MATH 355 LECTURE NOTES
CHAPTER 0: COUNTABLE SETS

KAMERYN J. WILLIAMS

1. Introduction

Set theory was birthed with a lemma in an 1874 article about transcendental numbers, those numbers which cannot be obtained from purely algebraic operations.[1] Building on his earlier ideas developed to study trigonometric series, German mathematician Georg Cantor (1845–1918) came upon a germ of an idea that would eventually grow to revolutionize mathematics. He would spend the rest of his career elaborating on this idea to found the branch of mathematics known as set theory. That is the topic of this course.

Set theory is the study of the mathematical concept of **set**, defined by Cantor as a multiplicity considered as a unity, with the emphasis on infinite sets. Despite the seeming simplicity of this concept, it is remarkably general and all of mathematics can be coded in terms of sets. In this class we will study the core concepts of Cantorian set theory—ordinals and cardinals—as well as some important concepts that were developed post-Cantor.

The definition in Cantor's 1874 lemma was that of a **countable** set, one whose elements can be listed in an infinite sequence. The lemma was that this definition is not trivial: There are sets which are not countable.

Lemma (Cantor 1874)**.** *The set $\mathbb{R}$ of real numbers is not countable.*

Proof. Let's see a proof Cantor came up with in 1891. His original proof is more complicated, and makes for a good problem set :)

Suppose toward a contradiction that there were an infinite sequence $\langle x_n : n \in \mathbb{N} \rangle$ of all real numbers. Look at the decimal expansions of these numbers.[2] We will define a new infinite sequence of digits $\langle d_n : n \in \mathbb{N} \rangle$ which will be the decimal expansion of a number y between 0 and 1. Namely, set $d_n = 5$ if the n-th digit of x_n past the decimal is 7, and set $d_n = 7$ otherwise.

Let's see that $y \neq x_n$ for any n. Suppose otherwise that $y = x_n$. Then, by construction, the n-th digit of y is 5 if and only if the n-th digit of x_n is 7. This is impossible. □

Much work in set theory can be understood as investigating how to generalize facts about countable sets to all infinite sets. Accordingly, our first task will be to understand countable sets. Once we have done that we will be ready to study ordinals and cardinals, two of the main objects of Cantorian set theory. And then we will move into ideas developed after Cantor.

Date: January 16, 2024.

[1] More precisely, a number is *algebraic* if it is a solution to a polynomial equations in rational coefficients and is *transcendental* if it is not.

[2] There is a small complication. Some numbers have two decimal expansions, not one. For example, $1.000\ldots = 0.999\ldots$. If a number has two decimal expansions, then let's only consider the one which ends in repeating 9s.

Instructions on how to read these lecture notes

Reading mathematics is an active affair. Mathematical writing is dense, and it takes time and effort to understand it. You should constantly be asking yourself questions like "why is this true?" or "do I believe this claim the author made?" or "what are the further implications of this idea?" Have paper with you that you can write on to fill in any gaps in the reasoning, work out an example, or so on. If you read passively you will find it difficult to internalize the concepts and will have a harder time with writing your own arguments.

Parts of these lecture notes will be assigned to you to read outside of class, and others we will go through in class together. Each section will end with a few exercises. These serve two purposes: to serve as a test of your understanding and to introduce yet more facts you should know. Do the exercises! All of them! They should be straightforward if you understood the section. If you are unable to solve an exercise that's telling you that you need to spend extra time on the material—reread the notes, talk to me in office hours, talk to classmates, or so on.

Separate from the exercises in different pdfs are problem sets for each chapter. These are more substantive and are what you will present in problem sessions and turn in as part of your portfolio for the class.

2. Sets

To say what a countable set is we first must say what a set is. Here is a proto-definition we will revisit later.

Proto-Definition 1. A *set* is a well-defined collection of objects, defined only by what objects are its *elements* or, synonymously, *members*. We write $x \in A$ to mean that the object x is an element of the set A.

The definition permits non-mathematical objects. It is sensible to talk, for example, about the set of people currently reading this document. Our focus, however, will be on sets of mathematical objects—numbers, points, etc.

Since sets are defined by their elements, to define a set it suffices to say exactly what its elements are. We will use curly braces as notation for this purpose. For example, $\{0, 1, 2, 3\}$ is the set whose elements are precisely the numbers 0, 1, 2, and 3. Note that sets don't come with a notion of order, since the only information is in or out. So $\{3, 2, 1, 0\}$ would be another way to write that same set.

This way of defining sets is not useful when the set has many elements, especially when it's infinite. Instead we will define a set by picking out a property which defines which objects are elements of the set. Let's see an example before the general definition.

Suppose we want to define the set of real numbers which are a rational multiple of π. There's a few ways one might write this. One would be to go purely symbolic:

$$\{x \in \mathbb{R} : \exists q \in \mathbb{Q}\ x = q\pi\}.$$

Or you might mix in natural language:

$$\{x \in \mathbb{R} : x = q\pi \text{ for some rational } q\}.$$

Either is correct, but it's considered bad style to be overly symbolic, especially with logical symbols like $\exists$.

Definition 2 (Set-builder Notation)**.** Let φ be a definite property, so that for an object x either $\varphi(x)$ holds or else $\varphi(x)$ doesn't. Then $\{x : \varphi(x)\}$ denotes the set of exactly all objects x for which $\varphi(x)$ holds. Often we're only interested in objects x from an extant set A and write $\{x \in A : \varphi(x)\}$. Another common variation is when we aren't interested in x itself but rather some other object $f(x)$ defined using x. For this case we write $\{f(x) : \varphi(x)\}$.

Here's an example of that last case. Suppose you want to define the set of integers which are odd perfect squares. One way to do this would be to write

$$\{n \in \mathbb{N} : n = k^2 \text{ for some odd } k \in \mathbb{N}\}.$$

But it's clearer to instead write

$$\{k^2 : k \in \mathbb{N} \text{ is odd}\}.$$

There will always be multiple ways to define a set. No matter how you define a set, however, what matters is what objects are elements of that set. For example, these two sets are the same: $\{n + n : n \in \mathbb{N}\}$ and $\{2n : n \in \mathbb{N}\}$ (why?).

In general, $X = Y$ is defined to mean that every element of X is an element of Y and every element of Y is an element of X. In other words, the two sets are *subsets* of each other: $X \subseteq Y$ and $Y \subseteq X$. Philosophers use the word *extensional* to refer to objects that have this property of being defined in terms of their elements. Other extensional objects in mathematics include functions—two functions are the same just in case the same inputs have the same outputs. You can have two properties which are different *intensionally* (they have different meanings) but they have the same extension (they designate exactly the same object or objects). The classic example used is the morning star Phosphorus versus the evening star Hesperus. They have different intensions but the same extension—namely, the planet Venus.

By extensionality there is a unique set with no elements. We call it the *empty set*, written $\emptyset$. It's defined by the property that $x \notin \emptyset$ no matter what x is.

Once you have multiple sets you can combine them. These basic set theoretic operations correspond to logical operations.

Definition 3. Let A and B be sets.

- The *union* of A and B is $A \cup B = \{x : x \in A \text{ or } x \in B\}$.
- The *intersection* of A and B is $A \cap B = \{x : x \in A \text{ and } x \in B\}$.
- The *difference* of A in B is $B \setminus A = \{x \in B : x \notin A\}$.[3]

There are also notions of union and intersection for more than two sets. Let $\mathcal{X}$ be a set of sets. (That's allowed!) Then:

$$\bigcap_{A \in \mathcal{X}} A = \{x : x \in A \text{ for all } A \in \mathcal{X}\}$$
$$\bigcup_{A \in \mathcal{X}} A = \{x : x \in A \text{ for some } A \in \mathcal{X}\}.$$

We will refer to these as *infinitary* intersection and union, as they are most interesting to us when $\mathcal{X}$ is infinite. Sometimes it's convenient to use the shorter $\cap\mathcal{X}$ or $\cup\mathcal{X}$ to refer to, respectively, the intersection of or union of all the sets in $\mathcal{X}$. A benefit of the more verbose notation is that it admits variations. Here's an example:

Proposition (Infinitary De Morgan's Laws)**.** *Let $\mathcal{X}$ be a set of sets and let C be a set. Then,*

$$C \setminus \bigcup_{A \in \mathcal{X}} A = \bigcap_{A \in \mathcal{X}} C \setminus A$$
$$C \setminus \bigcap_{A \in \mathcal{X}} A = \bigcup_{A \in \mathcal{X}} C \setminus A$$

Proof. Part of the problem set for this chapter :) □

If having sets whose elements are themselves sets seems odd, you will have ample time to get used to that in this class. We will study a lot of sets like this. One particularly important example of a set of sets is the powerset.

Definition 4. Let X be a set. The *powerset* of X is $\mathcal{P}(X) = \{Y : Y \subseteq X\}$ is the set of all subsets of X.

Here's a quick fact about powersets to get a feel for working with them.

Proposition 5. *Suppose $X = \{x_0, x_1, \ldots, x_{n-1}\}$ is a finite set with n elements. Then $\mathcal{P}(X)$ has 2^n elements.*

[3] Some authors write $B - A$. Ick.

Proof. Two subsets of X are the same if and only if they have the same elements, so we need to see how many ways there are to pick elements from X. We have n many choices to make—include x_n or not?—and each choice has two possible options. So in all there are $2 \times 2 \times \cdots 2 = 2^n$ many choices for how to get a subset of X. □

Exercises.

(1) You step outside of mathematics and define a set $\{x : x \text{ is a tall person}\}$. Is this a well-defined collection? Why or why not?

(2) Give two different definitions in set-builder notation for the set of integers which are multiples of 3.

(3) Check that $\cup$ and $\cap$ are associative and commutative:

$$(A \cap B) \cap C = A \cap (B \cap C)$$
$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$A \cap B = B \cap A$$
$$A \cup B = B \cup A$$

(4) Why is there not an analogous thing to check for the infinitary versions of union and intersection?

(5) Check the distributivity laws for union and intersection:

$$A \cap (B \cup C) = (A \cup C) \cap (B \cup C)$$
$$A \cup (B \cap C) = (A \cap C) \cup (B \cap C)$$

(6) Check the De Morgan laws for sets:

$$C \setminus (A \cup B) = (C \setminus A) \cap (C \setminus B)$$
$$C \setminus (A \cap B) = (C \setminus A) \cup (C \setminus B)$$

(7) What is $\cup\mathcal{P}(X)$? What is $\cap\mathcal{P}(X)$?

(8) Explicitly write out all elements of $\mathcal{P}(\emptyset)$ and $\mathcal{P}(\mathcal{P}(\emptyset))$. What about $\mathcal{P}(\mathcal{P}(\mathcal{P}(\emptyset)))$?

3. Countable sets

Before we can clear up what it means for a set to be countable, let's lay down some definitions about functions.

Recall that a *function* f from a set A to a set B is formalized to be a set of pairs $(a, f(a))$ with a in A and $f(a) \in B$, where each $a \in A$ appears in exactly one such pair. We will write this briefly as $f : A \to B$. Under this formalization, functions are by definition *total*: the *domain* of f, namely the set of inputs, is all of A. We reserve the term *partial function* for those whose domain is merely some of A. The *range* or *image* of f is the set $\{f(a) : a \in A\}$. Write $\operatorname{dom} f$ for the domain of a (possibly partial) function and $\operatorname{ran} f$ for the range.

Example 6. There is a unique function whose domain is the empty set. Under our formalization, this function is an empty set of ordered pairs. We call it the *empty function* and write $\emptyset$ to denote it.

A function $f : A \to B$ is *onto* B if $\operatorname{ran} f = B$. We also call these *surjections*, and say e.g. that f surjects A onto B. A function f is *one-to-one* if different inputs go to different outputs. In symbols: $a_0 \neq a_1$ implies $f(a_0) \neq f(a_1)$, where $a_0, a_1 \in \operatorname{dom} f$. We also call these *injections*. A function $f : A \to B$ is a *bijection* onto B if it is both one-to-one and onto B.

An important fact about injections is that they admit inverses: $f^{-1} : \operatorname{ran} f \to A$ is the function defined as $f^{-1}(b)$ is the unique a so that $f(a) = b$. In particular, if $f : A \to B$ is a bijection then $f^{-1} : B \to A$ is also a bijection. (Check this!)

A quirk of this formalization of function is that B isn't really part of what makes up a function. For example, the sine function can be thought of as a function $\mathbb{R} \to \mathbb{R}$ or $\mathbb{R} \to [-1, 1]$. Either way, it's the same function: the same inputs go to exactly the same outputs. In some contexts this is inconvenient and you want the *codomain* B to be part of the definition. (Under that definition $\sin : \mathbb{R} \to \mathbb{R}$ and $\sin : \mathbb{R} \to [-1, 1]$ would be different functions.) It's straightforward to translate from one formalization to the other, so ultimately it matters little which you choose. For the purposes of this class this is the more convenient formalization.

Here's a real-world example of a bijection that works for most people. Hold up your left hand, then your right hand. Touch each finger on the left hand to the corresponding finger on the right—thumb to thumb, index to index, etc. This gives a bijection from your left-hand fingers to your right-hand fingers.

Definition 7. A set X is *countable* if there is a bijection from X to a subset of $\mathbb{N}$.

For example, $\emptyset$ and $\mathbb{N}$ are countable. More generally, any subset of $\mathbb{N}$ is countable. Before we see less trivial examples, let's see some equivalent ways to formulate this property.

Lemma 8. *The following are equivalent for a set X.*

1. *X is countable;*
2. *There is an injection $f : X \to \mathbb{N}$;*
3. *(If X is nonempty.) There is a surjection $f : \mathbb{N} \to X$; and*
4. *(If X is nonempty.) There is a sequence $\langle x_n : n \in \mathbb{N} \rangle$ of all elements of X.*

Proof. ($1 \Rightarrow 2$) A bijection $f : X \to A \subseteq \mathbb{N}$ is an injection $f : X \to \mathbb{N}$.

($2 \Rightarrow 3$) Suppose you have an injection $f : X \to \mathbb{N}$. Fix $x \in X$. Define $g : \mathbb{N} \to X$ as $g(n) = f^{-1}(n)$ if $n \in \operatorname{ran} f$ and otherwise $g(n) = x$.

($3 \Rightarrow 4$) Given a surjection $f : \mathbb{N} \to X$ define a sequence $\langle x_n : n \in \mathbb{N} \rangle$ as $x_n = f(n)$. Done.

($4 \Rightarrow 1$) Given a sequence $\langle x_n : n \in \mathbb{N} \rangle$ of all elements of X define an injection $f : X \to \mathbb{N}$ as $f(x)$ is the least n so that $x = x_n$. □

Now let's look at some more interesting examples of countable sets.

Proposition 9. *$\mathbb{Z}$ is countable.*

Proof. The idea to enumerate $\mathbb{Z}$ in a sequence is to alternate between the positive and negative integers. Namely, set $x_{2n} = n$ and $x_{2n+1} = -n$ for $n \in \mathbb{N}$. Then $\langle x_n : n \in \mathbb{N} \rangle$ enumerates $\mathbb{Z}$. □

We can push this idea further.

Proposition 10. *Suppose A and B are countable. Then $A \cup B$ is also countable.*

Proof. Let $\langle a_n : n \in \mathbb{N} \rangle$ enumerate A and $\langle b_n : n \in \mathbb{N} \rangle$ enumerate B. Define a new enumeration $\langle x_n : n \in \mathbb{N} \rangle$ of $A \cup B$ as $x_{2n} = a_n$ and $x_{2n+1} = b_n$. □

You could prove this proposition using the injection characterization of being countable. But this adds a small complication if $A \cap B$ is nonempty—do you use the injection for A or the injection for B to decide where to send $x \in A \cap B$? The enumeration characterization lets us avoid this obstacle. Other times a different characterization of being countable is easier to check. It's nice to have options.

Definition 11. Let A and B be sets. Their *cartesian product* is

$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\}.$$

If n is a positive natural number let

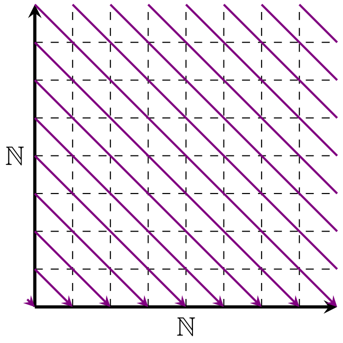
$$A^n = \{(a_0, a_1, \ldots, a_{n-1}) : a_i \in A \text{ for each } i < n\}.$$

And let $A^0 = \{\emptyset\}$.

That last part maybe looks a little strange, so let's understand it better. One way to think about A^n is to think of A as an alphabet, and then A^n is the set of length n words. How many words are there of length 0? Just one, the empty word. If this isn't convincing, when we look at cardinals later we will give a more general definition of exponentiation of sets and connect it to arithmetic operations. Then A^0 having 1 element corresponds to x^0 being 1 for any number x.

Proposition 12. *$\mathbb{N}^2$ is countable.*

Proof. Think of $\mathbb{N}^2$ as a two-dimensional grid of points. We enumerate the grid diagonally.



The violet lines give the enumeration, starting at $(0, 0)$ and moving to the northeast. □

Proposition 13. *$\mathbb{Q}$ is countable.*

Proof. It's enough to prove that the set $\mathbb{Q}^+$ of non-negative rationals is countable, by our earlier proposition that the union of two countable sets is countable.

Using the previous proposition we have an enumeration $(n_0, d_0), (n_1, d_1), \ldots$ of all the pairs of natural numbers. Create an enumeration of $\mathbb{Q}^+$ by setting $q_k = n_k/d_k$, or $q_k = 0$ if $d_k = 0$. Since every non-negative rational can be written as a ratio of two natural numbers, this gives an enumeration of all of $\mathbb{Q}^+$. □

We can generalize this.

Proposition 14. *Let A and B be countable sets. Then $A \times B$ is countable.*

Proof. Let $f : A \to \mathbb{N}$ and $g : B \to \mathbb{N}$ be injections. Then $h : A \times B \to \mathbb{N} \times \mathbb{N}$ defined as $h(a, b) = (f(a), g(b))$ is an injection. (Why?) Composing this injection with the bijection $\mathbb{N}\times\mathbb{N} \to \mathbb{N}$ given by the previous proposition gives an injection $A \times B \to \mathbb{N}$. □

Corollary 15. *Let A be a countable set. Then A^n is countable for any $n \in \mathbb{N}$.*

Proof. By induction on n. The key point is that A^{n+1} is in bijective correspondence with $A^n \times A$. (Why?) □

If A is a set of symbols, you can think of A^n as all length n words formed from the symbols in A. Can we push this even further to show there are only countably many words of any length?

Yes we can. Let's first do it for finite A. As a warm-up we need to see how big A^n is.

Lemma 16. *Suppose $A = \{a_0, \ldots, a_{k-1}\}$ is a finite set with k elements. Then A^n has k^n many elements.*

Proof. This is similar to the argument that the powerset of a k element set is 2^k. Namely, consider what an element of A^n looks like. It's an n-tuple of elements from a. To build such an n tuple we make n many independent choices, each with k options. So in all there are $k \times k \times \cdots \times k = k^n$ many ways to build up an element of A^n. □

For a set A, let

$$A^{<\omega} = \bigcup_{n\in\mathbb{N}} A^n$$

be the set of finite words with symbols from A.[4] For example, $\{0, 1\}^{<\omega}$ is the set of all finite binary strings.

Proposition 17. *Let A be a finite set. Then $A^{<\omega}$ is countable.*

Before we prove this observe that $A^{<\omega}$ will never be finite provided $A \neq \emptyset$, because it contains words of any finite length and there's infinitely many lengths.

Proof. Let's define a bijection $\mathbb{N} \to A^{<\omega}$ by first saying what gets mapped to the 0 length words, then the 1 length words, and so on. For this purpose fix a bijection $f : \{0, \ldots, k - 1\} \to A$ where $k \in \mathbb{N}$ is the size of A. For A^0 there's only one word: send 0 to the empty word, and we've used up 1 element of our domain. For A^1 there's k many words. For each $i < k$ send $1 + i$ to $(f(i))$. Now we've used up $1 + k$ elements of our domain. To cover the k^2 elements of A^2, for $i, j < k$ send $1 + k + (k * i + j)$ to the word $(f(i), f(j))$.

[4] An alternate notation one sees for this, especially in computer science, is the *Kleene star* A^*. I use the $A^{<\omega}$ notation because we will generalize it later after we've looked at ordinals.

At this point it's starting to be tedious to do exact counts, so let's abstract a bit. We're proceeding by induction on n to say where to map natural numbers to get A^n in the range. Assume we've done this below n, using up only a finite initial segment of $\mathbb{N}$, call it all the numbers $< N$. We need to continue by saying how to cover A^n. The set A^n has size k^n, which means there's a bijection $g : \{0, 1, \dots, k^n - 1\} \to A^n$. We'll use the the first k^n many unused numbers in the domain to handle these. Namely, for $i < k^n$ send $N + i$ to $g(i)$. Then we've used up all the numbers $< N + k^n$, so we still have space to continue.

It remains to confirm that the function we built up this way really is a bijection. It's one-to-one because we never sent a natural number to a word we'd already covered. And it's onto because by construction we cover every length n word for every n. So we're indeed done. □

You could work out an *ad hoc* argument to extend this to countable and infinite A, but let's be lazy and derive it from a powerful and important fact. And then I'll leave proving that fact as a problem for you to do.

Theorem 18. *A countable union of countable sets is countable. That is, if $A_0, A_1, \dots, A_n, \dots$ are all countable sets, then so is*

$$\bigcup_{n \in \mathbb{N}} A_n.$$

Proof. Do this as part of the problem set for this chapter :) □

Corollary 19. *Let A be any countable set, possible infinite. Then $A^{<\omega}$ is countable.*

Proof. From an earlier proposition we know that each A^n is countable. So $A^{<\omega}$ is a countable union of countable sets, whence it is countable. □

Exercises. The *composition* of two functions is obtained by doing them in succession. More precisely, if $f : A \to B$ and $g : B \to C$ then $g \circ f : A \to C$ is defined as $(g \circ f)(a) = g(f(a))$.

1. Check that if $f : A \to B$ and $g : B \to C$ are injections then their composition $g \circ f$ is an injection.
2. Check the analogous fact for surjections. Why does this imply the analogous fact for bijections?
3. Check that if $f : A \to B$ is a bijection then $f^{-1} : B \to A$ is a bijection.
4. Explain why there are only countably many books in the English language.
5. Explain why there are only countably many computer programs.

4. Induction and the well-order property

We've used induction a few times and it's an important feature about the natural numbers. Let's look at it more closely. We'll later generalize these ideas to a broader class of structures.

Fact 20 (Induction property of $\mathbb{N}$)**.** *Suppose that $X \subseteq \mathbb{N}$ is a set with the property that, for any $n \in \mathbb{N}$ if every $k < n$ is an element of X then $n \in X$. Then, $X = \mathbb{N}$.*

Some textbooks call this formulation of induction "strong induction". I'm just calling it induction because it's the formulation that's appropriate for generalization. More often used is this version.

Fact 21 ($+1$ induction property of $\mathbb{N}$)**.** *Suppose $X \subseteq \mathbb{N}$ is a set with the property that $0 \in X$ and, for any $n \in \mathbb{N}$, if $n \in X$ then $n + 1 \in X$. Then, $X = \mathbb{N}$.*

Proof from the induction property of $\mathbb{N}$. Fix $X \subseteq \mathbb{N}$ with the property that $0 \in X$ and if $n \in X$ then $n + 1 \in X$. I claim that if $n \in \mathbb{N}$ and every $k < n$ is in X then $n \in X$. To see this, consider such an n. If $n = 0$ then we already know $n \in X$. If $n > 0$, then by assumption $n - 1 \in X$. But then $(n - 1) + 1 = n \in X$, as desired. So by the induction property we get $X = \mathbb{N}$. □

This fact gives the proof technique known as *proof by induction.* To prove that something is true for all $n \in \mathbb{N}$, you prove it for 0 and you prove that if it's true at a number then it's true at the next number. This implies it's true for all natural numbers because if X is the set of n for which the property is true, then X satisfies the induction property and so X must be all of $\mathbb{N}$.

Sometimes induction is taken as a basic property. For example, the Dedekind–Peano axioms for natural number arithmetic take induction as one of the axioms. But you can see induction as coming from a more basic structural fact about the order on $\mathbb{N}$. First let's abstract a property of the order of $\mathbb{N}$, $\mathbb{Z}$, and other familiar structures.

Definition 22. A *linear order*[5] $(X, \leq)$ is a set X equipped with a binary relation $\leq$ on X satisfying:

1. $\leq$ is reflexive: $x \leq x$ for all x;
2. $\leq$ is transitive: $x \leq y \leq z$ implies $x \leq z$;
3. $\leq$ is antisymmetric if $x \leq y$ and $y \leq x$ then $x = y$; and
4. $\leq$ has trichotomy: for all $x, y \in X$ either $x \leq y$ or $y \leq x$.

If only the first three properties are satisfied then $(X, \leq)$ is called a *partial order.*

All of the familiar ordered number systems are linear orders: $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$.

Following common practice in mathematics, we will often just write X to refer to the linear order, not explicitly including the order $\leq$. It shouldn't be confusing to use the same symbol $\leq$ for different orders, but when it would improve clarity I'll write e.g. $\leq_X$.

On the topic of notation, it's convenient to introduce symbols $\geq$, $<$, and $>$ defined from $\leq$. This is just like how it works with number systems, but let me be explicit.

- $x \geq y$ means $y \leq x$;
- $x < y$ means $x \leq y$ but $x \neq y$; and
- $x > y$ means $y < x$.

All of these are inter-definable, so if you have one you can get the other three. As such, it doesn't matter which you use for the official definition of a linear order. You just have to make some small

[5] Also called a *total order.* Some authors also use phrases like "linear ordering" to have a very slightly different meaning—e.g. $(X, \leq)$ is a linear order but $\leq$ is a linear ordering of X. But I think that's confusing so I won't do it. Similar comments apply for other kinds of orders.

changes in how you phrase the axioms. For example, the *tri* in trichotomy makes the most sense if you formulate it in terms of $<$ and $>$.

Proposition 23. *Let $(X, \leq)$ be a partial order. Then $\leq$ has trichotomy if and only if for all $x, y \in X$ exactly one of the three possibilities $x < y$, $x = y$, or $x > y$ is true.*

Proof. ($\Rightarrow$) Fix x and y. We know that either $x \leq y$ or $x \geq y$. If both are true, then $x = y$. If $x \neq y$, then only one of the two can be true, by antisymmetry. So the two remaining options are $x < y$ or $x > y$, depending on which of the two cases we are in.

($\Leftarrow$) Fix x and y. If $x < y$ then $x \leq y$, and similarly if $x > y$ then $x \geq y$. And if $x = y$ then in particular $x \leq y$. So we have seen at least one of $x \leq y$ or $x \geq y$ is true. □

The order on $\mathbb{N}$ is special. It's not just a linear order, it's also a well-order.

Definition 24. A linear order $(X, \leq)$ is a *well-order* if every nonempty subset of X has a least element. That is, if $Y \subseteq X$ is not empty then there is $m \in Y$ so that $m \leq y$ for all $y \in Y$. When we want to refer to it by itself, we call this extra property *well-foundedness.*

Here's a few examples of well-orders.

(1) The linear order 2 consisting of two elements.
(2) The linear order 1 consisting of a single element.
(3) The linear order 0 consisting of zero elements.
(4) The order $\omega + \omega$ consisting of two copies of $\mathbb{N}$, with the elements of the second copy coming after all of the elements of the first copy.

This one is less obvious, so let's check it. I leave checking that $\omega + \omega$ is a linear order to you as an exercise, and only check the well-order property. Suppose $X \subseteq \omega + \omega$ is nonempty. If X contains an element from the first copy of $\mathbb{N}$, then it has a least element from that copy, by the fact that $\mathbb{N}$ is a well-order. This must be the least element of X. If X doesn't contain any elements from the first copy of $\mathbb{N}$, then it must contain elements from the second copy. It must have a least element from that copy, which must be the least element of X.

This idea can be generalized. If $(X, \leq)$ and $(Y, \leq)$ are linear orders then you can define a new linear order $X + Y$ as: the domain of $X + Y$ is

$$\{(0, x) : x \in X\} \cup \{(1, y) : y \in Y\}$$

with the order $\leq_{X+Y}$ defined as $(i, a) \leq_{X+Y} (j, b)$ if and only if $i < j$ or $i = j$ and $a \leq b$. (Here $\leq$ is whichever of $\leq_X$ or $\leq_Y$ makes sense.)

Proposition 25. *Suppose X and Y are linear orders. Then $X + Y$ is also a linear order. If X and Y are well-orders then so is $X + Y$.*

Proof. Do what we did with $\omega + \omega$. Fill in the details as an exercise. □

And here's some examples of linear orders which are not well-orders.

- $\mathbb{Z}$, because there's no smallest negative number.
- $\mathbb{Q}$, because there's there no smallest negative number.
- The set of reals ≥ 0, because there's no smallest positive real.

One reason why well-orders are an important concept in set theory is that they yield that induction is possible.

Theorem 26. *Suppose X is a well-order. Then X has the induction property: Suppose $Y \subseteq X$ has the property that for any $x \in X$ if every $y < x$ is an element of Y then $x \in Y$. Then, $Y = X$.*

Proof. Fix $Y \subseteq X$ so that for any $x \in X$ if every $y < x$ is in Y then so is x. Assume toward a contradiction that $Y \neq X$. Then, $X \setminus Y$ is nonempty. By well-foundedness, $X \setminus Y$ has a least element m. Note now that every $y < m$ is in Y, by leastness of m. But then $m \in Y$, a contradiction. □

Because $\mathbb{N}$ is a well-order, we get that induction works for the natural numbers. This theorem tells us that induction works for much more than just $\mathbb{N}$, but we will leave investigating so-called *transfinite induction* until we've looked at ordinals.

Exercises.

(1) Use induction to prove that, for any $n \in \mathbb{N}$,

$$\sum_{i=0}^{n} i = \frac{n(n+1)}{2}.$$

(2) Formulate a definition of linear order in terms of the strict order $<$ instead of $\leq$. Check that it's equivalent to the definition in terms of $\leq$.

(3) Check that $\omega + \omega$ is a linear order.

(4) Check that if X and Y are linear orders then so is $X + Y$.

(5) Check that if X and Y are well-orders then so is $X + Y$.

(6) Show that from the $+1$ version of induction on $\mathbb{N}$ you can prove that $\mathbb{N}$ has the induction property. (I.e. what some sources call strong induction.)

5. Recursive constructions

Closely related to the idea of induction is the idea of constructing objects by recursion. If you have some experience with programming, you should be familiar with using recursion to define functions. We also do this in mathematics.

Everyone's favorite first example of definitions by recursion is the factorial function, and it's mine too.

Example 27. Define a function $! : \mathbb{N} \to \mathbb{N}$ as: $0! = 1$ and $(n+1)! = (n+1) \cdot n!$.

You can check for yourself that, for example, $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$.

Why is this a valid definition of a function? We can see why using induction. Let X be the set of $n \in \mathbb{N}$ for which $n!$ makes sense. The base case definition $0! = 1$ tells us $0 \in X$ and the successor case for $(n+1)!$ tells us that if $n \in X$ then so is $n+1$. By induction, we get that X is all of $\mathbb{N}$.

Let's see another example of this.

Example 28. Define a function $\cdot : \mathbb{N}^2 \to \mathbb{N}$ by recursion on the second coordinate.

- $a \cdot 0 = 0$ for any $a \in \mathbb{N}$;
- $a \cdot (b+1) = (a \cdot b) + a$.

To verify this is a valid definition you do much the same as we did with the factorial.

When doing definitions by recursion, sometimes you want to have access to more than just one previous case. This is how, for example, the famous Fibonacci sequence is defined.

Example 29. The *Fibonacci sequence* is the sequence $\langle F_n : n \in \mathbb{N} \rangle$ of natural numbers defined recursively as:

- $F_0 = 0$ and $F_1 = 1$;
- $F_{n+2} = F_n + F_{n+1}$.

The first few numbers in this sequence are

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots$$

Because we need the two previous cases to define F_{n+2}, we need two base cases to get us started. And maybe it seems weird that we used recursion to define a sequence instead of a function. But a sequence is just a function whose domain is $\mathbb{N}$: you can think of $\langle x_n \rangle$ as a function x so that $x(n) = x_n$.

When you have an object defined by recursion, it naturally lends itself to proofs by induction. Let's see an example.

Proposition 30. *For any n we have*

$$\sum_{i=0}^{n} F_i^2 = F_n \cdot F_{n+1}.$$

Proof. The base case $n = 0$ is just the equality $0^2 = 0 \cdot 1$. For the successor case, assume

$$\sum_{i=0}^{n} F_i^2 = F_n \cdot F_{n+1}.$$

Then

$$\sum_{i=0}^{n+1} F_i^2 = \sum_{i=0}^{n} F_i^2 + F_{n+1}^2 = F_n \cdot F_{n+1} + F_{n+1}^2 = F_{n+1} \cdot (F_n + F_{n+1}) = F_{n+1} \cdot F_{n+2}. \qquad \square$$

We can unify all of these recursive definitions, and any other you might do, into a single definition. What is the general pattern? When defining what to do at stage n in the recursion, we want to know what happened at previous stages. Our examples only used one or two previous stages, but for full generality we want to allow the use of any previous stage. Of course, we can't use what happens at stage n or beyond, as that would make us fall prey to diagonalization.[6] We might also want to use the stage n itself in the recursion, as we did with the factorial. So at stage n we need to have both n and the partial function defined below n.

Let's get some notation before we state the theorem.

Definition 31. Let $f : X \to A$ be a function and suppose $Y \subseteq X$. The *restriction* of f to Y, written $f \upharpoonright Y$, is the function whose domain is Y and agrees with f on that domain. That is, $(f \upharpoonright Y)(x) = f(x)$ whenever $x \in Y$ and is undefined otherwise.

Definition 32. For any $n \in \mathbb{N}$, Write n to refer to the set $\{0, \dots, n-1\}$ of natural numbers $< n$. In particular, $0 = \emptyset$.[7]

Theorem 33 (Definitions by recursion)**.** *Let $G(n, f)$ be a function whose inputs are a natural number n and a function f whose domain is the set of natural numbers $< n$. Then there is a function F with domain $\mathbb{N}$ so that for all $n \in \mathbb{N}$ we have $F(n) = G(n, F \upharpoonright n)$.*

Before we see the proof, let's break this down into plain language. First, note that $0 = \emptyset$ and so $F \upharpoonright 0$ is always the empty function $\emptyset$—the unique function whose domain is the empty set. So $G(0, \emptyset)$ is saying what the base case is. Overall, what this function G is doing is unifying the base case(s) and the successor cases of the recursive definition into a single object. The theorem then says that making a recursive definition generates a genuine function F. So this is just a mathy way to say that definition by recursion is valid.

And now let's see how our previous examples fit into this framework.

Example 34. Let $G(n, f)$ be the function defined as $G(0, \emptyset) = 1$ and otherwise $G(n+1, f) = (n+1) \cdot f(n)$. Then the $F(n)$ produced by the theorem is the factorial function $n!$.

The multiplication example is a bit awkward, since we want a function with domain $\mathbb{N}^2$. But it also fits into this general pattern.

Example 35. Fix $a \in \mathbb{N}$. Let $G_a(n, f)$ be the function defined as $G(0, \emptyset) = 0$ and $G(n+1, f) = f(a) + a$. The theorem gives a function $F_a(n) = a \cdot n$. So we could define $a \cdot n$ as $F_a(n)$.

In effect, what we did here was do infinitely many recursive definitions simultaneously and then glued them together into a single function.

Example 36. Let $G(n, f)$ be the function defined as $G(0, \emptyset) = 0$, $G(1, f) = 1$, and $G(n+2, f) = f(n) + f(n+1)$. Then the theorem gives a function $F(n)$ which gives the n-th Fibonacci number.

Proof of theorem. The idea is this: To define $F(n)$ for any fixed n we only need to carry out the recursive construction a finite number of steps. So we will define $F(n)$ by saying there's a finite list of the steps done to compute it, and then argue that there cannot be a stage n for which this fails.

Here it is in more detail: Define $F(n) = y$ just in case there is a function f with domain $n+1$ so that $f(0) = G(0, \emptyset)$ and for all $k < n$ we have $f(k+1) = G(k+1, f \upharpoonright (k+1))$. To see that

[6] Consider the "definition" of a sequence $\langle x_n : n \in \mathbb{N} \rangle$ as $x_n = 0$ if and only if $x_n = 1$. Clearly this is not a valid definition.

[7] We'll come back to this funky overloading of the symbol n when we look at ordinals. For now, you can think of it as giving a convenient name for the set of stages below n.

$\operatorname{dom} F = \mathbb{N}$ first note that $0 \in \operatorname{dom} F$ because $F(0) = G(0, \emptyset)$. Now suppose that $n \in \operatorname{dom} F$, as witnessed by the function f with domain $n+1$. Letting $y = G(n+1, f)$ define a new function f' as $f \cup \{(n+1, y)\}$.[8] Then f' witnesses that $F(n+1)$ is defined. So $\operatorname{dom} F$ must be all of $\mathbb{N}$ by induction.

Finally, there's a small detail to check, namely that this definition gives a *function*. That is, we need to know that $F(n)$ is uniquely defined. This follows once we know that the f witnessing that $F(n)$ is defined. Again prove this by induction on n. There's a unique witness $F(0)$ is defined—the empty function—and to go from n to $n+1$ there's only one option for how to extend. □

Why did we go through so much effort in the proof to talk about these finite functions f rather than just going straight to F like we did with the examples? The point of the theorem is that definitions by recursion are valid, and it'd be circular to presume that F is well-defined. The seeming detour with the f's is what we need to do to confirm F really does exist.

To close out the section, let's see a fancier example of a definition by recursion.

Definition 37. Let $(X, \leq)$ and $(Y, \leq)$ be linear orders. An *embedding* of X into Y is a function $e : X \to Y$ so that $x_0 \leq_X x_1$ if and only if $f(x_0) \leq_Y f(x_1)$. If $\operatorname{ran} f = Y$ then we call f an *isomorphism* between X and Y. If there is an isomorphism between X and Y we say X and Y are *isomorphic* and write $X \cong Y$.

If we want to be extra clear that we are talking about orders, we talk about *order embeddings*, *order isomorphisms*, and so on.

Theorem 38 (Cantor)**.** *Any countable linear order embeds into $\mathbb{Q}$.*

Proof. Let X be a countable linear order. If X is finite this is easy: just send the least element to 0, the next to 1, and so on. So assume X is infinite. Fix an enumeration $x_0, x_1, \dots$ of the elements of X, where each element appears exactly once in the enumeration. We will define an embedding $f : X \to \mathbb{Q}$ using recursion along this enumeration.

To start, set $f(x_0) = 17/3$. (If your favorite rational number isn't 17/3 you can use it instead.) To proceed, assume we have already defined $f(x_0), f(x_1), \dots f(x_n)$ and we arranged it so that this partially constructed f embeds $\{x_0, x_1, \dots x_n\}$ into $\mathbb{Q}$. We need to define $f(x_{n+1})$ so that we preserve the order relations. Note that the rational numbers $f(x_0), f(x_1), \dots f(x_n)$ divide $\mathbb{Q}$ into $n+3$ many regions: the rationals below all of them, the rationals between the smallest and second smallest, etc., up to the rationals above all of them. Similarly, $x_0, x_1, \dots x_n$ divide X into up to $n+3$ many regions. (I say "up to", because e.g. x_7 might be the smallest element of X and so there's no region of elements below it. Or maybe there's no elements between x_1 and x_7.) The next element x_{n+1} must be in one of those $\leq n+3$ regions, say between x_i and x_j. Then set $f(x_{n+1})$ to be any rational number between $f(x_i)$ and $f(x_j)$.

After we've gotten through the entire enumeration we've defined a function $f : X \to \mathbb{Q}$. To see this function is an embedding, consider x_i and x_j. Without loss of generality, let $j > i$. Then at stage j in the construction we defined $f(x_j)$ so that $x_i \leq x_j$ if and only if $f(x_i) \leq f(x_j)$. □

Let's clear up a possible confusion about this argument. Unless X is just $\mathbb{N}$ itself, then you can't have that the order $\leq_X$ matches up with with the enumeration. That is, you shouldn't expect that $m < n$ implies that $x_m <_X x_n$. One way to think about this is, imagine if X is $\mathbb{Q}$. Then there's no smallest element of X, so x_0 cannot be the smallest element. So you'll get lots and lots of x_n's

[8] If this looks a bit weird: recall that functions are formalized as sets of ordered pairs. So we're extending f to have $n+1$ in its domain. This sort of notation is convenient when building up functions by hand.

which are smaller than x_0, even though they come later in the enumeration. More generally, there's always infinitely many rationals between rationals $a < b$, and so you'll have x_n's for arbitrarily large n which are between a and b. Any enumeration of $\mathbb{Q}$ has to dance around a lot.

Exercises.

(1) Give a recursive definition of exponentiation n^m on the natural numbers.
(2) Prove that the n-th Fibonacci number F_n is always less than 2^n.
(3) Give a recursive definition of iterated exponentiation (sometimes called *tetration*). Write this as $n \uparrow m$ to mean n iteratively exponentiated m many times.[9] What is the correct base case for this definition?
(4) For a natural number n and a set X, inductively define $\mathcal{P}^n(X)$:
  - $\mathcal{P}^0(X) = X$;
  - $\mathcal{P}^{n+1}(X) = \mathcal{P}(\mathcal{P}^n(X))$.

  Prove that $\mathcal{P}^n(\emptyset)$ has $2 \uparrow n$ elements.
(5) For a set X let $\mathcal{P}^\omega(X) = \bigcup_{n\in\mathbb{N}} \mathcal{P}^n(X)$. Explain why $\mathcal{P}^\omega(X)$ exists.

(Kameryn J. Williams) Bard College at Simon's Rock, 84 Alford Rd, Great Barrington, MA 01230
E-mail address: kwilliams@simons-rock.edu
URL: http://kamerynjw.net

[9] That is, $n \uparrow m = \underbrace{n^{n^{\cdot^{\cdot^{\cdot^{n^n}}}}}}_{m \text{ many}}$.