_V\}$. We will apply the uniqueness of 0_V. Let $x, y \in V$ such that $T(x) = T(y)$. Then

$$T(x) - T(y) = 0_V$$

By linearity of T,

$$T(x - y) = 0_V$$

Then, by the definition of the kernel of T, it follows that $x - y \in \ker(T)$. However, we know $\ker(T) = \{0_V\}$, so it must be that $x - y = 0_V$.

$$\begin{aligned} x - y &= 0_V \\ x &= y \end{aligned}$$

We have shown that $T(x) = T(y)$ implies that $x = y$ for all $x, y \in V$, which by definition means T is one-to-one. □

Here's a useful principle: *to prove that a function is injective (one-to-one) is a proof of uniqueness*, and *to prove that a function is surjective (onto) is a proof of existence.* Indeed, to prove that $f : X \to Y$ is injective is to prove that for each $y \in Y$, the $x \in X$ such that $f(x) = y$ is unique. On the other hand, to prove that f is surjective is to prove that for each $y \in Y$, the $x \in X$ such that $f(x) = y$ exists.

Discussion topic 1.77. A function $f : \mathbf{N} \to \mathbf{N}$ is said to be *increasing* if $f(m) < f(n)$ for any $m < n$. Show that if f is increasing, then f is injective, and that there is a unique f which is increasing and surjective. Are similar results true for $\mathbf{Z}$ or $\mathbf{R}$? What about $(\mathcal{P}(X), \subset)$ for some set X?

Discussion topic 1.78 (division algorithm, existence). Let $a, b \in \mathbf{N}$. Show that there are $q, r \in \mathbf{N}$ such that

$$a = bq + r.$$

Also, show that we can choose $r < b$.

Discussion topic 1.79 (division algorithm, uniqueness). Let $a, b \in \mathbf{N}$. Show that if there are $q, r \in \mathbf{N}$ such that $a = bq + r$ and $r < b$, then they are unique.

Discussion topic 1.80. Show that if X is an infinite set, then there is an infinite countable subset Y of X. Is it ever possible that Y is unique?

Discussion topic 1.81 (fundamental theorem of arithmetic). Show that every natural number $n \geq 2$ has a unique prime factorization. Since we already proved existence, Example 1.70, all that you have to prove is uniqueness. You might want to use Euclid's lemma, Exercise 1.10.

Chapter 2

Math 113: Fields

In this chapter we'll study operations like addition and multiplication, going back to when you first learned arithmetic. This material is similar to the latter half of Math 113.

2.1 What is a field?

Definition 2.1. Let X be a set. A *binary operation* $*$ on X is a function $X \times X \to X$, written $(a, b) \mapsto a * b$. We also say that $(X, *)$ is a *binary structure*.

So for example, addition and multiplication are binary operations on $\mathbf{R}$.

Definition 2.2. Let k be a set. Let $+$ (*addition*) and $\cdot$ (*multiplication*) be two binary operations on k, and let 0 and 1 be two elements of k. If the following conditions are satisfied, then we say that $(k, +, \cdot, 0, 1)$ is a *field*:

1. *Associativity*: For every $a, b, c \in k$, $a + (b + c) = (a + b) + c$ and $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.

2. *Commutativity*: For every $a, b \in k$, $a + b = b + a$ and $a \cdot b = b \cdot a$.

3. *Identity*: For every $a \in k$, $0 + a = a$ and $1 \cdot a = a$.

4. *Inversion*: For every $a \in k$, there is an element called $-a$ such that $a + (-a) = 0$ and an element called a^{-1} such that $a \cdot a^{-1} = 1$.

5. *Distributivity*: For every $a, b, c \in k$, $a \cdot (b + c) = a \cdot b + a \cdot c$.

6. *Nondegeneracy*: The cardinality $|k| \geq 2$.

If $+$, $\cdot$, 0, and 1 are understood, we simply say that k is a field.

We usually also will write ab to mean $a \cdot b$, and $a^n = aa \ldots a$ (n copies), $a - b = a + (-b)$, and $a/b = ab^{-1}$.

These are a lot of conditions to check, but somehow none of them should be a surprise; they're just the usual rules of addition and multiplication that you internalized long before taking this class. The only mildly puzzling condition here is nondegeneracy. This is needed for the proof of Proposition 2.7.

You'll notice immediately that **Q** and **R** are fields. So is **C**. On the other hand, **Z** is not a field, since not every element of **Z** has a multiplicative inverse.

Before we give more examples of fields, let's discuss some properties that they have.

Proposition 2.3. *The additive and multiplicative identities of a field are unique.*

Proof. Let k be a field. To prove uniqueness of 0, we do what we always do. Assume $x \in k$ is such that for every $y \in k$, $x + y = y$. Now we need to prove that $x = 0$. Since 0 is an identity, $x + 0 = x$, and since x is an identity, $0 + x = 0$. By commutativity, $x + 0 = 0 + x$, so we have $x = 0$.

The proof that 1 is unique is exactly the same, but in different notation. □

Proposition 2.4. *Let k be a field. For every $x \in k$, the additive and multiplicative inverses of x are unique.*

Proof. Let $y \in k$ be such that $x + y = 0$. We can add $-x$ to both sides of the equation, and $(-x) + x + y = 0 + (-x)$ gives $y = -x$.

The proof that x^{-1} is exactly the same. □

Proposition 2.5. *In any field k, for every $x \in k$, $0x = 0$.*

Proof. We have $0x + 0x = (0 + 0)x = 0x$ by distributivity and identity. By inversion, we can subtract $0x$ from both sides to get $0x = 0$. □

Proposition 2.6. *In any field k, for every $x, y \in k$, we have $x(-y) = -(xy)$. In fact, $-x = (-1)x$.*

Proof. By distributivity, $xy + x(-y) = x(y - y) = x0 = 0$. By inversion, we can subtract xy from both sides to get $x(-y) = -(xy)$.

For the second claim, by distributivity, $x + (-1)x = (1 - 1)x = 0x = 0$. So we can subtract x from both sides to get $(-1)x = -x$. □

Proposition 2.7. *In any field, $0 \neq 1$.*

Proof. To prove that two things are not equal, we assume that they are and derive a contradiction. So assume $0 = 1$. Then for every $x \in k$, $0 = 0x = 1x = x$ by identity. So $k = \{0\}$. Therefore k has only one element, which contradicts nondegeneracy. □

Proposition 2.8. *Let k be a field with $x, y \in k$. If $xy = 0$ then either $x = 0$ or $y = 0$.*

Proof. Assume towards contradiction that $xy = 0$ but x and y are nonzero. Then $x = xyy^{-1} = 0y^{-1} = 0$, which is a contradiction. □

Let's give two more examples of fields.

Example 2.9. Let k be a field, and let $k(x)$ denote the set of *rational functions* on k, functions of the form

$$x \mapsto \frac{a_0 + a_1x + \cdots + a_nx^n}{b_0 + b_1x + \cdots + b_mx^m}.$$

(Note that a rational function might not be defined on all of k, since the denominator might be zero for some $x \in k$.) For example, the function

$$x \mapsto \frac{2+3x}{x^2-1}$$

is a rational function on $\mathbf{Q}$, which is defined away from $\{1, -1\}$.

We can define 1 to be the function $x \mapsto 1$ and 0 to be the function $x \mapsto 0$. Besides, we define $(f+g)(x) = f(x)+g(x)$ and $(fg)(x) = f(x)g(x)$. These definitions make $k(x)$ a field.

Indeed, $((f+g)+h)(x) = (f(x)+g(x))+h(x) = f(x)+g(x)+h(x)$ (since $(k,+)$ is associative) $= f(x)+(g(x)+h(x)) = (f+(g+h))(x)$ and $(f+g)(x) = f(x)+g(x) = g(x)+f(x)$ (since $(k,+)$ is abelian) $= (g+f)(x)$. We also have $(0+f)(x) = 0+f(x) = f(x)$. Similarly for multiplication.

For inversion, let

$$f(x) = \frac{a_0 + a_1x + \cdots + a_nx^n}{b_0 + b_1x + \cdots + b_mx^m}.$$

Then let

$$-f(x) = \frac{-a_0 - a_1x - \cdots - a_nx^n}{b_0 + b_1x + \cdots + b_mx^m}.$$

So

$$\begin{aligned}(f+(-f))(x) &= \frac{a_0 + a_1x + \cdots + a_nx^n}{b_0 + b_1x + \cdots + b_mx^m} + \frac{-a_0 - a_1x - \cdots - a_nx^n}{b_0 + b_1x + \cdots + b_mx^m} \\ &= \frac{a_0 + a_1x + \cdots + a_nx^n - a_0 - a_1x - \cdots - a_nx^n}{b_0 + b_1x + \cdots + b_mx^m}\end{aligned}$$

since k is distributive

$$= \frac{a_0 - a_0 + a_1x - a_1 + \cdots + a_nx^n - a_nx^n}{b_0 + b_1x + \cdots + b_mx^m}$$

since k is commutative

$$= \frac{0}{b_0 + b_1x + \cdots + b_mx^m} = 0.$$

Similarly, we put

$$f^{-1}(x) = \frac{b_0 + b_1x + \cdots + b_mx^m}{a_0 + a_1x + \cdots + a_nx^n}$$

so that

$$ff^{-1}(x) = \frac{a_0 + a_1x + \cdots + a_nx^n}{b_0 + b_1x + \cdots + b_mx^m}\frac{b_0 + b_1x + \cdots + b_mx^m}{a_0 + a_1x + \cdots + a_nx^n} = 1.$$

Finally, $(f(g+h))(x) = f(x)(g(x)+h(x)) = f(x)g(x)$ since k is distributive, whence $f(g+h) = fg+fh$. Therefore $k(x)$ is a field.

Here's another example of a field, where the addition acts a little strange...

Example 2.10. Let p be a prime number. If $n \in \mathbf{N}$, we define n modulo p to be the remainder of dividing n by p. The remainder exists (Exercise 1.78) and is unique (Exercise 1.79) by the division algorithm (and is the number r appearing in the statement of the division algorithm). For example, 16 modulo 3 is 1, since 16 divided by 3 is 5 with a remainder of 1. Notice that n modulo p always lies in $\{0, 1, \ldots, p-1\}$.

Let $\mathbf{F}_p$ be the set $\{0, 1, \ldots, p-1\}$. We define field operations on $\mathbf{F}_p$, by taking addition and multiplication modulo p. That is, if $x, y \in \mathbf{F}_p$, then $x + y$ is taken modulo p and xy is taken modulo p. These operations are known as *modular arithmetic*, and make $\mathbf{F}_p$ into a field.

Before we can prove that $\mathbf{F}_p$ is a field, we need to prove a famous theorem about modular arithmetic.

Theorem 2.11 (Fermat's little theorem)**.** *If $x \in \mathbf{N}$ and p is a prime number, then $x^{p-1} = 1$ modulo p.*

The proof of this requires some setup. Before we begin, recall Euclid's lemma, Lemma 1.10, which we will use a lot here.

Lemma 2.12. *If $ux = uy$ modulo p for some $u, x, y \in \mathbf{N}$, and u is not divisible by p, then $x = y$ modulo p.*

Proof. We know that p divides $ux - uy$ (doing arithmetic in $\mathbf{N}$), so p divides $u(x-y)$ (since arithmetic in $\mathbf{N}$ is distributive). But then p divides $x - y$ since p does not divide u, by Euclid's lemma. So $x = y$ modulo p. □

Proof of Theorem 2.11. There are two cases to consider. First, if x is divisible by p, say $x = np$, then $x^p - x = (np)^p - np = n^p p^p - np = (n^p p^{p-1} - n)p$ is divisible by p already.

So now we assume x is not divisible by p. Let us write down the sequence of numbers $x, 2x, \ldots, (p-1)x$. Assume towards contradiction that some jx is divisible by p. Since $j < p$, j does not divide p. So Euclid's lemma implies that j divides x. But this is a contradiction. Therefore every jx is not divisible by p, so is nonzero modulo p. So jx modulo p is one of $1, \ldots, p-1$ (since it is not 0).

Moreover, if $jx = kx \bmod p$, then by Lemma 2.12, $j = k$ modulo p, and since $j < p$ and $k < p$, this implies $j = k$. Therefore each of the jx is sent to a unique one of the $1, \ldots, p-1$. That is, the function $\{x, 2x, \ldots, (p-1)x\} \to \{1, 2, \ldots, p-1\}$ given by reduction modulo p is a bijection. That is, $x, 2x, \ldots, (p-1)x$ modulo p is a rearrangement of the sequence $1, 2, \ldots, p-1$.

Since multiplication modulo p is commutative, it follows that

$$x \cdot 2x \cdot 3x \cdot \cdots \cdot (p-1)x = 1 \cdot 2 \cdot \cdots \cdot p - 1$$

modulo p. In other words, $x^{p-1}(p-1)! = (p-1)!$ modulo p. Therefore $x^{p-1} = 1$ modulo p. □

Proof that $\mathbf{F}_p$ is a field. Associativity, commutativity, distributivity, and existence of identities are easy and tedious to check (which we leave to you as Homework 2.20.) For example, we'll show commutativity of addition. If $x, y \in \mathbf{F}_p$, then $x +_p y$ is the remainder of $x +_{\mathbf{N}} y = y +_{\mathbf{N}} x$, and $y +_p x$ is the remainder of $y +_{\mathbf{N}} x$. So $x +_p y = y +_p x$.

For inversion of addition, if $x \in \mathbf{F}_p$, we let $-x$ denote $p -_{\mathbf{N}} x$. (For example, if $x = 1$ and $p = 3$, $-x = 2$. Then $x +_{\mathbf{N}} (-x) = p$, and p modulo p is 0. So $x +_p (-x) = 0$. Therefore $-x$ is the inverse of x.

The hard one is inversion of multiplication. Let $x^{-1} = x^{p-2}$ modulo p. We claim that $x^{-1} = 1$ modulo p

By Fermat's little theorem, $xx^{p-2} = x^{p-1} = 1$ modulo p. Therefore x^{-1} is an inverse of x. So $\mathbf{F}_p$ admits inversion of multiplication. We conclude that $\mathbf{F}_p$ is a field. □

But $\mathbf{F}_p$ is a little weird. $1 + 1 + \cdots + 1$ (p copies) is 0. This didn't happen in any of the fields we discussed before (except $\mathbf{F}_p(x)$, of course!)

Definition 2.13. The *characteristic* of a field k, denoted $\operatorname{char} k$ is the least number n such that $1 + 1 + \cdots + 1$ (n copies) is 0, or 0 if no such number exists.

So $\mathbf{Q}$ is a characteristic 0 field, while $\mathbf{F}_p$ is a characteristic p field.

Here's a definition that we'll use in a few exercises. Holomorphic functions are the main objects of study in complex analysis (Math 185).

Definition 2.14. A function $f : \mathbf{C} \to \mathbf{C}$ is said to be a *holomorphic function* if the derivative $f'(z)$ exists for every $z \in \mathbf{C}$.

Discussion topic 2.15. A *meromorphic function* is a function f/g, where f and g are holomorphic functions on $\mathbf{C}$ and there are only countably many $z \in \mathbf{C}$ such that $g(z) = 0$. (So a meromorphic function might not be defined everywhere, because g might have a zero; but it is defined at "most" points of $\mathbf{C}$, since $\mathbf{C}$ is uncountable by Cantor's diagonal argument.) An example of a meromorphic function is tan, since

$$\tan z = \frac{\sin z}{\cos z},$$

and $\cos z = 0$ iff $z = \pi + 2n$ for some $n \in \mathbf{Z}$ (and $\mathbf{Z}$ is countable).

Let $\mathcal{M}(\mathbf{C})$ denote the set of all meromorphic functions on $\mathbf{C}$. Show that $\mathcal{M}(\mathbf{C})$ is a field. (First, you'll have to define what the field operations are...)

Discussion topic 2.16. Here's one of the main techniques for proving that something is a field.

Let F be a field and $k \subseteq F$ be a nonempty set. Say that k is *closed under field operations*; that is, k satisfies the following conditions:

1. *Closed under addition*: If $x, y \in k$, then $x + y \in k$.
2. *Closed under multiplication*: If $x, y \in k$, then $xy \in k$.
3. *Closed under inversion*: If $x \in k$, then $-x \in k$ and $x^{-1} \in k$.

Show that k is a field if and only if k is closed under field operations.

Discussion topic 2.17. Say that a *algebraic number* is an element $x \in \mathbf{C}$ such that there is a polynomial $p : \mathbf{C} \to \mathbf{C}$, with

$$p(y) = a_0 + a_1 y + \cdots + a_n y^n$$

with every $a_j \in \mathbf{Z}$, such that $p(x) = 0$. Let $\mathbf{A}$ denote the set of algebraic numbers. Show that $\mathbf{A}$ is a field.

Discussion topic 2.18 (freshman's dream). Here's a useful result which will come up, for example, when we study the Frobenius (Homework 2.72). Let k be a field of characteristic p, p a prime number. Show that for every $x, y \in k$,

$$(x+y)^p = x^p + y^p.$$

Why is this equation called the *freshman's dream*?

Discussion topic 2.19 (quadratic formula). Recall the *quadratic formula*: for $b, c \in \mathbf{C}$, if $f(x) = ax^2 + bx + c$ and $a \neq 0$, then $f(x) = 0$ iff

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

(Notice that without loss of generality, we can assume $a = 1$ – why?)

Prove the quadratic formula in any field of characteristic $\neq 2$. What goes wrong with the proof in characteristic 2?

Homework 2.20. Fill in more details of the proof given in Example 2.10. Specifically, show that $\mathbf{F}_p$ is distributive, and addition and multiplication have identities.

Homework 2.21. What day of the week will it be $10^{10^{100}}$ days after a Sunday?

Homework 2.22. Show that for any $n \in \mathbf{N}$ $n^{37} - n$ is divisible by 383838. (You're allowed to look up the prime factorization of 383838.)

Homework 2.23. Let k be a field of characteristic 0. Recall that if $f(x) = ax^2 + bx + c$ is a quadratic polynomial over k (i.e. with $a, b, c \in k$), then the *discriminant* of f is $\Delta(f) = b^2 - 4ac$. Assume that $f = gh$, for some linear polynomials g, h over k. Show that $\Delta(f) = 0$ if and only if we can choose g so that $f = g^2$.

2.2 The category of fields

When faced with a new definition in mathematics (here, that of a *field*), it is useful to come up with lots of examples (as we have done), including a "prototype" (here, $\mathbf{Q}$, the simplest example). Once that is done, the next job is to see how multiple examples of the definition relate to each other.

Definition 2.24. Let k and F be fields. A *morphism of fields* is a function $\iota : k \to F$ such that:

1. *Preservation of structure*: For every $x, y \in k$, $\iota(x+y) = \iota(x) + \iota(y)$ and $\iota(xy) = \iota(x)\iota(y)$.
2. *Nondegeneracy*: The image $\iota(k)$ has cardinality at least 2.

We shall use the following result, which we leave to the peanut gallery to prove.

Proposition 2.25. *If ι is a morphism of fields $k \to F$, then:*

1. Preservation of identity*: $\iota(0) = 0$ and $\iota(1) = 1$.*

2. Preservation of inverses*: For every $x \in k$, $\iota(-x) = -\iota(x)$ and $\iota(x^{-1}) = \iota(x)^{-1}$.*

3. Compositionality*: If $\jmath : F \to E$ is a morphism of fields, then the composition $\jmath \circ \iota$ is a morphism of fields $k \to E$.*

The "stupid" example of a morphism of fields is in case $k \subseteq F$ and the map is $x \mapsto x$. For example, the identity on $\mathbf{Q} \subset \mathbf{R}$. In some sense, there is no other example.

Proposition 2.26. *If ι is a morphism of fields, then ι is injective.*

Proof. We must show that there are no $x, y \in k$ such that $\iota(x) = \iota(y)$ but $x \neq y$; since we are proving nonexistence, we approach by contradiction. Assume that ι is not injective, so there are $x \neq y$ with $\iota(x) = \iota(y)$.

Under these assumptions, we have $\iota(x) = \iota(y)$, so $\iota(x) - \iota(y) = 0$. Therefore $\iota(x - y) = 0$. So $1 = \iota(1) = \iota((x-y)^{-1}(x-y)) = \iota((x-y)^{-1})\iota(x-y) = \iota((x-y)^{-1}) \cdot 0 = 0$. This contradicts Proposition 2.7. □

Let $\iota : k \to F$ be a morphism of fields. We want to think k as "the same as" $\iota(k)$, in a way that we will make more precise later. To give an idea of why we would want to do this, consider the following annoying problem.

Assume we have a field k with (say) four elements, say

$$k = \{\text{one}, \text{two}, \text{three}, \text{four}\}.$$

Now Aidan's grandparents might not like this field, and might prefer a field k^{ES} whose elements are named uno, dos, tres, and cuatro. Someone else might prefer yet another field k^{JP} whose elements are named ichi, ni, san, and shi. Of course, any sane person would recognize that while these fields have different elements, they are the "same field" in some sense.

A more down-to-earth example is as follows.

Example 2.27. Let k be a field. By Example 2.9, the set $k(x)$ of rational functions on k is a field. For $x \in k$, we let $\iota(x)$ be the function $k \to k$ defined by $y \mapsto x$ for every $y \in k$ (so $\iota(x)$ is a constant function).

We claim that ι is a morphism of fields $k \to k(x)$. Indeed, let $x, y, z \in k$. Then $\iota(x)(z) + \iota(y)(z) = x + y = \iota(x+y)(z)$. Similarly $\iota(x)(z)\iota(y)(z) = xy = \iota(xy)(z)$.

Moreover, the image $\iota(k)$ is precisely the set of constant functions in $k(x)$. It's pretty clear that k and $\iota(k)$ are the same in all but name: there's an obvious bijection between them, and their operations are preserved by the bijection. But $\iota(k)$ is not the same set as k, since elements of $\iota(k)$ are rational functions, while elements of k are whatever elements of k are.

Let's now make all this finagling more precise. The key definition is that of an "isomorphism of fields."

Definition 2.28. An *isomorphism of fields* is a bijective morphism of fields. If there is an isomorphism of fields $k \to F$, we say that k and F are *isomorphic fields*.

So to prove that a morphism is an isomorphism, we just have to show that it is surjective, since it already is injective by Proposition 2.26.

Proposition 2.29. *The inverse of an isomorphism of fields is an isomorphism of fields.*

Proof. Let $\iota : k \to F$ be an isomorphism. Clearly ι^{-1} is a bijection, so we just need to show that ι^{-1} is a morphism of fields. We'll prove that it preserves addition; the preservation of multiplication is similar.

Let $x, y \in F$. We have $\iota(\iota^{-1}(x)\iota^{-1}(y)) = \iota(\iota^{-1}(x))\iota(\iota^{-1}(y)) = xy = \iota(\iota^{-1}(xy))$. Since ι is injective, it follows that $\iota^{-1}(x)\iota^{-1}(y) = \iota^{-1}(xy)$.

Since k has at least 2 elements and ι^{-1} is surjective, $\iota^{-1}(F)$ has at least 2 elements. □

Proposition 2.30. *If ι is a morphism of fields $k \to F$, then its image $\iota(k)$ is a field, and ι is an isomorphism $k \to \iota(k)$.*

Proof. In light of Exercise 2.16, we just have to show that $\iota(k)$ is closed under field operations.

Since ι is injective, ι is an isomorphism $k \to \iota(k)$, so we have an isomorphism $\iota^{-1} : \iota(k) \to k$. Then if $x, y \in \iota(k)$, $\iota^{-1}(x) + \iota^{-1}(y) \in k$, so $x + y = \iota(\iota^{-1}(x) + \iota^{-1}(y)) \in \iota(k)$. Therefore $\iota(k)$ is closed under addition. The proofs of the other conditions are similar. □

Notice that the key step in all these proofs is to "pull back" the claim from $\iota(k)$ to k, which we already know a lot of wonderful properties about.

Note that isomorphic structures might not look like each other at all! [1] But in general, it's not too hard to show that two things are isomorphic: just come up with an isomorphism. It's not so easy to show that two things are not isomorphic: what if there was a crazy isomorphism that we never thought of?

Definition 2.31. Let k be a field. An *invariant* of k is any property of k which is preserved by every isomorphism.

For example, having the element "san" is not an invariant, since our "field" k^{JP} had "san" but was isomorphic to k^{ESP}, where that element was replaced with "tres." This is our strategy to prove that two fields k and F are not isomorphic: find an invariant of k which is not true of F.

An example of an invariant is cardinality. If k and F do not have the same cardinality, then there are no bijections $k \to F$ and so we have no hope of finding an isomorphism.

But if k and F have the same cardinality, then we need to use the field structure itself to find an invariant. One idea is to try to find an equation whose solution sets in k and F cannot be bijected.

Example 2.32. The fields $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{C}$ are not isomorphic. First, $\mathbf{Q}$ is not isomorphic to either $\mathbf{R}$ or $\mathbf{C}$, because both $\mathbf{R}$ and $\mathbf{C}$ are uncountable and so there can be no bijections between $\mathbf{Q}$ and either $\mathbf{R}$ or $\mathbf{C}$.

[1] Amusing aside, for those who have taken Math 54: we can define vector spaces whose vectors are elements of $\mathbf{Q}$ instead of $\mathbf{R}$. For example, $\mathbf{Q}^3$ is the set of three-dimensional vectors whose coordinates are rational numbers. We can define vector spaces to have any cardinality as their dimension, not just elements of $\mathbf{N}$. Then as vector spaces, $\mathbf{R}$ and $\mathbf{C}$ are both isomorphic to the vector space $\mathbf{Q}^{\mathfrak{c}}$ for a certain cardinality $\mathfrak{c}$, so $\mathbf{R}$ and $\mathbf{C}$ are isomorphic to each other as $\mathbf{Q}$-vector spaces.

Second, we must show that $\mathbf{R}$ and $\mathbf{C}$ are not isomorphic. The equation $x^4 = 1$ has two solutions in $\mathbf{R}$, namely $x = 1$ and $x = -1$, but four in $\mathbf{C}$, since $x = i$ and $x = -i$ also have two solutions. So if $\iota : \mathbf{C} \to \mathbf{R}$ is an isomorphism, then $i^4 = 1$, so $\iota(i)^4 = 1$. Therefore $\iota(i) = -1$ or $\iota(i) = 1$. But $\iota(-1) = -1$ and $\iota(1) = 1$. So ι is not injective, a contradiction.

We have already generalized the notion of two sets being equal to two fields being isomorphic: they're not literally the same, but they're "close enough for our purposes." Now we generalize the notion of subset, keeping in mind that if $\iota : k \to F$ is a morphism of fields, then $\iota(k)$ is isomorphic to k, so $\iota(k)$ is "close eonugh for our purposes" to a subset of F.

Definition 2.33. Let k and F be fields. If there is a morphism of fields $k \to F$, we say that F is a *extension field* of k, or that k is a *subfield* of F.

So $\mathbf{R}$ is an extension field of $\mathbf{Q}$. So are $\mathbf{A}$, $\mathbf{C}$ and $\mathbf{Q}(x)$. This isn't a coincidence.

Lemma 2.34. *Let k be a field of characteristic 0. Then k is an extension field of $\mathbf{Q}$.*

Proof. Let us define n' to be $1+1+\cdots+1$ (n times) in k. Then the map $n \mapsto n'$ is injective: if not, say $n' = m'$, then $n' - m' = 0$, so $(n-m)' = 0$, so k is characteristic p for some $0 < p \leq n - m$, which is a contradiction. Therefore we can treat the elements n' of k as though they were natural numbers.

We define $\iota : \mathbf{Q} \to k$ in stages. First we define it on $\mathbf{N}$: if $n \in \mathbf{N}$ then $\iota(n) = n'$. This forces us to define $\iota(-n) = -(1+1+\cdots+1)$ (and $\iota(0) = 0$). So ι is defined on $\mathbf{Z}$. Then if $n/m \in \mathbf{Q}$, we have no choice but to define $\iota(n/m) = \iota(n)/\iota(m)$. This is typical in the definition of a morphism in lots of different branches of math: once we define it on a small subset of the domain, we have no choice but to define it on the rest of the domain in a certain way. (Actually, in this case, we had to define $\iota(n) = n'$.)

Now we show that ι is a morphism of fields. Let $n/m, j/\ell \in \mathbf{Q}$. Then

$$\begin{aligned}\iota\left(\frac{n}{m} + \frac{j}{\ell}\right) &= \iota\left(\frac{n\ell + jm}{m\ell}\right) = \frac{\iota(n\ell + jm)}{\iota(m\ell)} \\ &= \frac{(n\ell + jm)'}{(m\ell)'} = \frac{(n\ell)' + (jm)'}{(m\ell)'} \\ &= \frac{n'}{m'} + \frac{j'}{\ell'} = \iota\left(\frac{n}{m}\right) + \iota\left(\frac{j}{\ell}\right).\end{aligned}$$

The argument is similar for multiplication, and we leave it as Homework 2.44. So ι is a morphism of fields. □

Something similar happens if k is NOT of characteristic 0.

Lemma 2.35. *Let k be a field of characteristic $p \neq 0$. Then p is a prime number, and k is an extension field of $\mathbf{F}_p$.*

Proof. Once we have shown that p is prime, we can just copy and paste the argument from Lemma 2.34. We then leave the remaining details to you in Homework 2.45.

So we must rule out the case that p is composite, which we do by contradiction. Assume ℓ is a prime number which divides p, such that $\ell \neq p$ (which exists by the fundamental theorem of arithmetic). So $\ell < p$, and so $\ell \in \mathbf{F}_p$. Also, $p/\ell \in \mathbf{F}_p$. So $p\ell \cdot \ell = p = 0$ modulo p, so $p\ell \cdot \ell = 0$ in k. This is a contradiction of Proposition 2.8. Therefore p is not composite. □

Definition 2.36. The fields $\mathbf{Q}$ and $\mathbf{F}_p$ for p prime are called *prime fields*.

Summarizing the above discussion:

Theorem 2.37. *Every field is an extension of a prime field.*

Definition 2.38. Say that a field $\overline{k}$ is *algebraically closed* if for every polynomial $p : \overline{k} \to \overline{k}$, there is a root $x \in \overline{k}$, i.e. $p(x) = 0$.

Discussion topic 2.39. If there is a morphism of fields $k \to F$, show that $\operatorname{char} k = \operatorname{char} F$.

Notice that this implies that characteristic is an invariant. But this is actually stronger than that, since we didn't assume that the morphism in question was surjective.

Discussion topic 2.40. Prove Proposition 2.25.

Discussion topic 2.41. Let p be a prime number and $q = p^2$. Let $\mathbf{F}_q$ denote a set with q elements containing $\mathbf{F}_p$. Define elements on $\mathbf{F}_p$ which make $\mathbf{F}_p$ a field. (Hint: first try when $p = 2$.) If you are brave, try to do this for $q = p^k$, where $k \in \mathbf{N}$.

Discussion topic 2.42. Let k be a prime field and let $\iota : k \to F$ be a morphism of fields and let P the prime field of k. Show that ι is constant on P.

Discussion topic 2.43. Show that algebraic closure is an invariant.

Homework 2.44. Complete the proof of Lemma 2.34 by showing that if $\iota : \mathbf{Q} \to k$ is as above, then ι preserves multiplication.

Homework 2.45. Complete the proof of Lemma 2.35 by defining a function $\iota : \mathbf{F}_p \to k$, for k as above, and showing that ι is a morphism of fields.

Homework 2.46. Let k, F, E be fields. Show that:

1. k is isomorphic to k;

2. if k is isomorphic to F then F is isomorphic to k;

3. and if k is isomorphic to F and F is isomorphic to E then k is isomorphic to E.

So isomorphism is an equivalence relation on the category of all fields.

Homework 2.47. Let k be a field, which is an extension of F and of E. Show that $F \cap E$ has more than one element if and only if $F \cap E$ is a field.

Homework 2.48. Let $\mathbf{Q}(\sqrt{2})$ be the subfield of $\mathbf{R}$ consisting of all elements $x \in \mathbf{R}$ such that we can write

$$x = a + b\sqrt{2},$$

for $a, b \in \mathbf{Q}$. Show that there are exactly two morphisms of fields $\mathbf{Q}(\sqrt{2}) \to \mathbf{Q}(\sqrt{2})$.

2.3 Existence of algebraic closures, and Zorn's lemma

Here's an optional section which will demonstrate a useful proof technique that every math major should see at least once in their lives: proof of existence by Zorn's lemma.

Our goal is the following theorem.

Theorem 2.49. *If k is any field, then there is an algebraically closed field $\overline{k}$ and a morphism of fields $k \to \overline{k}$. Besides, $\overline{k}$ is unique.*

To do this, we'll need a very powerful property known as *Zorn's lemma*. To do this, we'll need to develop a branch of math known as *order theory* a little.

Definition 2.50. Let X be a set. A *partial ordering* $\leq$ of X is a binary relation on X such that for every $x_1, x_2, x_3 \in X$,

1. *Reflexivity*: $x_1 \leq x_1$.
2. *Antisymmetry*: If $x_1 \leq x_2$ and $x_2 \leq x_1$, then $x_1 = x_2$.
3. *Transitivity*: If $x_1 \leq x_2$ and $x_2 \leq x_3$ then $x_1 \leq x_3$.

We say that $(X, \leq)$ is a *poset*.

Definition 2.51. Let X be a poset. A *chain* Y in X is a subset $Y \subseteq X$ such that for every $y_1, y_2 \in Y$, either $y_1 \leq y_2$ or $y_2 \leq y_1$.

Definition 2.52. Let X be a poset and $Y \subseteq X$. If there is a $y^* \in X$ such that for every $y \in Y$, $y \leq y^*$, we say that y^* is an *upper bound* for Y.

Definition 2.53. Let X be a poset and $x^* \in X$. If the only $x \in X$ such that $x^* \leq x$ is x^*, then we say that x^* is *maximal* for X.

Axiom 2.54 (Zorn's lemma)**.** *Let X be a nonempty poset, such that for every chain Y in X, Y has an upper bound. Then X has a maximal element.*

The proof of Zorn's lemma uses transfinite induction, so we won't discuss it except in the additional practice, Homework **??**.[2]

In what follows, we use $k[x]$ to denote the set of all polynomials.

Theorem 2.55 (Kronecker)**.** *Let k be any field and $p \in k[x]$. Then there is an extension F of k and an element $z \in F$ such that $p(z) = 0$.*

The proof of Kronecker's theorem is difficult, and we only need it to show that there is an extension of k satisfying certain conditions. We could skip it for lack of time, though it does demonstrate some "categorical" proof techniques used in advanced algebra classes, because it has nothing to do with Zorn's lemma.

[2] It turns out that Zorn's lemma is true if and only if the axiom of choice, Axiom **??** is. So you can use Zorn's lemma to prove the existence of nonmeasurable sets.

Proof. We first note that without loss of generality we can assume that p is an *irreducible polynomial*: if $f, g \in k[x]$ have $fg = p$, then $f = p$ and $g = 1$. Otherwise, we can factor p and use induction on the number of factors.

We define an equivalence relation $\sim$ on $k[x]$, by saying that $f \sim g$ assuming that there is a $h \in k[x]$ such that

$$f(x) = g(x) + h(x)p(x).$$

We choose a subset F of $k[x]$, consisting of exactly one f from each equivalence class. So if $f \in F$ and $f \sim g$, then $f = g$; and for every $g \in k[x]$, there is an $f \in F$ such that $f \sim g$. Elements of F are known as *residue classes* with respect to p.

Notice that there is a surjective function $\pi : k[x] \to F$ to its residue class. That is, if $f \in k[x]$, let $\pi(f)$ be the unique $g \in F$ such that $f \sim g$.

For $f, g \in F$ we define $f + g$ to be $\pi(f +_{k[x]} g)$, $fg = \pi(f \cdot_{k[x]} g)$, $0 = \pi(0_{k[x]})$, and $1 = \pi(1_{k[x]})$. We claim that F is a field, unimaginatively known as the *residue class field* of p.

We leave it to you to show that π preserves field operations, except inversion of multiplication (Exercise 2.57). This implies that F satisfies associativity, commutativity, identity, distributivity, and inversion of addition. Later we will see that F has at least the cardinality of k, so F is nondegenerate.

For inversion of multiplication, let $f \in F$. We need to find $g, h \in k[x]$ such that for every $x \in k$,

$$f(x)g(x) = 1 + h(x)p(x).$$

We leave it to you to show that this is possible assuming that p is irreducible (Exercise 2.58).

Let $\jmath : k \to k(x)$ be the usual morphism of fields. Notice that the image of $\jmath$ is contained in $k[x]$, so it makes sense to define $\iota = \pi \circ \jmath$. By Exercise 2.57, it follows that ι is a morphism of fields $k \to F$. Therefore F is an extension of k.

Now let $f \in k[x]$ be given by $f(x) = x$, and let $z = \pi(f)$. We claim that $p(z) = 0$. First, $0 = 0 + 0p$, so $0 \sim p$. Therefore $\pi(p) = 0$. Also, $p(f(x)) = p(x)$, and since π preserves operations, $p(z) = p(\pi(f)) = \pi(p)(f) = 0(f) = 0$. □

Now we use Zorn's lemma. Though this proof feels unusual, it is essentially par for the course for a proof by Zorn's lemma.

Definition 2.56. An extension F of k is an *algebraic extension* if for every algebraically closed field $E \supseteq k$, E is an extension of F.

Let $\mathcal{F}$ be the set of all algebraic extensions of k not isomorphic to k. We want to use Zorn's lemma on $\mathcal{F}$, and a proof by Zorn's lemma has four steps:

1. Show that $\mathcal{F}$ is nonempty.

2. Show that $\mathcal{F}$ admits a partial ordering.

3. Show that every chain in $\mathcal{F}$ has an upper bound.

4. Show that the maximal element of $\mathcal{F}$ that exists by Zorn's lemma has the desired condition.

Proof of Theorem 2.49. If k is not algebraically closed, then there is a $p \in k[x]$ without a root in k, but by Kronecker's theorem, there is an extension F of k such that p has a root in F. Existence of roots is an invariant, so F is not isomorphic to k. If F is not algebraic over k, then $\overline{k} \subseteq F$ and we're done. So we can assume without loss of generality that F is algebraic; then $F \in \mathcal{F}$, so $\mathcal{F}$ is nonempty.

We now say that $F \leq E$ if E is an extension of F. Then $\leq$ is a partial ordering of $\mathcal{F}$: the identity shows $F \leq F$; if $F \leq E$ and $E \leq K$, then we can find a morphism of fields $F \to K$ by compositionality, which proves transitivity; for antisymmetry, if $\iota : F \to E$ and $\jmath : E \to F$ are morphisms, then $\jmath \circ \iota$ is an isomorphism.

So now let $\mathcal{C}$ be a chain in $\mathcal{F}$. We need to find a field K which is an upper bound to $\mathcal{C}$. To do this, notice that if $E, F \in \mathcal{C}$, then either $E \leq F$ or $F \leq E$, so we can assume without loss of generality $F \leq E$. Then if $\iota : F \to E$ is a morphism of fields, $\iota(F) \subseteq E$. So we can assume without loss of generality that $F \subseteq E$, since F is isomorphic to $\iota(F)$.

Therefore we can assume that if $F, E \in \mathcal{C}$ and $F \leq E$, then $F \subseteq E$. Let $K = \bigcup \mathcal{C}$. For $x, y \in K$, there is a $E \in \mathcal{C}$ such that $x, y \in E$, and since E is a field, $x + y$, xy, $-x$, and x^{-1} are defined and satisfy the usual conditions. Therefore K is a field. Just like with the proof that $\mathcal{F}$ is nonempty, we can assume K is an algebraic extension of k, so $K \in \mathcal{F}$. If $F \in \mathcal{C}$, then $F \subseteq K$, so $F \leq K$. Therefore K is an upper bound for $\mathcal{C}$.

So by Zorn's lemma, there is a field $\overline{k} \in \mathcal{F}$ which is maximal. By assumption, $\overline{k}$ is an algebraic extension of k. Assume that $\overline{k}$ is not algebraically closed. Then there is a polynomial $p \in \overline{k}[x]$ such that for every $x \in \overline{k}$, $p(x) \neq 0$, so by Kronecker's theorem there is an algebraic extension F over $\overline{k}$ which is not isomorphic to $\overline{k}$. Then F is an algebraic extension of k, which contradicts maximality of $\overline{k}$. □

Discussion topic 2.57. Fill in details of the proof of Kronecker's theorem, Theorem 2.55, by showing that the map π has $\pi(f + g) = \pi(f) + \pi(g)$, $\pi(0) = 0$, $\pi(1) = 1$, and $\pi(-f) = -f$.

Discussion topic 2.58. Fill in details of the proof of Kronecker's theorem, Theorem 2.55, by showing that if $p \in k[x]$ is an irreducible polynomial, and $f \in k[x]$, then there are $g, h \in k[x]$ such that

$$f(x)g(x) = 1 + h(x)p(x).$$

2.4 Finite fields

We can use the above machinery to do something quite powerful: *classify all finite fields*! This is the next thing to do after learning a new definition, finding the key examples, and finding the relationships between them: figure out how to classify them. This will be perhaps the first "deep" theorem we cover, and should give an idea how depeer theorems in upper-division math classes are taught. This will be much harder than previous sections, and is more like "bonus" material if we have enough time at the end of the semester, and enough students have taken Math 54.

Specifically, our goal is the following:

Theorem 2.59. *Let k be a field of cardinality $q \in \mathbf{N}$ and characteristic p. Then there is a $n \in \mathbf{N}$ such that $q = p^n$.*

To do this, we'll need some lemmata.

Lemma 2.60. *Let k be a finite field and let F be a subfield of k. If $\alpha \in k$, then there is a $n_F(\alpha) \in \mathbf{N}$ and $a_0, \dots, a_{n_F(\alpha)} \in F$ such that*

$$a_0 + a_1\alpha + \cdots + a_{n_F(\alpha)}\alpha^n = 0$$

and at least one of the a_j is nonzero.

Students of Math 54 will recognize this as meaning that $1, \alpha, \dots, \alpha^n$ is linearly dependent.

Proof. Suppose not. Then for every m, if

$$a_0 + a_1\alpha + \cdots + a_m\alpha^m = 0,$$

we have $a_j = 0$ for every a_j. So any element of the set ("span", for students of Math 54)

$$F(\alpha) = \{a_0 + a_1\alpha + \cdots + a_m\alpha^m \in k : m \in \mathbf{N}, a_j \in F\}$$

can be uniquely written as $a_0 + a_1\alpha + \cdots + a_m\alpha^m$; if not, say $a_0 + a_1\alpha + \cdots + a_m\alpha^m = b_0 + b_1\alpha + \cdots + b_m\alpha^m$, then we consider

$$a_0 + a_1\alpha + \cdots + a_m\alpha^m - b_0 - b_1\alpha - \cdots - b_m\alpha^m = (a_0 - b_0) + (a_1 - b_1)\alpha + \cdots + (a_m - b_m)\alpha^m,$$

so every $a_j - b_j = 0$, a contradiction. (For students of Math 54, $1, \dots, \alpha_m, \dots$ is a basis).

We now consider the set $F^{<\omega}$ of all finite sequences of elements of F. This means that if $a \in F^{<\omega}$, there is a m such that $a = (a_0, \dots, a_m)$ for some $a_j \in F_p$. Clearly $F^{<\omega}$ is infinite, and since every element of $F(\alpha)$ can be uniquely written in terms of an element of $F^{<\omega}$, it follows that the function

$$a \mapsto a_0 + a_1\alpha + \cdots + a_m\alpha^m$$

is injective $F^{<\omega} \to F(\alpha)$. (In other words, $F(\alpha)$ is infinite-dimensional.) Therefore $F(\alpha)$ is an infinite subset of the finite set k, a contradiction. □

Definition 2.61. Let $k \subseteq F$ be fields and let $\alpha \in F$. The set

$$k(\alpha) = \{a_0 + a_1\alpha + \cdots + a_m\alpha^m \in F : m \in \mathbf{N}, a_j \in k\}$$

is called the *simple extension* of k given by α.

Lemma 2.62. *Let $k \subseteq F$ be fields and let $\alpha \in k$. The simple extension $k(\alpha)$ is a field.*

(Note for students of Math 54: In fact a similar phenomenon happens for any field k, as long as it is finite-dimensional as a vector space over its prime field. For example, we could define the simple extension $\mathbf{Q}(\sqrt{2})$, which would be 2-dimensional over $\mathbf{Q}$, and would be a field.)

Proof. We must show that $k(\alpha)$ is closed under addition and multiplication, and their inversions, and has 0 and 1. In fact by taking $a_1 = a_2 = \cdots = a_{n_\alpha} = 0$ and using $a_0 \in k$, we see that $k \subseteq k(\alpha)$, so $k(\alpha)$ contains 0 and 1 since k does. We leave this to the peanut gallery as Exercise 2.68.

The hard one is multiplicative inverses. First notice that the α^j have multiplicative inverses, since they satisfy a polynomial over k. Let $\beta \in k(\alpha)$. So we can write $\beta = a_0 + a_1\alpha + \cdots + a_m\alpha^m$ for some $m \in \mathcal{M}$ and $a_j \in k$. We can then rationalize the denominator to get a multiplicative inverse. From there it follows that $k(\alpha)$ is a field. □

Definition 2.63. Let $k \subseteq F$ be finite fields. The *degree* of F over k is the unique n such that there are elements $\alpha_1, \ldots, \alpha_n \in F$, called a *basis* of F over k, such that for every $\alpha \in F$, we can find unique $a_1, \ldots, a_n$ such that

$$\alpha = a_1\alpha_1 + \cdots + a_n\alpha_n,$$

if such a number n exists and is unique.

(Or, in Math 54 terminology, F is dimension n over k.)

Lemma 2.64. *Let $k \subseteq F \subseteq E$ be fields. If F is degree n over k and E is degree m over F, then E is degree mn over k.*

Proof. Let $\alpha_1, \ldots, \alpha_n$ be a basis of F over k, and let $\beta_1, \ldots, \beta_m$ be a basis of E over F.

Let $\gamma \in E$. We need to find $a_{ij} \in k$ such that $\gamma = \sum_{i,j} a_{ij}\alpha_i\beta_j$. There are b_j such that $\gamma = \sum_j b_j\beta_j$, and a_{ij} such that $\beta_j = \sum_i a_{ij}\alpha_i$. Then

$$\gamma = \sum_j \sum_i a_{ij}\alpha_i\beta_j = \sum_{i,j} a_{ij}\alpha_i\beta_j.$$

Discussion topic 2.65. Show that the α_{ij} are unique.

Therefore $\{\alpha_i\beta_j : i \leq n, j \leq m\}$ is a basis of E over k. □

Lemma 2.66. *Let $k \subseteq F$ be finite fields and k have cardinality q. If F has degree d over k, then F has cardinality q^d.*

We leave this lemma as Exercise 2.69.

Proof of Theorem 2.59. By Theorem 2.37, the field $F_1 = \mathbf{F}_p$ is a subfield of k, with degree $d_1 = 1$. Now we iterate the following process: assume F_j is defined and has degree d_j over $\mathbf{F}_p$, and choose $\alpha_j \in k \setminus F_j$. Let $F_{j+1} = F_j(\alpha_j)$. Then by Lemma 2.62, F_{j+1} is a subfield of k, and by Lemma 2.60, F_{j+1} has degree $n_{F_j}(\alpha_j) + 1$ over F_j. Therefore F_{j+1} has degree $d_{j+1} = d_j n_{F_j}(\alpha_j)$ over $\mathbf{F}_p$ by Lemma 2.64. Since there are finitely many elements in k, this process must eventually stop, say at $F_J = k$. Then k has degree d_J over $\mathbf{F}_p$, so by Lemma 2.66, k has cardinality p^{d_J}. □

In fact the following is true:

Theorem 2.67. *Let $q = p^n$ for some prime p and $n \in \mathbf{N}$. There is a unique field $\mathbf{F}_q$ of cardinality q.*

However, the proof is significantly harder. It might be taught in Math 114.

Discussion topic 2.68. Complete the proof of Lemma 2.59: Show that $k(\alpha)$ is closed under addition, multiplication, and taking additive inverses.

Discussion topic 2.69. Prove Lemma 2.66.

Discussion topic 2.70. Let $\mathbf{Q}(2^{1/3})$ be the simple extension associated to $2^{1/3}$; so $\mathbf{Q}(2^{1/3})$ consists of elements of $\mathbf{C}$ of the form $a_0 + a_1 2^{1/3} + a_2 2^{2/3}$ for $a_j \in \mathbf{Q}$. Show that there is only one morphism of fields $\mathbf{Q}(2^{1/3}) \to \mathbf{Q}(2^{1/3})$. (Hint: If ι is a morphism of fields and $x^3 = 2$, then prove that $\iota(x)^3 = 2$, using the fact that 2 is contained in a prime field. If ι is not constant, then prove that a real number is sent to a number which is not real. So prove that $\mathbf{Q}(2^{1/3}) \subseteq \mathbf{R}$.)

Definition 2.71. Let k be a field of characteristic p. If p is a prime number, then define the *Frobenius* ϕ of k to be the function $k \to k$ given by $\phi(x) = x^p$. Otherwise, define ϕ to be the identity.

Discussion topic 2.72. Show that the Frobenius is a morphism of fields. (Hint: Exercise 2.18.) Moreover, if k is a finite field, then the Frobenius of k is a bijection. But show that if k is a field of characteristic $p \neq 0$, then the Frobenius of $k(x)$ is not surjective.

Discussion topic 2.73. Show that if k is a finite field of characteristic p and cardinality p^n and m divides n, then k has a subfield of cardinality p^m. (Hint: Consider the fields F_j appearing in the proof of Theorem 2.59.)

Chapter 3

Math 104: Rigorous calculus

In calculus, you learn 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 $\mathbf{R}$ using the machinery we've developed so far: we have learned how to write down symbols, not draw pictures (though pictures *can* be a useful tool in developing proofs!)

If $f : \mathbf{R} \to \mathbf{R}$ is "continuous in the pencil sense," then the graph doesn't have any "jumps," which happens if whenever x and y are "close," that $f(x)$ and $f(y)$ are "close."

3.1 Continuity

Before we give the definition of continuity, we'll give some useful notation. If a property $P(x)$ holds for all x in a set X, we write $\forall x \in X\ P(x)$ (read "for all $x \in X$, $P(x)$). If a property holds for at least one x, we write $\exists x \in X\ P(x)$ (read "there exists an $x \in X$ such that $P(x)$.") The symbols $\forall$ and $\exists$ are called *quantifiers*.

Definition 3.1. A *continuous function* is a function $f : \mathbf{R} \to \mathbf{R}$ such that for each $x \in \mathbf{R}$ and $\varepsilon > 0$ there is a $\delta > 0$ such that for every $y \in \mathbf{R}$ such that $|y - x| < \delta$,

$$|f(y) - f(x)| < \varepsilon.$$

In symbols, one has

$$\forall x \in \mathbf{R}\ \forall \varepsilon > 0\ \exists \delta > 0\ \forall y \in \mathbf{R}\ |y - x| < \delta \implies |f(y) - f(x)| < \varepsilon.$$

The game of continuity goes like this. If we want our function to be continuous at $x \in \mathbf{R}$, we fix an $\varepsilon > 0$ – it doesn't matter how small. For this choice of ε, we need to find the corresponding $\delta > 0$ such that the above condition holds. In other words, our δ is a function of ε and x.

For this reason, *quantifier order is important*! If we had instead said that there is a $\delta > 0$ such that for every $\varepsilon > 0$... then *one* δ would have to work for every ε. This would imply that the range of the function had length at most δ.

Example 3.2. Let $f(x) = x$. We claim that f is continuous.

Indeed, let $\varepsilon > 0$ and $x \in \mathbf{R}$. We must find a $\delta > 0$ such that if $y \in \mathbf{R}$ and $|y - x| < \delta$, $|f(y) - f(x)| < \varepsilon$. But this just means that we need to find a δ such that if $|y - x| < \delta$, then $|y - x| < \varepsilon$. So let $\delta = \varepsilon$. Since we could find such a δ, it follows that f is continuous.

Discussion topic 3.3. Show that if $z \in \mathbf{R}$, the constant function $f(x) = z$ is continuous.

Another useful characterization of continuity is in terms of sequences.

Definition 3.4 (sequences)**.** Let X be a set. A *sequence* in X is a function $x : \mathbf{N} \to X$.

Since sequences are "special" functions, we usually write $x(n)$ as x_n, and write x as (x_n).

Definition 3.5 (convergence)**.** A sequence of real numbers (x_n) *converges* to $x \in \mathbf{R}$ in a metric space if

$$\forall \varepsilon > 0\ \exists N > 0\ \forall n \geq N\ |x_n - x| < \varepsilon.$$

We write this as $x_n \to x$ or

$$\lim_{n\to\infty} x_n = x.$$

The definition of a continuous function has several inequalities. We'll often need to use the triangle inequality (Homework 1.6) to break up a "difficult" inequality into lots of easier inequalities. For convenience, we restate the conclusion:

Theorem 3.6 (triangle inequality)**.** *Let $x, y, z \in \mathbf{R}$. Then*

$$|x - z| \leq |x - y| + |y - z|.$$

Theorem 3.7. *A function $f : \mathbf{R} \to \mathbf{R}$ is continuous if and only if for all $x \in \mathbf{R}$ and all sequences $x_n \to x$,*

$$\lim_{n\to\infty} f(x_n) = f(x).$$

Proof. First we fix $x \in \mathbf{R}$.

If f is continuous and (x_n) is a sequence converging to x, then there are really three different εs floating around:

1. We are given a sequence $x_n \to x$, meaning

$$\forall \varepsilon > 0\ \exists N > 0\ \forall n \geq N |x_n - x| < \varepsilon.$$

2. We know f is continuous at x, meaning

$$\forall \delta > 0\ \exists \rho > 0\ \forall y \in \mathbf{R},\ |y - x| < \delta \implies |f(y) - f(x)| < \varepsilon.$$

3. We want to show $f(x_n) \to f(x)$, meaning

$$\forall \varepsilon > 0\ \exists N > 0\ \forall n \geq N\ |f(x_n) - f(x)| < \varepsilon.$$

Let $\varepsilon > 0$. Since f is continuous, there is a $\delta > 0$ which is allowed to depend on ε such that if $|y - x| < \delta$ then $|f(y) - f(x)| < \varepsilon$. Since $x_n \to x$, there is a $N > 0$ which is allowed to depend on δ such that $\forall n \geq N$, $|x_n - x| < \delta$. But then $\forall n \geq N$, $|f(x_n) - f(x)| < \varepsilon$, which completes the proof.

For the converse, we'll proceed by contradiction. So now we need to talk about how to negate quantifiers. Notice that if $\neg$ is the symbol for "not", then $\neg\forall$ is the same as $\exists\neg$:

if it is not true that every x has some property, then there must be some x without that property. Similarly, $\neg\exists$ is the same as $\forall\neg$.

Since we are proceeding by contradiction, we assume that f is not continuous at x. So

$$\exists\varepsilon > 0\ \forall\delta > 0\ \exists y \in \mathbf{R}\ |y - x| < \delta \text{ and } |f(y) - f(x)| \geq \varepsilon.$$

So, let $\delta_\varepsilon = \frac{1}{n}$. As we vary n, we get a sequence x_n s.t. $|x_n - x| < \delta_\varepsilon = \frac{1}{n}$. In particular as $n \to \infty$, we see that $x_n \to x$ and for every n, $|f(x_n) - f(x)| \geq \varepsilon$. In other words, we have produced a sequence $\{x_n\}$ converging to x but $f(x_n) \not\to f(x)$ as desired. □

Proposition 3.8. *If $f, g : \mathbf{R} \to \mathbf{R}$ are continuous, then $g \circ f : X \to Z$ is continuous.*

Proof. Consider a sequence converging to p, $x_n \to p$.

Then, since the function f is continuous (using the limit function of continuity or the $\varepsilon - \delta$ definition) applying f to the sequence means $f(x_n) \to f(p)$. Let $y_n = f(x_n)$ and $q = f(p)$, since f maps $X \to Y$. Then $y_n \to q$.

Then applying g to $y_n \to q$ means $g(y_n) \to g(q)$ since g is also a continuous function. This implies $g \circ f(x_n) = g(f(x_n)) \to g(q)$ where $g(q) \in Z$. □

Example 3.9. Show that $f : \mathbf{R} \to \mathbf{R}$ where

$$f(x) = \begin{cases} \sin(\frac{1}{x}) & \text{if } x \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

is not continuous at 0.

Proof. Consider the sequence

$$x_k = \frac{1}{\frac{\pi}{2} + 2\pi k}.$$

Then $f(x_k) = \sin(\frac{\pi}{2} + 2\pi k) = 1$, because $\forall k \in \mathbf{Z}$,

$$\sin(\frac{\pi}{2} + 2\pi k) = 1.$$

Thus, $\lim_{k\to\infty} f(x_k) = 1 \neq 0 = f(0)$. □

Example 3.10. Show that $f : \mathbf{R} \to \mathbf{R}$ where

$$f(x) = \left\{ \begin{array}{cc} x\ \sin(\frac{1}{x}) & \text{if } x \neq 0 \\ 0 & \text{otherwise} \end{array} \right.$$

is continuous at 0.

Proof. Take $\varepsilon > 0$. We want to show that $|f(x)| < \varepsilon$ when $|x| < \delta$.
Assume $|x| < \delta$ and let $\delta = \varepsilon$. This is trivially true when $x = 0$. Thus, we want to show this is true when $x \neq 0$.

We know $-1 \leq \sin(\theta) \leq 1$ because sin oscillates between ±1. Let $\theta = \frac{1}{x}$. Then $-1 \leq \sin(\frac{1}{x}) \leq 1$.

$$\begin{aligned} |f(x)| &= |x \sin(\frac{1}{x})| \\ &= |x|\,|\sin(\frac{1}{x})| \\ &\leq |x| < \varepsilon \end{aligned}$$

□

From the above examples, we can see that to show an explicit f is continuous, it is usually easier to use (ε, δ) definition of continuity.

But to show an explicit f is not continuous it is usually easier to use the sequence definition of continuity: just find any sequence for which the convergence property fails, and f will then not be continuous.

Discussion topic 3.11. Show that if f and g are continuous, $f+g$, fg, and $|f|$ are continuous as well.

Discussion topic 3.12. Show that $\mathbf{Q}$ is *dense* in $\mathbf{R}$ in the sense that for every $x \in \mathbf{R}$, there is a sequence of rational numbers x_n such that $x_n \to x$.

Let $f, g : \mathbf{R} \to \mathbf{R}$ be continuous functions. Assume that for every $q \in \mathbf{Q}$, we have $f(q) = g(q)$. Now show that for every $x \in \mathbf{R}$, $f(x) = g(x)$.

Discussion topic 3.13. Let $f : \mathbf{Z} \to \mathbf{R}$ be any function. Show that f is continuous (so for every $x \in \mathbf{Z}$ and $\varepsilon > 0$ there is a $\delta > 0$ such that for every $y \in \mathbf{Z}$ with $|y - x| < \delta$, $|f(y) - f(x)| < \varepsilon$.

Discussion topic 3.14. Let $\alpha \in \mathbf{R}$. Say that a function $f : \mathbf{R} \to \mathbf{R}$ is α-*Hoelder* if there is a $C > 0$ such that for every $x, y \in \mathbf{R}$,

$$|f(y) - f(x)| \leq C|y - x|^{\alpha}.$$

Let f be α-Hoelder. Show that if $\alpha > 0$, then f is uniformly continuous. Show that if $\alpha > 1$ then f is constant.

Discussion topic 3.15. Let

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbf{Q} \\ 0 & \text{otherwise} \end{cases}$$

Show f is not continuous $\forall x \in \mathbf{R}$.

Homework 3.16. Show that if f is a polynomial, then f is continuous. (Hint: Induct on Homework 3.11.)

3.2 Continuous functions on compact intervals

Recall from calculus that the *closed interval* $[a, b]$ is defined by

$$[a, b] = \{x \in \mathbf{R} : a \leq x \leq b\}.$$

(For reasons we shall soon see, we also call this the *compact interval*.) In contrast to the compact interval, we also can consider the *open interval*

$$(a, b) = \{x \in \mathbf{R} : a < x < b\}.$$

For open intervals we allow $a = -\infty$ and $b = \infty$. So for example, $(a, \infty) = \{x \in \mathbf{R} : x > a\}$ while $(-\infty, \infty) = \mathbf{R}$.

In this section we'll prove the following three theorems about continuous functions:

Theorem 3.17 (intermediate value theorem)**.** *Let $f : [a, b] \to \mathbf{R}$ be continuous. Then $\forall y \in f([a, b])\ \exists x \in [a, b]$ such that $f(x) = y$.*

Theorem 3.18 (extreme value theorem)**.** *Let $f : [a, b] \to \mathbf{R}$ be continuous. Then there exist $c, d \in [a, b]$ such that for every $x \in [a, b]$, $f(x) \in [f(c), f(d)]$.*

Definition 3.19. A function $f : [a, b] \to \mathbf{R}$ is *uniformly continuous* if for every $\varepsilon > 0$ there is a $\delta > 0$ such that for every $x, y \in [a, b]$ with $|y - x| < \delta$, $|f(y) - f(x)| < \varepsilon$.

Notice that uniform continuity is stronger than continuity: δ does not depend on the point x, only on ε. A function whose derivative blows up will not be uniformly continuous. So you can think of uniform continuity as being related to "bounded derivatives." Besides, uniform continuity illustrates the importance of quantifier order: formally, a function f is uniformly continuous if

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall x, y \in \mathbf{R}\ |y - x| < \delta\ |f(y) - f(x)| < \varepsilon.$$

Note the similarity to the definition of continuity, the only difference being that we swapped quantifier order.

Theorem 3.20 (Heine-Cantor)**.** *Let $f : [a, b] \to \mathbf{R}$ be continuous. Then f is uniformly continuous.*

These results are not true about $\mathbf{Q}$, or even (a, b), or even $\mathbf{R}$!

Discussion topic 3.21. Show that there is a continuous function $\mathbf{Q} \to \mathbf{R}$ which violates the intermediate value theorem. (Hint: use Example 1.39.)

Discussion topic 3.22. Show that there is a continuous function $(a, b) \to \mathbf{R}$ with $-\infty < a < b < \infty$ which violates the extreme value theorem. (Hint: make the function surjective.)

Discussion topic 3.23. Show that there is a continuous function $\mathbf{R} \to \mathbf{R}$ which violates the Heine-Cantor theorem. (Hint: make the derivative surjective.)

So we must use a very special property of $[0, 1]$, to prove these theorems.

Definition 3.24. Let $\mathcal{U}$ be a family of open intervals (a, b) (where we allow $a = -\infty$ or $b = \infty$). If $\bigcup \mathcal{U} \supseteq A$, we say that $\mathcal{U}$ is a *open cover* of A.

Axiom 3.25 (compact-connected principle)**.** *Let $\mathcal{U}$ be an open cover of $[a, b]$. Then:*

1. *(Compactness) There are* finitely many *open intervals $U_1, \dots, U_n \in \mathcal{U}$ such that*
$$\mathcal{U}^\flat = \{U_1, \dots, U_n\}$$
is an open cover of $[a, b]$.

2. *(Connectedness) There are* two distinct *open intervals $U_1, U_2 \in \mathcal{U}$ such that $U_1 \cap U_2$ is nonempty.*

We will take the compact-connected principle on faith. It can be proven to be *equivalent to* other properties of $[0, 1]$ or $\mathbf{R}$ (the most famous being the Cauchy completion axiom, and the least-upper-bound axiom), but one cannot prove it completely from scratch. This is similar to the well-ordering principle of $\mathbf{N}$: there is no question of proving the compact-connected axiom. (In Math 104, you might prove that *there is an ordered field k* whose unit interval $[0, 1]$ satisfies compact-connected principle and simply declaring by fiat that $\mathbf{R} = k$. For example, see Chapter 1 of Pugh.)

At least the compact-connected principle is intuitive. Compactness means that you cannot fit infinitely many intervals into the finite region $[a, b]$. On the other hand, connectedness means that $[a, b]$ cannot be torn into two distinct intervals.

Discussion topic 3.26. Show that compactness fails for $\mathbf{R}$. That is, find an open cover $\mathcal{U}$ of $\mathbf{R}$ such that any finite subset $\mathcal{U}^\flat \subset \mathcal{U}$ is not an open cover of $\mathbf{R}$.

Discussion topic 3.27. Show that connectedness fails for $[0, 1] \cup [2, 3]$. That is, find an open cover $\mathcal{U}$ of $[0, 1] \cup [2, 3]$ such that for every distinct $U_1, U_2 \in \mathcal{U}$, $U_1 \cap U_2$ is empty.

Proving results using the compactness half, which is also known as the *Heine-Borel covering lemma*, is an example of a *compactness argument*: we have infinitely many open intervals to work with, which is rather difficult. So, to avoid the hassle, we replace them with a more "compact" alternative: only finitely many open intervals, which is much more manageable. (More generally, a set satisfying a variant of the Heine-Borel theorem is called a *compact set*, but discussing this would take us too far afield.)

Notice the analogy with induction: $\mathbf{N}$ had a certain wonderful property, the well-ordering principle, which gave us a powerful proof technique which reduced an infinite process to a process with only two steps. Here, $[a, b]$ also has a wonderful property, the compact-connected principle, which gives a proof technique which reduces an infinite process to a process with only finitely many steps.

Before proving these theorems, we'll also prove the following lemmata.

Lemma 3.28. *Let $f : \mathbf{R} \to \mathbf{R}$ be a function. Then f is continuous iff for every $a < b$, the preimage of (a, b),*
$$f^{-1}((a, b)) = \{x \in \mathbf{R} : f(x) \in (a, b)\},$$
can be written as
$$f^{-1}((a, b)) = (x_1, y_1) \cup (x_2, y_2) \cup (x_3, y_3) \cup \cdots$$
for some disjoint $(x_1, y_1), (x_2, y_2), \dots$.

Proof. First assume f is continuous. Let $a < b \in \mathbf{R}$. Either $U = f^{-1}((a,b))$ is empty or it isn't. If U is empty, then we're done. So we can assume that there is a $x \in U$.

If we choose $\varepsilon > 0$ to be small enough, then $(f(x)-\varepsilon, f(x)+\varepsilon) \subseteq (a,b)$. So $f^{-1}((f(x)-\varepsilon, f(x)+\varepsilon)) \subseteq U$. So by continuity, there is a $\delta > 0$ such that if $|y-x| < \delta$, $f(y) \in (a,b)$ and so $y \in U$. The set of all such y is $(x-\delta, x+\delta)$, so $(x-\delta, x+\delta) \subseteq U$. So x is contained in an open interval contained in U.

Let $\mathcal{U}(x)$ denote the set of all open intervals containing x which are contained in U. All these open intervals intersect at x, and the union of overlapping open intervals is an open interval, say $\bigcup \mathcal{U}(x) = (\alpha(x), \beta(x))$. Since x was arbitrary, we conclude that we can write U as a union of open intervals $(\alpha(x), \beta(x))$. Besides, if $y \in (\alpha(x), \beta(x))$, then the largest open interval containing y contained in U is $(\alpha(x), \beta(x))$, so $\alpha(x) = \alpha(y)$ and $\beta(x) = \beta(y)$. We conclude that that y lies in no other open interval. So the open intervals are disjoint.

Discussion topic 3.29. Prove the converse. (Hint: for every $x \in \mathbf{R}$ and $\varepsilon > 0$, use the fact that $(f(x)-\varepsilon, f(x)+\varepsilon)$ is an open interval.)

□

Lemma 3.30. *If $K \subseteq \mathbf{R}$ satisfies the Heine-Borel covering lemma, then K has a minimum and a maximum.*

In other words, assume that for every open cover $\mathcal{U}$, there are finitely many open intervals $U_1, \ldots, U_n \in \mathcal{U}$ such that

$$\mathcal{U}^\flat = \{U_1, \ldots, U_n\}$$

is also an open cover of K. If this is true, then there is an element $x \in K$ such that for every $y \in K$, $f(y) \leq f(x)$ and an element $x \in K$ such that for every $y \in K$, $f(y) \geq f(x)$.

Proof. We first claim that K is bounded: in other words, there is a $R > 0$ such that $K \subseteq [-R, R]$. To do this, we need to use a compactness argument.

For every $x \in K$, let $U(x) = (x-1, x+1)$ and let $\mathcal{U} = \{U(x) : x \in K\}$. Then $x \in U(x) \subseteq \mathcal{U}$, so $\mathcal{U}$ is an open cover. By compactness, it follows that there is a subset of $\mathcal{U}$, $\mathcal{U}^\flat = \{U(x_1), \ldots, U(x_n)\}$, which is also an open cover. So let

$$R = \max(|x_1|, |x_2|, \ldots, |x_n|) + 1.$$

Then if $x \in K$, there is a $i \leq n$ such that $x \in U(x_i)$. Moreover, $U(x_i) \subseteq [-R, R]$. So $K \subseteq [-R, R]$. This proves the claim.

We will now prove that K has a maximum. To do this, we claim that the set K' of $x \in \mathbf{R}$ such that for every $y \in K$, $x > y$, is an open interval, say (α, ∞). From this, we will see that $\max K = \alpha$. Again we need to use a compactness argument.

Let $x \in K'$ and for every $y \in K$ let $\omega(y) = (x-y)/2$. (So $\omega(y) > 0$ since $x > y$.) Let $U(y) = (y - \omega(y), y + \omega(y))$ and let $\mathcal{U}$ be the set of all such $U(y)$. Also, let $V(y) = (x - \omega(y), x + \omega(y))$. Then $\mathcal{U}$ is an open cover, and let's say $\mathcal{U}^\flat = \{U(y_1), \ldots, U(y_n)\}$ is also an open cover. Let $W = \bigcup \mathcal{U}^\flat$ and let $S = V(y_1) \cap \cdots \cap V(y_n)$.

We will now prove that $S \cap W$ is empty. To prove that a set is empty, we proceed by contradiction. Let $z \in S \cap W$. Then $z \in V(y_i)$ for every $i \leq n$, but also there is a $j \leq n$ such that $z \in U(y_j)$. So $z \in U(y_j) \cap V(y_j)$. Therefore

$$|y_j - x| \leq |z - x| + |y_j - z| < \omega(y_j) + \omega(y_j) = 2\omega(y_j) = |y_j - x|.$$

Therefore $|y_j - x| < |y_j - x|$, a contradiction. So $S \cap W$ is empty.

Since $S \cap W$ is empty, then by choosing $\delta(x) = \min(\omega(y_1), \ldots, \omega(y_n))/2$, $(x - \delta(x), x + \delta(x)) \subset K'$. Since x was arbitrary, it will follow that $K' = \bigcup_{x \in K'}(x - \delta(x), x + \delta(x))$, so K' is an open interval.

Discussion topic 3.31. Prove that K has a minimum.

□

With all this setup done, the proof of these three famous calculus theorems turns out to be quite easy!

Proof of Theorem 3.17. Intuitively, if the intermediate value theorem failed, then a continuous function would be able to "rip $[a, b]$ in two." This clearly would violate connectedness. So we proceed by contradiction.

Suppose that there is a $y \in f([a, b])$ such that for every $x \in [a, b]$, $f(x) \neq y$. We now consider the open intervals $(-\infty, y)$ and (y, ∞). Then $[a, b] = f^{-1}((-\infty, y)) \cup f^{-1}((y, \infty))$.

By Lemma 3.28 $f^{-1}((-\infty, y))$ and $f^{-1}((y, \infty))$ are disjoint unions of open intervals in $[a, b]$, so we can write $[a, b]$ as a disjoint union of open intervals, say $[a, b] = \bigcup \mathcal{U}$ for some collection of disjoint, open intervals $\mathcal{U}$. Therefore there are $\alpha < \beta$ such that $[a, b] = (\alpha, \beta) \cup U$ where $U = \bigcup \mathcal{U} \setminus (\alpha, \beta)$, and the union is disjoint. This contradicts connectedness. Therefore the intermediate value theorem is true. □

Proof of Theorem 3.18. Let $K = f([a, b])$. We claim that K satisfies the Heine-Borel covering lemma. Then Lemma 3.30 will finish the proof.

Let $\mathcal{U}$ be an open cover of K. For each U in $\mathcal{U}$, we let $\mathcal{V}(U)$ be the collection of open intervals given by Lemma 3.28. Then $\mathcal{W} = \bigcup_U \mathcal{V}(U)$ is an open cover of $[a, b]$: if $x \in [a, b]$, then there is a $y \in K$ such that $f(x) = y$, and a $U \in \mathcal{U}$ such that $y \in U$, so a $V \in \mathcal{V}(U)$ such that $x \in V$, and $V \in \mathcal{W}$. So there is a subset $\mathcal{W}^\flat = \{W_1, \ldots, W_n\}$ of $\mathcal{W}$ which is also an open cover, by compactness of $[a, b]$.

Each of the W_j was contained in some $f^{-1}(U)$, so let U_j be the $U \in \mathcal{U}$ such that $f^{-1}(U_j)$ contained W_j, and let $\mathcal{U}^\flat = \{U_1, \ldots, U_n\}$. Then if $y \in K$, there is a $x \in [a, b]$ and a $j \leq n$ such that $x \in W_j$ and $f(x) = y$. So $y \in U_j$. So $\mathcal{U}^\flat$ is an open cover of K. □

Proof of Theorem 3.20. Let $\varepsilon > 0$. To prove the Heine-Cantor theorem we must find a $\delta > 0$ such that if $|y - x| < \delta$, then $|f(y) - f(x)| < \varepsilon$.

Since f is continuous, for every $x \in [a, b]$ there is a $\omega(x) > 0$ such that if $|y - x| < \omega(x)$, $|f(y) - f(x)| < \varepsilon/2$. We thus have infinitely many $\omega(x)$'s to work with! This seems problematic, so we construct an open cover where each open interval corresponds to a $\omega(x)$, and then use a compactness argument to show that there are really only finitely many $\omega(x)$'s to worry about.

Let

$$U(x) = \left(x - \frac{\omega(x)}{2}, x + \frac{\omega(x)}{2}\right).$$

Then $|f(y) - f(x)| < \varepsilon/2$ as long as $y \in U(x)$. If

$$\mathcal{U} = \{U(x) : x \in [a, b]\},$$

then for every $x \in [a,b]$, $x \in U(x) \subseteq \bigcup \mathcal{U}$, so $\mathcal{U}$ is an open cover of $[a,b]$. Therefore there are finitely many $x_1, \dots, x_n$ such that $\mathcal{U}^\flat = \{U(x_1), \dots, U(x_n)\}$ is an open cover of $[a,b]$.

Now let

$$\delta = \frac{\min(\omega(x_1), \dots, \omega(x_n))}{2}.$$

Since we are taking the minimum of a finite set and every $\omega(x_i) > 0$, one has $\delta > 0$. To finish the proof we just need to show that if $|y - x| < \delta$, then $|f(y) - f(x)| < \varepsilon$.

Let $|y - x| < \delta$. Since $\mathcal{U}^\flat$ is an open cover of $[a,b]$, there is a $i \leq n$ such that $x \in U(x_i)$. Then by definition of $U(x_i)$,

$$|x_i - x| < \frac{\omega(x_i)}{2}$$

so $|f(x_i) - f(x)| < \varepsilon/2$. Moreover, we can use the triangle inequality, Theorem 3.6, to prove that

$$|y - x_i| \leq |x_i - x| + |y - x| < \frac{\omega(x_i)}{2} + \delta \leq \frac{\omega(x_i)}{2} + \frac{\omega(x_i)}{2} = \omega(x_i).$$

It follows that $|f(y) - f(x_i)| < \varepsilon/2$. So

$$|f(y) - f(x)| \leq |f(y) - f(x_i)| + |f(x_i) - f(x)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

again by the triangle inequality. □

Homework 3.32. Let $f : [0,1] \to \mathbf{Z}$ be a continuous function. Show that f is constant.

Homework 3.33. Let $K = [0,5] \cap \mathbf{Q}$ and show that there are two open intervals U_1, U_2 such that $U_1 \cap U_2$ is empty and $U_1 \cup U_2 = K$.

Homework 3.34. Find an explicit example of an open cover $\mathcal{U}$ of $(0,1)$ such that for every finite subset $\mathcal{U}^\flat \subset \mathcal{U}$, $\mathcal{U}^\flat$ is not an open cover of $(0,1)$.

Homework 3.35. Let $f : [0,1] \to \mathbf{R}$ be continuous and assume that for every $q \in [0,1] \cap \mathbf{Q}$ we have $f(q) = 0$. Show that for every $x \in \mathbf{R}$, $f(x) = 0$.

Homework 3.36. Let $C([0,1])$ be the set of all continuous functions $[0,1] \to \mathbf{R}$. For $f \in C([0,1])$, define $|f|_\infty = \max_{x \in [0,1]} |f(x)|$, which makes sense by the extreme value theorem and Homework 3.11. Show that a version of the triangle inequality holds, namely

$$|f - h|_\infty \leq |f - g|_\infty + |g - h|_\infty.$$

Homework 3.37. Let $f : [0,1] \to [0,1]$ be a continuous, strictly increasing function (that is, if $x < y$, then $f(x) < f(y)$), such that $f(0) = 0$ and $f(1) = 1$. Show that f is a bijection, using the intermediate value theorem (which you should not prove; with the tools we've given you so far, this would be very hard):

3.3 Differentiability

You may recall the formal definitions of the *derivative*. Informally, of course, the derivative represents the slope of the tangent line to the graph at the point, as approximated better and better by slopes of secant lines.

First, we need the notion of a limit of a function.

Definition 3.38. Let $f : X \to Y$ be a function between metric spaces, $x_0 \in X$, and $y_0 \in Y$. Suppose that for each $\varepsilon > 0$, there is a $\delta > 0$ such that if $d(x, x_0) < \delta$, then $d(f(x), y_0) < \varepsilon$. Then y_0 is the *limit* of $f(x)$ as $x \to x_0$, written

$$\lim_{x \to x_0} f(x) = y_0.$$

Discussion topic 3.39. Show that if $f : (a, b) \to \mathbf{R}$ is continuous and $x_0 \in (a, b)$, then $\lim_{x \to x_0} f(x) = f(x_0)$.

We're now ready to define the derivative formally.

Definition 3.40 (Derivative, Differentiable)**.** Let $f : (a, b) \to \mathbf{R}$, and let $x_0 \in (a, b)$. Consider the limit

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

If this limit exists, then it is the *derivative* of f at x_0, denoted $f'(x_0)$. We also say that f is *differentiable* at x_0. If f is differentiable at every $x_0 \in (a, b)$, then f is simply *differentiable.*

In addition, note that the following limits are equivalent.

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

Proposition 3.41. *Let $f : (a, b) \to \mathbf{R}$ be differentiable at $x_0 \in (a, b)$. Then f is continuous at x_0.*

Proof. Since f is differentiable at x_0, the limit

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

exists. Observe that as $x \to x_0$, $(x - x_0) \to 0$, and so

$$\lim_{x \to x_0} \left(\frac{f(x) - f(x_0)}{x - x_0} \cdot (x - x_0) \right) = 0$$

(We are implicitly using the fact that the limit of a product is the product of the limits, if the limits exists. The proof of this is similar to a previous homework problem and left as an exercise.) Then

$$\lim_{x \to x_0} (f(x) - f(x_0)) = 0$$

Finally, $f(x_0)$ is a constant, so

$$\lim_{x \to x_0} f(x) = f(x_0)$$

This means f is continuous at x_0. □

Discussion topic 3.42. Show that if f is a constant function then $f' = 0$.

Proposition 3.43. *Let $f, g : (a, b) \to \mathbf{R}$ be differentiable at x_0 and $c, d \in \mathbf{R}$. Then:*

1. *(Linearity) $cf + dg$ is differentiable at x_0 and $(cf + dg)'(x_0) = cf'(x_0) + dg'(x_0)$.*
2. *(Leibniz rule) fg is differentiable at x_0 and $(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$.*
3. *(chain rule) For g differentiable at $f(x_0)$, $f \circ g$ is differentiable at x_0 and $(f \circ g)'(x_0) = f'(g(x_0))g'(x_0)$.*

Proof. 1. We have

$$\begin{aligned}\lim_{x\to x_0} \frac{(cf + dg)(x) - (cf + dg)(x_0)}{x - x_0} &= \lim_{x\to x_0} \frac{cf(x) - cf(x_0)}{x - x_0} + \frac{dg(x) - dg(x_0)}{x - x_0} \\ &= cf'(x_0) + dg'(x_0).\end{aligned}$$

2. Notice that

$$\begin{aligned}(fg)(x) - (fg)(x_0) &= f(x)g(x) - f(x_0)g(x_0) \\ &= f(x)g(x) - f(x_0)g(x_0) + f(x_0)g(x) - f(x_0)g(x) \\ &= (f(x) - f(x_0))g(x) + f(x_0)(g(x) - g(x_0)).\end{aligned}$$

Since g is continuous, as $x \to x_0$,

$$\frac{(fg)(x) - (fg)(x_0)}{x - x_0} = \frac{f(x) - f(x_0)}{x - x_0} g(x) + f(x_0)\frac{g(x) - g(x_0)}{x - x_0} \to f'(x)g(x) + f(x)g'(x).$$

3. Let $y_0 = f(x_0)$ and $y = f(x)$. Since g is differentiable, the function

$$\rho(y) = \frac{g(y) - g(y_0)}{y - y_0} - g'(y)$$

has $\rho(y) \to 0$ as $y \to y_0$. Define $\rho(0) = 0$. Then $g(y) - g(y_0) = (g'(y) + \rho(y))(y - y_0)$, even if $y = y_0$. Since f is continuous, $y - y_0 \to 0$ as $x \to x_0$. So

$$\frac{g(y) - g(y_0)}{x - x_0} = (g'(y) + \rho(y))\frac{y - y_0}{x - x_0} \to g'(y_0)f'(x_0).$$

□

Proposition 3.44 (first derivative test)**.** *Let $f : (a, b) \to \mathbf{R}$ be differentiable and f let have a* local *maximum or minimum at $x_0 \in (a, b)$ (so there is a $\delta > 0$ such that $|x - x_0| < \delta$, $f(x) \le f(x_0)$). Then $f'(x_0) = 0$.*

Proof. We prove this for when f has a local maximum. The proof for local minima is similar.

Since f has a maximum, if $\rho \in (-\delta, \delta)$, then $f(x_0) \ge f(x_0 + \rho)$.

Discussion topic 3.45. Show that we can assume $\rho > 0$.

Since $\rho > 0$,

$$\frac{f(x_0 + \rho) - f(x_0)}{\rho} \le 0.$$

This remains true even as $\rho \to 0$. By continuity of f, the claim holds. □

Theorem 3.46 (mean value theorem)**.** *Let $f : [a, b] \to \mathbf{R}$ be continuous and let f be differentiable on (a, b). Then there is a $x \in (a, b)$ such that*

$$f'(x) = \frac{f(b) - f(a)}{b - a}.$$

Proof. Let

$$S = \frac{f(b) - f(a)}{b - a}.$$

So S is the slope of the secant line of f through b and a. Then $g(x) = f(x) - Sx$ is differentiable, hence continuous. Also, $g(a) = g(b)$, namely

$$g(a) = g(b) = \frac{bf(a) - af(b)}{b - a}.$$

By the extreme value theorem, g has a maximum and a minimum.

If g is a constant, then $g' = 0$ and we can choose any $x \in (a, b)$ since then $f' = S$. Otherwise, $\min g < \max g$ and since $g(a) = g(b)$, if $\min g = g(a)$ then $\max g$ is attained somewhere in (a, b). Similarly if $\max g = g(b)$ then $\min g$ is attained somewhere in (a, b). So there is a point $x \in (a, b)$ where g attains its minimum or maximum. Therefore g has a local maximum or local minimum at x. So $g'(x) = 0$ by the first derivative test. □

Discussion topic 3.47 (Cauchy's mean value theorem). Let $f, g : [a, b] \to \mathbf{R}$ be continuous and differentiable on (a, b). Show that there is a $x \in (a, b)$ such that

$$(f(b) - f(a))g'(x) = f'(x)(g(b) - g(a)).$$

The proof is basically the same as the mean value theorem.

Theorem 3.48 (l'Hospital's rule)**.** *Let $f, g : (a, b) \to \mathbf{R}$ be differentiable such that $f(x) \to 0$ and $g(x) \to 0$ as $x \to b$. Assume that for all $x \in \mathbf{R}$, $g(x) \neq 0$, and that there is a $L \in \mathbf{R}$ such that*

$$\lim_{x \to b} \frac{f'(x)}{g'(x)} = L.$$

Then

$$\lim_{x \to b} \frac{f(x)}{g(x)} = L.$$

Proof. Let $\varepsilon > 0$. We must find a $\delta > 0$ such that $|f(x)/g(x) - L| < \delta$ provided that $x > b - \delta$. Actually, we do have a δ such that if $x > b - \delta$ then

$$\left| \frac{f'(x)}{g'(x)} - L \right| < \frac{\varepsilon}{2}.$$

For each $x > b - \delta$, let $y(x) > b - \delta$ be so close to b that

$$|f(y(x))| + |g(y(x))| < \frac{g(x)^2 \varepsilon}{4|f(x)| + 4|g(x)|}$$

and $2|g(y(x))| < |g(x)|$. Such a $y(x)$ exists because $f \to 0$ and $g \to 0$. Then

$$\begin{aligned}\left|\frac{f(x)}{g(x)} - L\right| &= \left|\frac{f(x)}{g(x)} - \frac{f(x)-f(y)}{g(x)-g(y)} + \frac{f(x)-f(y)}{g(x)-g(y)} - L\right| \\ &\leq \left|\frac{g(x)f(y) - f(x)g(y)}{g(x)(g(x)-g(y))}\right| + \left|\frac{f'(z)}{g'(z)} - L\right| < \varepsilon\end{aligned}$$

for some $z \in (x, y)$ given by Cauchy's mean value theorem. □

For these homeworks, it's okay to use the standard derivative rules $(e^x)' = e^x$, $\sin' = \cos$, and $\cos' = -\sin$ (along with the linear, Leibniz, and chain rules we proved above). You might also use Euler's formula, Example 1.4.

Homework 3.49. Show that the derivative of a polynomial $a_0 + a_1 x + \cdots + a_n x^n$ is $a_1 + 2a_2 x + \cdots + n a_n x^{n-1}$.

Homework 3.50. Show that there is a continuous function which is not differentiable.

Homework 3.51 (second derivative test). Let $f : (a, b) \to \mathbf{R}$ and assume f' is differentiable and $f'(x_0) = 0$. Show that if $f''(x_0) > 0$ then f has a local minimum at x_0 and if $f''(x_0) < 0$ then f has a local maximum at x_0.

Homework 3.52. Let $f, g : (a, b) \to \mathbf{R}$ be differentiable at x_0 and assume $g(x_0) \neq 0$. Show then that

$$\frac{f}{g}'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}.$$

Homework 3.53. Show that $\sqrt{x+1} - \sqrt{x} \to 0$ as $x \to \infty$.

Homework 3.54. Show that the function $f(x) = e^{x^3+1}$ is a bijection $\mathbf{R} \to (0, \infty)$.

Definition 3.55. Let $f : \mathbf{R} \to \mathbf{R}$ be a function. We say that f is a *smooth function* if for every $x \in \mathbf{R}$, the higher derivative $f^{(n)}(x) = f''^{\ldots\prime}(x)$ (n primes) exists.

Homework 3.56. Let

$$f(x) = \exp\left(-\frac{1}{1-x^2}\right)$$

if $x \in (-1, 1)$ or $f(x) = 0$ otherwise. Show that f is smooth – this isn't obvious at $x = -1$ or $x = 1$, but you can use l'Hospital's rule and induction to pull it off.

Definition 3.57. Let $\mathcal{U}$ be an open cover of $[0, 1]$. A *partition of unity* subordinate to $\mathcal{U}$ is a family of smooth functions f_U, for each $U \in \mathcal{U}$, such that:

1. For every $x \in [0, 1]$, $\sum_{U \in \mathcal{U}} f_U(x) = 1$.

2. For every $x \in [0, 1]$ and every $U \in \mathcal{U}$, if $x \notin U$, then $f_U(x) = 0$.

Homework 3.58. Show that for any open cover $\mathcal{U}$ of $[0, 1]$, there is a partition of unity subordinate to $\mathcal{U}$.

Homework 3.59 (Cauchy-Riemann equations). This exercise is for students of Math 53. The definition of differentiability makes perfect sense for functions $f : \mathbf{C} \to \mathbf{C}$ and points $z \in \mathbf{C}$ (though z might approach z_0 from any direction, not just along the real line). If $f : \mathbf{C} \to \mathbf{C}$ satisfies the above condition, then we say that f is a *holomorphic function*. Let $f(x + iy) = u(x, y) + iv(x, y)$, where $u : \mathbf{R}^2 \to \mathbf{C}$ and $v : \mathbf{R}^2 \to \mathbf{C}$ are functions, and $f : \mathbf{C} \to \mathbf{C}$ is holomorphic. Show that then

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$$
$$\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

Chapter 4

Pathology

Here we will discuss some interesting counterexamples, concerning the lengths of subsets of $\mathbf{R}$. The moral to take away here is that your intuition will often betray you.

4.1 The length of Q and ε-management

In previous proofs, we've had to sum up a finite number of εs. In order to prove that some quantity $x < \varepsilon$, where $x = x_1 + \cdots + x_n$ and each of the x_j could be arbitrarily small, we put $x_j < \varepsilon/n$ for each j. But this is not always so easy – for example, if we have to work with *infinite* sums, as we will have to in this example.

Suppose that $U \subseteq \mathbf{R}$ consists of a countable union of open intervals (a_j, b_j), and by drawing a picture we can see that the length μ of U satisfies

$$\mu(U) \leq \sum_{j=1}^{\infty} b_j - a_j.$$

Indeed, the length of a single open interval (a, b) is $b - a$, and while the intervals that make up U might overlap, this just means that we will overestimate the length of U by summing their lengths.

In fact, if $A \subseteq U$ is *any* set, then $\mu(A) \leq \mu(U)$.

Definition 4.1. Let $\mathcal{U}$ be a collection of open intervals $(a_j, b_j) \subseteq \mathbf{R}$. If $A \subseteq \bigcup \mathcal{U}$, then we say that $\mathcal{U}$ is an *open cover* of A.

Summarizing the above discussion,

Lemma 4.2 (countable subadditivity). *If $A \subseteq \mathbf{R}$ and $(a_1, b_1), (a_2, b_2), \ldots$ is a countable open cover of A, then*

$$\mu(A) \leq \sum_{j=1}^{\infty} \mu((a_j, b_j)) = \sum_{j=1}^{\infty} b_j - a_j.$$

From this, it follows that a wealth of sets will have no length.

Lemma 4.3. *Let $A \subseteq \mathbf{R}$ and assume that for every $\varepsilon > 0$, there is an open cover $U_1, \ldots, U_j, \ldots$ of A such that the total length*

$$\sum_{j=1}^{\infty} \mu(U_j) < \varepsilon.$$

Then $\mu(A) = 0$.

Proof. By definition of length, $\mu(A) \geq 0$. Assume that $\mu(A) > 0$. Then we have an open cover $U_1, \ldots, U_j, \ldots$ such that $\sum_j \mu(U_j) < \mu(A)$ (taking $\varepsilon = \mu(A)$). So $\mu(A) \leq \sum_j \mu(U_j) < \mu(A)$ which is a contradiction. □

Theorem 4.4. *Let $A \subseteq \mathbf{R}$ be countable. Then $\mu(A) = 0$.*

Proof. The trick is to "give ourselves a ε of room," which we "spread into infinitely many pieces" using a convergent sum. We can't spread ε evenly, since if we divide it up as ε/n then $\sum_j \varepsilon/n = \infty$.

Let $\varepsilon > 0$, and let $A = \{x_1, x_2, \ldots\}$. For each x_j, define $a_j = x_j - \varepsilon/2^{j+1}$ and $b_j = x_j + \varepsilon/2^{j+1}$. Then $b_j - a_j = \varepsilon/2^j$.

Let us take for granted that

$$\sum_{j=1}^{\infty} \frac{1}{2^j} = 1.$$

Since $(a_1, b_1), \ldots, (a_j, b_j), \ldots$ is clearly an open cover,

$$\mu(A) < \sum_{j=1}^{\infty} b_j - a_j = \varepsilon \sum_{j=1}^{\infty} \frac{1}{2^j} = \varepsilon.$$

So $\mu(A) = 0$. □

Recall that $\mathbf{Q}$ is countable, so we're done. The fact that there is a rational number between every point on the line, yet the total length of the set of rational numbers is 0, is rather disturbing, no?

Homework 4.5. Show that

$$\sum_{j=1}^{\infty} \frac{1}{2^j} = 1,$$

as follows. First, show that

$$\sum_{j=1}^{n-1} \frac{1}{2^j} = \frac{1}{2}\left(\frac{1-(1/2)^n}{1-1/2}\right) = S_{n-1}.$$

Then, show that $S_n \to 1$ as $n \to \infty$.

4.2 The Cantor set

Here's a rather advanced example. Whenever considering this example, it's important to draw lots of pictures.

Let us define a subset of $\mathbf{R}$, known as the *Cantor set*, which was clearly named after its discoverer, Henry John Smith.

Start with the unit interval, $[0,1]$, which we shall write as $C_0 = [0,1]$. Remove its middle third $(1/3, 2/3)$ to get a set $C_1 = [0,1/3] \cup [2/3,1]$. Remove the middle thirds of each of the parts of C_1, $(1/9,2/9) \cup (7/9,8/9)$, to get a set $C_2 = [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1]$. Iterate this process forever to get a sequence of sets $C_0, C_1, C_2 \ldots$ such that

$$C_0 \supset C_1 \supset C_2 \supset C_3 \supset \ldots.$$

Definition 4.6. The *Cantor set* is $C = \bigcap_n C_n$.

Let us now derive the following seemingly paradoxical family of properties of C. First, let us check that the definition isn't completely silly.

Proposition 4.7. *C is nonempty.*

Proof. $0 \in C_j$ for every j, since 0 is always an endpoint and so is never part of the deleted middle third, so $0 \in C$. □

Then let us show that there are gaps between every point in the Cantor set, yet points in the Cantor set come arbitrarily close to each other. This should remind you somewhat of the behavior of points in $\mathbf{Q}$. To emphasize this bizarre behavior, Charles Pugh and other authors have referred to C as "Cantor dust."

Proposition 4.8. *If $x \in C$ and $\varepsilon > 0$, then there is a $y \in \mathbf{R}$ such that $y \notin C$ and $|y-x| < \varepsilon$.*

Proof. Since $x \in C$, if $j \geq 1$ then there is an interval $[a,b] \subseteq C_j$ such that $x \in [a,b]$. On the other hand, if we take $[a,b]$ as large as possible, then $b-a = (1/3)^j$, as can be checked by induction: if $j = 0$ we have $[a,b] = [0,1]$ so $b - a = 1 = (1/3)^0$, while if intervals of C_j have length $(1/3)^j$, then intervals of C_{j+1} have length $(1/3)^{j+1}$, since they are a third of their "parent" interval in C_j.

So if j is large enough, $b - a < \varepsilon/2$, so there is a point $y \in \mathbf{R}$ such that $x - \varepsilon < y < a \leq x \leq b$. If y is large enough, then y will not fall into the next interval before $[a,b]$. So $y \notin C_j$, so $y \notin C$. □

Proposition 4.9. *If $x \in C$ and $\varepsilon > 0$, then there is a $y \in C$ such that $x \neq y$ and $|y-x| < \varepsilon$.*

Proof. Using the same notation as the previous proof, take j so large that $b - a < \varepsilon/2$, and put $y = a$. Then $|y - x| < \varepsilon$. Since y is an endpoint of an interval, $y \in C_k$ for every $k \geq j$, and since $y \in C_j$ we already have $y \in C_k$ for every $k \leq j$. So $y \in C$. □

Proposition 4.10. *There does not exist an open interval (a,b) such that $(a,b) \subseteq C$.*

Proof. Since we are being asked to prove that something does not exist, we proceed by contradiction.

Suppose $(a,b) \subseteq C$. Then if $x = (b-a)/2$, there is a $y \in \mathbf{R}$ such that $y \notin C$ (so $y \notin (a,b)$) and $|y - x| < (b-a)/4$. But then $a < y < b$ so $y \in (a,b)$, which is a contradiction. □

But in spite of being "dust," spread sparsely throughout $[0, 1]$, C is uncountable.

Proposition 4.11. *C is uncountable.*

Proof. Let us identify C with another object: the set 2^ω of all infinite binary strings. An *infinite binary string* is a sequence of 0's and 1's, running off forever. For example, the string

$$\beta = 0110101010100101010010100101010\dots$$

could be the start of an infinite binary string. If α is an infinite binary string, we write α_j to mean the jth entry in the binary string. For example, in the above string $\beta_4 = 0$.

Given an infinite binary string α, we find a point $x(\alpha)$ of C, as follows. If $\alpha_1 = 0$, then we start in the left interval $[0, 1/3]$. Otherwise $\alpha_1 = 1$ and we start in the right interval $[2/3, 1]$. Then if $\alpha_2 = 0$, we go left, either to $[0, 1/9]$ if we were already in $[0, 1/3]$ or $[2/3, 7/9]$ if we were in $[2/3, 1]$. If $\alpha_2 = 1$, we go right, to $[2/9, 1/3]$ or $[8/9, 1]$.

We repeat this process forever: if $\alpha_j = 0$, we go left in C_j, and if $\alpha_j = 1$, we go right in C_j. Since we never leave the Cantor set, we end up with a point $x(\alpha)$ of C. On the other, if $\beta \neq \alpha$, say $\beta_j \neq \alpha_j$, then we take a different turn at C_j, so $x(\beta) \neq x(\alpha)$. So x is an injection $2^\omega \to C$.

Discussion topic 4.12. Prove that 2^ω is uncountable.

If we could count C, then we could use the inverse of x to count 2^ω, which is a contradiction. So C is uncountable. □

Homework 4.13. If you know the Schröder-Bernstein theorem (Homework 5.16), and you are brave enough, prove that there is a bijection $C \to \mathbf{R}$. It's not hard to show that there is an injection $C \to \mathbf{R}$, so the hard part is to show the opposite direction, but the binary string trick will help with that.

Homework 4.14. Show that the Cantor set C has $\mu(C) = 0$, even though C is uncountable. (On the other hand, you might read on Wikipedia about the *fat Cantor set F*, which behaves very similarly to the Cantor set but has $\mu(F) > 0$.)

4.3 A nonmeasurable set

Is μ defined for every $A \subseteq \mathbf{R}$? This turns out to be a rather touchy question.

The function μ, if it should agree with our intuitive notion of length, should satisfy certain properties:

1. $\mu(A) \geq 0$. (So a set should not have negative length.)

2. Assume that $\mathcal{F}$ is a countable family of disjoint sets. Then if $A = \bigcup \mathcal{F}$, $\mu(A) = \sum_{F \in \mathcal{F}} \mu(F)$. (So the length of A consists of the lengths of its parts.)

3. Assume that $x \in \mathbf{R}$, and write $A + x = \{a + x : a \in A\}$ (so $A + x$ is just A, but shifted to the left by x). Then $\mu(A) = \mu(A + x)$. (So the length of A is not changed by shifting it.)

If $\mu(A)$ can be defined, then A is said to be a *measurable set*.

Let's work on the circle S. If $\alpha \in \mathbf{R}$, then we can define a map $R_\alpha : S \to S$ by rotating S by $2\pi\alpha$.

Definition 4.15. The *orbit* of a point $x \in S$ by R_α is

$$O_\alpha(x) = \{R_{n\alpha}(x) : n \in \mathbf{Z}\}.$$

The orbit of x is the set of all points that x will end up at if we repeated rotate the circle by α or $-\alpha$.

Discussion topic 4.16. If α is irrational, then:

1. $O_\alpha(x)$ is countable and infinite.

2. The family of sets
$$\mathcal{F} = \{O_\alpha(x) : x \in S\}$$
is uncountable and disjoint, and $\bigcup \mathcal{F} = S$.

Theorem 4.17. *If $P \subset S$ contains exactly one point from each element of $\mathcal{F}$, then P is nonmeasurable.*

Proof. Suppose that P is measurable, so $\mu(P)$ is defined. Let $R_{n\alpha}(P)$ denote the rotation of P by $R_{n\alpha}$, so by the assumptions on μ we have $\mu(R_{n\alpha}(P)) = \mu(P)$. By the above exercise,

$$\mu(S) = \sum_{n=-\infty}^{\infty} R_{n\alpha}(P) = \sum_{n=-\infty}^{\infty} \mu(P).$$

If $\mu(P) = 0$, then $\mu(S) = 0$, which is a contradiction. Otherwise $\mu(P) > 0$, but then $\mu(S) = \infty$, which is also a contradiction. $\square$

But does P exist? Consider the following reasonable-sounding statement:

Axiom 4.18 (axiom of choice)**.** *Let $\mathcal{F}$ be a family of sets. There is a set Q such that, for each $F \in \mathcal{F}$, there is exactly one $f \in F$ such that $f \in Q$.*

The set Q was constructed by "choosing" an f from each F, and putting it in Q. This statement cannot be proven from the properties of sets we discussed early on, and is a bit contentious. But most mathematicians are content to assume that the axiom of choice is true, and we've even stealthily been using it all along!

Chapter 5

Additional practice

Here are some more fun exercises that we might do in class or assign as homework, which use multiple techniques from the above sections. They're a bit harder than what we've done so far, though.

5.1 Transfinite induction

Let's generalize the principle of induction. Before doing so, though, you might want to review the proof of Theorem 1.57, which said that $\mathbf{N}$ admitted induction precisely because $\mathbf{N}$ was well-ordered.

Throughout this section, we take $0 \in \mathbf{N}$.

Definition 5.1. A *chain* is a set X equipped with a binary relation $<$, such that for every $x_1, x_2, x_3 \in X$:

1. If $x_1 \neq x_2$, either $x_1 < x_2$ or $x_2 < x_1$.
2. If $x_1 < x_2$ and $x_2 < x_3$, then $x_1 < x_3$.
3. It is not true that $x_1 < x_1$.

Homework 5.2. Show that $\mathbf{Q}$ is a chain with $<$, but if X has at least 2 elements, then $\mathcal{P}(X)$ is not a chain with $\subset$. Come up with another example of a chain, and another non-example.

Definition 5.3. An *ordinal* is a chain α such that for every subset A of α, A has a least element.

Homework 5.4. Show that neither $\mathbf{Q}$ nor $\mathcal{P}(X)$ is an ordinal. But show that $\mathbf{N}$ is an ordinal, and the set $\mathbf{N} \cup \{\omega\}$, where ω is an element which is greater than every $n \in \mathbf{N}$, is an ordinal.

Homework 5.5. Show that if α is an ordinal and $\beta \in \alpha$, the set of all $\gamma < \beta$ is an ordinal as well.

So if $\beta \in \alpha$, we also write β to mean the set of all $\gamma < \beta$. For example, ω is an ordinal (and in fact $\omega = \mathbf{N}$). Similarly, 6 is an ordinal, consisting of $0, 1, \ldots, 5$.

Definition 5.6. Let α be an ordinal and $\beta \in \alpha$.

1. If there is a $\gamma < \beta$ such that β is the smallest element of α which is greater than γ, we call β the *successor* of γ and write $\beta = \gamma + 1$.

2. If β is the least element of α, we say that $\beta = 0$.

3. Otherwise, we say that β is a *limit*.

Homework 5.7. Let $2\omega = \mathbf{N} \cup \{\omega, \omega + 1, \omega + 2, \dots\}$. What are the limits of 2ω?

Homework 5.8 (principle of transfinite induction). Let α be an ordinal, and let $P(\beta)$ be a property that elements $\beta \in \alpha$ can have. Suppose that:

1. (Zero stage) $P(0)$ is true.

2. (Successor stages) If $P(\beta)$ is true, then $P(\beta + 1)$ is true.

3. (Limit stages) If β is a limit, and for every $\gamma < \beta$, $P(\gamma)$ is true, then $P(\beta)$ is true.

Show that for every $\beta \in \alpha$, $P(\beta)$ is true.

Definition 5.9. Fix an ordinal κ. Define the *beth numbers*

1. $\beth_0$ is the cardinality of $X_0 = \mathbf{N}$.

2. If $\beth_\alpha$ is the cardinality of a set X_α, let $\beth_{\alpha+1}$ denote the cardinality of the set $X_{\alpha+1} = \mathcal{P}(X_\alpha)$.

3. If α is a limit and for every $\beta < \alpha$, $\beth_\beta$ is the cardinality of a set X_β, let $X_\alpha = \bigcup_{\beta<\alpha} X_\beta$ and let $\beth_\alpha$ be the cardinality of X_α.

Write $\beth_\beta < \beth_\alpha$ iff X_β has a smaller cardinality than β_α.

Homework 5.10. Show that if $\alpha, \beta \in \kappa$, $\beta < \alpha$ iff $\beth_\beta < \beth_\alpha$. (Hint: Transfinite induction. You'll have to use Cantor's diagonal argument for successor stages, but you'll want to take unions at limit stages.)

5.2 Cardinality

Homework 5.11. Let $\mathcal{F}$ be a countable family of countable sets. Show that the union of all the sets in $\mathcal{F}$ is countable.

Definition 5.12. Let $P(\mathbf{Z})$ denote the set of polynomials with integer coefficients, i.e. functions $f : \mathbf{C} \to \mathbf{C}$ given by

$$f(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n,$$

for some $n \in \mathbf{N}$ (which we call the *order* of f) and some $a_0, \dots, a_n \in \mathbf{Z}$. We let $P_n(\mathbf{Z})$ be the set of polynomials of order n.

Homework 5.13. Show that there is a bijection $P_n(\mathbf{Z}) \to \mathbf{Z}^{n+1}$.

Definition 5.14. Let $\mathbf{A}$ denote the set of *algebraic numbers*, i.e. numbers $z \in \mathbf{C}$ such that there is a $f \in P(\mathbf{Z})$ such that $f(z) = 0$.

Homework 5.15. Show that $\mathbf{A}$ is countable.

Homework 5.16 (Schröder-Bernstein theorem). Let $f : X \to Y$ and $g : Y \to X$ be injective functions. Prove that X and Y have the same cardinality. If you are brave, you can try to prove it on your own, but you might also try filling in the following proof by Kőnig.

Proof sketch. There are sets $\tilde{X}$ and $\tilde{Y}$ and bijections $X \to \tilde{X}$ and $Y \to \tilde{Y}$, such that $\tilde{X} \cap \tilde{Y} = \emptyset$. Then there are injections $\tilde{f} : \tilde{X} \to \tilde{Y}$ and $\tilde{g} : \tilde{Y} \to \tilde{X}$. Replacing X with $\tilde{X}$ and Y with $\tilde{Y}$, we can assume that $X \cap Y = \emptyset$.

Now consider the sequence, possibly infinite in both directions, given by

$$\cdots \mapsto f^{-1}(g^{-1}(x)) \mapsto g^{-1}(x) \mapsto x \mapsto f(x) \mapsto g(f(x)) \mapsto f(g(f(x))) \mapsto \cdots .$$

If we go back far enough, the preimages might stop being well-defined (why?) so the sequence might stop to the left on an element of X or an element of Y.

There is a family $\mathcal{F}$ of sets such that if $A, B \in \mathcal{F}$, $A \cap B = \emptyset$, and $\bigcup \mathcal{F} = X \cup Y$. Because of properties of the infinite sequence above, for each $A \in \mathcal{F}$ there is a bijection $A \cap X \to A \cap Y$. So there is a bijection $X \to Y$. □

Homework 5.17. Let X be any set. Show that X is infinite iff X has the same cardinality as $X \times X$.

You should not attempt the remaining exercises until you have completed the section on transfinite induction.

Homework 5.18. Let κ be any ordinal, and $\alpha \in \kappa$. Show that there is a set of cardinality $\beth_\alpha$. (Proof: transfinite induction.)

You might wonder if every infinite set has cardinality $\beth_\alpha$ for some ordinal α. This problem is known as the (generalized) *continuum hypothesis*, and it turns out that it is impossible to prove this either true or false from the assumptions most mathematicians make.

Definition 5.19. Let $k \in \mathbf{N}$. The *infinite k-ary tree* is the tree T_k such that every vertex of T_k has exactly k children.

Definition 5.20. Let $C^0 = [0, 1]$. Assume that C^n is already defined, and define C^{n+1} by removing the middle thirds of each segment in C^n; so $C^1 = [0, 1/3] \cup [2/3, 1]$, $C^2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$, and so on. The *Cantor set* is the intersection $C = \bigcap_n C^n$.

Homework 5.21. Show that the following sets all have cardinality $\beth_1$.

1. The real numbers $\mathbf{R}$.
2. The complex numbers $\mathbf{C}$.
3. The set of all paths through the infinite k-ary tree T_k, for every $k \geq 2$.
4. The Cantor set C.

5. The power set $\mathcal{P}(X)$, for any countable, infinite set X.

(Hint: First show that T_2, C, and $\mathcal{P}(\mathbf{N})$ all have the same cardinality. By definition, the cardinality of $\mathcal{P}(\mathbf{N})$ is $\beth_1$. There's an easy bijection $\mathbf{R} \times \mathbf{R} \to \mathbf{C}$. Show that the choices of k and X in the definitions of T_k and $\mathcal{P}(X)$ don't affect cardinality. Then use the Schröder-Bernstein theorem on injections $\mathbf{R} \to \mathcal{P}(\mathbf{N})$ and $C \to \mathbf{R}$.)

5.3 Combinatorics

Combinatorics is the study of counting quantities, especially along graphs and trees.

Recall that

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Homework 5.22. Show that for every $n, k \in \mathbf{N}$, $\binom{n}{k}$ is the number of subsets of cardinality k of a set of n elements.

Homework 5.23 (Pascal's formula). Show that for every $n, k \in \mathbf{N}$,

$$\binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k}.$$

Homework 5.24 (binomial theorem). Show that for any $x, y \in \mathbf{R}$ and $n \in \mathbf{N}$,

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Homework 5.25. Suppose that six people are at a party. If they are meeting for the first time, we'll call them *strangers*; if they've met them before, we'll call them *friends.* Show that for each individual at the party, either he must have three friends or three strangers.

Homework 5.26 (Ramsey's friendship theorem). Suppose that six people are at a party. Show that at least three of the partygoers are strangers to each other, or otherwise at least three of the partygoers are mutual friends.

Homework 5.27 (Hall's marriage theorem). Suppose that the king wants to marry off each of his sons to a princess, and each of the princes only wants to marry a princess he has already met. Prove that the following are equivalent:

1. The king can marry each of his sons to an acquaintance.
2. For any set P of princesses, the number of princes at least one of the princesses in P has met is as least as big as the number of princesses, $|P|$.

If you get stuck, you might try rephrasing this into a problem about injective functions. One of the directions is easy, and for the other, you can use induction.

5.4 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. This section should not be attempted until you have completed the chapter on advanced calculus.

Definition 5.28 (metric space)**.** A *metric space* is an ordered pair (X, d), where X is a set and $d : X \times X \to \mathbf{R}$ is a *metric*, which is a function that satisfies the following properties:

1. $d(x, y) \geq 0$
2. $d(x, y) = 0 \iff x = y$
3. $d(x, y) = d(y, x)$
4. $d(x, z) \leq d(x, y) + d(y, z)$

Property (1) asserts that distances should be nonnegative, while property (2) asserts that the distance from a point to itself should be zero. Property (3) is called symmetry, while property (4) is called the Triangle Inequality.

Discussion topic 5.29. Show that property (1) is unnecessary: properties (2), (3), and (4) of a metric space already imply property (1).

Discussion topic 5.30. Show that $\mathbf{R}^n = \mathbf{R} \times \cdots \times \mathbf{R}$ (n copies) is a metric space with

$$d(x, y) = \sqrt{|y_1 - x_1|^2 + \ldots |y_n - x_n|^2}.$$

Discussion topic 5.31. Show that $(0, \infty)$ is a metric space with $d(x, y) = |\log(y/x)|$.

Discussion topic 5.32. Show that for any set X, X is a metric space with the metric $d(x, y) = 1$ if $y \neq x$ and $d(x, x) = 0$.

Definition 5.33. Let (X, d) be a metric space, $x \in X$ and $\varepsilon > 0$. The *open ball* around x of radius ε is

$$B(x, \varepsilon) = \{y \in X : d(x, y) < \varepsilon\}.$$

The *closed ball* is

$$\overline{B}(x, = \{y \in X : d(x, y) \leq \varepsilon\}.$$

Definition 5.34. Let (X, d) be a metric space and $A \subseteq X$. Say that $x \in A$ is a *limit point* of A if for every $\varepsilon > 0$, the open ball $B(x, \varepsilon)$ contains a point $y \in A$ such that $x \neq y$.

Discussion topic 5.35. Let $X = \{1/n : n \in \mathbf{N}\}$. Show that 0 is a limit point of X.

Discussion topic 5.36. If x is a limit point of $A \subseteq X$, then show that every open ball around x contains infinitely many points of A.

Discussion topic 5.37. Show that a finite set has no limit points.

Definition 5.38. Let $A \subseteq X$. Say that $x \in A$ is a *interior point* of A if there is a $\delta > 0$ such that the open ball $B(x, \delta) \subseteq A$.

Definition 5.39. Let $A \subseteq X$. Say that A is *open* if every point of A is an interior point of A. Say that A is *closed* if every limit point of A is a point of A.

Discussion topic 5.40. Show that every open ball is an open set.

Discussion topic 5.41. Show that $E \subseteq X$ is open iff $X \setminus E$ is closed. (Hint: contradiction in both directions.)

5.5 p-adic analysis

These exercises should not be attempted by a student who has not learned about metric spaces and fields yet. Let p be a prime number.

Homework 5.42. For any nonzero $x \in \mathbf{Q}$, show that there is a unique $n \in \mathbf{Z}$ such that $x = p^n(a/b)$ where neither a nor b are divisible by p.

Definition 5.43. The *p-adic norm* is defined as follows: $|0|_p = 0$, and if $x \neq 0$ is rational, then $|x|_p = p^{-n}$, where n is given by Homework 5.42.

Homework 5.44. For $x \in \mathbf{Q}$, show that there are primes $p_1, \dots, p_n$ and integers $a_1, \dots, a_n$ such that $x = p_1^{a_1} \dots p_n^{a_n}$.

Definition 5.45. An *ultrametric* on a set X is a metric d such that for every $a, b, c \in X$,

$$d(x, z) \leq \max(d(x, y), d(y, z)),$$

the *isoceles triangle inequality*.

Homework 5.46. For $x, y \in \mathbf{Q}$, let $d_p(x, y) = |x - y|_p$. Show that d_p is an ultrametric.

Definition 5.47. A *Cauchy sequence* is a sequence of rational numbers $x_1, \dots$ such that for every $\varepsilon > 0$ there is a $N > 0$ such that if $n_1, n_2 > N$ then $d(x_{n_1}, x_{n_2}) < \varepsilon$. Two Cauchy sequences $x_1, \dots$ and $y_1, \dots$ are *Cauchy equivalent* if for every $\varepsilon > 0$ there is a $N > 0$ such that if $n > N$ then $d(x_n, y_n) < \varepsilon$.

Homework 5.48. Show that a constant sequence is Cauchy. Find an example of a Cauchy sequence which does not converge to any element of $\mathbf{Q}$.

Homework 5.49. Show that Cauchy equivalence is a equivalence relation, and that there is a set $\mathbf{Q}_p$ containing exactly one Cauchy sequence from each Cauchy equivalence class.

Definition 5.50. An element of $\mathbf{Q}_p$ is called a *p-adic number*.

Homework 5.51. Think of a way to define field operations on $\mathbf{Q}_p$ and a morphism of fields $\mathbf{Q} \to \mathbf{Q}_p$. (Do not actually prove that $\mathbf{Q}_p$ is a field.)

5.6 Computability

Here are some exercises that would be best for someone with programming experience.

Definition 5.52. Say that $x \in \mathbf{R}$ is a *computable number* if there is a computer program which takes in a number $n \in \mathbf{N}$ and returns the nth digit of x. Otherwise, say that x is *uncomputable*.

Homework 5.53. Show that there is an uncomputable number.

Homework 5.54. Prove that $\mathbf{Q}$ is countable using computability. Which proof of the countability of $\mathbf{Q}$ do you prefer?

Definition 5.55. Let $f : X \to Y$ be a function. f is said to be a *computable function* if there is a computer program which takes elements $x \in X$ as input and returns $f(x)$ as output.

To show that a function is computable, all you have to do is write a computer program which computes it. But to show that a function is not computable, a useful strategy is contradiction. It's especially common to show that if a function were computable, then it would be possible to solve the halting problem, which we already showed is false.

Homework 5.56. For each $n \in \mathbf{N}$, let $P(n)$ be the prime factorization of n. Show that P is computable.

Homework 5.57. Recall that for any pair $(n, p) \in \mathbf{N} \times \mathbf{N}$, there is a pair $(k, \ell) \in \mathbf{N} \times \mathbf{N}$ such that $n^k - n^\ell$ is divisible by p. Let $F(n, p)$ be the function which returns the least such k and ℓ. Is F computable?

Definition 5.58. Given a string F, define the *Kolmogorov complexity* of F, $K(F)$, to be the file size of the code of the shortest program (in a given programming language) which outputs F.

Homework 5.59. Show that for every $N \in \mathbf{N}$, there is a string $S(N)$ such that $K(S(N)) \geq N$. Moreover, show that if K is computable, then S is computable.

Homework 5.60. Show that it cannot be the case that both K and S are computable; conclude that K is not computable.

5.7 Fermat's last theorem

This section should not be attempted until you've learned about fields. It's probably a lot harder than the other sections!

Fermat's little theorem should not be confused with the following theorem.

Theorem 5.61 (Fermat's last theorem)**.** *Let $a, b, c, n \in \mathbf{N}$. If $a, b, c \geq 1$ and $n \geq 3$, then $a^n + b^n \neq c^n$.*

Even though its statement is so simple, Fermat's last theorem is considered one of the most difficult theorems to prove in mathematics, and was finally proven by Sir Andrew Wiles, building on the work of Ken Ribet and many others.

Let's prove a *very special case* of Fermat's last theorem: when $n = 4$. In this case, rewrite the equation as $c^4 - b^4 = (a^2)^2$. So we just need to prove that the equation $x^4 - y^4 = z^2$ has no solutions if $x, y, z \geq 1$. Actually, it will be more convenient to rewrite this equation as

$$(x^2 + y^2)(x^2 - y^2) = z^2.$$

Homework 5.62. Show that without loss of generality, we can assume that $\gcd(x, y) = 1$. Under this assumption, show that either $\gcd(x^2 + y^2, x^2 - y^2) = 1$, or x and y are odd while z is even.

Homework 5.63. If x and y are odd while z is even, show that there are numbers $0 < e < d < x$ such that $x^2y^2 = d^4 - e^4$ and that $\gcd(d, e) = 1$. Why does this cause a contradiction?

Homework 5.64. If $\gcd(x^2+y^2, x^2-y^2) = 1$, show that there are odd numbers s and t such that $s^2 = x^2+y^2$ and $t^2 = x^2-y^2$ and such that $\gcd(s,t) = 1$.

Homework 5.65. Let $u = (s+t)/2$ and $v = (s-t)/2$. Show that u, v are integers and $\gcd(u,v) = 1$. Use this to show that there are integers g, h such that $g^2 = u$ and $h^2 = v$.

Homework 5.66. Find an integer k such that $g^4 - h^4 = k^2$ or else $h^4 - g^4 = k^2$; and such that $0 < g < d < x$, causing a contradiction.

5.8 Fundamental theorem of algebra

This section shouldn't be attempted until completing the chapters on fields and on advanced calculus.

You can use the following theorem without proof. It will be proven in Math 185.

Theorem 5.67 (Liouville)**.** *Let $f : \mathbf{C} \to \mathbf{C}$ be a holomorphic function (c.f. Definition 2.14). If there is a $C > 0$ such that for every $z \in \mathbf{C}$, $|f(z)| \leq C$, then f is a constant function.*

Homework 5.68. Show that if $p : \mathbf{C} \to \mathbf{C}$ is any polynomial, then there is a $z_0 \in \mathbf{C}$ such that for every $z \in \mathbf{C}$, $|p(z_0)| \leq |p(z)|$.

Homework 5.69 (fundamental theorem of algebra). Show that $\mathbf{C}$ is algebraically closed.

Homework 5.70. Show that if $p : \mathbf{C} \to \mathbf{C}$ is a polynomial, then there are linear polynomials $p_1, \dots, p_n$ such that $p = p_1 \dots p_n$. (Hint: induction.)

Homework 5.71. Show that $\mathbf{C}$ is a simple extension of $\mathbf{R}$.

Definition 5.72. Let k be a field, and let F be a simple extension of k. If, for every simple extension E of k of finite degree, $E \subseteq F$, we say that F is an *algebraic closure* of k.

Homework 5.73. Show that if k has an algebraic closure, then it is unique up to isomorphism. (So we are justified in talking about "the" algebraic closure, $\overline{k}$, of k.)

Homework 5.74. Show that $\mathbf{C}$ is the algebraic closure of $\mathbf{R}$.

Homework 5.75. If you've taken Math 54, show that if $A \in \mathbf{R}^{n \times n}$ is a matrix, then there is a $\lambda \in \mathbf{C}$ and a nonzero vector $v \in \mathbf{C}^n$ such that $Av = \lambda v$, i.e. an eigenvector. (Hint: look at the characteristic polynomial $\chi_A(z) = \det(A - zI)$.)

5.9 Binary structures

This shouldn't be attempted until completing the chapter on fields.

Definition 5.76. Let $(X, \sharp)$ and $(Y, \flat)$ be binary structures. An *isomorphism of binary structures* is a bijection $\varphi : X \to Y$ such that

$$\varphi(x_1) \flat \varphi(x_2) = \varphi(x_1 \sharp x_2)$$

for every $x_1, x_2 \in X$. If an isomorphism exists, then we say that $(X, \sharp)$ and $(Y, \flat)$ are *isomorphic*.

Homework 5.77. Show that $(\mathbf{R}, +)$ and $((0, \infty), \cdot)$ are isomorphic.

Homework 5.78. Define a binary structure $(\mathbf{N}, *)$ by $a * b = a^{2b}$. Is $*$ commutative or associative?

Definition 5.79. Let $(X, *)$ be a binary structure. An *idempotent* of $*$ is an $x \in X$ such that $x * x = x$.

Homework 5.80. Let $(X, *)$ be a binary structure. Let Y be the set of all idempotents of $*$ in X. Is $(Y, *)$ a substructure?

Definition 5.81. The *cross product* of two vectors $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$ in $\mathbf{R}^3$ is given by

$$x \times y = (x_2y_3 - x_3y_2, x_1y_3 - x_3y_1, x_1y_2 - x_2y_1).$$

Homework 5.82. Show that × is neither commutative nor associative, but does satisfy *anti-commutativity*, namely $x \times y + y \times x = 0$, and satisfies *Jacobi's identity*,

$$x \times (y \times z) + y \times (x \times z) + z \times (x \times y) = 0.$$

Definition 5.83. Define the *special orthogonal algebra* $\mathfrak{so}(3)$ to be the binary structure of all 3×3 matrices of the form

$$\begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix}$$

under the operation $[A, B] = AB - BA$.

Homework 5.84. Show that the function

$$\mathfrak{so}(3) \to \mathbf{R}^3$$
$$\begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix} \mapsto (a, b, c)$$

is an isomorphism $(\mathfrak{so}(3), [\cdot, \cdot]) \to (\mathbf{R}^3, \times)$.

Homework 5.85. Define the *Heisenberg group*, the binary structure $(H, \cdot)$ of all matrices of the form

$$\begin{bmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix}$$

under multiplication. Show that $(H, \cdot)$ is isomorphic to the binary structure $(\mathbf{R}^3, *)$, where $\mathbf{R}^3 = \mathbf{R} \times \mathbf{R} \times \mathbf{R}$ and $(a, b, c) * (x, y, z) = (a + x, b + y, c + z + ay)$. Is $(H, \cdot)$ isomorphic to $(\mathfrak{so}(3), [\cdot, \cdot])$?

Definition 5.86. Let $(X, \sharp)$ be a binary structure. Say that $(X, \sharp)$ is *cyclic* if there is an element $x_0 \in X$ such that for every $x \in X$, we can write $x = x_0 \sharp x_0 \sharp \dots \sharp x_0$ (for some number of copies of x_0).

Homework 5.87. Show that cyclicity is an invariant: if $(X, \sharp)$ is cyclic and $(X, \sharp)$ is isomorphic to $(Y, \flat)$, then $(Y, \flat)$ is cyclic.

Homework 5.88. Show that $(\mathbf{Q}, +)$ is not isomorphic to $(\mathbf{Z}, +)$.

Chapter 6

Old stuff

This material was previously taught in MUSA 74 and is kept as a curiosity.

6.1 Introduction to groups

One common example of a group you will see throughout Math 113 is the set of integers modulo some natural n. We will use the notation $\mathbf{Z}_n$ to refer to the set of integers modulo n. For example, $\mathbf{Z}_5 = \{0, 1, 2, 3, 4\}$. You may see this written as $\mathbf{Z}/n\mathbf{Z}$ in other places, and you'll see why in Example 6.83.

Consider the following tables.

$\mathbf{Z}_5$

$+$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\mathbf{Z}_5^*$

$\times$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

$\mathbf{Z}_4$

$+$	0	1	3	2
0	0	1	3	2
1	1	2	0	3
3	3	0	2	1
2	2	3	1	0

The first table displays the addition table of numbers modulo 5, the second table displays the multiplication table of numbers modulo 5, and the third table displays the addition table of numbers modulo 4. Do you see any patterns, similarities, or differences?

One pattern is that each number appears exactly once in each row and column. Why is that? Well, it turns out that those numbers along with the associated operation have a special structure on it, namely that of a group.

Definition 6.1 (Group)**.** A group $(G, *)$ is a set G with an operation $* : G \times G \to G$ such that the following axioms are satisfied:

1. $\forall a, b, c \in G$, we have $(a * b) * c = a * (b * c)$ (Associativity)
2. $\exists e_G \in G$ such that $\forall x \in G$, we have $e_G * x = x * e_g = x$ (Identity)
3. $\forall a \in G$, $\exists b \in G$ such that $a * b = b * a = e_G$ (Inverse)

Note that if $a \in G$, $n \in \mathbf{Z}$, we write a^{-1} to mean the inverse of a, and a^n to mean a multiplied by itself n times (where multiplication by a^{-1} counts as multiplication by -1 times).

The order of these axioms presented is important because you need associativity to talk about identity and identity to talk about inverse.

Let us see a few examples of this in practice.

Example 6.2. Let us show that $(\mathbf{Z}_n, +)$ is a group, where the operation is usual addition of integers.

1. Associativity: Associative because + is associative and closed under mod.

2. Identity: 0 is the identity.

3. Inverse: $k + (n - k) = 0 \implies (n - k)$ is the inverse of k.

Usually associativity is immediate, and there is only one sensible choice for the identity, so the inversion condition is the interesting one.

Example 6.3. Let us show that $(\mathbf{Z}_n^*, \times)$ is a group, where $\mathbf{Z}_n^*$ is the set of elements of $\mathbf{Z}_n$ that are relatively prime to n, and the operation is usual multiplication.

1. Associativity: Associative because $\times$ is associative and closed under mod.

2. Identity: 1 is the identity.

3. Inverse: Use Euclid's Division Algorithm to prove.

Example 6.4. Let us show that $(\mathbf{Q} \setminus 0,\ a * b = ab/2)$ is a group.

1. Associativity: Assume this is true.

2. Identity: We are trying to find an element of $\mathbf{Q}$ such that $a * b = a$. This happens when $b = 2$. Thus, 2 is the identity.

3. Inverse: To find the inverse, fix a. We are trying to find an element of $\mathbf{Q}$ such that $a * b = 2$. We know $a * b = ab/2$. This means we are trying to a find a value for b such that $ab/2 = 2$. This occurs when $b = 4/a$. Thus, $b = 4/a$ is an inverse.

Example 6.5. Let us show that $(M(n, \mathbf{R}), +)$, the set of all $n \times n$ matrices, is a group.

1. Associativity: Associative because + for $n \times n$ matrices is associative.

2. Identity: $A + 0 = A \rightarrow 0$ is the identity.

3. Inverse: $A + -A = 0 \rightarrow -A$ is the inverse.

At the beginning of the chapter we said that a group was a collection of symmetries of some object. The following example will make this more clear.

Example 6.6. Let us show that (GL($n, \mathbf{R}$), $*$), the set of all $n \times n$ invertible matrices, is a group.

1. Associativity: Associative because $*$ for $n \times n$ invertible matrices is associative.
2. Identity: $A * I = A \to I$ is the identity.
3. Inverse: $A * A^{-1} = I \to A^{-1}$ is the inverse.

Furthermore, GL($n, \mathbf{R}$) can be thought of the group of all symmetries of $\mathbf{R}^n$: they are the ways one can transform $\mathbf{R}^n$, while respecting its vector space structure. The group operation is composition of transformations. All the properties of a group are necessary for this to make sense:

1. Associativity: Transformations are functions, and transformations of functions are associative.
2. Identity: "Doing nothing" should be a symmetry.
3. Inverse: It should be possible to undo a symmetry.

Whenever you see a group, you should think about what transformations it could possibly represent.

Example 6.7. As you can check, $(\mathbf{R}, +)$ is a group. It is the collection of translations of $\mathbf{R}^2$ in the horizontal direction. That is, we can think of a point $x \in \mathbf{R}$ as an instruction to translate the vector $(v_1, v_2) \in \mathbf{R}^2$ to $(v_1 + x, v_2)$.

Proposition 6.8. *Suppose $(G, *)$ is a group. Then the identity of G is unique. Furthermore, $\forall g \in G$, g^{-1} is unique.*

Proof: Identity is unique. Let there be an identity e_g for all $x \in G$ where $e_G {*} x = x$. Suppose there exists another identity e_g' for all $x \in G$, where $e_g' * x = x$.

$$\begin{aligned} e_g * x &= e_g' * x \\ (e_g * x) * x^{-1} &= (e_g' * x) * x^{-1} \\ e_g = e_g * (x * x^{-1}) &= e_g' * (x * x^{-1}) = e_g' \end{aligned}$$

Notice how we used associativity in the last step, so associativity is actually useful! □

Proof: Inverse is unique. Fix a. Suppose b, b' are inverses of a. Then $a {*} b = e$ and $a {*} b' = e$, where e is the identity element of G.

$$\begin{aligned} a * b &= a * b' \\ (a^{-1} * a) * b &= (a^{-1} * a) * b^{-1} \\ b &= b' \end{aligned}$$

□

Example 6.9. Let us show that $(a * b)^{-1} = b^{-1} * a^{-1}$.

$$\begin{aligned} e &= e \\ (a * b) * (a * b)^{-1} &= e \\ (a^{-1} * a) * b * (a * b)^{-1} &= a^{-1} * e \\ b * (a * b)^{-1} &= a^{-1} \\ (b^{-1} * b) * (a * b)^{-1} &= b^{-1} * a^{-1} \\ (a * b)^{-1} &= b^{-1} * a^{-1} \end{aligned}$$

Example 6.10. Prove that $(\mathbf{Z}, -)$ is not a group.

We can show this through associativity. $a - (b - c) = (a - b) + c \neq (a - b) - c$.

Now look back at the tables, which we can now call *group tables*. What does it mean mathematically that an element g appears once in a row? It means that fixing $a \in G$, we can find $b \in G$ such that $a * b = g$. But because we have inverses of elements in a group, we see that such an element b necessarily exists, namely $b = a^{-1} * g$. Furthermore, suppose we had another element b' such that $a * b' = g$. But then

$$\begin{aligned} a * b &= a * b' \\ a^{-1} * a * b &= a^{-1} * a * b' \\ b = e * b &= e * b' = b' \end{aligned}$$

So we see that an element g appears exactly once in a row of the group table, and the proof for columns proceeds similarly. Thus, we see that in a group table, every element shows up exactly once in any given row or column.

Now suppose we have a group structure on a set of 3 elements. What can the group table possibly look like? Well, we know that one of those elements is the identity element e and the effect of multiplication by e so it must look something like,

*	e	a	b
e	e	a	b
a	a		
b	b		

Using what we have learned, complete the table. There should only be one possibility, and in fact this means there is *only one group of order 3, up to isomorphism*[1]. In general we can play the same game with more elements, and it turns into a sudoku of sorts.

Homework 6.11. Show that $(\mathbf{R} \setminus \{-1\},\ a * b = a + b + ab)$ is a group.

Homework 6.12. Show that $(\mathbf{R}^{\geq 0},\ a * b = \sqrt{ab})$ is not a group.

Homework 6.13. Show that $(\mathbf{Z} \setminus \{0\}, \times)$ is not a group.

Homework 6.14. Show that if V is a vector space, then $(V, +)$ is a group.

We write $x^n = x * x * \cdots * x$, where there are n copies of x.

[1] We will learn what this means later on.

Homework 6.15. Let $(G, *)$ be a group. We say that G is *abelian* if for every $x, y \in G$, $x * y = y * x$.

Show that if G is abelian, then $(x * y)^n = x^n * y^n$. (Hint: if $n = 1$ then this is easy.)

Homework 6.16. Suppose that $(G, *)$ is a finite group (this means that $(G, *)$ is a group, and G is finite when viewed as a set) with identity e. Show that for each $x \in G$, there exists $n \in \mathbf{N}$ such that $x^n = e$.

6.2 Group homomorphisms

In linear algebra, we learned that the "correct" thing to study wasn't vector spaces: it was a special family of functions between vector spaces, that we called linear transformations or matrices.

In group theory, the same principle will hold. The analogue of a linear transformation is a homomorphism.

Definition 6.17 (Group homomorphism)**.** Let $(G, *)$, $(H, \circ)$ be groups. A map $\varphi : G \to H$ such that

$$\varphi(x * y) = \varphi(x) \circ \varphi(y)$$

is called a *group homomorphism.*

Let us see a few examples of this in practice.

Example 6.18. $\pi : \mathbf{Z} \to \mathbf{Z}_5$, both under addition, where $\pi(a)$ is defined to be $a \pmod 5$.

Proof.

$$(a + b) \pmod 5 = \pi(a + b) = \pi(a) + \pi(b) = (a \pmod 5) + (b \pmod 5).$$

□

Example 6.19. $L_c : (\mathbf{R}, +) \to (\mathbf{R}, +)$, $L_c(x) = cx$, $c \in \mathbf{R}$

Proof.

$$L_c(a + b) = c(a + b) = ca + cb = L_c(a) + L_c(b).$$

□

Example 6.20. Let D be the group of smooth (infinitely differentiable) functions $\mathbf{R} \to \mathbf{R}$, under addition. Then the map $D \to D$ given by $f \mapsto f'$ is a group homomorphism.

Proof.

$$\frac{d}{dx}(a + b) = (a + b)' = a' + b' = \frac{d}{dx}(a) + \frac{d}{dx}(b).$$

□

Example 6.21. $\Phi : (\mathbf{R}, +) \to (\mathbf{R}^+, \times)$, $\Phi(x) = e^x$.

Proof.

$$\Phi(x,y) = e^{x+y} = e^x \times e^y = \Phi(x) \times \Phi(y).$$

□

Just like with linear transformations, we want to define a special kind of homomorphism φ such that if $\varphi : G \to H$, then G and H are "the same group."

Definition 6.22 (isomorphism)**.** A map $\varphi : G \to H$ is called an *isomorphism* if

1. φ is a group homomorphism, and
2. φ is a bijection (one-to-one and onto):
   (a) One-to-one (injective): $\forall g\,, g' \in G$, $\varphi(g) = \varphi(g') \implies g = g'$
   (b) Onto (surjective): $\forall h \in H$, $\exists g \in G$ such that $f(g) = h$.

If these conditions hold, we write $G \cong H$ (i.e. G is *isomorphic* to H).

Now look back at the group tables from Lecture 2. The homomorphism property is essentially the same as "If we replace x with $\varphi(x)$ the group tables are the same." We can see from the tables that we have the isomorphism $(\mathbf{Z}_5)^* \cong \mathbf{Z}_4$. In other words, multiplication modulo 5 behaves the same as addition modulo 4!

Example 6.23. The map Φ defined in Example **??** is an isomorphism.

Proof. We know Φ is a group homomorphism.

To show Φ is one-to-one, assume $\Phi(x) = \Phi(y)$. Then $\Phi(x) = e^x$, and $\Phi(y) = e^y$. Since the two are equal, $e^x = e^y$, which means $x = \log(e^x) = \log(e^y) = y$[2]. This proves that $x = y$.

To show Φ is onto, let y be some number in $\mathbf{R}^+$. Since $y \in \mathbf{R}^+$, there exists some $x = \log(y)$. By the definition of log, that means $\log(e^x) = \log(y)$ which means that $e^x = y$. Since $\Phi(x) = e^x$, $\Phi(x) = y$. □

Lemma 6.24. *Let $f : G \to H$ be a group homomorphism. Then:*

1. $f(e_G) = e_H$
2. $\forall x \in G,\ f(x^{-1}) = f(x)^{-1}$
3. $\forall n \in \mathbf{Z},\ x \in G,\ f(x^n) = f(x)^n$

Proof.
1. Let e_G be the identity element of G. By definition of group homomorphism, $f(e_G * e_G) = f(e_G)f(e_G)$. We also know that $f(e_G * e_G) = f(e_G)$. Thus $f(e_G)f(e_G) = f(e_G * e_G) = f(e_G)$. Multiplying by $f(e_G)^{-1}$ on each side we have $f(e_G) = e_H$.
2. We know $f(e_G) = e_H$ (see above). Since $e_G = x * x^{-1}\,\forall\, x \in G$, $f(e_G) = f(x * x^{-1}) = f(x)f(x^{-1}) = f(x)f(x)^{-1} = e_H$.

[2] in Math 104, log will mean logarithm base e

3. We will prove this by induction. The base case, $n = 1$ is trivially true ($f(x^1) = f(x) = f(x)^1$). Let us assume the induction hypothesis, for $n = k$, $f(x^k) = f(x)^k$. We will try to prove that for $n = k + 1$, $f(x^{k+1}) = f(x)^{k+1}$. By the definition of group homorphism, $f(x^{k+1}) = f(x * x^k) = f(x)f(x^k)$. Thus, from the induction hypothesis, $f(x^{k+1}) = f(x)f(x^k) = f(x)f(x)^k = f(x)^{k+1}$.

□

Definition 6.25. A group G is *abelian* or *commutative* if $\forall a, b \in G$, $a * b = b * a$.

Unsurprisingly, this property is preserved by isomorphism.

Lemma 6.26. *Suppose G is abelian, and $G \cong H$. Then H is abelian.*

Proof. Since $G \cong H$, $\exists f : G \to H$ that is one-to-one, onto, and homomorphic.

$\forall\, z_1,\, z_2 \in\, H\; \exists x_1, x_2 \in G$ such that $f(x_1) = z_1$ and $f(x_2) = z_2$. Since G is abelian, we know that $x_1 * x_2 = x_2 * x_1$. This means $z_1z_2 = f(x_1)f(x_2) = f(x_1 * x_2) = f(x_2 * x_1) = f(x_2)f(x_1) = z_2z_1$ □

Homework 6.27. Let $\phi : G \to H$ and $\gamma : H \to K$ be group homomorphisms. Show $\gamma \circ \phi : G \to K$ is also a group homomorphism.

Homework 6.28. Let $g \in G$ be fixed. Prove $\iota_g : G \to G$ defined by $\iota_g(x) = gxg^{-1}$ is an isomorphism of G into itself. (We call ι_g "conjugation by g.)

Homework 6.29. Prove that $\Phi : G \to G$ where $\Phi(g) = g^2$ is a homomorphism if and only if G is abelian.

6.3 Subgroups

Definition 6.30. Let G be a group. A subset $H \subseteq G$ is called a *subgroup* if H is closed under products and inverses. That is, $x, y \in H \implies x^{-1} \in H$, $xy \in H$.

Example 6.31.

$$(\mathbf{Z}, +) \subseteq (\mathbf{Q}, +) \subseteq (\mathbf{R}, +) \subseteq (\mathbf{C}, +)$$

All the inclusions above are subgroups. Let us show one of them. We will show $\mathbf{Q}$ is a subgroup of $(\mathbf{R}, +)$. Take $x \in \mathbf{Q}$. The inverse of x in the group $(\mathbf{R}, +)$ is $-x$ which is also in $\mathbf{Q}$ as x is in $\mathbf{Q}$. Similarly given any two elements of $\mathbf{Q}$, their product is also in $\mathbf{Q}$.

Example 6.32. The set $\{0, 2\}$ is a subgroup of $\mathbf{Z}_4$. The inverse of 2 is just itself, and one can check that adding any combination of 0 and 2 returns back 0 and 2. The subset $\{0, 1, 2\}$ is not a subgroup! The inverse of 1 is 3 which is not in the subset.

The example above should convince you that subgroups of a group G are special subsets of the group. Notice that elements of $\{0, 2\}$ are evenly spaced from 0 to 4 while elements of $\{0, 1, 2\}$ are not. In fact, given a divisor d of n, the subset $\{0, d, 2d, \ldots (\frac{n}{d} - 1)\, d\}$ will always be a subgroup of $\mathbf{Z}_n$ by pretty much the same analysis above.

Example 6.33. G and $\{e\}$ are always subgroups of G.

Proposition 6.34 (Subgroup criterion)**.** *A subset H of a group G is a subgroup $\iff \forall x, y \in H$, $xy^{-1} \in H$.*

Proof. We first claim that the conditions of a subgroup imply that any subgroup H contains the identity e of the bigger group G. Take $x \in H$, then $x^{-1} \in H$ and thus $e = xx^{-1} \in H$ as desired. The forward direction follows from the definition of a subgroup. For the backwards direction, we know that $e \in H$ so letting $x = e$ in the condition above shows that for any $y \in H$, we have $y^{-1} \in H$. Now let $y = y^{-1}$ in the condition and we see that $xy \in H$ for any $x, y \in H$ as desired. □

Proposition 6.35. *Let $f : G \to H$ be a group homomorphism. Then* $\ker(f) = \{x \in G|\, f(x) = e_H\}$ *is a subgroup of G and* $\operatorname{Im}(f) = \{y \in H|\, \exists\, x \in G \text{ s.t. } f(x) = y\}$ *is a subgroup of H.*

Proof. Let $x \in \ker(f)$. Then $f(x^{-1}) = f(x)^{-1} = e^{-1} = e$ so $x^{-1} \in \ker(f)$. Similarly, let $x, y \in \ker(f)$. Then $f(xy) = f(x)f(y) = e \cdot e = e$ so $xy \in \ker(f)$ and thus $\ker(f)$ is a subgroup.

Suppose $y \in \operatorname{Im}(f)$. By definition there exists $x \in G$ s.t. $f(x) = y$. Then $y^{-1} = f(x)^{-1} = f(x^{-1})$ so that $x^{-1} \in \operatorname{Im}(f)$. Now suppose $y_1, y_2 \in \operatorname{Im}(f)$. Then we have x_1, x_2 s.t. $f(x_1) = y_1$ and $f(x_2) = y_2$. Then $y_1y_2 = f(x_1)f(x_2) = f(x_1x_2)$ and thus $y_1y_2 \in \operatorname{Im}(f)$. □

Lemma 6.36. *If $f : G \to H$ be a group homomorphism, then f is one-to-one if and only if* $\ker(f) = \{e\}$*.*

Proof. Suppose f is one-to-one and suppose $x \in \ker(f)$. Then $f(x) = e_H$. But we also know that $f(e_G) = e_H$. Because f is one-to-one, we see that $x = e_G$, as desired.

Suppose $\ker(f) = \{e\}$ and suppose $f(x) = f(y)$. We want to use the fact that the kernel is trivial, so we want an expression that equals e. Thus, we will multiply both sides by $f(y)^{-1}$ and we then have $f(x)f(y)^{-1} = e \implies f(xy^{-1}) = e$. This shows $xy^{-1} \in \ker(f)$ and therefore by assumption we have $xy^{-1} = e \implies x = y$. □

A homomorphism with a big kernel is "very non-injective" and a homomorphism with a big image is "very non-surjective." If it has a big kernel, lots of things in the domain get sent to the same place; if it has a big image, lots of things in the codomain aren't in the image.

Example 6.37. Let us now show $f(e_G) = e_H$ for $f : G \to H$ a group homomorphism, but in a different way than before. We want to show that $f(e_G)$ is the identity for the group H. That means that we want to show $f(e_G)y = y$, $\forall y \in H$. This seems hard, however what if $y = f(x)$? Then we have

$$f(e_G)f(x) = f(e_G * x) = f(x)$$

so we see $f(e_G)$ is the identity of $\operatorname{Im}(f)$. But $\operatorname{Im}(f)$ is a subgroup of H! Moreover, from above we know that $e_H \in \operatorname{Im}(f)$. But recall from earlier that we showed the identity of a group is unique. Applying this to the subgroup $\operatorname{Im}(f)$, we see that $f(e_G) = e_H$. Summarizing,

1. We first showed that $f(e_G)$ was the identity element of $\operatorname{Im}(f)$.

2. $\operatorname{Im}(f)$ is a subgroup and also contains e_H.

3. Identities in groups are unique.

Notice how we broke up the original problem into easier pieces. This is how you should approach a hard problem: break it up into smaller, easier ones and use these as stepping stones/tools to solve the tougher one at hand.

Homework 6.38. Show $\{z \in \mathbf{C} | \, |z| = 1\}$ is a subgroup of $(\mathbf{C}^{\times}, \cdot)$.

Homework 6.39. Show $(2\mathbf{Z}, +) \subseteq (\mathbf{Z}, +)$.

Homework 6.40. Show $(\mathrm{GL}(n, \mathbf{R}), +)$ is NOT a subgroup of $(M_n(\mathbf{R}), +)$.

Homework 6.41. Find the kernel of the map $\phi : \mathbf{C}^{\times} \to \mathbf{R}^{\times}$ where $\phi(a + bi) = a^2 + b^2$.

6.4 Generators and Relations

Definition 6.42. Given a set of elements $A = \{a_1, \dots, a_k\}$ of a group G, the subgroup generated by A denoted by $\langle A \rangle$ is defined to be

$$\langle A \, \rangle := \{a_{i_1}^{\varepsilon_1} \dots a_{i_k}^{\varepsilon_k} | \; a_{ij} \in A, \varepsilon_i = \pm 1\}$$

If $\langle A \, \rangle = G$, then we say that G is *generated* by the elements $a_1, \dots, a_k$.

Definition 6.43. A group G is *cyclic* if there exists $a \in G$ such that $\langle \{a\} \rangle = G$. In other words, every element $g \in G$ may be written as $g = a^k$ for some $k \in \mathbf{N}$.

You should think of elements of $\langle A \, \rangle$ as words in the alphabet $\{a_1, \dots, a_k, a_1^{-1}, \dots, a_k^{-1}\}$.

Example 6.44. Elements of $\langle a, b \rangle$ look like $a^n, b^m, a^n b^m, b^2 a^{-1}, ab^{-1}ab^2$, etc.

Example 6.45. $(\mathbf{Z}, +)$ is generated by 1, as any positive integer n can be written as $n = 1 + \dots + 1$ and any negative integer can be written as $-n = -1 - \dots - 1$.

Example 6.46. $\mathbf{Z}_6$ is generated by $\{2, 3\}$. To show this, we just need to show any element in $\mathbf{Z}_6$ can be written as the sum/difference of 2's and 3's.

$$0 = 2 - 2, \; 1 = 3 - 2, \; 2 = 2, \; 3 = 3, \; 4 = 2 + 2, \; 5 = 3 + 2 \equiv 2 - 3 \pmod 6$$

One can think of generators of a group as a "basis" for the group. In linear algebra, a linear transformation is completely determined by its action on a basis. In group theory, generators of a group play the same role, namely

Lemma 6.47. *Suppose that the set $A = \{a_1, \dots, a_n\}$ generates G. Then any group homomorphism $f : G \to H$ is uniquely determined by its values on $A = \{a_1, \dots, a_n\}$.*

Proof. As $\langle A \rangle = G$, we can write any element $g \in G$ as $g = a_{i_1}^{\varepsilon_1} \dots a_{i_k}^{\varepsilon_k}$. But by the homomorphism property we see that

$$f(g) = f(a_{i_1}^{\varepsilon_1} \dots a_{i_k}^{\varepsilon_k}) = f(a_{i_1})^{\varepsilon_1} \dots f(a_{i_k})^{\varepsilon_k}$$

We know the value of f on each of the a_i and it follows that we now know the value of f on g. If two homomorphisms have the same values on A then it is clear they are equal throughout. □

Example 6.48. Suppose $\phi : \mathbf{Z} \to \mathbf{Z}_6$ is a group homomorphism such that $\phi(1) = 4$. Let us compute $\phi(10)$. Notice that $10 = 1 + \cdots + 1$, so we may apply the group homomorphism property to see that

$$\phi(10) = \phi(1 + \cdots + 1) = \phi(1) + \cdots + \phi(1) = 4 + \cdots + 4 \equiv 4 \pmod 6$$

One consequence of the above lemma is that we have a "bijection" between the following two sets

$$\{f : G \to H | \ f \text{ is a group homomorphism}\} \longleftrightarrow \{\text{set of values } f(a_i) \in H\}$$

as it shows that the map sending a group homomorphism f to the set of values $\{f(a_i)\}$ is injective. Conversely, given a set of values $f(a_i)$ on the generators, we can turn this data into a group homomorphism by defining $f(g) = f(a_{j_1})^{e_1} \cdots f(a_{j_k})^{e_k}$ when $g = a_{j_1}^{e_1} \cdots a_{j_k}^{e_k}$. By construction, this f will satisfy the homomorphism property by writing out any two elements in terms of the generators. Thus our map is one to one and onto and thus a bijection. Or is it? Can $f(a_i) \in H$ be any element in H? NO! What goes wrong? Well suppose $a_1^{e_1} a_2^{e_2} = a_3^{e_3} a_4^{e_4}$[3] Then we better have

$$f(a_1)^{e_1} f(a_2)^{e_2} = f(a_1^{e_1} a_2^{e_2}) = f(a_3^{e_3} a_4^{e_4}) = f(a_3)^{e_3} f(a_4)^{e_4}$$

Definition 6.49. A *relation* in a group G is a product of elements of the group such that the product is equal to the identity.

With this definition, let us rephrase the condition we discovered above. Bring all the terms in $a_1^{e_1} a_2^{e_2} = a_3^{e_3} a_4^{e_4}$ and $f(a_1)^{e_1} f(a_2)^{e_2} = f(a_3)^{e_3} f(a_4)^{e_4}$ to one side using inverses. Then we see that $a_4^{-e_4} a_3^{-e_3} a_1^{e_1} a_2^{e_2} = e$ is a relation in G, and that $f(a_4)^{-e_4} f(a_3)^{-e_3} f(a_1)^{e_1} f(a_2)^{e_2} = e$ is a relation in H. We now see that the value of each $f(a_i)$ must be so that the relations in G are preserved under the map f. In other words, given an "equation" in G, it must also be an "equation" in H when we replace a_i with $f(a_i)$. Any homomorphism must preserve relations by the homomorphism property, and conversely, if a map $f : G \to H$ between two groups preserves all the relations in G, then this is sufficient for f to be a group homomorphism. Let $\{a_1, \ldots, a_k\}$ generate G. The real bijection now reads,

$$\{f : G \to H | \ f \text{ is a group homomorphism}\} \longleftrightarrow \{\text{relations preserving values } f(a_i) \in H\}$$

Example 6.50. Let us find all the group homomorphisms $\phi : \mathbf{Z}_5 \to \mathbf{Z}_{10}$. 1 is a generator for $\mathbf{Z}_5$ so by the bijection above, we just want to find the set of values in $\mathbf{Z}_{10}$ where 1 can be sent. In $\mathbf{Z}_5$ we have the relation $1 + 1 + 1 + 1 + 1 = 0$, and this is the only relation constraining 1. Thus, the set of values b s.t. $b + b + b + b + b \equiv 0 \pmod{10}$ is $\{0, 2, 4, 6, 8\}$, and $\phi(1) = 0, 2, \ldots$ gives all the homomorphisms from $\mathbf{Z}_5$ to $\mathbf{Z}_{10}$

Example 6.51. Let us find all the group homomorphisms $\phi : \mathbf{Z}_5 \to \mathbf{Z}_9$. Again, we just want to find the number of x such that $5x \equiv 0 \pmod 9$. But 5 has a multiplicative inverse modulo 9 because 5 and 9 are coprime. Mutliply both sides of the equation by 2 and we see that $x \equiv 0 \pmod 9$, and so there is only one group homomorphism from $\mathbf{Z}_5 \to \mathbf{Z}_9$, namely the map sending all elements of $\mathbf{Z}_5$ to the identity in $\mathbf{Z}_9$.

[3] For example, in example **??**, you can see that we expressed the element 5 in two different ways.

In general, there is always at least one group homomorphism from any group G to any other group H where we send all the elements of G to the identity element of H.

Example 6.52. Let us find all group homomorphisms from $\psi : \mathbf{Z}_6 \to (\mathbf{C}^\times, \cdot)$. We want to find where we can send 1 to. In $\mathbf{Z}_6$, 1 only has the relation $1+1+1+1+1+1 = 0$. In $\mathbf{C}$ we must then have $\psi(1) * \psi(1) * \psi(1) * \psi(1) * \psi(1) * \psi(1) = \psi(1)^6 = 1$. Using polar form, we let $\psi(1) = re^{i\theta}$ so $r^6 e^{6i\theta} = 1 \implies r = 1, 6i\theta = 2\pi ik$ meaning $\theta = 2\pi k/6$. If we were to graph these points in the complex plane we would get six evenly spaced points on the circle of raidus 1 centered at 0, also known as the $6th$ roots of unity. Each point gives us a different homomorphism, and these are all the homomorphisms of $\mathbf{Z}_6$ into $\mathbf{C}^\times$.

Example 6.53. Let us do the same problem but replace $(\mathbf{C}^\times, \cdot)$ with $(\mathbf{R}^\times, \cdot)$. What changes? Not all six points above are in $\mathbf{R}^\times$! Only two points are: namely 1 and -1. Therefore, there are only two homomorphisms from $\mathbf{Z}_6$ to $\mathbf{R}^\times$.

We know that, given generators $\{a_1, \ldots, a_k\}$ of a group G, we have certain restrictions on what $f(a_1), \ldots, f(a_k)$ can be if we want $f : G \to H$ to be a group homomorphism. What if we had no restrictions, or no relations in G? Suppose there existed a group F generated by $\{x_1, \ldots, x_n\}$ with no relations. We will call F a "free group." Then given *any* n elements $\{h_1, \ldots, h_n\}$ in *any* group H, there will exist a group homomorphism $\Phi : F \to H$ such that $\Phi(x_i) = h_i$ because we do not have any relations! How do we construct such a group F? It is not necessary to know the exact construction, but the general idea is to use words in the alphabet $\{x_1, \ldots, x_n\}$.

Homework 6.54. Check that $\langle A \rangle$ is a subgroup.

Homework 6.55. Show $(\mathbf{Q}^{>0}, \cdot)$ is generated by $\left\{\frac{1}{p} | \, p \text{ is a prime}\right\}$.

Homework 6.56. Suppose H_1, H_2 are subgroups of G. Show that $H_1 \bigcap H_2$ is also a subgroup of G.

Homework 6.57. Let A be a subset of a group G. Show that $C_G(A) = \{g \in G | \, gag^{-1} = a \; \forall a \in A\}$ is a subgroup of G.

6.5 Quotient groups

Subgroups allowed us to study groups by "focusing in" on a special part of the group. Now we'll consider the notion of a quotient group, which will allow us to "zoom out" and look at the group from on high, considering two elements of the group to be "basically the same" if they have certain properties in common.

Definition 6.58 (Coset)**.** Let H be a subgroup of G. Then the subset $gH = \{gh \mid h \in H\}$ is the *left coset* of H containing g. The subset $Hg = \{hg \mid h \in H\}$ is the *right coset* of H containing g.

g is always an element of gH because H is a subgroup and must contain an identity element. Therefore we can take $ge_G = g \in gH$.

Example 6.59. $G = \mathbf{Z}_8 = \{0, 1, 2, 3, 4, 5, 6, 7\}$ and $H = \{0, 4\}$. We will now list all the cosets of H.

$$\begin{aligned} 0H = \{0, 4\} &\iff 4H = \{0, 4\} \\ 1H = \{1, 5\} &\iff 5H = \{1, 5\} \\ 2H = \{2, 6\} &\iff 6H = \{2, 6\} \\ 3H = \{3, 7\} &\iff 7H = \{3, 7\} \end{aligned}$$

Notice how the cosets $4H, 5H, 6H, 7H$ are the same as $0H, 1H, 2H, 3H$ respectively as a set and each coset is the same size.

Lemma 6.60. $g_1H = g_2H \iff g_2^{-1}g_1 \in H$

Proof. First let us assume that $g_1H = g_2H$, meaning that the cosets are equal as sets. Therefore $\exists h \in H$ such that

$$\begin{aligned} g_1h_1 &= g_2h \\ g_2^{-1}g_1h_1 &= h \\ g_2^{-1}g_1 &= hh_1^{-1} \in H \end{aligned}$$

Now we want to show that $g_1H = g_2H$.

To prove $g_1H \subseteq g_2H$, take $g_1h_1 \in g_1H$. There $\exists h \in H$ such that

$$g_2^{-1}g_1 = h.$$
$$g_1h_1 = g_2hh_1 \text{ where } hh_1 \in H$$

Because $g_1h_1 \in g_2H$ then it must be that $g_1H \subseteq g_1H$.

To prove the other direction, that $g_2H \subseteq g_1H$. Take $g_2 = g_1h^{-1}$. Then we have $g_2h' = g_1h^{-1}h'$ where $h^{-1}h' \in H$. So it must be that $g_2h' \in g_1H$ and $g_2H \subseteq g_1H$. This proves that $g_1H = g_2H$. □

Homework 6.61. Prove the analogous statement: $Hg_1 = Hg_2 \iff g_2g_1^{-1} \in H$

Lemma 6.62. *Define a relation $g_1 \sim g_2$ iff $g_2^{-1}g_1 \in H$. Then $\sim$ is an equivalence relation on G.*

Proof.

1. **Reflexivity**
   $g_1^{-1}g_1 = e \in H$ so $g_1 \sim g_1$
2. **Symmetry**
   $g_1^{-1}g_2 \in H \iff g_2^{-1}g_1 \in H$ so $g_1 \sim g_2 \iff g_2 \sim g_1$
3. **Transitivity**
   We have $g_1 \sim g_2$ equal to $g_1^{-1}g_2 \in H$ and $g_2 \sim g_3$ equal to $g_2^{-1}g_3 \in H$. We can multiply these to get $g_1^{-1}g_2g_2^{-1}g_3 = g_1^{-1}g_3 \in H$. Therefore $g_1 \sim g_2, g_2 \sim g_3 \implies g_1 \sim g_1$.

□

Lemma 6.63. *If we have an equivalence relation $\sim$ on a set X, then $\sim$ partitions X. That is, the subsets $S_a = \{x \in X \mid a \sim x\}$ are either equal or disjoint and $X = \bigsqcup_{a \in X} S_a$.*

Proof. Suppose $a \sim b$. First we claim $S_a = S_b$. Take $x \in S_a \implies a \sim x$. We know since $a \sim b$ then $b \sim a$. Then by transitivity, $b \sim a$ and $a \sim x \implies b \sim x$.

The next case to prove is that if $a \nsim b$ then $S_a \cap S_b = \{\emptyset\}$. Suppose there is a $z \in S_a \cap S_b$. Therefore $a \sim z$ and $b \sim z$. This implies that $z \sim b$ and then $a \sim b$ which is a contradiction of our original assumption that $a \nsim b$.

Therefore the subsets $S_a = \{x \in X \mid a \sim x\}$ are either equal or disjoint. We know that $b \in S_b$ since $b \sim b$. From the previous statements, we know that either $S_b = S_b'$ or that S_b and S_b' are disjoint. Therefore $X = \bigsqcup_{a \in X} S_a$. □

Lemma 6.64. $|H| = |gH|$

Proof. Consider a mapping $\phi : H \to gH$ where $\phi(h) = gh$. We will proceed to prove that ϕ is a bijection.

$$\begin{aligned} gh_1 = \phi(h_1) = \phi(h_2) &= gh_2 \\ g^{-1}gh_1 &= g^{-1}gh_2 \\ h_1 &= h_2 \end{aligned}$$

Therefore ϕ is one to one. $\phi(h) = gh$ for all $h \in H$ proves that ϕ is onto. Therefore, because ϕ is a bijection between H and gH, then it must be that $|H| = |gH|$ □

Theorem 6.65 (Lagrange)**.** *Let H be a subgroup of a finite group G. Then the size of H divides the size of G.*

Proof. By lemma2, $g_1 \sim g_2$ is an equivalence relation. By lemma3, $G = \bigsqcup_{g \in G} S_g$ and using lemma1 we have $S_g = gH$ and $|gH| = |H|$ by lemma4. Therefore $|G| = \sum |gH| = \sum |H| = r|H|$ where $r \in \mathbf{Z}$. This proves that the size of H does divide the size of G. □

Proposition 6.66. *Any group G of prime order is cyclic.*

Proof. We know from Lagrange's Theorem that $|\langle z \rangle| r = |G| = p$ where p is prime for a cyclic group generated by $z \in G$. Because p is prime, the order of $\langle z \rangle$ is either 1 or p. Since $\langle z \rangle$ is a subgroup it must contain $\{e, z\}$ so $|\langle z \rangle| \geq 2$. This means that $|\langle z \rangle| = p$. □

Homework 6.67. Find all cosets of $4\mathbf{Z}$ in $\mathbf{Z}$.

Homework 6.68. Suppose H, K are finite subgroups of G s.t. $(|H|, |K|) = 1$. Prove that $H \cap K = \{e\}$

Proposition 6.69. *Let $|G| = n$. Then $a^n = e$ for $\forall a \in G$.*

Proof. Consider $\langle a \rangle$. By Lagrange's theorem, we know $|\langle a \rangle| r = n$. Because $\{1, a, a^2, ..., a^k, ...\} \subseteq \langle a \rangle$ and G is finite we must have $a^k = a^j \implies a^{k-j} = e$ for some $k > j$. As soon as a^m is e the elements a^m repeat Therefore, $|\langle a \rangle|$ is the smallest k such that $a^k = e$. We then have

$$a^n = a^{|\langle a \rangle| r} = \left(a^{|\langle a \rangle|}\right)^r = e^r = e$$

□

Corollary 6.70. $a^{p-1} \equiv 1 \pmod 9$ *for some prime p.*

Proof. Take the group $\mathbf{Z}/p\mathbf{Z}^*$. $|\mathbf{Z}/p\mathbf{Z}^*| = p-1$. Therefore $a^{p-1} = e \implies a^{p-1} \equiv 1 \bmod p$. □

Definition 6.71. A *normal subgroup* $N \subseteq G$ is a subgroup such that $\forall n \in G, \forall g \in G, gng^{-1} \in N$.

Example 6.72. $\{e\}$ and G are always normal subgroups.

Proof.

1. $geg^{-1} = gg^{-1} = e \in \{e\}$
2. G is normal because it is closed.

□

Example 6.73. Let $f : G \to H$ be a group homomorphism. Then $\ker f$ is normal.

Proof. Let $z \in \ker f$. We want to show that $gzg^{-1} \in \ker f$.

$$\begin{aligned} f(gzg^{-1}) &= f(g)f(z)f(g)^{-1} \\ &= f(g)f(g)^{-1} \\ &= f(e) \in \ker f. \end{aligned}$$

□

Proposition 6.74. *A subgroup N is normal if and only if $gN = Ng, \forall g \in G$.*

Proof. For the forward direction, suppose N is normal, we want to show that $gN = Ng$.
Take $g_1n_1 \in gN$, we know $g_1n_1g_1^{-1} = n_2 \in N$. This implies $g_1n_1 = n_2g_1 \in Ng$, so $gN \subseteq Ng$.
Now take a $n_1g_1 \in Ng$. We know $g_1n_1g_1^{-1} = n_2 \in N$. This implies $g_1n_1 = n_2g_1 \in gN$, so $gN \subseteq Ng$.
Therefore $gN = Ng$.
For the other direction, now suppose that $gN = Ng$ $\forall g \in G$. We want to show that $gng^{-1} \in N$ for $g \in G, n \in N$.
Notice that $gn \in gN = Ng$. This implies $gn = n'g$ for some $n' \in N$. We can multiply both sides by g^{-1} to obtain $gng^{-1} = n'$. Therefore $gng^{-1} = n'$. □

Definition 6.75. The *center* of a group G, written $Z(G)$, is the set

$$Z(G) = \{z \in G : \forall h \in G\ zh = hz\}.$$

Example 6.76. Show that $Z(G)$ is a normal subgroup.

Proof. We want to show that for $\forall z \in Z(G)$, $gzg^{-1}h = hgzg^{-1}$.

$$(gzg^{-1})h = zgg^{-1}h = zh = hz = hgg^{-1}z = h(gzg^{-1})$$

□

Example 6.77. A commutator in a group is an element of the form $aba^{-1}b^{-1}$ for some $g \in G$. Let $C(G)$ be the subgroup generated by all commutators. Show $C(G)$ is a normal subgroup.

Proof. To prove this we will use induction.

Base Case: $n = 1$. We want to show $ga_1b_1a_1^{-1}b_1^{-1}g^{-1} \in C(G)$. Let $z_{ga_1,b_1} := ga_1b_1a_1^{-1}g^{-1}b_1^{-1}$ and $z_{b_1,g} := b_1gb_1^{-1}g^{-1}$. Then notice

$$ga_1b_1a_1^{-1}b_1^{-1}g^{-1} = ga_1b_1a_1^{-1}(g^{-1}b_1^{-1}b_1g)b_1^{-1}g^{-1} = z_{ga_1,b_1}z_{b_1,g} \in C(G)$$

So $ga_1b_1a_1^{-1}b_1^{-1}g^{-1} \in C(G)$.

Inductive Hypothesis: $g(a_1b_1a_1^{-1}b_1^{-1})...(a_{n-1}b_{n-1}a_{n-1}^{-1}b_{n-1}^{-1})g^{-1} \in C(G)$

Inductive Step:

$$g(a_1b_1a_1^{-1}b_1^{-1})...(a_nb_na_n^{-1}b_n^{-1})g^{-1} = g(a_1b_1a_1^{-1}b_1^{-1})...(a_{n-1}b_{n-1}a_{n-1}^{-1}b_{n-1}^{-1})g^{-1}g(a_nb_na_n^{-1}b_n^{-1})g^{-1}.$$

By the inductive hypothesis,

$$g(a_1b_1a_1^{-1}b_1^{-1})...(a_{n-1}b_{n-1}a_{n-1}^{-1}b_{n-1}^{-1})g^{-1} \in C(G)$$

and

$$g(a_nb_na_n^{-1}b_n^{-1})g^{-1} \in C(G).$$

Therefore, the entire expression above is an element of $C(G)$. □

Homework 6.78. Compute $3^{31} \pmod 7$.

Recall that coset of a subgroup H partitions the group G into subsets of the same size. What if we treated cosets gH not as sets but as elements of a group? What would the group operation be?

A basic definition would be $g_1H * g_2H := g_1g_2H$. However, a problem with this definition is that it is not always well defined (i.e. if $g_1H = g_1'H$, then $g_1H * g_2H = g_1g_2H = g_1'g_2H = g_1' * g_2H$).

We want our operation to be well-defined, meaning that if $g_1H = g_1'H$, $g_2'H = g_2H$, then $g_1H * g_2H = g_1'H * g_2'H$.

Definition 6.79. The *quotient group* G/H is the set of left cosets of H in G. It is often verbally referred to as "G mod H."

Proposition 6.80. *If H is a normal subgroup of G, then the operation $* : G/H \times G/H \to G/H$ where $(g_1H) * (g_2H) = g_1g_2H$ is well-defined.*

Proof. Suppose $g_1H = g_1'H$, $g_2H = g_2'H$. We want to show that $g_1H * g_2H = g_1'H * g_2'H$.

Since $g_1H = g_1'H$, $g_2H = g_2'H$, then $g_1 = g_1'h_1$ and $g_2 = g_2'h_2$ where $h_1, h_2 \in H$.

Then:

$$\begin{aligned} g_1'H * g_2'H &= g_1'g_2'H \\ &= g_1h_1{-}1g_2h_2^{-1}H \end{aligned}$$

Let $h_3 = g_2^{-1}h_1^{-1}g_2$ where $h_3 \in H$ because when conjugating the element h_1 by g_2, the end result, h_3, must be in H since H is, by assumption, a normal subgroup. Then $h_1^{-1}g_2 = g_2h_3$.

$$\begin{aligned} g_1'H * g_2'H &= g_1g_2h_3h_2^{-1}H \\ &= g_1g_2H \\ &= g_1 * g_2H \end{aligned}$$

□

In fact, if H is normal, $\forall g_1h \in g_1H$ and $\forall g_2h' \in g_2H$, then $(g_1h)(g_2h') \in g_1g_2H$. If $g_1h \in g_1H$, then h is said to be a representative of the coset of H.

Proposition 6.81. *Let N be a normal subgroup of G. Then the set of left cosets of N with the operation $(g_1N)(g_2N) = g_1g_2N$ forms a group G/N called the quotient (factor) group of G by N.*

Proof.
1. The identity is N or eN.
   $(g_1N)(eN) = g_1eN = g_1N$
2. $(g_1N)(g_1{-}1N) = eN$
3. Since the operation is defined for g_1 and for g_1^{-1} and the original group operations is associate, the new one must be associative too.

□

Example 6.82. For any n, notice that $n\mathbf{Z}$ is a coset of $\mathbf{Z}$. Show that $g(n\mathbf{Z}) = (n\mathbf{Z})g$

Proof. Let $na \in n\mathbf{Z}$

$$\begin{aligned} g(n\mathbf{Z}) &= \{g + na\} \\ &= na + g \\ &= (n\mathbf{Z})g \end{aligned}$$

□

Since $n\mathbf{Z}$ is normal, $\mathbf{Z}/n\mathbf{Z}$ is a group.

Example 6.83. $\mathbf{Z}/n\mathbf{Z} \cong (\mathbf{Z}_n, +)$

Proof. We know $\Phi : \mathbf{Z}/n\mathbf{Z} \to \mathbf{Z}/n$ and $\Phi(g(n\mathbf{Z})) = g \bmod n$.

Next, we show that Φ is a homomorphism:

$$\begin{aligned} \Phi((g_1 + g_2)n\mathbf{Z}) &= \Phi(g_1n\mathbf{Z}) + \Phi(g_2n\mathbf{Z}) \\ (g_1 + g_2) \pmod n &= g_1 \pmod n + g_2 \pmod n \end{aligned}$$

To show $\mathbf{Z}/n\mathbf{Z}$ is $1 - 1$, it suffices to show $\ker\Phi = \{n\mathbf{Z}\} = e$. Indeed, $\Phi(g(n\mathbf{Z})) = 0 \pmod n$ so $g = na$ for some a, so $g(n\mathbf{Z}) = n\mathbf{Z}$.

Then, to prove $\mathbf{Z}/n\mathbf{Z}$ is surjective, we observe that given $g \pmod n$, $\Phi(g(n\mathbf{Z})) = g \pmod n$. □

Lemma 6.84. *If G is an abelian group, then for all subgroups H, for all $g \in G$, $gH = Hg$.*

Proof. Take $gh \in gH$. Since it is abelian, $gh = hg \in Hg$. Then take $hg \in Hg$, so $hg = gh \in gH$. □

Corollary 6.85. *All subgroups of an abelian group are normal.*

Proof. We have that for $g \in G$ and $h \in H$ where $H \leq G$, that $ghg^{-1} = gg^{-1}h = eh = h \in H$ so H is normal. □

Homework 6.86. Let $V \subseteq W$ be a subspace of a vector space W. Show that:

1. V is a normal subgroup of W.
2. V/W is a vector space.
3. If V is finite-dimensional then
$$\dim(V/W) = \dim V - \dim W.$$

Example 6.87. What is the quotient group of $\{0, 4\}$ in $\mathbf{Z}/8\mathbf{Z}$?

Proof. We want to prove that $\mathbf{Z}/8\mathbf{Z}/\{0, 4\} \cong \mathbf{Z}/4\mathbf{Z}$. Elements of the quotient group are the left cosets of $\{0, 4\}$ in $\mathbf{Z}/8\mathbf{Z}$. For example,

$$H = \{0, 4\}, 1H = \{1, 5\}, 2H = \{2, 6\}, 3H = \{3, 7\}.$$

Thus, we can define a natural map $\phi : \mathbf{Z}/4\mathbf{Z} \to \mathbf{Z}/8\mathbf{Z}/\{0, 4\}$ where we send each element x to the coset x is in, that is $\phi(a) = aH$. For example, 2 gets mapped to $\{2, 6\}$. ϕ is a homomorphism as $\phi(a+b) = (a+b)H = aH+bH = \phi(a)+\phi(b)$. This is one to one because $Ker(e) = 0$, and onto by construction essentially. So this is an isomorphism. □

Lemma 6.88. *Let N be a normal subgroup of G, and let $k = |G/N|$. Prove $a^k \in N$ for all $a \in G$.*

Proof. Let $[a] \in G/N$. Then $[a]^k = e \in G/N$ because every element of G/N has order dividing k by Lagrange's theorem. The identity in G/N is the identity coset N, so we have $a^k N = N$, in other words, a^k lands in $N = eN$. □

Theorem 6.89 (first isomorphism theorem)**.** *If $f : G \to H$ is a homomorphism, then $G/\ker(f) \cong \operatorname{Im}(f)$.*

Proof...? Let $\Phi : G/\ker(f) \to \operatorname{Im}(f)$ be defined by $\Phi(x \ker(f)) = f(x)$.
First we must prove this is a homomorphism. We see that

$$\Phi((x \ker(f))(y \ker(f))) = \Phi(xy \ker(f)) = f(xy) = f(x)f(y) = \Phi(x \ker(f))\Phi(y \ker(f))$$

Next we must show it is one to one. Observe that

$$\begin{aligned} \Phi(x \ker(f)) &= \Phi(y \ker(f)) \\ f(x) &= f(y) \\ f(x)f(y)^{-1} &= f(y)f(y)^{-1} \\ f(xy^{-1}) &= e \end{aligned}$$

So $xy^{-1} \in \ker(f)$ and therefore $x \ker(f) = y \ker(f)$.

Lastly, we must show that Φ is onto. For all $y \in \operatorname{Im}(f)$, there exists a $x \in G/\ker(f)$ such that $\Phi(x) = y$.

So thre is indeed an isomorphism between $G/\ker(f)$ and $\operatorname{Im}(f)$. □

The above argument is cheating. The coset $x \ker f$ can be represented in many ways, so if $x \ker f = y \ker f$, then we should have $\Phi(x \ker f) = \Phi(y \ker f)$ if Φ is a map from $G/\ker(f) \to H$. For this to be true we must then have $f(x) = f(y)$ by the defintiion of Φ, but it isn't immediately clear why this will happen.

More generally, suppose we want to define a homomorphism $\overline{f} : G/N \to H$. Our only real hope is to first define a homomorphism $f : G \to H$ and then hope it passes to the quotient. The homomorphisms that pass to the quotient are precisely the ones *compatible* with the subgroup N. What do I mean by this? Well, a homomorphism and a normal subgroup (f, N) are said to be compatible if $f(n) = e \in H$, $\forall n \in N$, that is N is a subgroup of the kernel of f.

Example 6.90. Here is an example of a non-compatible pair. Suppose we have $\phi : \mathbf{Z}/8\mathbf{Z} \to \mathbf{Z}/8\mathbf{Z}$ where $\phi(1) = 3$. Now let $N = \{0, 4\}$. We calculate that $\phi(0) = 0$ and $\phi(4) = 4$, which is **bad** because it shows ϕ isn't constant on the subgroup N. In other words, they are both elements of the same coset but do not map to the same element under ϕ. So this is why it is necessary to show that a mapping is well defined.

Lemma 6.91. *$y \in xH$ if and only if $yH = xH$.*

Proof. Take $yh_1 \in yH$ and $y = xh$. Then $yh_1 = xhh_1 \in xH$. Conversely, take $xh_2 \in xH$. Then $xh_2 = yh^{-1}h_2 \in yH$.
So we can see that $yh_1 = xh_2$ and therefore $y = xh_2h_1^{-1} \in xH$ □

Proposition 6.92. *Let $f : G \to H$ be a homomorphism and $N \subset G$ a normal subgroup. If (f, N) is a compatible pair, meaning $N \subseteq \ker f$, then it gives a well-defined map $\overline{f} : G/N \to H$ defined by $\overline{f}(gN) = f(g)$ and vice versa.*

Proof. Suppose $yN = xN$. By the lemma above we then have that $y = xn,\ n \in N$. Now

$$\overline{f}(yN) = f(y) = f(xn) = f(x)f(n) = f(x) = \overline{f}(xN)$$

showing $\overline{f}$ is a well defined map on G/N. □

This shouldn't be too surprising; recall that the intuition behind G/N is that we identify each coset of N in G into one single element, with N being the identity element. Thus, for f to descend to a map on G/N we should have f be constant on N. Moreover, as $f(e) = e$ for any group homomorphism, we should then have $f(n) = e\ \forall n \in N$.

*Proof of Theorem **??**, take 2.* It remains to show that Φ is well defined which by the preceding propositon is the same as showing $N \subseteq Ker(f)$. Since $N = \ker(f)$ in this case, the proof is complete. □

Example 6.93. Use the first isomorphism theorem to prove $\mathbf{Z}/8\mathbf{Z}/H \cong \mathbf{Z}/4\mathbf{Z}$ where $H = \{0, 4\}$.

Proof. First we want to define a homomorphism $f : \mathbf{Z}/8\mathbf{Z} \to \mathbf{Z}/4\mathbf{Z}$. So let $f(1) = 1 \pmod 4$.

Next we want to show that $\ker(f) = H$ and $\text{Im}(f) = \mathbf{Z}/4\mathbf{Z}$. We observe that $\ker(f) = \{0, 4\} = H$ and $\text{Im}(f) = \mathbf{Z}/4\mathbf{Z}$. □

Homework 6.94. Suppose N is a normal subgroup of G and $H \subseteq G$. Show $N \cap H$ is normal in H.

Example 6.95. In linear algebra, if $T : V \to V$ is a linear map, then the *rank theorem* says $\dim(V) = \dim(\ker(T)) + \dim(\operatorname{Im}(T))$. This actually follows from the first isomorphism theorem and Homework **??**.

Proof. We know that $V/\ker(T) \cong \operatorname{Im}(T)$, so in fact

$$\dim(V/\ker(T)) = \dim(\operatorname{Im}(T)).$$

But as you proved on the homework,

$$\dim(V/W) = \dim V - \dim W$$

so

$$\dim V = \dim \ker T + \dim \operatorname{Im} T.$$

□

There is a duality between multiplication in G/N and multiplication in G:

1. If we know something about the multiplication in G/N, then we can obtain information about multiplication in G by passing to a representative in G.
2. If we know something about how multiplication in G behaves, then we can obtain information about multiplication in G/N by choosing a representative and going up to a coset.

Example 6.96. If $G/Z(G)$ is cyclic, then G is abelian, where $Z(G) = \{g \in G | gh = hg \forall h \in G\}$

Proof. There exists a coset $aZ(G)$ such that all elements of $G/Z(G)$ can be written as $a^k Z(G)$. We can take $g_1 g_2 \in G$. We now that g_1 is in a coset of $Z(G)$, so for $b_1 \in Z(G)$ and $b_2 \in Z(G)$,

$$\begin{aligned} g_1 &= gb_1 \\ &= a^{k_1} b_1 \\ g_2 &= a^{k_2} b_2 \end{aligned}$$

Which means that $g_1 g_2 = a^{k_1} b_1 a^{k_2} b_2 = a^{k_1} a^{k_2} b_1 b_2 = a^{k_2} a^{k_1} b_1 b_2 = a^{k_2} a^{k_1} b_2 b_1 = a^{k_2} b_2 a^{k_1} b_1 = g_2 g_1$. So G is abelian. □

6.6 The classical Fourier transform

In Math 54, you learn about inner-product spaces. If $(V, \langle \cdot, \cdot \rangle)$ is an inner-product space, then we define the *norm* (or length) of $v \in V$ by $|v| = \sqrt{\langle v, v \rangle}$. Then V is a metric space, with $d(v_1, v_2) = |v_1 - v_2|$.

Homework 6.97 (Pythagorean theorem). If $V = \mathbf{R}^n$ with its usual dot product, then (V, d) is isometric to $\mathbf{R}^n$ with its usual euclidean metric.

You also learn that if $e \in V$ is a unit vector and $v \in V$, then the projection of v onto the line ℓ spanned by e, $p = \langle v, e\rangle e$, is the unique vector in ℓ such that for every $q \in \ell$, $|p - v| \leq |q - v|$.

Homework 6.98. ℓ is closed in V. (So if $v \notin \ell$, then v is not a limit point of ℓ. Therefore this best approximation makes sense.)

Let $L^2([-\pi, \pi])$ be the set of all functions $f : [-\pi, \pi] \to \mathbf{C}$ such that

$$\int_{-\pi}^{\pi} |f(x)|^2 \, dx < \infty.$$

Homework 6.99 (Schwarz' inequality). $L^2([-\pi, \pi])$ is an inner-product space, with inner product

$$\langle f, g\rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\overline{g(x)} \, dx$$

(where $\overline{z}$ means the complex conjugate of z, so $\overline{x + iy} = x - iy$ if $x, y \in \mathbf{R}$).

(Hint: first prove

$$\left|\int_{-\pi}^{\pi} f(x)\overline{g(x)} \, dx\right|^2 \leq \int_{-\pi}^{\pi} |f(x)|^2 \, dx \int_{-\pi}^{\pi} |g(x)|^2 \, dx.)$$

Homework 6.100. Given $n \in \mathbf{Z}$, let $f_n(x) = e^{inx}$. Then the functions $\{f_n : n \in \mathbf{Z}\}$ are *orthonormal* in $L^2([0, 2\pi])$ (i.e. $|f_n| = 1$ and $\langle f_n, f_m\rangle = 0$ if $n \neq m$).

Now the e^{inx} are *waves*, as proven by the following theorem from calculus.

Homework 6.101 (Euler's identity). If $x \in \mathbf{R}$, then

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

(Hint: Write out the Taylor series of both sides.)

Theorem 6.102. *If X denotes the span of the f_n, then its closure (viewed as a subset of $L^2[-\pi, \pi]$) is*

$$\overline{X} = \{f \in L^2([-\pi, \pi]) : f(-\pi) = f(\pi)\},$$

functions which are periodic with period 2π.

So if $f \in \overline{X}$, we can decompose f into projections

$$f(x) = \frac{1}{2\pi} \sum_{n \in \mathbf{Z}} e^{inx} \int_{-\pi}^{\pi} f(\xi) e^{-in\xi} \, d\xi.$$

This decomposition is called the *Fourier series* of f, which you might've seen in Math 54. The idea is that the functions e^{inx} are really easy to understand from the point of view of linear algebra – they're just eigenvectors of the derivative operator $Df = f'$, with eigenvalues n. So you can think of this decomposition as "diagonalizing" D.

We want to do this with functions with period P. Actually, it'll be more useful to work with the frequency $\xi = 1/P$. So we can use some trigonometry to write

$$f(x) = \xi \sum_{n \in \mathbf{Z}} e^{2\pi n \xi i x} \int_{-P/2}^{P/2} f(y) e^{-2\pi i n \xi y}\, dy.$$

Since $y \mapsto e^{2\pi i n \xi y}$ is a wave of frequency $n\xi$, we've basically decomposed f into waves of frequency $n\xi$.

So far this has all been review of Math 54. But what if f isn't periodic at all? Then we can think of it having *infinite* period, and taking the limit as $P \to \infty$, we have

Theorem 6.103 (Fourier's inversion theorem)**.** *Let $f \in L^2$. If all the integrals in the below equation are finite, then*

$$f(x) = \int_{-\infty}^{\infty} e^{2\pi i \xi x} \int_{-\infty}^{\infty} f(y) e^{-2\pi i y \xi}\, dy\, d\xi.$$

We write

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(y) e^{-iy\xi}\, dy$$

for the *Fourier transform* of f. It is the "part" of f which oscillates with frequency ξ.

The Fourier transform is useful in differential equations because of the following result.

Homework 6.104.

$$\widehat{f'}(\xi) = 2\pi i \xi \hat{f}(\xi).$$

So differentiation, which is hard to work with, becomes multiplication by a frequency, which is easy to work with. (This is what we meant earlier by "diagonalizing" the derivative.)

It's also useful in signal processing. Your radio receives lots of signals of different frequencies, but if you want to listen to 89.3 FM, your radio will take the Fourier transform of the signal f, then take the inverse Fourier transform. That way, the function

$$x \mapsto \int_{-\infty}^{\infty} \chi(\xi) \hat{f}(\xi) e^{2\pi i x \xi}\, d\xi = \int_{89.3-\varepsilon}^{89.3+\varepsilon} \hat{f}(\xi) e^{2\pi i x \xi}\, d\xi$$

is just the signal transmitted by 89.3 FM.

Another example: in quantum mechanics, particles P are described by wavefunctions f. $f(x)$ roughly tells you the "probability that P is close to x". Meanwhile the Fourier transform $\hat{f}(\xi)$ roughly tells you the "probably that P has momentum close to ξ." We have

Theorem 6.105 (Heisenberg's uncertainty principle)**.**

$$\left(\int_{-\infty}^{\infty} (x - x_0)^2 |f(x)|^2\, dx\right) \left(\int_{-\infty}^{\infty} (\xi - \xi_0)^2 \left|\hat{f}(\xi)\right|^2\, d\xi\right) \geq \frac{1}{16\pi^2}$$

The integral

$$\left(\int_{-\infty}^{\infty} (x - x_0)^2 |f(x)|^2\, dx\right)$$

is the "standard deviation in position" and

$$\left(\int_{-\infty}^{\infty} (\xi - \xi_0)^2 \left|\hat{f}(\xi)\right|^2 \, d\xi\right)$$

is the "standard deviation in momentum". So this inequality says that both cannot be "small" at the same time, which is Heisenberg's uncertainty principle: if you know where a particle is, you don't know how fast it's moving, and vice versa.

6.7 The abstract transform

Actually, the Fourier transform is so useful we want to use it not just on $\mathbf{R}$, but on lots of abelian groups.

A *topological group* is a group equipped with a metric such that the group multiplication and the inversion are continuous maps.

Homework 6.106. The following groups are also topological groups:

1. If G is a finite group, then G under the discrete metric.
2. If H is a subgroup of a topological group G, then H under the metric induced by G.
3. $\mathbf{R}$ is a topological group under its usual metric and under addition.
4. $\mathbf{C}^*$ (nonzero complex numbers) under its usual metric and under multiplication.

For example, the subgroup of $\mathbf{C}^*$, T, of elements of norm 1 is an (abelian) topological group as well. Elements of T are *rotations* of the plane, since they are all of the form $e^{i\theta}$ for some $\theta \in [0, 2\pi)$.

We usually call T the *circle group*.

Now let's assume that G is an *abelian* group.

Definition 6.107. A *character* on G is a continuous group homomorphism $G \to T$. The *dual group* of G, written $\hat{G}$, is the group of all characters on G under multiplication: if $g \in G$, then we define

$$(\varphi\psi)(g) = \varphi(g)\psi(g).$$

Since multiplication in T is commutative, multiplication in $\hat{G}$ is as well. So $\hat{G}$ is an abelian group. Besides this, if $\psi \in \hat{G}$ and $g \in G$, then $\psi^{-1}(g) = \overline{\psi(g)}$, the complex conjugate of $\psi(g)$.

Theorem 6.108. *The following abelian topological groups are dual:*

1. $\hat{\mathbf{Z}} \cong T$. *For each* $e^{i\theta} \in T$, *write* $\psi_\theta(n) = e^{in\theta}$.
2. $\hat{T} \cong \mathbf{Z}$. *For each* $n \in \mathbf{Z}$, *write* $\psi_n(e^{i\theta}) = e^{in\theta}$.
3. $\hat{\mathbf{R}} \cong \mathbf{R}$. *For each* $x \in \mathbf{R}$, *write* $\psi_x(\xi) = e^{2\pi i x\xi}$.
4. *If* G *is a finite abelian group, then* $\hat{G} \cong G$.

(4) is worth trying to prove after taking Math 114. The proof uses roots of unity and the classification of finitely generated abelian groups.

Theorem 6.109 (Pontraygin duality theorem)**.** *For any "nice" abelian topological group G, $\hat{\hat{G}} \cong G$.*

(If you've taken Math 104 and know what compactness is, when we say "nice", we mean that for each $x \in G$, there is an open set $U \ni x$ such that $\overline{U}$ is compact. If this is true of some metric space X, we say that X is a locally compact space.

If you've taken Math 110 and know what naturality is, the isomorphism $\hat{\hat{G}} \cong G$ is natural, but the isomorphism $\hat{G} \cong G$ is unnatural.)

Now let's assume that our abelian group has an "integration" operation. All the abelian groups mentioned in the statement of Theorem **??** have a natural "integration":

1. For a function $f : \mathbf{Z} \to \mathbf{C}$, we define
$$\int_{\mathbf{Z}} f(n)\ dn = \sum_{n \in \mathbf{Z}} f(n).$$

2. For a function $f : T \to \mathbf{C}$, we define integration to be the average line integral around T of f; that is,
$$\int_T f(z)\ dz = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(e^{i\theta})\ d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(e^{i\theta})\ d\theta.$$

3. For a function $f : \mathbf{R} \to \mathbf{C}$, we define integration to be the improper integral of f; that is,
$$\int_{\mathbf{R}} f(x)\ dx = \lim_{R \to \infty} \int_{-R}^{R} f(x)\ dx.$$

4. If G is a finite abelian group, for a function $f : G \to \mathbf{C}$, we define
$$\int_G f(g)\ dg = \sum_{g \in G} f(g).$$

Theorem 6.110 (Haar)**.** *Any "nice" abelian topological group has an integration notion, called the* Haar integral.

The Haar and Pontraygin theorems are highly nontrivial to prove.

Homework 6.111. Try to come up with integration notions on other familiar topological groups. In particular, what's integration on the general linear group $\mathrm{GL}(V)$?

For each of the abelian groups G, we let $L^2(G)$ denote the vector space of all functions $f : G \to \mathbf{R}$ such that
$$\int_G |f(g)|^2\ dg < \infty.$$

Definition 6.112. Let $f \in L^2(G)$. The *Fourier transform* of f, written $\hat{f}$, is the map $\hat{G} \to \mathbf{R}$ defined by

$$\hat{f}(\psi) = \int_G f(g)\psi^{-1}(g)\, dg,$$

if this exists.

Homework 6.113. The Fourier transform on T is the same thing as the Fourier series for periodic functions on $[-\pi, \pi]$, where we think of $\theta \in [-\pi, \pi]$ as the same thing as $e^{i\theta} \in T$. (This makes sense because we assume $f(-\pi) = f(\pi)$.)

Besides, the Fourier transform on $\mathbf{R}$ is just the same thing as the classical Fourier transform defined above.

If you've taken or are taking CS 170, the fast Fourier transform is an algorithm for rapidly computing the Fourier transform on $\mathbf{Z}/n\mathbf{Z}$. Viewed in frequency space, it becomes much easier to multiply polynomials, which is valuable in countless applications.

Index