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 Friday lectures on groups: groups, homomorphisms, subgroups, generators and relations, and quotients

3. 9 Monday lectures on metric spaces: metric spaces, limit and interior points, open and closed sets, and continuity and convergence

4. 2 lectures on Fourier transforms

Chapter 1

Introduction

1.1 What are proofs?

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.

1.2 Direct Proofs

In a direct proof, you should analyze every piece of information you are given in the problem. Direct proofs often involve starting from definitions and drawing connections between given facts.

Example 1.1. Let n be a natural number. Prove that if n is even, then n^2 is even.

Proof. We will use a direct proof. Recall the definition of *even*: if n is even, then it is divisible by 2. That is, 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. □

We pretty much always will start a direct proof by writing down a key definition or theorem. We always want to write down the key players so we have them laid out in front of us.

Let's see a more advanced example. The following is a classic proof in linear algebra.

Example 1.2. 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. Again, we will use a direct 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 is accomplished in much the same way as above; take an arbitrary vector $x \in \ker(T)$ and a scalar c, and show that $cx \in \ker(T)$.

Homework 1.3. Verify that, in fact, $\ker(T)$ is closed under scalar multiplication.

□

The proof of 1.2 looks a bit long, but it really had three sub-theorems, each its own proof:

1. $\ker(T)$ is a subset of V.
2. $\ker(T)$ is closed under addition.

3. $\ker(T)$ is closed under scalar multiplication.

This is standard fare in mathematics, with multiple pieces of a proof all working together. In fact 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 the proof of 1.2, we checked to make sure we used all our assumptions. Indeed, if you ever complete a proof without using all your assumptions, one of the three things went wrong:

1. You assumed too much. In this case, the statement of the theorem you were trying to prove is wrong, and 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!

Homework 1.4. Show that if $a|x$ (that is, x is divisible by a) and $b|y$ then $ab|xy$.

1.3 Proof by Cases

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.

Here is an interesting proof by cases.

Example 1.5. 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. □

One unique aspect of this proof is that it 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.

For the following problems, you should think about what the cases are.

Homework 1.6. Prove that for all real numbers x, $x + |x - 1| \geq x$.

Homework 1.7 (real triangle inequality)**.** Prove that for any real numbers x and y, $|x+y| \leq |x| + |y|$.

Homework 1.8 (Euclid's lemma)**.** Show that if p is prime and $p|ab$ then $p|a$ or $p|b$.

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 1.9. Prove that $\sqrt{2}$ 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 *uncountably infinite*, which means there is no enumeration of the elements of the set. (More formally, there is no bijection between $\mathbf{N}$ and the set.)

Example 1.10 (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. □

Homework 1.11. Use a proof by contradiction to show there are infinitely many prime numbers.

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:

1. 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.

2. Therefore, $2^{2^5}+1$ 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.

Proposition 1.12 (principle of induction). *Let $P(n)$ be a statement indexed by $n \in \mathbf{N}$, the set of all natural numbers. (We're assuming that the first natural number is 1, but some writers will take $0 \in \mathbf{N}$. Be aware!)*

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

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

Let us see an example of this in practice.

Example 1.13. 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 1.13, 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.

Proposition 1.14 (principle of induction)**.** *Let $S \subseteq \mathbf{N}$ (that is, S is a subset of $\mathbf{N}$). Suppose that*

1. *$1 \in S$; that is, 1 is an element of 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. □

Here's another statement you might have heard sometime in your life.

Proposition 1.15 (well-ordering principle)**.** *Any non-empty subset of $\mathbf{N}$ has a least element.*

By a *least element*, we mean that if $S \subseteq \mathbf{N}$ is nonempty, there is some $n \in S$ such that if $m < n$, then $m \notin S$.

Now here's something that is somewhat surprising.

Theorem 1.16. *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.*

To prove an "if and only if" type statement ("P if and only if Q", or "P implies Q and Q implies P"), it suffices to:

1. Assume P, and prove Q.

2. Assume Q, and prove P.

Often it's easier to *prove two directions separately* than try to take on both at once.

Proof of Theorem 1.16, first try. First we'll assume the well-ordering principle.

Given a set S satisfying (i) and (ii) of 1.14 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? □

In these scenarios when the problem is hard, there is one method that pull us back from the brink: proof by contradiction. We'll have to assume that POI is false even though WOP is true.

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 1.16, 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.

Homework 1.17. Here is a fun example. Let

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

and prove that

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

Proceed by induction. If you get stuck, remember that you're trying to prove $\Gamma(n + 1) = n\Gamma(n)$, and use integration by parts.

Fool's Theorem 1.18. *For all n, $\frac{d}{dx}(x^n) = 0$.*

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

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

$\square$

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.

Homework 1.19. Show that

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

Homework 1.20. Show that

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

Homework 1.21. Show a polynomial of degree n has at most n roots.

Homework 1.22. Show that

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

Homework 1.23 (the Fibonacci sequence)**.** Let F_n be the sequence defined by $F_n = F_{n-1} + F_{n-2}$ and $F_0 = 0, F_1 = 1$, and show

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

Homework 1.24 (Euclid's lemma again)**.** Prove that if p is prime and $p|a_1 \ldots a_n$ then there is some $i \in \{1, \ldots, n\}$ such that $p|a_i$.

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.

Recall that 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.

Example 1.25 (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. Let's see an example from discrete math where it's impossible to give such instructions, while still proving existence.

Example 1.26 (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.

Proof. Let v_1 be the root of the tree.

We want to prove the following lemma:

Lemma 1.27. *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 1.26, 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. You may recall the following proof from linear algebra, and here you should focus on the structure of the proof.

Example 1.28. 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 1.29. Let G be the *infinite binary tree*: the tree such that every node has exactly two children. Then the infinite path guaranteed by Example 1.26 is NOT unique.

Proof. We just need to show that there are two infinite paths through G which are distinct. By Example 1.26, there is a path $v_1 \dots v_n \dots$ 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 $v_1 \dots v_n \dots$, any infinite path containing w_2 is not equal to $v_1 \dots v_n \dots$, disproving uniqueness.

We apply Example 1.26 again, to the tree consisting of w_2 and all its descendents, to get a path $w_2 \dots w_n \dots$. Then $v_1 w_2 w_3 \dots w_n \dots$ is an infinite path through G which is certainly not equal to $v_1 v_2 v_3 \dots v_n \dots$. So infinite paths through G are not unique. □

Homework 1.30 (fundamental theorem of arithmetic)**.** Show that every natural number $n \geq 2$ has a unique prime factorization.

Recall that we already proved existence, 1.25, so all that you have to prove is uniqueness. You might want to use Euclid's lemma, Homework 1.8.

Here are some more examples.

Homework 1.31. Show that the inverse of a matrix A is unique. (Why is this claim trivial whenever A is not invertible?)

Homework 1.32. Show that a finite-dimensional vector space has a basis. Is it unique?

Homework 1.33. If $f : [0,1] \to \mathbf{R}$ is a function, recall that the *antiderivative* of f is a function $F : [0,1] \to \mathbf{R}$ such that $F' = f$. For which f is the antiderivative F unique?

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.

Homework 1.34. 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 (i.e. both injective and surjective).

You can use the *intermediate value theorem* as a black box:

Theorem 1.35 (intermediate value theorem)**.** *If $f : [0,1] \to [0,1]$ is continuous, and both x and y are in the image of f (that is, $x \in f([0,1])$, and $y \in f([0,1])$), then if $z \in [x,y]$, $z \in f([0,1])$.*

1.7 If-and-Only-If Proofs

Often you will be asked to show that some property A holds *if and only if* B holds. This typically means you must split your proof into two sections: the "forward" direction $A \implies B$, and the "reverse" direction $B \implies A$. Recall the contrapositive method of proof we discussed earlier. To show $B \implies A$, it is equivalent (and occasionally easier) to show $(\neg A) \implies (\neg B)$.

The following is another proof from linear algebra.

Example 1.36. 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. □

Always remember to prove *both* directions of if-and-only-if statements! Particularly stern graders may award less than half credit for incomplete proofs. For the sake of clarity, it is recommended that you structure your if-and-only-if proofs as above, with ($\Longrightarrow$) and ($\Longleftarrow$) marking the sections of the proof.

Often times one direction will be easy and the other one will be harder. For example, proving that if there is an injection then A is infinite in the following homework problem will probably be easier than provign the other direction.

Homework 1.37. Suppose that A is a set. Prove that A is infinite if and only if there is an injection $f : \mathbf{N} \to A$.

If A is infinite, is f unique?

Chapter 2

Math 104: Real Analysis

Analysis is the study of inequalities and limits. Unsurprisingly, *real* analysis deals with limits of functions of real numbers.

There are three main parts to Math 104:

1. Developing the theory of the set of real numbers, **R**.

2. Developing the topology of metric spaces.

3. Developing the theory of calculus.

We'll focus in on (2).

2.1 Metric spaces

Metric spaces are an extension of the notion of a set, which allows us to talk about the distance between points in the set.

Definition 2.1 (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.

But it turns out that (1) is actually unnecessary.

Let's show $(2), (3), (4) \implies (1)$.

Let $x = z$, and y be arbitrary. Then, using the Triangle Inequality,

$$\begin{aligned} d(x,z) &\leq d(x,y) + d(y,z) \\ d(z,z) &\leq d(z,y) + d(y,z) \\ 0 &\leq d(y,z) + d(y,z) \\ 0 &\leq 2d(y,z). \end{aligned}$$

So, regardless of our choice of y, $d(y,z) \geq 0$. Thus, it suffices to prove (2), (3), and (4), to show that d is a metric.

Example 2.2. Show that ($\mathbf{R}$, $d(x,y) = |x-y|$) is a metric space.

It should be easy to see that (2) is satisfied, and one can show the triangle inequality by casework or drawing a picture. In general, for most of these examples, we'll be checking if the Triangle Inequality is satisfied.

One of the most useful proof techniques more generally is to *draw a picture.*

Example 2.3. (Euclidean metric) Show that $\left(\mathbf{R}^n,\ d(\vec{p},\vec{q}) = \sqrt{(q_1-p_1)^2 + ... + (q_n-p_n)^2}\right)$ is a metric space.

Let us first try to verify the triangle inequality explicitly as is. Let $\vec{p} = (p_1, \ldots, p_n), \vec{q} = (q_1, \ldots, q_n)$ and $\vec{s} = (s_1, \ldots, x_n)$. We then want to prove the inequality

$$\sqrt{(q_1-p_1)^2 + \ldots + (q_n-p_n)^2} \leq \sqrt{(s_1-p_1)^2 + \ldots + (s_n-p_n)^2} + \sqrt{(q_1-s_1)^2 + \ldots + (q_n-s_n)^2}$$

We square both sides and but we soon find that there are too many variables. So instead let's solve an easier problem. What if $\vec{s} = \vec{0}$? Then we want to verify the inequality

$$\begin{aligned} \sqrt{(q_1-p_1)^2 + \ldots + (q_n-p_n)^2} &\leq \sqrt{p_1^2 + \ldots + p_n^2} + \sqrt{q_1^2 + \ldots + q_n^2} \\ (q_1-p_1)^2 + \ldots + (q_n-p_n)^2 &\leq \left(\sqrt{p_1^2 + \ldots + p_n^2} + \sqrt{q_1^2 + \ldots + q_n^2}\right)^2 \\ q_1^2 - 2q_1p_1 + p_1^2 + \ldots + q_n^2 - 2q_np_n + p_n^2 &\leq \\ p_1^2 + \ldots + p_n^2 + q_1^2 + \ldots + q_n^2 &+ 2\sqrt{(p_1^2 + \ldots + p_n^2)(q_1^2 + \ldots + q_n^2)} \\ -2q_1p_1 - \ldots - 2q_np_n &\leq 2\sqrt{(p_1^2 + \ldots + p_n^2)(q_1^2 + \ldots + q_n^2)} \\ (-q_1p_1 - \ldots - q_np_n)^2 &\leq (p_1^2 + \ldots + p_n^2)(q_1^2 + \ldots + q_n^2) \end{aligned}$$

The last equation above is the Cauchy-Schwarz inequality from Math 110. Steps are reversible so we have proved the triangle inequality holds when $\vec{s} = \vec{0}$. Now check that the Euclidean metric is invariant under translations, that is $d(\vec{x},\vec{y}) = d(\vec{x}-\vec{s},\vec{y}-\vec{s})$. With this, we have

$$d(\vec{p},\vec{q}) = d(\vec{p}-\vec{s},\vec{q}-\vec{s}) \leq d(\vec{p}-\vec{s},\vec{s}-\vec{s}) + d(\vec{s}-\vec{s},\vec{q}-\vec{s}) = d(\vec{p},\vec{s}) + d(\vec{s},\vec{q}).$$

Example 2.4. Show that $(\mathbf{R}^+, d(x,y) = |\log(y/x)|)$ is a metric space.

We need to show that

$$|\log(z/x)| \leq |\log(y/x)| + |\log(z/y)|$$

Using the properties of logarithms, this is the same as

$$|\log(z) - \log(x)| \leq |\log(y) - \log(x)| + |\log(z) - \log(y)|.$$

If we use the following substitution

$$\begin{aligned} z' &= \log(z) \\ x' &= \log(x) \\ y' &= \log(y) \end{aligned}$$

then we see that this is the same as Example 1, which we have shown is a metric space. So, $(\mathbf{R}^+, d(x,y) = |\log(y/x)|)$ is a metric space.

Example 2.5 (discrete metric)**.** Show $(X, d(x,y))$ is a metric space, where X is any set and

$$d(x,y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{otherwise} \end{cases}$$

Again, we need to show that the Triangle Inequality holds. Let's look closer at this distance formula. The distance between two points is 0 if the two points are the same and 1 if the two points are different. Knowing this, we'll prove the Triangle Inequality holds via cases.

$$d(x,z) \leq d(x,y) + d(y,z)$$

Case 1: When $x = z$,

$$\begin{aligned} d(z,z) &\leq d(z,y) + d(y,z) \\ 0 &\leq 2d(z,y) \end{aligned}$$

This inequality holds whether $y = z$ or $y \neq z$.

Case 2: When $x \neq z$,

$$1 \leq d(x,y) + d(y,z)$$

If $x = y$, then $y \neq z$, so

$$\begin{aligned} 1 &\leq 0 + 1 \\ 1 &\leq 1 \end{aligned}$$

If $x \neq y$,

$$1 \leq 1 + d(y,z)$$

This inequality holds whether $y = z$ or $y \neq z$.

Since both Case 1 and Case 2 holds, the Triangle Inequality holds, so $(X, d(x,y))$ is a metric space.

Example 2.6 (Hamming distance)**.** Let $X = \{\vec{x} = (x_1, ..., x_n) \mid x_i \in \mathbf{R}\}$ be the set of sequences of length n. We claim that the following definition makes X into a metric space.

$$d(\vec{x}, \vec{y}) = \#\{n \in \mathbf{N} |\ x_n \neq y_n\}$$

We want to show the triangle inequality, that is, $d(\vec{x}, \vec{z}) \leq d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z})$. It suffices to show that if $x_k \neq z_k$ then we must have $x_k \neq y_k$ or $y_k \neq z_k$. Suppose otherwise. Then we have $x_k = y_k$ and $y_k = z_k$ implying $x_k = z_k$ which is a contradiction.

This metric space shows up in computer science in error-correcting codes where a message transmitted over a noisy channel can be recovered if the ”distance” between transmitted message and original message is small and in biology when detecting mutations in genomes.

Homework 2.7. Suppose d_1, d_2 are metrics on a space X. Show that the following are also metrics.

1. $d_1 + d_2$
2. $\max\{d_1, d_2\}$
3. $\min\{d_1, d_2\}$
4. $d_1 d_2$
5. $\frac{1}{3}d_1 + \frac{2}{3}d_2$

Homework 2.8. For $p, q \in [0, \pi/2)$ let

$$d(p, q) = \sin |p - q|.$$

Is d a metric on $[0, \pi/2)$?

Homework 2.9. Show that if (X, d) is a metric space, and $Y \subseteq X$, then (Y, d) is a metric. (When d is viewed as a metric on Y, it is called the subspace metric or the metric induced by X.)

Homework 2.10. If (X, d) and (Y, e) are metric spaces, we say that they are *isometric* if there is a bijection $f : X \to Y$ such that for each $x_1, x_2 \in X$, $e(f(x_1), f(x_2)) = d(x_1, x_2)$. (Then we think of them as the “same metric space”).

Recall that $\mathbf{R}$ is a metric space, with $d(x, y) = |x - y|$. Is $[0, 1]$ isometric to $[0, 2]$ (with the subspace metric induced by $\mathbf{R}$)? Is $[0, 1]$ isometric to $[100, 101]$? Prove it.

2.2 Open and closed sets

In high school, you learned about open and closed intervals, which were intervals (a, b) and $[a, b]$ in $\mathbf{R}$ (so that open intervals didn’t contain their endpoints, but closed intervals did).

Now we’ll generalize this notion to metric spaces. We could just as well have thought of the open interval $(0, 1)$ as the set of all real numbers strictly within $1/2$ of $1/2$, and the closed interval $[0, 1]$ as the set of real numbers within or equal to $1/2$ of $1/2$.

Our development of the theory of open and closed sets follows Rudin’s “Principles of Mathematical Analysis,” Chapter 2.

Definition 2.11 (Neighborhoods)**.** The *neighborhood* or *open ball* of a point x of radius $r > 0$ in a metric space is

$$N_r(x) := \{y \in X \mid d(x,y) < r\}$$

The *closed ball* of x and r is

$$\overline{N_r(x)} := \{y \in X \mid d(x,y) \leq r\}$$

It's often helpful to think of a metric space as a horribly twisted version of (a subset of) $\mathbf{R}^2$, so that it can be drawn on a piece of paper. Then, a ball $N_r(x)$ is literally a circle of radius r, centered on a point x.

Definition 2.12 (Limit Points)**.** Let $E \subseteq X$, where X is a metric space. Then, $p \in X$ is called a *limit point* for the set E if for all neighborhoods N at p, N contains a point $q \in E$ such that $q \neq p$.

Example 2.13. Let $E = \left\{\frac{1}{n} \mid n \in \mathbf{Z}^+\right\}$. Show that 0 is a limit point.

Let $N_r(0)$ be a neighborhood at 0. Then

$$d\left(0, \frac{1}{n}\right) = \frac{1}{n}$$
$$\frac{1}{n} < r$$

From this, we see that if $n > \frac{1}{r}$, then $d(0, \frac{1}{n}) < r$ so $\frac{1}{n} \in N_r(0)$.

What does it mean for a point p to not be a limit point of the set E? If we negate the definition, then for a point p to not be a limit point of the set E, there exist a neighborhood about p such that the neighborhood does not contain a point q in E where $q \neq p$.

Theorem 2.14. *If p is a limit point of a set E, then every neighborhood of p contains infinitely many points in E.*

Proof. Assume that p is a limit point of E, but there exists a neighborhood $N^*(p)$ such that $N^*(p)$ contains finitely many points in E. Now consider $N_r^*(p)$, where r is less than the shortest distance between p and the finitely many points in E contained in $N^*(p)$. By construction, $N_r^*(p)$ would contain no points in E, but this contradicts the definition of a p being a limit point. Therefore, our assumption is false and every neighborhood of p contains infinitely many points in E. □

Theorem 2.15. *A finite set has no limit points.*

Proof. The proof comes directly from the previous theorem. Suppose p is a limit point of E. This implies that $N_r(p)$ contains infinitely many points of E. However, E is finite, so p cannot be a limit point of E. □

Definition 2.16 (Interior points)**.** Let $E \subseteq X$, where X is a metric space. Then $p \in X$ is called an *interior point* if there exists a neighborhood N of p such that $N \subseteq E$.

Interior points of E are necessarily in E, since $x \in N \subseteq E$.

Definition 2.17 (Open sets)**.** A set E is open in X if every point in E is an interior point of E.

Continuing the metaphor with $\mathbf{R}$, open sets are those which do not contain their own boundary (just as the open interval $(0, 1)$ does not contain its boundary points 0 and 1).

Theorem 2.18. *Every open ball $N_r(p)$ in a metric space is an open set.*

Proof. Let $x \in N_r(p)$. We want to show that x is an interior point. We claim that $N_s(x) \leq N_r(p)$, where $s = r - d(p, x)$. Take $z \in N_s(x)$. This implies $d(z, x) < s$. We now want to show $d(p, z) < r$. So,

$$\begin{aligned} d(p, z) &< d(p, x) + d(x, z) \\ &< d(p, x) + r - d(p, x) \\ &< r. \end{aligned}$$

□

Homework 2.19. Let $E \subseteq X$, define $E^\circ = \{x \in E : x \text{ is an interior point}\}$. Show E° is an open set in X.

Homework 2.20. Prove $E \subseteq X$ is open if and only if $E^\circ = E$.

Definition 2.21 (closed set)**.** A set $S \subseteq X$ is *closed* if S contains all its limit points; that is, if $p \in X$ is a limit point of the set S, then $p \in S$. We denote $S^* = \{\text{limit points of } S\}$

Closed sets are dual to open sets in the sense that closed sets "contain their own boundary," just as the closed interval $[0, 1]$ contains its boundary points 0 and 1. However, most sets are neither open nor closed, but rather "ajar". For example, $[0, 1)$ contains some of its boundary points, but not all of them. On the other hand, the empty set and the entire metric space are both open and closed, because neither have a boundary at all.

Proposition 2.22. *The* closure $\overline{S} = S \cup S^*$ *is a closed set.*

Before we begin the proof, there is a confusing matter we should resolve. A set is closed if it contains it's limit points and $\overline{S}$ is constructed by adding in all limit points of S to S, so aren't we done? Well no, because we want to show that the set $\overline{S}$ is closed, not S. In the process of adding in the limit points of S we possibly created new points to be "close" to our set and these points *a priori* may not be in $\overline{S}$. However, we will show that this is the case.

Proof. Let z be a limit point of $\overline{S}$. We want to show that $z \in S \cup S^*$.

Let $N_r(z)$ be a neighborhood of z. By definition $\exists\, a_r \in N_r(z)$ such that $a_r \in \overline{S}$. If $a_r \in S$ then $N_r(z)$ contains a point of S. Otherwise, $a_r \in S^*$ and is a limit point of S. However, if a_r is a limit point of S then *for any* neighborhood about a_r, $\exists q \in N$ such that $q \in S$. We notice that because $a_r \in N_r(z)$, and neighborhoods are open sets, then $\exists N$ completely contained in $N_r(z)$. We now have a $q \in S$ s.t. $q \in N_r(z)$, aka z is a limit point of S. □

Corollary 2.23. $\overline{\overline{S}} = \overline{S}$.

Recall that a standard trick to prove "P if and only if Q" is to prove both directions separately. Here we'll do the same thing: to prove an equality of sets, *prove both are subsets of each other.*

Proof. First, $\overline{S} \subseteq \overline{\overline{S}}$ by definition.

Also, $\overline{\overline{S}} \subseteq \overline{S}$. Indeed, take $z \in \overline{S}^*$, so that z is a limit point of $\overline{S}$. Therefore, $z \in \overline{S}$ by 2.22.

Therefore $\overline{\overline{S}} = \overline{S}$. □

Theorem 2.24. *If F is a closed set and $S \subset F$ then $\overline{S} \subset F$.*

Proof. Take $z \in \overline{S}$. Either $z \in S$ or $z \in S^*$.

If $z \in S$, then $S \subset F$ implies $z \in F$.

If $z \in S^*$ then z is a limit point of S, meaning it is also a limit point of F. However, since F is closed, this must imply that $z \in F$. □

Since closed sets are "those which contain their boundary" and open sets are "those which do not contain their boundary", it makes sense that if E is an open set, all of its boundary must be in its complement, $X \setminus E$ – and since the boundary of E should also be the boundary of $X \setminus E$, it follows that $X \setminus E$ contains its boundary and is closed.

Let's prove this.

Theorem 2.25. *$E \subseteq X$ is open iff its complement $E^c = X \setminus E$ is closed.*

Proof. Let E be open and z be a limit point of E^c. We want to show $z \in E^c$. Assume $z \notin E^c$, so $z \in E$ which is an open set. Therefore, there is some $r' > 0$ such that $N_{r'}(z) \in E$. This means that $N_r'(z)$ contains no points of E^c, which contradicts our original assumption that z is a limit point.

For the other direction, suppose E^c is closed but E is not open. So $\exists p \in E$ such that $\forall r > 0$, $N_r(p) \not\subset E$. This means that if $r > 0$ then $N_r(p)$ has elements in E^c. Thus p is a limit point of E^c, but E^c is closed and therefore $p \in E^c$, which is a contradiction to the fact that $p \in E$. □

Here's some facts that might be somewhat surprising.

Homework 2.26. Recall that $\mathbf{R}^2$ is a metric space, with the "euclidean" or "ℓ^2" metric $d((x_1, x_2), (y_1, y_2)) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$.

Show that $\mathbf{R}$ with its usual metric is isometric to $\mathbf{R}$ with the metric induced by $\mathbf{R}^2$ (so $\mathbf{R}$ is the "same metric space" whether we think of it as a metric space in its own right or just as a line in the plane).

But show that $(0, 1)$ is an open set when we think of it as a subset of $\mathbf{R}$, but *not* an open set when we think of it as a subset of $\mathbf{R}^2$.

Show also that $(0, 1)$ is a closed set when we think of it as a subset of itself (with the metric induced by $\mathbf{R}$), but not when we think of it as a subset of $\mathbf{R}$ or $\mathbf{R}^2$.

Homework 2.27. Since $\mathbf{R}$ is a metric space, the set of all rational numbers $\mathbf{Q}$ is also a metric space (with the metric induced by $\mathbf{R}$). But $\mathbf{Q}$ is rather badly behaved, both as a metric space and as a set in $\mathbf{R}$.

If you don't know it already, first show that if $x < y$ are real numbers, then there exists $q \in \mathbf{Q}$ such that $x < q < y$. (Because this happens, we say that $\mathbf{Q}$ is a *dense subset* of $\mathbf{R}$. But beware – the word "dense" has a few other meanings in topology and analysis!) This will make the rest of the problem much easier.

Show that $\mathbf{Q}$ is neither closed nor open when we view it as a subset of $\mathbf{R}$.

But also show that if $U = (-\sqrt{2}, \sqrt{2})$, then $U \cap \mathbf{Q}$ is both closed and open when we think of it as a subset of $\mathbf{Q}$. (Hint: If $K = [-\sqrt{2}, \sqrt{2}]$, then $K \cap \mathbf{Q}$ is both closed and open. Why?)

2.3 Continuity

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. Worse, it doesn't seem like you could generalize this well to arbitrary metric spaces.

But this isn't so hard to fix. 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."

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 $P(x)$.")

Definition 2.28 (continuity)**.** Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is *continuous* at a point $x_1 \in X$ if $\forall \varepsilon > 0$, $\exists \delta > 0$ such that $\forall x_2 \in X$, if $d_X(x_1, x_2) < \delta$, then $d_Y(f(x_1), f(x_2)) < \varepsilon$.

The game of continuity goes like this. If we want our function to be continuous at a given point, we fix an ε – it doesn't matter how small. For this choice of ε, we need to find the corresponding δ such that the above condition holds. In other words, our δ is a function of ε.

To make this easier to remember, here is a "training wheels" definition for continuity. A function $f : X \to Y$ is continuous at a point $x_1 \in X$ if $\forall \varepsilon > 0 \ \exists \delta_\varepsilon > 0$ such that $\forall x_2$,

$$d_X(x_1, x_2) < \delta_\varepsilon \implies d_Y(f(x_1), f(x_2)) < \varepsilon.$$

Example 2.29. In a metric space X, the identity function $id : X \to X$, $id(x) = x \ \forall x \in X$ and constant functions $c_y : X \to Y$, $c_y(x) = y \ \forall x \in X$ are always continuous at all points $x \in X$.
For the identity function, fix $\varepsilon > 0$ and some point $p \in X$. Then,

$$\begin{aligned} d_x(id(x), id(p)) &< \varepsilon \\ d_x(x, p) &< \varepsilon \end{aligned}$$

Thus $\forall \varepsilon > 0$, let $\delta = \varepsilon$. Then the above calculation implies

$$d_x(x,p) < \delta \implies d_x(id(x), id(p)) < \varepsilon$$

Therefore, $id(x)$ is continuous $\forall x \in X$.

For the constant function, fix $\varepsilon > 0$ and some point $p \in X$. Then,

$$\begin{aligned} d_x(c_y(x), c_y(p)) &< \varepsilon \\ d_x(y,y) &< \varepsilon \\ \varepsilon &> 0 \end{aligned}$$

Let $\delta > 0$. Then $\forall \varepsilon > 0\ \exists \delta > 0$ such that

$$d_x(x,p) < \delta \implies d_x(c_y(x), c_y(p)) < \varepsilon$$

Therefore, $c_y(x)$ is continuous $\forall x \in X$.

Another useful characterization of continuity is in terms of sequences.

Definition 2.30 (sequences)**.** Let X be a metric space. 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 2.31 (convergence)**.** A sequence of points (x_n) converges to a point p in a metric space if $\forall \varepsilon > 0\ \exists N \in \mathbf{N} > 0$ such that $\forall n > N$

$$d(x_n, p) < \varepsilon.$$

We write this as $x_n \to p$ or $\lim_{n\to\infty} x_n = p$.

The limit $\lim_{n\to\infty} x_n = p$ implies p is a limit point of the set $\{x_n \mid n \in \mathbf{N}\}$. Indeed, consider $N_r(p)$. Let $\varepsilon = r$. Then $\exists N > 0$ such that $\forall n \geq N$,

$$d(x_n, p) < r \implies x_n \in N_r(p).$$

Theorem 2.32. *A function $f : X \to Y$ is continuous at p if and only if for all sequences $x_n \to p$,*

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

Proof. The "training wheels" definition of continuity will help us in this proof as there are essentially three ε's floating around. We want to first show that a function continuous at a point p implies that for all sequences $x_n \to p$, we have $f(x_n) \to f(p)$. In other words,

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

$$\forall \varepsilon_{conv(p)} > 0\ \exists N_{\varepsilon_{conv(p)}} > 0 \text{ such that } \forall n \geq N_{\varepsilon_{conv(p)}},\ d(x_n, p) < \varepsilon_{conv(p)}.$$

2. We know f is continuous at p, meaning

$$\forall \varepsilon > 0\ \exists \delta_\varepsilon > 0 \text{ such that } \forall x \in X,\ d(x,p) < \delta_\varepsilon \implies d(f(x), f(p)) < \varepsilon.$$

3. We want to show $\lim_{n\to\infty} f(x_n) = f(p)$, meaning

$$\forall \varepsilon > 0 \; \exists N_\varepsilon > 0 \text{ such that } \forall n \geq N_\varepsilon, \; (f(x_n), f(p)) < \varepsilon.$$

From (2) we see that if $d(x_n, p) < \delta_\varepsilon$, then $d(f(x_n), f(x)) < \varepsilon$. From (1), this is possible by letting $\varepsilon_{conv(p)} = \delta_\varepsilon$. Then, we get a N_{δ_ε} such that $\forall n \geq N_{\delta_\varepsilon}$, $d(x_n, p) < \delta_\varepsilon$. But now (2) implies $d(f(x_n), f(p)) < \varepsilon$. So, we have produced a N_ε, namely N_{δ_ε} such that $\forall n \geq N_{\delta_\varepsilon}$, $d(f(x_n), f(p)) < \varepsilon$, meaning $f(x_n) \to f(p)$.

For the converse, we'll proceed by contradiction. Assume that $f : X \to Y$ is not continuous at p. That is to say, $\exists \varepsilon > 0 \; \forall \delta_\varepsilon > 0 \; \exists x$ such that

$$d(x, p) < \delta_\varepsilon \text{ and } d(f(x), f(p)) \geq \varepsilon.$$

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

Theorem 2.33 (Intermediate Value Theorem). *Let $f : [a, b] \to \mathbf{R}$ be continuous. Then $\forall y \in [f(a), f(b)] \; \exists x \in [a, b]$ such that $f(x) = y$.*

We won't prove the intermediate value theorem, as it uses some properties of $\mathbf{R}$ (namely the so-called "least-upper-bound axiom" or "completeness axiom") which we won't be discussing in MUSA 74. It's pictorially not hard to see, in any case.

The set of all rational numbers is written $\mathbf{Q}$. It turns out that *between every pair of real numbers is a rational number* (if $a < b$, just truncate the decimal expansion of b, late enough to be $> a$, to get a rational number in $(a, b]$) – and *between every pair of rational numbers is an irrational number* (if $a < b$, add a nonrepeating sequence of digits to a, chosen $\leq b$, to get an irrational number in $(a, b]$). Because of this, $\mathbf{Q}$ can govern the behavior of continuous functions $\mathbf{R} \to \mathbf{R}$.

Lemma 2.34. *If $f : \mathbf{R} \to \mathbf{R}$ is continuous and $\forall x \in \mathbf{R} \; f(x) \in \mathbf{Q}$, then f is constant.*

Proof. Proceed by contradiction.

Suppose f is not constant. Then that means that $\exists\, x, y$ such that $f(x) < f(y)$. Furthermore, since $f(x), f(y) \; \in \mathbf{Q}$ and $f(x)$ is continuous, there must $\exists\, c \in \mathbf{R} - \mathbf{Q}$ such that $f(x) < c < f(y)$.

By the IVT, that means $\exists\, x_1 \in [x, y]$ such that $f(x_1) = c$. But $\forall x$, $f(x) \in \mathbf{Q}$. Thus, there is a contradiction: $\nexists\, x, y$ such that $f(x) < f(y)$. □

Lemma 2.35. *Let $f : [0, 2] \to \mathbf{R}$ be continuous with $f(0) = f(2)$. Then $\exists x\, , y \in [0, 2]$ with $|x - y| = 1$ such that $f(x) = f(y)$.*

Proof. Without loss of generality, let $y = x - 1$ and let $g(x) = f(x) - f(x - 1)$.

Then

$$g(1) = f(1) - f(0)$$
$$g(2) = f(2) - f(1) \implies g(2) = f(0) - f(1)$$

since f(0) = f(2). Therefore

$$\begin{aligned} g(1) + g(2) &= f(1) - f(0) + f(0) - f(1) \\ &= 0 \end{aligned}$$

so that $g(1) = -g(2)$. Thus, without loss of generality, assume $g(1) < 0 < g(2)$ so by the IVT, $\exists x \in [1, 2]$ such that $g(x) = 0$.

Thus, since $g(x) = f(x) - f(x-1)$,

$$\begin{aligned} 0 &= f(x) - f(x-1) \\ f(x) &= f(x-1) \\ f(x) &= f(y) \end{aligned}$$

□

where $y = x - 1$.

Proposition 2.36. *If $f : X \to Y$ and $g : Y \to Z$ 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 2.37. 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 2.38. 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.

Homework 2.39. 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 2.40. Consider $\mathbf{R}$ as a metric space with its usual metric. Show that for all n, $f(x) = x^n$ is continuous.

Homework 2.41. Consider $\mathbf{R}$ as a metric space with its usual metric. Show that if f and g are continuous, $f + g$ is continuous as well.

Homework 2.42. Consider $\mathbf{R}$ as a metric space with its usual metric. Show that if f is a polynomial, then f is continuous. (Hint: use 2.40 and 2.41, and induction).

2.4 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 2.43. 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

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

Homework 2.44. Show that if f is continuous at x_0, then $\lim_{x\to x_0} f(x) = f(x_0)$.

We're now ready to define the derivative formally.

Definition 2.45 (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.

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 2.46. *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. □

Homework 2.47. Does the converse of the previous proposition hold? If so, give a proof. If not, give a counterexample.

Chapter 3

Math 113: Algebra

In Math 113, one considers three kinds of algebraic structures:

1. *Groups*, which are the collection of symmetries on some object. The classic example of a group is the set of all permutations (i.e. ordered ways of choosing) n objects.

2. *Rings*, which are structures with an addition, subtraction, and multiplication operation. The classic example is the set of all integers $\mathbf{Z}$.

3. *Fields*, which are rings that also have division. The classic example is the set of all rational numbers $\mathbf{Q}$.

We'll focus in on groups.

3.1 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 3.83.

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 3.1 (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 3.2. 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 3.3. 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 3.4. 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 3.5. 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 3.6. Let us show that $(\mathrm{GL}(n, \mathbf{R}), *)$, 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: $A * I = A \rightarrow I$ is the identity.

3. Inverse: $A * A^{-1} = I \rightarrow A^{-1}$ is the inverse.

Furthermore, $\mathrm{GL}(n, \mathbf{R})$ 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 3.7. 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 3.8. *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 $e_g' * x = x$.

$$e_g * x = e_g' * x$$
$$(e_g * x) * x^{-1} = (e_g' * x) * x^{-1}$$
$$e_g = e_g * (x * x^{-1}) = e_g' * (x * x^{-1}) = e_g'$$

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 3.9. 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 3.10. 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.

[1] We will learn what this means later on.

Homework 3.11. Show that ($\mathbf{R} \setminus \{-1\}$, $a * b = a + b + ab$) is a group.

Homework 3.12. Show that ($\mathbf{R}^{\geq 0}$, $a * b = \sqrt{ab}$) is not a group.

Homework 3.13. Show that ($\mathbf{Z} \setminus \{0\}, \times$) is not a group.

Homework 3.14. 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.

Homework 3.15. 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 3.16. 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 $x^n = e$.

3.2 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 3.17 (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 3.18. $\pi : \mathbf{Z} \to \mathbf{Z}_5$, 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 3.19. $L_c : (\mathbf{R}, +) \to (\mathbf{R}, +)$, $L_c(x) = cx$, $c \in \mathbf{R}$

Proof.

$$L_c(a + b) = c(a + b) = ca + cb = L_c(a) + L_c(b).$$

□

Example 3.20. 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 3.21. $\Phi : (\mathbf{R}, +) \to (\mathbf{R}^+, \times)$, $\Phi(x) = e^x$.

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 3.22 (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 3.23. The map Φ defined in Example 3.21 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 3.24. *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. $\forall n \in \mathbf{Z}$, $x \in G$, $f(x^n) = f(x)^n$

[2] in Math 104, log will mean logarithm base e

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 $e_G = x * x^{-1}\,\forall\, x \in G$, $f(e_G) = f(x * x^{-1}) = f(x)f(x^{-1}) = f(x)f(x)^{-1} = e_H$.

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 3.25. 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 3.26. *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 $z_1 z_2 = f(x_1)f(x_2) = f(x_1 * x_2) = f(x_2 * x_1) = f(x_2)f(x_1) = z_2 z_1$ □

Homework 3.27. 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 3.28. 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 3.29. Prove that $\Phi : G \to G$ where $\Phi(g) = g^2$ is a homomorphism if and only if G is abelian.

3.3 Subgroups

Definition 3.30. 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 3.31.

$$(\mathbf{Z}, +) \subseteq (\mathbf{Q}, +) \subseteq (\mathbf{R}, +) \subseteq (\mathbf{C}, +)$$

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 3.32. 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 $\{0, d, 2d, \dots (\frac{n}{d} - 1)\, d\}$ will always be a subgroup of $\mathbf{Z}_n$ by pretty much the same analysis above.

Example 3.33. G and $\{e\}$ are always subgroups of G.

Proposition 3.34 (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 3.35. *Let $f : G \to H$ be a group homomorphism. Then $\ker(f) = \{x \in G \mid f(x) = e_H\}$ is a subgroup of G and $\mathrm{Im}(f) = \{y \in H \mid \exists\, x \in G \text{ s.t. } f(x) = y\}$ 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 $y \in \mathrm{Im}(f)$. By definition there exists $x \in G$ s.t. $f(x) = y$. Then $y^{-1} = f(x)^{-1} = f(x^{-1})$ so that $x^{-1} \in \mathrm{Im}(f)$. Now suppose $y_1, y_2 \in \mathrm{Im}(f)$. Then we have x_1, x_2 s.t. $f(x_1) = y_1$ and $f(x_2) = y_2$. Then $y_1 y_2 = f(x_1)f(x_2) = f(x_1 x_2)$ and thus $y_1 y_2 \in \mathrm{Im}(f)$. □

Lemma 3.36. *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 3.37. 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 $\mathrm{Im}(f)$. But $\mathrm{Im}(f)$ is a subgroup of H! Moreover, from above we know that $e_H \in \mathrm{Im}(f)$. But recall from earlier that we showed the identity of a group is unique. Applying this to the subgroup $\mathrm{Im}(f)$, we see that $f(e_G) = e_H$. Summarizing,

1. We first showed that $f(e_G)$ was the identity element of $\mathrm{Im}(f)$.
2. $\mathrm{Im}(f)$ 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 3.38. Show $\{z \in \mathbf{C} | \, |z| = 1\}$ is a subgroup of $(\mathbf{C}^\times, \cdot)$.

Homework 3.39. Show $(2\mathbf{Z}, +) \subseteq (\mathbf{Z}, +)$.

Homework 3.40. Show $(\mathrm{GL}(n, \mathbf{R}), +)$ is NOT a subgroup of $(M_n(\mathbf{R}), +)$.

Homework 3.41. Find the kernel of the map $\phi : \mathbf{C}^\times \to \mathbf{R}^\times$ where $\phi(a + bi) = a^2 + b^2$.

3.4 Generators and Relations

Definition 3.42. Given a set of elements $A = \{a_1, \ldots, a_k\}$ of a group G, the subgroup generated by A denoted by $\langle A \rangle$ is defined to be

$$\langle A \,\rangle := \{a_{i_1}^{\varepsilon_1} \ldots a_{i_k}^{\varepsilon_k} | \; a_{ij} \in A, \varepsilon_i = \pm 1\}$$

If $\langle A \,\rangle = G$, then we say that G is *generated* by the elements $a_1, \ldots, a_k$.

Definition 3.43. 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 $\{a_1, \ldots, a_k, a_1^{-1}, \ldots, a_k^{-1}\}$.

Example 3.44. 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 3.45. $(\mathbf{Z}, +)$ is generated by 1, as any positive integer n can be written as $n = 1 + \cdots + 1$ and any negative integer can be written as $-n = -1 - \cdots - 1$.

Example 3.46. $\mathbf{Z}_6$ 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 3.47. *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 3.48. 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 3.49. 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 $\{a_1, \dots, a_k\}$ generate G. The real bijection now reads,

$$\{f : G \to H |\ f \text{ is a group homomorphism}\} \longleftrightarrow \{\text{relations preserving values } f(a_i) \in H\}$$

[3] For example, in example 3.46, you can see that we expressed the element 5 in two different ways.

Example 3.50. 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 3.51. 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 $\mathbf{Z}_9$.

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 3.52. 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 3.53. 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 3.54. Check that $\langle A \rangle$ is a subgroup.

Homework 3.55. Show $(\mathbf{Q}^{>0}, \cdot)$ is generated by $\left\{ \frac{1}{p} | \, p \text{ is a prime} \right\}$.

Homework 3.56. Suppose H_1, H_2 are subgroups of G. Show that $H_1 \bigcap H_2$ is also a subgroup of G.

Homework 3.57. 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.

3.5 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 3.58 (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 3.59. $G = \mathbf{Z}_8 = \{0, 1, 2, 3, 4, 5, 6, 7\}$ 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 3.60. $g_1H = g_2H \iff g_2^{-1}g_1 \in H$

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 3.61. Prove the analogous statement: $Hg_1 = Hg_2 \iff g_2g_1^{-1} \in H$

Lemma 3.62. *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 3.63. *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 3.64. $|H| = |gH|$

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 3.65 (Lagrange)**.** *Let H be a subgroup of a finite group G. Then the size of H divides the size of G.*

Proof. By lemma 3.62, $g_1 \sim g_2$ is an equivalence relation. By lemma 3.63, $G = \bigsqcup_{g \in G} S_g$ and using lemma 3.60 we have $S_g = gH$ and $|gH| = |H|$ by lemma 3.64. 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 3.66. *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 3.67. Find all cosets of $4\mathbf{Z}$ in $\mathbf{Z}$.

Homework 3.68. Suppose H, K are finite subgroups of G s.t. $(|H|, |K|) = 1$. Prove that $H \cap K = \{e\}$

Proposition 3.69. *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 3.70. $a^{p-1} \equiv 1 \pmod 9$ *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 3.71. A *normal subgroup* $N \subseteq G$ is a subgroup such that $\forall n \in G, \forall g \in G, gng^{-1} \in N$.

Example 3.72. $\{e\}$ and G are always normal subgroups.

Proof.

1. $geg^{-1} = gg^{-1} = e \in \{e\}$
2. G is normal because it is closed.

□

Example 3.73. 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 3.74. *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 $g_1 n_1 \in gN$, we know $g_1 n_1 g_1^{-1} = n_2 \in N$. This implies $g_1 n_1 = n_2 g_1 \in Ng$, so $gN \subseteq Ng$.
Now take a $n_1 g_1 \in Ng$. We know $g_1 n_1 g_1^{-1} = n_2 \in N$. This implies $g_1 n_1 = n_2 g_1 \in gN$, 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 3.75. The *center* of a group G, written $Z(G)$, is the set

$$Z(G) = \{z \in G : \forall h \in G \; zh = hz\}.$$

Example 3.76. 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 3.77. 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 3.78. 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 3.79. 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 3.80. *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 3.81. *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 3.82. 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 3.83. $\mathbf{Z}/n\mathbf{Z} \cong (\mathbf{Z}_n, +)$

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:

$$\Phi((g_1 + g_2)n\mathbf{Z}) = \Phi(g_1 n\mathbf{Z}) + \Phi(g_2 n\mathbf{Z})$$
$$(g_1 + g_2) \pmod{n} = g_1 \pmod{n} + g_2 \pmod{n}$$

To show $\mathbf{Z}/n\mathbf{Z}$ is $1 - 1$, it suffices to show $\ker \Phi = \{n\mathbf{Z}\} = e$. Indeed, $\Phi(g(n\mathbf{Z})) = 0$ (mod 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 (mod n), $\Phi(g(n\mathbf{Z})) = g$ (mod n). □

Lemma 3.84. *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 3.85. *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 3.86. 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 3.87. 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 3.88. *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 3.89 (first isomorphism theorem)**.** *If $f : G \to H$ is a homomorphism, then $G/\ker(f) \cong \mathrm{Im}(f)$.*

Proof...? Let $\Phi : G/\ker(f) \to \mathrm{Im}(f)$ 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 $y \in \mathrm{Im}(f)$, there exists a $x \in G/\ker(f)$ such that $\Phi(x) = y$.

So thre is indeed an isomorphism between $G/\ker(f)$ and $\mathrm{Im}(f)$. □

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 3.90. 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 3.91. *$y \in xH$ 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 3.92. *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 3.89, 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 3.93. 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 $\mathrm{Im}(f) = \mathbf{Z}/4\mathbf{Z}$. We observe that $\ker(f) = \{0, 4\} = H$ and $\mathrm{Im}(f) = \mathbf{Z}/4\mathbf{Z}$. □

Homework 3.94. Suppose N is a normal subgroup of G and $H \subseteq G$. Show $N \cap H$ is normal in H.

Example 3.95. In linear algebra, if $T : V \to V$ is a linear map, then the *rank theorem* says $\dim(V) = \dim(\ker(T)) + \dim(\mathrm{Im}(T))$. This actually follows from the first isomorphism theorem and Homework 3.86.

Proof. We know that $V/\ker(T) \cong \mathrm{Im}(T)$, so in fact

$$\dim(V/\ker(T)) = \dim(\mathrm{Im}(T)).$$

But as you proved on the homework,

$$\dim(V/W) = \dim V - \dim W$$

so

$$\dim V = \dim\ker T + \dim \mathrm{Im}\, T.$$

□

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 3.96. 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. □

Chapter 4

Conclusion: The Fourier transform

Let's wrap up the class by giving a cute example of the interplay among the theories developed in Math 110, Math 113, Math 104, and Math 185, which has far-reaching applications throughout the sciences, and might even pop up in computer science classes. This application will give some examples of metrics and groups that arise naturally in the real world.

The theorems in this section will be left unproven, since it's the end of the class and you can in principle prove them yourself. Some of them are quite difficult, though, so while attempting them is good practice you shouldn't worry if you can't solve them all. (The ones which are definitely too hard are denoted Theorem rather than Homework. In any case, we won't be assigning them as problems, but if you wanted to work on them and ask us for feedback on your solutions we'd be happy to provide.)

4.1 The classical 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 4.1 (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 4.2. 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 4.3 (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 4.4. 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 4.5 (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 4.6. *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

$$f(x) = \xi\sum_{n\in\mathbf{Z}} e^{2\pi n\xi ix}\int_{-P/2}^{P/2} f(y)e^{-2\pi in\xi y}\,dy.$$

Since $y \mapsto e^{2\pi in\xi y}$ 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 4.7 (Fourier's inversion theorem)**.** *Let $f \in L^2$. If all the integrals in the below equation are finite, then*

$$f(x) = \int_{-\infty}^{\infty} e^{2\pi i\xi x} \int_{-\infty}^{\infty} f(y)e^{-2\pi iy\xi}\, dy\, d\xi.$$

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 4.8.

$$\widehat{f'}(\xi) = 2\pi i\xi\hat{f}(\xi).$$

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

$$x \mapsto \int_{-\infty}^{\infty} \chi(\xi)\hat{f}(\xi)e^{2\pi ix\xi}\, d\xi = \int_{89.3-\varepsilon}^{89.3+\varepsilon} \hat{f}(\xi)e^{2\pi ix\xi}\, d\xi$$

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 4.9 (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.

4.2 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 4.10. 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 4.11. 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 4.12. *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 4.13 (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 4.12 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 4.14 (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 4.15. 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 4.16. 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 4.17. 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