NOTES ON SET THEORY

1. Introduction

Throughout mathematics, sets are used as one of several fundamental types of mathematical objects, along with numbers, ordered pairs, functions, etc. But it turns out that sets are special, in that every other type of mathematical object can be "compiled" into sets. For example:

- A function $f : X \to Y$ can be "compiled" into the set of ordered pairs $\{(x, f(x)) \mid x \in X\}$, sometimes called its **graph**; see Definition 2.42 and Remark 2.47.
- An ordered pair (x, y) can be "compiled" into the set $\{\{x\}, \{x, y\}\}$ (among many other possibilities); see Definition 2.30 and Exercise 2.31.
- The natural number 3 can be "compiled" into $\{0, 1, 2\}$, where $2 := \{0, 1\}$, $1 := \{0\}$, and $0 := \varnothing = \{\}$; thus
$$3 = \{\{\}, \{\{\}\}, \{\{\}, \{\{\}\}\}\}.$$

In this role, set theory serves as the "machine language" (or if you prefer, "assembly language") underlying the higher-level language of ordinary math. Part of the goal of this course is to introduce this "machine language" and the "compilation" process from higher-level math.

Aside from serving as a low-level foundations for the rest of math, set theory also studies several mathematical concepts of fundamental importance in their own right, such as induction, cardinality, and choice. You've surely encountered instances of these concepts already in other areas; in this course, we will define them and develop their basic theory in full generality.

2. Axioms

Informally, a **set** A is a collection of objects. Given A and some other object x, you are allowed to ask whether or not x is in the collection A, denoted
$$x \in A.$$
Moreover, this is the *only* feature of a set: it is completely determined by what all of its elements are. This is captured by the

Axiom of Extensionality 2.1. For two sets A, B,
$$A = B \iff \forall x\, (x \in A \iff x \in B).$$

The word "axiom" means that this assertion is *assumed*, rather than *proved* as a theorem would be. Every theorem in math must be proved from simpler assertions; we must necessarily start somewhere, with some basic statements we consider so intuitively unobjectionable that we're willing to take them on faith, hence declare them to be *axioms*.

Similarly, we must take some basic mathematical concepts as undefined in terms of simpler ones. Recall also that in set theory, all other mathematical objects are defined from sets. Thus, formally:

Definition 2.2. The word **set** is a synonym for "mathematical object", and is left undefined.

There is a binary relation $\in$ between sets, also undefined. That is, for any sets (i.e., mathematical objects) x, A, we can connect these two "nouns" via the "verb" $\in$ into the "complete sentence"
$$x \in A.$$
This complete sentence does not "mean" anything; the only thing we know about it is that the Axiom of Extensionality holds (not because of any justification, but only because we said so).

2.A. **Comprehension.** Conceptually, the Axiom of Extensionality tells us that sets turn *assertions*, i.e., "complete sentences", into *objects*, i.e., "nouns". In math, as in English, these are two entirely distinct grammatical categories:

- "It snowed a lot this winter" is a complete sentence.
- "That it snowed a lot this winter" is *not* a complete sentence, but rather a noun (phrase).
- "It is true that it snowed a lot this winter" is again a complete sentence, with the same meaning as the first sentence.
- "It is false that it snowed a lot this winter" is also a complete sentence, with an entirely different meaning.
- "I know that it snowed a lot this winter" is also a complete sentence, with a third meaning.

Similarly:

- $\mathbb{R}$ (the set of real numbers) is a noun.
- "$x \in \mathbb{R}$" is a complete sentence (that depends on the variable x).
- "$x \notin \mathbb{R}$" is a complete sentence with a different meaning.

The Axiom of Extensionality tells us that a set A (noun) is completely determined by the meaning of the assertion "$x \in A$". What about the reverse procedure, the mathematical analog of the English word "that", to turn an assertion (depending on a variable) into a set?

Axiom of Comprehension 2.3. For any mathematical assertion $\phi(x)$ depending on a variable x, there is a (unique, by Extensionality) set A such that

$$\forall x\,(x \in A \iff \phi(x)).$$

This set A is denoted

$$\{x \mid \phi(x)\}.$$

Here, by a "mathematical assertion", we mean an assertion that can be expressed using the basic binary relation $\in$, as well as the basic equality relation $=$, using the usual logical operations of "and", "or", "not", $\exists$, and $\forall$. The variable x is allowed to appear in this expression, as are any previously known mathematical objects (i.e., sets).[1]

Example 2.4. $\varnothing$ is an abbreviation for $\{x \mid \text{false}\}$, where "false" is a nullary "or", or if you prefer, some arbitrary trivially false statement, such as "$x \neq x$".

Similarly, for finitely many objects $x_1, \ldots, x_n$, let $\{x_1, \ldots, x_n\} := \{x \mid x = x_1 \text{ or } \cdots \text{ or } x = x_n\}$.

Example 2.5. For a set X and assertion $\phi(x)$, define the abbreviation

$$\{x \in X \mid \phi(x)\} := \{x \mid x \in X \text{ and } \phi(x)\}.$$

Example 2.6. For two sets A, B, define the abbreviation

$$A \subseteq B \;:\!\iff \forall x\,(x \in A \implies x \in B).$$

Then for a set X, its **powerset** is

$$\mathcal{P}(X) := \{A \mid A \subseteq X\} = \{A \mid \forall x\,(x \in A \implies x \in X)\}.$$

[1]Formally, ϕ should be a first-order formula in the signature of set theory $\{\in\}$, with some free variables, and with other sets assigned to all variables except for x. That is, if $y_1, \ldots, y_n$ are the other free variables except x appearing in ϕ, then the "Axiom of Comprehension" is really an axiom schema, consisting of the sentence

$$\forall y_1 \cdots \forall y_n\, \exists A\, \forall x\,(x \in A \iff \phi)$$

for each such formula ϕ.

Example 2.7. If $\mathcal{A}$ is a set of sets (this allows us to avoid having to define what an "indexed collection of sets $(A_i)_{i\in I}$" means, for now; see Definitions 2.59 and 2.60), then define

$$\bigcup \mathcal{A} := \{x \mid \exists A \in \mathcal{A}\,(x \in A)\},$$
$$\bigcap \mathcal{A} := \{x \mid \forall A \in \mathcal{A}\,(x \in A)\},$$

where as usual,

$$\exists A \in \mathcal{A}\,(\cdots) \iff \exists A\,(A \in \mathcal{A} \text{ and } \cdots),$$
$$\forall A \in \mathcal{A}\,(\cdots) \iff \forall A\,(A \in \mathcal{A} \implies \cdots).$$

In particular, if $\mathcal{A} = \{A, B\}$ (per Example 2.4),

$$A \cup B := \bigcup\{A, B\},$$
$$A \cap B := \bigcap\{A, B\}.$$

Definition 2.8. Naive Set Theory consists of the Axioms of Extensionality and Comprehension.

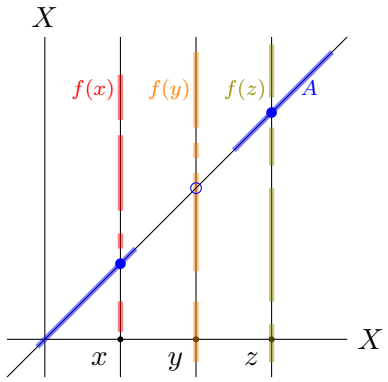
The above examples, along with the brief descriptions from the Introduction of how other standard mathematical notions may be "compiled", should help to convince you that all of "normal" math, i.e., outside of set theory, may be "compiled" into Naive Set Theory. Unfortunately, Naive Set Theory is too powerful for its own good:

2.B. **Cantor's theorem and Russell's paradox.**

Theorem 2.9 (Cantor)**.** Let X be a set, $f : X \to \mathcal{P}(X)$ be a function. Then f is not surjective, i.e., there is an $A \in \mathcal{P}(X)$ such that for all $x \in X$, $f(x) \neq A$.

Of course, we have not yet reduced the notion of "function" to sets – see Definition 2.42. Thus, for now, functions should be understood in the informal sense you're used to from "ordinary" math.

Before giving the one-line proof, we first explain the idea. We want to find a subset $A \subseteq X$ which does not equal any $f(x)$, which by Extensionality means that $A, f(x)$ must differ on the membership of at least one element. Luckily for us, we have just enough elements of X to allocate an element for $A, f(x)$ to differ on for each x: namely, we may allocate x itself. Here is a picture:



We visualize the set X as a (horizontal) line, and each of the subsets $f(x) \subseteq X$ as a subset of the same (vertical) line, so that the entire function f is represented as a subset of the plane X^2. The set A is defined as the subset of the (diagonal) line consisting of precisely the elements not on each vertical line; thus it cannot equal any of the vertical lines. This proof technique is therefore called **diagonalization**.

Proof. Let $A := \{x \in X \mid x \notin f(x)\}$. Then for all $x \in X$, $x \in A \iff x \notin f(x)$, so $A \neq f(x)$. □

Corollary 2.10 (Russell's paradox)**.** Naive Set Theory is inconsistent (self-contradictory).

Proof. Let $V = \{x \mid \text{true}\}$ be the set of all sets (where as in Example 2.4, "true" is a nullary "and", or if you prefer, some trivially true statement such as $x = x$). Note that $V = \mathcal{P}(V)$ (since all objects are sets). Thus $\text{id} : V \to V = \mathcal{P}(V)$ is a surjection, contradicting Cantor's theorem. □

If we "plug in" the above proof of Cantor's theorem into this proof, we get:

Proof. Let $A := \{x \mid x \notin x\}$. Then $A \in A \iff A \notin A$, a contradiction. □

Note that this latter proof shows that Comprehension, rather than Extensionality, is the problem. Namely, Comprehension is too "absolutist": there are general principles of logic[2] which tell us that in any reasonable formalized "mathematical universe", there will always be informal "meta-concepts" that our universe cannot "see". In set theory, this takes the form of "properties" $\phi(x)$: each such property does define an informal "meta-collection" of mathematical objects; but Russell's paradox says that this collection cannot itself always be an object in the mathematical universe.

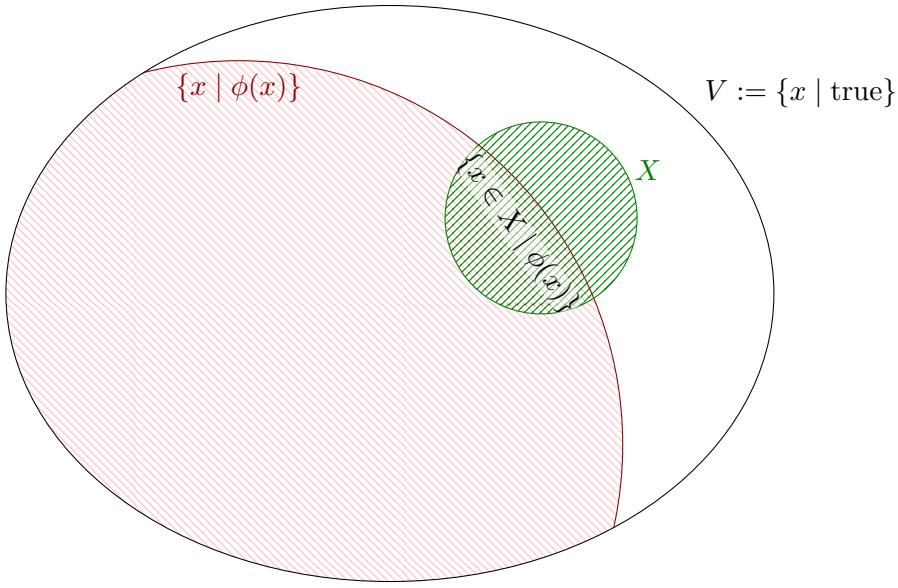
Definition 2.11. A **class** is an informal collection $\{x \mid \phi(x)\}$ defined by a property $\phi(x)$. That is, "class" is roughly synonymous with "property"/"mathematical assertion"/"first-order formula" $\phi(x)$, except that we think of it as the collection defined by $\phi(x)$, rather than the expression $\phi(x)$ itself.[3]

We say that a class $\{x \mid \phi(x)\}$ **is a set** if that instance of Comprehension holds, i.e., there is a (unique, by Extensionality) set A such that $\forall x\,(x \in A \iff \phi(x))$.

A class which is not a set is called a **proper class**. For example, the class in the second proof of Russell's paradox above is a proper class.

2.C. **The theory $\mathsf{ZF}^- - \mathsf{Infinity}$.** The most common way[4] out of Russell's paradox is to restrict the Axiom of Comprehension so that only "sufficiently small" classes form sets.

(2.12)



Intuitively speaking, we allow ourselves to build new sets whose "sizes are bounded" in terms of preexisting ones. For example,

[2] e.g., the Gödel incompleteness theorems, and Tarski's undefinability of truth

[3] Warning: one can easily formalize these "expressions" into mathematical objects, e.g., finite strings of symbols such as $\wedge, \vee, \exists, \in$, etc. But it is then impossible to define, within the language of set theory itself, what such a formalized expression $\phi(x)$ means; this is known as Tarski's undefinability of truth.

[4] Two other approaches, which we will not discuss in detail, are to (a) declare the formula "$x \notin x$" appearing in Russell's paradox to be invalid, because the elements of a set should always be "simpler" than the set itself, leading to a theory called Quine's New Foundations; or (b) disallow the formula "$x \notin x$" because it mentions *negation* without restricting the size of the defined class, leading to a theory called Positive Set Theory.

Axiom of Powerset 2.13. For any set X, $\mathcal{P}(X) = \{A \mid A \subseteq X\}$ from Example 2.6 is a set.

This comprehension is allowed, because even though the size of $\mathcal{P}(X)$ will always be bigger than that of X (formally, by Cantor's theorem; see Theorem 5.33), the size only grows by a "controlled" amount. Similarly,

Axiom of Union 2.14. For any set $\mathcal{A}$, $\bigcup \mathcal{A} = \{x \mid \exists A \in \mathcal{A}\,(x \in A)\}$ from Example 2.7 is a set.

Axiom of Finite Sets 2.15. For any $x_1, \ldots, x_n$, $\{x_1, \ldots, x_n\}$ from Example 2.4 is a set.[5]

As is typical throughout math, instead of assuming an n-ary "combining" operation, it is enough to assume the nullary and binary cases:

Axiom of Empty Set 2.16. $\varnothing = \{x \mid \text{false}\}$ (Example 2.4) is a set.

Axiom of Pairing 2.17. For any x, y, $\{x, y\} = \{z \mid x = z \text{ or } y = z\}$ is a set.

Proof of Finite Sets from Union, Empty Set and Pairing. By induction on n.[6] For $n = 0$ this is by Empty Set. If $\{x_1, \ldots, x_n\}$ is a set, then $\{x_1, \ldots, x_n, x_{n+1}\} = \bigcup\{\{x_1, \ldots, x_n\}, \{x_{n+1}, x_{n+1}\}\}$. □

The preceding axioms all allow us to build new sets that are slightly bigger than existing ones. We now introduce two axiom schemas that say directly that a class smaller than a set is a set.

Axiom of Restricted Comprehension/Separation 2.18. A class contained in a set is a set.

That is, for any property $\phi(x)$ (as in the original Comprehension 2.3) and set X, if $\{x \mid \phi(x)\} \subseteq X$, meaning $\forall x\,(\phi(x) \implies x \in X)$, then $\{x \mid \phi(x)\}$ is a set.

Equivalently, for any $\phi(x)$ and set X, the intersection $X \cap \{x \mid \phi(x)\} = \{x \in X \mid \phi(x)\}$ from Example 2.5 is a set; this is depicted in the above picture (2.12).

Proof that these two axioms are equivalent. Assuming that any class contained in X is a set, then

$$\{x \in X \mid \phi(x)\} = \{x \mid x \in X \text{ and } \phi(x)\}$$

is a class contained in X, hence is a set.

Conversely, assuming $\{x \in X \mid \phi(x)\}$ is always a set, we have

$$\begin{aligned} \{x \mid \phi(x)\} \subseteq X &\iff \forall x\,(\phi(x) \implies x \in X) && \text{by definition of } \subseteq \\ &\iff \forall x\,(\phi(x) \iff x \in X \text{ and } \phi(x)) \\ &\iff \{x \mid \phi(x)\} = \{x \in X \mid \phi(x)\} && \text{which is a set.} \end{aligned}$$

□

Example 2.19. $V = \{x \mid \text{true}\}$ is not a set. If it were, then every Comprehension $\{x \mid \phi(x)\}$ would reduce to the Restricted Comprehension $\{x \in V \mid \phi(x)\}$, recovering in particular Russell's paradox.

Example 2.20. For any nonempty set $\mathcal{A}$, $\bigcap \mathcal{A} = \{x \mid \forall A \in \mathcal{A}\,(x \in A)\}$ from Example 2.7 is a set.

Proof. Fix $A_0 \in \mathcal{A}$. Then

$$\bigcap \mathcal{A} = \{x \in A_0 \mid \forall A \in \mathcal{A}\,(x \in A)\},$$

since for any x,

$$\forall A \in \mathcal{A}\,(x \in A) \iff x \in A_0 \text{ and } \forall A \in \mathcal{A}\,(x \in A).$$

□

Remark 2.21. For $\mathcal{A} = \varnothing$, the same definition of $\bigcap \mathcal{A}$ would yield the entire universe V.

In mathematical practice, one typically only intersects subsets $A \subseteq X$ of a fixed, context-dependent ambient set X (e.g., closed subsets of a topological space, subgroups of a group, ...). In such contexts, the "right" convention is to define the nullary intersection $\bigcap \varnothing := X$.

[5] This would be an axiom schema.

[6] Formally, this induction is taking place in the metatheory, i.e., this is really a *proof schema*: for each n, we get a different proof of the corresponding axiom in the axiom schema of Finite Sets 2.15.

While Restricted Comprehension says that any subclass of a set is a set, one might expect more generally that a class which "injects" into a set ought also be a set. Relatedly, one might also expect that a class which admits a "surjection" from a set ought also be a set. One needs to be careful about what this "injection"/"surjection" means: if we assume it is given by a function which is already a set, then that more-or-less defeats the purpose, since this function will already be an "upper bound" on its domain/range (see Exercise 2.40). Hence, we need to work once again with "meta-collections", i.e., properties, this time of pairs:

Axiom of Replacement 2.22. Let $\phi(x, y)$ be a property of *two* variables x, y (and possibly depending on other known objects). For any set X, if

$$\forall x \in X \underbrace{\forall y \forall z \, (\phi(x, y) \text{ and } \phi(x, z) \implies y = z)}_{\text{"}\exists \text{ at most one } y \text{ s.t. } \phi(x,y)\text{"}},$$

then $\{y \mid \exists x \in X \, \phi(x, y)\}$ is a set.

This axiom is quite powerful:

Exercise 2.23. Prove Restricted Comprehension from Replacement and no other axioms (except Extensionality).

Exercise 2.24. Another common version of Replacement uses "$\exists!$" instead of "$\exists$ at most one".

(a) Prove Restricted Comprehension from this version of Replacement and Empty Set.
(b) Prove that the two versions of Replacement are equivalent, using only Empty Set.
(c) Prove yet another version of Replacement that uses "$\exists$ at most a set of":

$$\forall x \in X \, \exists A \, \forall y \, (\phi(x, y) \implies y \in A).$$

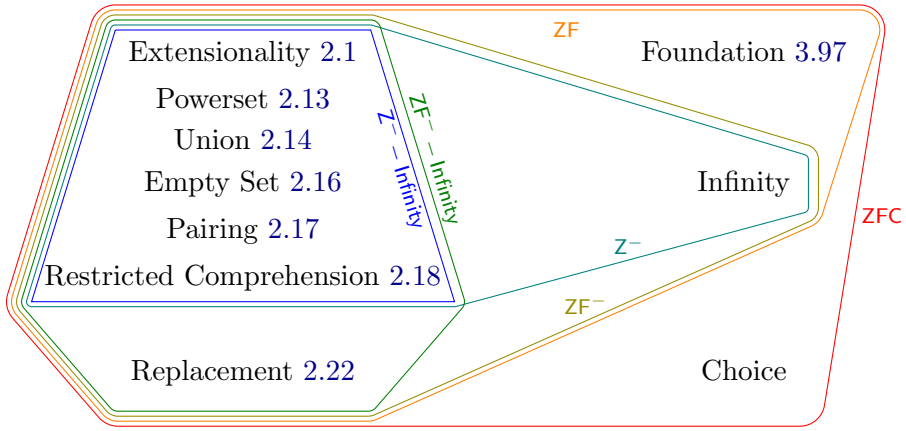
(You may use Extensionality and all axioms from this subsection.)

Exercise 2.25. Prove Pairing from Replacement, Empty Set, and Powerset.

Definition 2.26. The set theory $\mathsf{ZF}^- - \mathsf{Infinity}$ consists of the Axioms of Extensionality 2.1, Powerset 2.13, Union 2.14, Empty Set 2.16, Pairing 2.17, and Replacement 2.22; we thus also have Restricted Comprehension 2.18 by Exercise 2.23.

The awkward name with two minus signs is because this theory is lacking two important axioms, Foundation and Infinity, from the theory known as ZF that will be introduced later; see 3.97 and 3.138. Adding the further Axiom of Axiom 4.2 yields the set theory known as ZFC, which is widely accepted as the "standard" foundations for mathematics.

(2.27)



2.D. **Ordered pairs, Cartesian products, relations, functions.** We now begin to discuss the process of "compiling" other types of commonly used mathematical objects into sets. Broadly speaking, this goes as follows. For the given type of object, we formulate some axioms for it as if it were primitive, that capture everything we need to know when using this type of object in mathematical practice. We then "encode" this type of object into sets, and then prove the desired axioms from the set theory axioms. There may be many reasonable such "encodings", in which case it doesn't matter which one we pick: once we've proved the axioms, we know all we need to use the definition in practice, and never need to think about the encoding again.[7]

For ordered pairs, we need to know two things about them in practice:

(2.28) For any mathematical objects x, y, there is another object called the pair (x, y).

(2.29) ("Extensionality for pairs") The *only* feature of an ordered pair is its two coordinates:

$$(a, b) = (c, d) \iff a = c \text{ and } b = d.$$

Definition 2.30 (Kuratowski)**.** For any x, y, let $(x, y) := \{\{x\}, \{x, y\}\}$.

Proof of (2.28). This is indeed a set, by the Axiom of Pairing 2.17 applied thrice. □

Proof of (2.29). $\Longleftarrow$ is obvious (note: unlike in Extensionality 2.1 for sets, where $\Longrightarrow$ was the obvious direction[8]). Now suppose $(a, b) = (c, d)$. Then $\{a\} \in (a, b) = (c, d) = \{\{c\}, \{c, d\}\}$, whence $\{a\} = \{c\}$ or $\{c, d\}$, both of which contain c, whence $c \in \{a\}$, whence $c = a$. So

$$\{\{a\}, \{a, b\}\} = (a, b) = (c, d) = (a, d) = \{\{a\}, \{a, d\}\}.$$

If $a = b$, then the LHS is $\{\{a\}\}$, hence so is the RHS, hence $\{a, d\} = \{a\}$, hence $d = a = b$. Otherwise, $\{a, b\}$ in the LHS must equal $\{a, d\}$ in the RHS (since it is not $\{a\}$ which does not contain b), hence $b \in \{a, b\} = \{a, d\}$, hence $b = d$ (since $b \neq a$). □

Exercise 2.31. Which of the following encodings also work, i.e., also satisfy (2.28) and (2.29)?

(a) $(x, y) := \{x, y\}$
(b) $(x, y) := \{x, \{y\}\}$
(c) $(x, y) := \{\{0, x\}, \{1, y\}\}$
(d) $(x, y) := \{x, \mathcal{P}(y)\}$
(e) $(x, y) := \{\mathcal{P}(x), \mathcal{P}(y) \setminus \{\varnothing\}\}$
(f) $(x, y) := \{x, \{x, y\}\}$ [Hint: this depends on whether the Axiom of Foundation 3.97 holds.]

Definition 2.32. For two classes X, Y, their **Cartesian product** is

$$X \times Y := \{(x, y) \mid x \in X \text{ and } y \in Y\} = \{p \mid \exists x \in X\, \exists y \in Y\, (p = (x, y))\}.$$

Proposition 2.33. If X, Y are sets, then so is $X \times Y$.

Proof. For each x, for each y, we have a set (x, y); thus by Replacement 2.22 (applied to the function $\phi(y, p) :\Longleftrightarrow p = (x, y)$), we have a set $\{x\} \times Y = \{p \mid \exists y \in Y\, (p = (x, y))\}$; thus by Replacement again (applied to $\psi(x, s) :\Longleftrightarrow s = \{p \mid \exists y \in Y\, (p = (x, y))\}$), we have a set

$$\{\{p \mid \exists y \in Y\, (p = (x, y))\} \mid x \in X\};$$

now take Union. □

[7] Again, a computer analogy is helpful: the only type of data on (modern) computers is bytes, i.e., strings of 8 bits. On my computer, the letter 'M' is encoded as the byte 01001101_2, while on yours it may be 11010100_2; in programming practice (that's not super-low-level, e.g., hardware drivers), we never need to think about these encodings.

[8] There is a philosophical distinction between the notions of sets vs. pairs (other than that only the former can serve as a foundation for mathematics). Pairs are known as a *positive type* of object, in that they are originally specified by how they are *created*: by combining two other objects (2.28). Thus, the nontrivial direction of Extensionality for pairs says that if two pairs are the same, then they must have been created the same way. By contrast, sets (in the set-theoretic sense) are a *negative type* of object, being specified by how they may be *used*: by asking if some x is $\in$ it. The nontrivial direction of Extensionality says that if two sets look the same when used, then they are the same.

Exercise 2.34. Give a different proof that $X \times Y$ is a set, using Powerset instead of Replacement, that however has the disadvantage of depending on our specific chosen encoding of pairs.

Definition 2.35. As indicated above, if $F(x)$ is a mathematical *expression* (rather than assertion) that depends on a variable x, and X is a set, we use the shorthand

$$\{F(x) \mid x \in X\} := \{y \mid \exists x \in X\, (F(x) = y)\},$$

which is a set by Replacement. Here, by "mathematical expression", we really mean a "meta-function", i.e., its graph is a "meta-relation" defined by a property $\phi(x, y)$ as in the statement of Replacement 2.22.

Definition 2.36. A set (or class) R is a **binary relation** if each of its elements is an ordered pair (x, y), in which case we write

$$x\ R\ y \iff (x, y) \in R.$$

Conversely, if $\bowtie$ is a symbol that already denotes some binary relation, then we abuse notation by also using $\bowtie$ to denote the class defined by the above. For example,

$$\in = \{(x, y) \mid x \in y\}.$$

Exercise 2.37. Show that this is a proper class.

Definition 2.38. The **domain** and **range** of a binary relation R are

$$\mathrm{dom}(R) := \{x \mid \exists y\, ((x, y) \in R)\},$$
$$\mathrm{rng}(R) := \{y \mid \exists x\, ((x, y) \in R)\}.$$

Proposition 2.39. If R is a set, then so are $\mathrm{dom}(R), \mathrm{rng}(R)$.

Proof. By Replacement: $\mathrm{dom}(R) = \{x \mid \exists p \in R\, \exists y\, (p = (x, y))\}$, and for each p, there is at most one x such that $\exists y\, (p = (x, y))$, by "Extensionality for pairs" (2.29); similarly for $\mathrm{rng}(R)$. □

Exercise 2.40. Give another proof using Union instead of Replacement, assuming the Kuratowski encoding of pairs (cf. Exercise 2.34).

Corollary 2.41. If R is a binary relation and also a set, then it is a subset of $X \times Y$ for some sets X, Y. In that case, we call R a **binary relation between** X, Y.

Proof. $X := \mathrm{dom}(R)$, $Y := \mathrm{rng}(R)$ works. □

Definition 2.42. A relation f is a **function** if for each x, there is at most one y such that $x\ f\ y$. If such unique y exists, then we denote it by $f(x)$.

If f is a function, $\mathrm{dom}(f) = X$, and $\mathrm{rng}(f) \subseteq Y$, then we say that f is a **function from** X **to** Y, denoted $f : X \to Y$, and call Y a **codomain** of f.

Outside of set theory, functions are usually treated as a primitive type of object, distinct from sets, much as pairs are. The following axioms dictate how we use functions in practice:[9]

(2.43) If $f : X \to Y$ is a function, and $x \in X$, then we get an object $f(x) \in Y$.

(2.44) ("Extensionality for functions") For two functions $f, g : X \to Y$, we have

$$f = g \iff \forall x \in X\, (f(x) = g(x)).$$

(2.45) ("Comprehension for functions") To define a function $f : X \to Y$, specify for each $x \in X$ a unique $f(x) \in Y$. That is, specify a property $\phi(x, y)$ such that $\forall x \in X\, \exists! y \in Y\, \phi(x, y)$.

Exercise 2.46. Verify that the encoding of functions as sets of pairs satisfies these axioms.

[9] The form of these axioms shows that functions are a *negative type*, like sets; cf. Footnote 8.

Remark 2.47. Unlike with pairs (see Exercise 2.31), this standard encoding of functions subjectively feels fairly "canonical", and does not involve the same level of trickery as the encoding of pairs.

Nonetheless, we should still keep in mind the distinction between the *concept* of a function, which is still best thought of as primitive, and its *encoding* as a set of pairs. To emphasize this distinction, people usually define the **graph** of a function $f : X \to Y$ to mean

$$\{(x, f(x)) \mid x \in X\},$$

which formally is the same as f under the standard encoding, but explicitly indicates that we are thinking of f as a set of pairs rather than a function.

Definition 2.48. For two classes X, Y,

$$Y^X := \{f \mid f \text{ is a function } X \to Y\}.$$

This is an abuse of notation: there are several other operations denoted the same way in set theory (see Remark 2.67, Exercise 3.163, Remark 5.30). Less ambiguous notations people sometimes use include ${}^X Y, \mathrm{Fun}(X, Y)$. We think these are too ugly and/or verbose, and so will depend on context for clarity.

Corollary 2.49 (of Definition 2.32)**.** For sets X, Y, so is Y^X.

Proof. Y^X is a set of sets of pairs, i.e., $Y^X \subseteq \mathcal{P}(X \times Y)$. □

We assume you are familiar with other standard concepts related to functions, and will have no difficulties formalizing them into set theory:

Definition 2.50. For relations $R \subseteq X \times Y$ and $S \subseteq Y \times Z$, their **composition** is

$$S \circ R := \{(x, z) \in X \times Z \mid \exists y \in Y\, (x\ R\ y\ S\ z)\}$$
$$\text{(shorthand for } \{p \in X \times Z \mid \exists x \in X\, \exists y \in Y\, \exists z \in Z\, (p = (x, z) \text{ and } x\ R\ y \text{ and } y\ S\ z)\}).$$

(As usual, the order is "wrong", ultimately so that we can write $f(x)$ rather than $(x)f$.)

Exercise 2.51. Prove that if $f : X \to Y$ and $g : Y \to Z$ are functions, then so is $g \circ f : X \to Z$.

Definition 2.52. The **identity function** is (as a class of pairs) the same as the equality relation $=$. The **identity function on** X is its restriction to X, i.e., intersection with $X \times X$.

Exercise 2.53. Prove that relation composition is associative and has id as identity element.

Example 2.54. The **inverse** of a binary relation R is

$$R^{-1} := \{(y, x) \mid (x, y) \in R\}.$$

Exercise 2.55. Let $R \subseteq X \times Y$ be a binary relation.

(a) What does $R^{-1} \circ R \subseteq \mathrm{id}_X$ mean?
(b) What does $R^{-1} \circ R \supseteq \mathrm{id}_X$ mean?
(c) Prove that R is a function $X \to Y$ iff $R^{-1} \circ R$?? id_X and $R \circ R^{-1}$?? id_Y, where each ?? is either $\subseteq$ or $\supseteq$ (which?).
(d) Conclude that for a function $f : X \to Y$, the relation $f^{-1} : Y \to X$ is also a function iff $\forall y \in Y\, \exists! x \in X\, (f(x) = y)$, i.e., f is a **bijection**.
(e) Show that a function $f : X \to Y$ is **injective**, resp., **surjective** (defined the usual way), iff one of the other inclusions in (c) above holds (which?).

Exercise 2.56. For a relation $R \subseteq X \times Y$, define the **image** $R[A] \subseteq Y$ of a subset $A \subseteq X$, specializing to the case when R is a function; $R^{-1}[B] \subseteq X$ is then the **preimage** of $B \subseteq Y$.

Show that taking image under a relation preserves arbitrary unions (first write what this means), and preserves arbitrary intersections iff $R = f^{-1}$ for a function $f : Y \to X$.

2.E. **Independence of encoding, indexed products and (disjoint) unions.** Bijections provide one way of formalizing the idea that the choice of encoding of ordered pairs, functions, etc., is irrelevant:

Proposition 2.57. Let $(,)$ and $(,)'$ be two ways of encoding ordered pairs, both obeying the axioms (2.28) and (2.29). Then there is a bijection F (between the classes of ordered pairs encoded either way) converting between these encodings, namely

$$F(x,y) := (x,y)'.$$

In particular, for any sets (or classes) X, Y, letting $\times, \times'$ denote the Cartesian products defined using either encoding, the above bijection between *all* pairs restricts to a bijection

$$\begin{aligned} F : X \times Y &\longrightarrow X \times' Y \\ (x,y) &\longmapsto (x,y)'. \end{aligned}$$

Proof. We may certainly define the relation F by the above formula, i.e.,

$$F := \{(p,p') \mid \exists x,y\,(p = (x,y) \text{ and } p' = (x,y)')\}.$$

To check that F is a function, we need to know

$$(p,p'_1),(p,p'_2) \in F \implies p'_1 = p'_2.$$

From $(p,p'_1) \in F$, we get that $p = (x_1,y_1)$ and $p'_1 = (x_1,y_1)'$ for some x_1, y_1, while from $(p,p'_2) \in F$, we get that $p = (x_2,y_2)$ and $p'_2 = (x_2,y_2)'$ for some (*a priori* different) x_2, y_2; but by the extensionality axiom (2.29) for $(,)$, from $(x_1,y_1) = p = (x_2,y_2)$ we get $x_1 = x_2$ and $y_1 = y_2$, whence $p'_1 = (x_1,y_1) = (x_2,y_2) = p'_2$. Similarly, F^{-1} is a function. □

Exercise 2.58. Similarly, for any two ways of encoding functions obeying (2.43), (2.44) and (2.45), show that we have a bijection $Y^X \cong Y^{X\prime}$ between the respective sets of functions, for any two sets X, Y.

Even if we accept the standard (Kuratowski) encoding of pairs, note that there are two obvious ways to build triples (and higher n-tuples) from pairs:

$$\begin{aligned} (x,y,z)_1 &:= ((x,y),z), \\ (x,y,z)_2 &:= (x,(y,z)). \end{aligned}$$

More generally, in areas such as real analysis we would want to consider "∞-tuples", i.e., infinite sequences $(x_0, x_1, \dots)$; in fact, we may as well consider arbitrary indexed families $(x_i)_{i\in I}$.

Definition 2.59. An **indexed family** $(x_i)_{i\in I}$ over a set or class I is another name for a function f with domain I, where x_i is another name for $f(i)$.

Definition 2.60. For an indexed family of sets $(A_i)_{i\in I}$, define the **indexed union**

$$\textstyle\bigcup_{i\in I} A_i := \bigcup\{A_i \mid i \in I\}$$

(constructed via the Axioms of Union and Replacement).

Exercise 2.61. Show that the concepts of indexed union and union of a set of sets are interchangeable: conversely, for a set of sets $\mathcal{A}$,

$$\textstyle\bigcup \mathcal{A} = \bigcup_{A\in\mathcal{A}} A.$$

Definition 2.62. For an indexed family of sets $(X_i)_{i\in I}$, its **indexed Cartesian product** $\prod_{i\in I} X_i$ is the set of all indexed families $(x_i)_{i\in I}$ where each $x_i \in X_i$.

Proposition 2.63. If $(X_i)_{i\in I}$ is a family of sets indexed over a set I, then $\prod_{i\in I} X_i$ is a set.

Proof. $\prod_{i\in I} X_i \subseteq (\bigcup_{i\in I} X_i)^I$. □

We now have several ways of encoding n-tuples:

Definition 2.64 (preliminary; see Infinity 3.138)**.**

$$\begin{aligned} 0 &:= \varnothing, \\ 1 &:= \{0\}, \\ 2 &:= \{0, 1\}, \\ 3 &:= \{0, 1, 2\}, \\ &\vdots \end{aligned}$$

Exercise 2.65. Let $n \geq 2$. We may encode n-tuples as

$$(x_0, \ldots, x_{n-1}) := (((x_0, x_1), x_2), \ldots x_{n-1})$$

(or any other way of writing the parentheses). We may also regard the tuple as an indexed family over the domain n. Show that there is a canonical bijection converting between these encodings. For example, for $n = 3$, for any sets X_0, X_1, X_2, we have bijections

$$(X_0 \times X_1) \times X_2 \cong \textstyle\prod_{i \in 3} X_i \cong X_0 \times (X_1 \times X_2).$$

Remark 2.66. Of course, we could not have originally defined ordered pairs via indexed families, since functions were defined in terms of ordered pairs. But the above encoding still works for $n = 2$, *given* the concept of function, yielding *another* encoding of ordered pairs.

Remark 2.67. When $(X)_{i \in I}$ is a constant family of sets, note that our definition of $\prod_{i \in I} X$ agrees with the set of functions X^I (Definition 2.48). In particular, X^n is the set of functions $n \to X$, which is in canonical bijection with (not equal to) $((X \times X) \times \cdots) \times X$.

For "canonical" bijections such as those above, it is common in informal mathematical practice to treat them as equalities, by "identifying" elements in both sets. An important feature of *actual* equality is (one direction of) Extensionality: equal things should be interchangeable in all contexts. Of course, the Axiom of Extensionality tells us that this literally holds only for actually equal sets. But for many constructions used in practice, sets in bijection are also "interchangeable":

Definition 2.68. An operation on sets $F(X_0, \ldots, X_{n-1})$, e.g., $\times, \mathcal{P}, \cap$, is called **functorial** (on bijections[10]) if it comes equipped with, for each bijections $f_i : X_i \cong Y_i$, an *induced bijection* $F(f_0, \ldots, f_{n-1}) : F(X_0, \ldots, X_{n-1}) \cong F(Y_0, \ldots, Y_{n-1})$. These induced bijections should respect composition in the f_i: if we have another family of bijections $g_i : Y_i \cong Z_i$, then we require

$$F(g_0, \ldots, g_{n-1}) \circ F(f_0, \ldots, f_{n-1}) = F(g_0 \circ f_0, \ldots, g_{n-1} \circ f_{n-1}).$$

Exercise 2.69. Show that this implies $F(\mathrm{id}_{X_0}, \ldots, \mathrm{id}_{X_{n-1}}) = \mathrm{id}_{F(X_0, \ldots, X_{n-1})}$ and $F(f_0^{-1}, \ldots, f_{n-1}^{-1}) = F(f_0, \ldots, f_{n-1})^{-1}$.

Example 2.70. $\times$ is a functorial binary operation: for $f_0 : X_0 \cong Y_0$ and $f_1 : X_1 \cong Y_1$, we have

$$\begin{aligned} X_0 \times X_1 &\cong Y_0 \times Y_1 \\ (x_0, x_1) &\mapsto (f_0(x_0), f_1(x_1)), \end{aligned}$$

and it is easily seen that this preserves composition in the f_i.

Example 2.71. "Exponentiation", i.e., sets of functions, is functorial: for f_0, f_1 as above, we have

$$\begin{aligned} X_1^{X_0} &\cong Y_1^{Y_0} \\ h &\mapsto f_1 \circ h \circ f_0^{-1} : Y_0 \to X_0 \to X_1 \to Y_1. \end{aligned}$$

[10]The general context for this concept is the area of math called *category theory*, which we will not go into.

Exercise 2.72. Verify that this preserves composition in the f_i.

Exercise 2.73. Show that $\mathcal{P}$ (powerset) is a functorial unary operation on sets.

Example 2.74. $\bigcup$ (union) is *not* a functorial unary operation. For example, $\{\varnothing\} \cong \{\{\varnothing\}\}$, but $\bigcup\{\varnothing\} = \varnothing \ncong \{\varnothing\} = \bigcup\{\{\varnothing\}\}$.

Exercise 2.75. Show that $\cup$ is not a functorial binary operation either.

This reflects the fact that in mathematical practice, it is unusual to take the union of two (or more) sets without knowing something about how they are related. Usually, we only take union of subsets *of a given ambient set*; or else, we take a *disjoint union* of unrelated sets. This latter concept is again defined up to a choice of encoding:

Definition 2.76. For a family of sets $(X_i)_{i\in I}$ indexed over a set I, its **disjoint union** $\bigsqcup_{i\in I} X_i$ is a set equipped with an indexed family of injections $\iota_i : X_i \to \bigsqcup_{j\in I} X_j$ whose images are disjoint and cover $\bigsqcup_{j\in I} X_j$. In other words:

(2.77) For each $i \in I$ and $x \in X_i$, we have a corresponding element $\iota_i(x) \in \bigsqcup_{i\in I} X_i$.

(2.78) Each $y \in \bigsqcup_{i\in I} X_i$ is equal to such an $\iota_i(x)$ for a unique i and $x \in X_i$.

One ("standard") encoding is given by

$$\bigsqcup_{i\in I} X_i := \{(i, x) \in I \times \bigcup_{i\in I} X_i \mid x \in X_i\},$$
$$\iota_i(x) := (i, x).$$

Exercise 2.79. Show that all encodings of disjoint union obeying these axioms are in canonical bijection with each other. Moreover, $\bigsqcup_{i\in I}$ is a functorial "I-ary operation" (define what this means).

We also mention various other "canonical" bijections commonly used throughout math. These are perhaps not all thought of as converting between different "encodings" of the same concept; nonetheless, one frequently abuses notation/terminology by treating them as equalities.

Example 2.80. For any set X, there is a bijection between subsets of X and their **indicator** (or **characteristic**) functions:

$$\mathcal{P}(X) \cong 2^X$$
$$A \mapsto \begin{pmatrix} \chi_A : X \to 2 = \{0, 1\} \\ x \mapsto \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{else} \end{cases} \end{pmatrix}$$
$$f^{-1}[\{1\}] \leftarrow\!\shortmid f.$$

Example 2.81. For any sets X, Y, Z, we have bijections

$$Z^{X\times Y} \cong (Z^X)^Y$$
$$f \mapsto (y \mapsto (x \mapsto f(x, y)))$$
$$(g(y)(x) \leftarrow\!\shortmid (x, y)) \leftarrow\!\shortmid g,$$

and similarly $Z^{X\times Y} \cong (Z^Y)^X$.

Exercise 2.82. Give a bijection $\mathcal{P}(X \times Y) \cong \mathcal{P}(X)^Y$.

Exercise 2.83. For an indexed family of sets $(X_i)_{i\in I}$ and a set Y, give a bijection

$$Y^{\bigsqcup_{i\in I} X_i} \cong \prod_{i\in I} Y^{X_i}.$$

Exercise 2.84. In particular, $\mathcal{P}(\bigsqcup_{i\in I} X_i) \cong \prod_{i\in I} \mathcal{P}(X_i)$.

3. Induction

We turn now to developing the general theory of induction. In contrast to Section 2, which was largely about set theory as the low-level "machine code" of math, here we are concerned with ideas which are very widely used in everyday math outside of logic. The low-level aspects will reenter the picture eventually (especially once we turn to naturals and ordinals); but for the most part, our discussion will feel much more like "normal" abstract math, akin to algebra, analysis, etc.

Our approach is to treat induction *axiomatically*: rather than answering "what is induction", we will define what can constitute *a* notion of induction. This is similar to how topology generalizes from *the* notion of limit in calculus to spaces equipped with *a* notion of limit, or how rings generalize from *the* arithmetic of numbers to *a* notion of numbers equipped with arithmetic operations, etc. The general idea is: we have a set X of elements that we're inducting on, and a way of "deriving" new elements from previous ones; we say a *principle of induction* holds if everything can eventually be derived. The original induction on $\mathbb{N}$ will be a (very) special case of the general notion.

3.A. Monotone set operators and the Knaster–Tarski fixed point theorem.

Definition 3.1. A **monotone set operator** $T : \mathcal{P}(X) \to \mathcal{P}(X)$ on a set X is a function obeying

$$\forall A, B \in \mathcal{P}(X)\,(A \subseteq B \implies T(A) \subseteq T(B)).$$

A subset $A \subseteq X$ is **T-closed** if $T(A) \subseteq A$.

There are many possible interpretations of this simple definition. For the purposes of induction, we think of T as specifying, for each subset $A \subseteq X$, the set of new elements $T(A)$ which can be "derived" from A. Being T-closed means that all elements "derivable" from A are already in A.

Example 3.2. We have a monotone set operator on $X = \mathbb{N}$ given by

$$T(A) := \{0\} \cup \{n + 1 \mid n \in A\}.$$

In other words, we start with 0 (the base case), and can derive $n + 1$ from n (the inductive case). The only T-closed subset of $\mathbb{N}$ is all of $\mathbb{N}$ (this will be taken as the *definition* of $\mathbb{N}$; see Infinity 3.138).

Example 3.3. We have another monotone set operator on $\mathbb{N}$, given by

$$\begin{aligned} T(A) &:= \{n \in \mathbb{N} \mid \forall m < n\,(m \in A)\} \\ &= \{n \in \mathbb{N} \mid n \subseteq A\} \quad \text{(recalling Definition 2.64; see also Infinity 3.138).} \end{aligned}$$

This says that n can be derived once we know everything smaller, and corresponds to the principle of "strong induction"; see Example 3.11.

Example 3.4. Let X be any set, and let $(f_i : X^{N_i} \to X)_{i \in I}$ be a family of "N_i-ary operations" on X, where the N_i are arbitrary sets. The set X equipped with such a family $(f_i)_{i \in I}$ is sometimes called an **algebra**, or more verbosely, a *first-order structure over a functional signature*. Examples:

(a) $\mathbb{R}$ equipped with $+ : \mathbb{R}^2 \to \mathbb{R}$, $\cdot : \mathbb{R}^2 \to \mathbb{R}$, $- : \mathbb{R} \to \mathbb{R}$, $0 : \mathbb{R}^0 \to \mathbb{R}$, and $1 : \mathbb{R}^0 \to \mathbb{R}$, or a subset thereof, e.g., only $+, -, 0$.

(b) $\mathbb{R}^n$ equipped with $+ : (\mathbb{R}^n)^2 \to \mathbb{R}^n$ (vector addition), $\vec{0} : (\mathbb{R}^n)^0 \to \mathbb{R}^n$ (zero vector), and for each $a \in \mathbb{R}$, the unary operation $a \cdot (-) : \mathbb{R}^n \to \mathbb{R}^n$ (scalar multiplication).

(c) $\mathcal{P}(X)$ for an arbitrary set X, equipped with $\cap, \cup, \neg, \varnothing, X$ (where $\neg A := X \setminus A$).

(d) $[-\infty, \infty]$ equipped with $\limsup : [-\infty, \infty]^{\mathbb{N}} \to [-\infty, \infty]$ (or lim, if we allow partial functions).

(e) $\mathbb{N}$ equipped with $0 : \mathbb{N}^0 \to \mathbb{N}$ and $\mathrm{S} : \mathbb{N}^1 \to \mathbb{N}$ where $\mathrm{S}(n) := n + 1$ (**successor**).

We may then define the monotone set operator

$$T(A) := \{f_i(\vec{x}) \mid i \in I \text{ and } \vec{x} \in A^{N_i}\}.$$

A T-closed set is then one closed under the operations. For example, in (b), a T-closed set is a vector subspace of $\mathbb{R}^n$. In (d), T-closed means topologically closed. (e) recovers T from Example 3.2.

Example 3.5. Let X be an arbitrary set, and define the monotone set operator T on X^2 by

$$\begin{aligned} T(A) &:= \{(x,x) \mid x \in X\} \cup \\ &= \{(y,x) \mid (x,y) \in A\} \cup \\ &= \{(x,z) \mid (x,y),(y,z) \in A\}. \end{aligned}$$

Then $A \subseteq X^2$ is T-closed iff A is reflexive, symmetric, and transitive, i.e., an equivalence relation on X (see Definition 3.44).

Theorem 3.6 (Knaster–Tarski fixed point)**.** Let $T : \mathcal{P}(X) \to \mathcal{P}(X)$ be a monotone set operator. For every $A \subseteq X$, there is a smallest T-closed $\overline{T}(A) \supseteq A$, called the **$T$-closure** of A, or sometimes the T-closed subset **generated by** A. Moreover, $T(\overline{T}(\varnothing)) = \overline{T}(\varnothing)$.

Proof. The first claim follows from combining the following two facts, which are useful on their own:

Lemma 3.7. For any monotone set operator $T : \mathcal{P}(X) \to \mathcal{P}(X)$, the T-closed sets are closed under arbitrary intersections, i.e., if $\mathcal{A} \subseteq \mathcal{P}(X)$ is a set of T-closed sets, then so is $\bigcap \mathcal{A}$.

(This includes the case $\bigcap \mathcal{A} = \varnothing$ from Remark 2.21.)

Proof. For each $A \in \mathcal{A}$, we have $T(\bigcap \mathcal{A}) \subseteq T(A) \subseteq A$ by monotonicity, whence $T(\bigcap \mathcal{A}) \subseteq \bigcap \mathcal{A}$. □

Proposition 3.8. For any set X and family of subsets $\mathcal{A} \subseteq \mathcal{P}(X)$, the following are equivalent:

(i) $\mathcal{A}$ is closed under intersections (including $\bigcap \varnothing = X$ from Remark 2.21).
(ii) For every $A \subseteq X$, there is a smallest $\overline{A} \in \mathcal{A}$ such that $A \subseteq \overline{A}$.

(Such an $\mathcal{A} \subseteq \mathcal{P}(X)$ is sometimes called a **closure system**.)

Proof. (i) $\implies$ (ii) Let $\overline{A} := \bigcap\{B \in \mathcal{A} \mid A \subseteq B\}$. Then $\overline{A} \in \mathcal{A}$ since $\mathcal{A}$ is closed under intersections, and every other $B \in \mathcal{A}$ such that $A \subseteq B$ is one of the sets we're intersecting, hence contains $\overline{A}$.
(ii) $\implies$ (i) This follows from the preceding lemma, since $A \mapsto \overline{A}$ is easily monotone: if $A \subseteq B$, then $A \subseteq B \subseteq \overline{B} \in \mathcal{A}$, whence $\overline{A} \subseteq \overline{B}$. □

Finally, to show $T(\overline{T}(\varnothing)) = \overline{T}(\varnothing)$: $\subseteq$ is because $\overline{T}(\varnothing)$ is T-closed; then by monotonicity, $T(T(\overline{T}(\varnothing))) \subseteq T(\overline{T}(\varnothing))$, whence $T(\overline{T}(\varnothing))$ is T-closed, and contains $\varnothing$, whence $\overline{T}(\varnothing) \subseteq T(\overline{T}(\varnothing))$. □

Definition 3.9. We call a monotone set operator $T : \mathcal{P}(X) \to \mathcal{P}(X)$ **inductive** if $\overline{T}(\varnothing) = X$, i.e., the only T-closed subset of X is the entirety of X.

The following is merely a restatement of the definition of $\overline{T}(\varnothing)$:

Principle of T-induction 3.10. Let $T : \mathcal{P}(X) \to \mathcal{P}(X)$ be an inductive monotone set operator.

(a) For any $A \subseteq X$, if $T(A) \subseteq A$, then $A = X$.
(b) Equivalently, for any property $\phi(x)$ of elements $x \in X$, if
- $\forall x \in X\, (\underbrace{x \in T(\{y \in X \mid \phi(y)\})}_{\text{IH}} \implies \phi(x))$ (inductive case),

then $\forall x \in X$, $\phi(x)$.
(c) Equivalently, any $\varnothing \neq B \subseteq X$ contains a "T-minimal element" $x \in T(X \setminus B)$.

These statements are trivially equivalent: (a) $\iff$ (b) by taking $A := \{x \mid \phi(x)\}$ and $\phi(x) :\iff x \in A$; and (a) $\iff$ (c) by taking A, B to be complements of each other and taking the contrapositive. The statement (b) is intended to resemble a conventional statement of the ordinary principle of (weak or strong) induction for $\mathbb{N}$; the "induction hypothesis" (IH) above says "x can be derived from y for which ϕ is known". The statement (c) is intended to resemble the "well-ordering principle"; the conclusion $x \in T(X \setminus B)$ means "$x \in B$ can be derived from only the elements in the complement of B", or that "x does not depend on any other elements of B".

Example 3.11. For T from Example 3.2 which closes under 0 and successor S, the above becomes

(a) For any $A \subseteq \mathbb{N}$, if $0 \in A$ and $\{n+1 \mid n \in A\} \subseteq A$, then $A = \mathbb{N}$.
(b) For any property $\phi(x)$ of natural numbers $x \in \mathbb{N}$, if
   - $\forall x \in \mathbb{N}\,((x = 0 \text{ or } x = y+1 \text{ for some } y \text{ s.t. } \phi(y)) \implies \phi(x))$,

   then $\forall x \in \mathbb{N}$, $\phi(x)$.

For the T corresponding to "strong induction" from Example 3.3, we instead get

(a) For any $A \subseteq \mathbb{N}$, if $\forall n\,(n \subseteq A \implies n \in A)$, then $A = \mathbb{N}$.
(b) For any property $\phi(x)$ of natural numbers x, if
   - $\forall x \in \mathbb{N}\,((\forall y < x,\ \phi(y)) \implies \phi(x))$,

   then $\forall x \in \mathbb{N}$, $\phi(x)$.
(c) Any $\varnothing \neq B \subseteq \mathbb{N}$ contains some x such that $\forall y < x\,(y \in \mathbb{N} \setminus B)$, i.e., $x \in B$ is minimal.

This last statement (c) is usually known as the "well-ordering principle" (for $\mathbb{N}$).

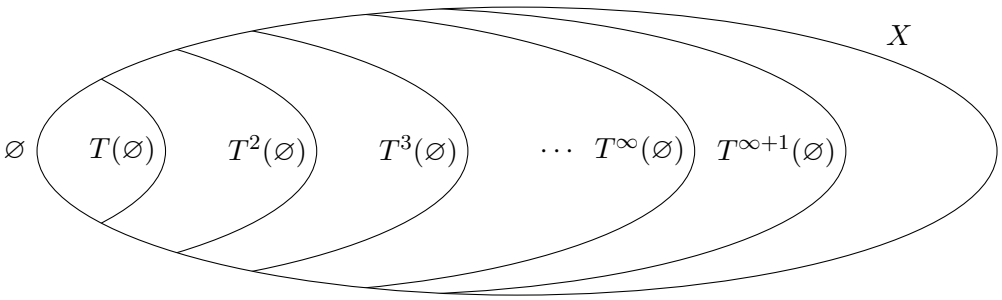
Exercise 3.12. What is the contrapositive statement (c) for the T corresponding to ordinary induction from Example 3.2?

Remark 3.13. If a monotone $T : \mathcal{P}(X) \to \mathcal{P}(X)$ as in Knaster–Tarski 3.6 is not inductive, then we may restrict it to the subset $\overline{T}(\varnothing) \subseteq X$ to get an inductive operator $T : \mathcal{P}(\overline{T}(\varnothing)) \to \mathcal{P}(\overline{T}(\varnothing))$. We thus call $\overline{T}(\varnothing) \subseteq X$ the **inductive part** of X (equipped with T).

Remark 3.14. The proof of the Knaster–Tarski Theorem 3.6 is a "top-down" or (in philosophical terminology) *impredicative* construction: in order to build the *smallest* set obeying some condition, we had to look at *all possible* such sets. In other words, to build a simple thing, we had to look at everything more complicated than it. This is technique is very powerful, but a bit unsatisfying, since it tells us basically nothing about what the simple thing actually looks like.

A perhaps more satisfying "bottom-up" construction is to start with $\varnothing$ (nothing), then add everything derivable from that, yielding $T(\varnothing)$, then add everything derivable from that, yielding $T^2(\varnothing) = T(T(\varnothing))$, etc. After infinitely many steps, we're done if everything derivable from $T^\infty(\varnothing) := \bigcup_{n\in\mathbb{N}} T^n(\varnothing)$ can already be derived from a finite stage; this will be true if the notion of "derivation" defined by T is "finitary" in nature, e.g., if we're closing under finitary operations such as $+, \cdot$ in Example 3.4(a). But if we're closing under operations such as $\limsup$ (Example 3.4(d)) that can take an infinite sequence as input, then it's possible that the sequence includes a term in $T^n(\varnothing)$ for each n. So we have to keep going: $T^{\infty+1}(\varnothing) := T(T^\infty(\varnothing))$, etc.

(3.15)



This description is only informal at this stage, because this "transfinite" process is an *instance* of the general inductive processes we're aiming to formalize; see Section 3.J.

Remark 3.16. For a monotone $T : \mathcal{P}(X) \to \mathcal{P}(X)$, more generally starting from any subset $A \subseteq X$, we have a "relative" induction principle for $\overline{T}(A)$: for any $B \subseteq X$, if

- $A \subseteq B$, and
- $T(B) \subseteq B$,

then $\overline{T}(A) \subseteq B$.

Example 3.17. For $T : \mathcal{P}(\mathbb{R}^n) \to \mathcal{P}(\mathbb{R}^n)$ which closes under vector operations from Example 3.4(b), this says that to prove that a subset $B \subseteq \mathbb{R}^n$ contains the linear span of some vectors $\vec{v}_i$, it suffices to check that B contains each $\vec{v}_i$ and is itself a linear subspace.

For example, this is how one usually proves that $\operatorname{span}(A) \subseteq A^{\perp\perp}$, for every $A \subseteq \mathbb{R}^n$: clearly $A \subseteq A^{\perp\perp}$; and the orthogonal complement $B^\perp$ of every subset is a linear subspace.

However, this "relative" principle of induction is in some sense not needed, since we can always reduce it to the "absolute" form:

Exercise 3.18. Let $T : \mathcal{P}(X) \to \mathcal{P}(X)$ be a monotone set operator, and fix $A \subseteq X$.

(a) Verify that $T_A(B) := A \cup T(B)$ is also a monotone set operator on X.
(b) Show that $\overline{T_A}(\varnothing) = \overline{T}(A)$, and that the principle of induction 3.10 for $\overline{T_A}(\varnothing)$ agrees with the "relative" principle of induction for $\overline{T}(A)$ from Remark 3.16.
(c) Conclude that $\overline{T}(A) = A \cup T(\overline{T}(A))$.

Exercise 3.19. One other slightly odd feature of the general setup of Knaster–Tarski is that a general monotone $T : \mathcal{P}(X) \to \mathcal{P}(X)$ need not obey $A \subseteq T(A)$ for all $A \subseteq X$, e.g., the T corresponding to ordinary induction (Example 3.2). In other words, we should think of $T(A)$ as those elements "derivable" from A *in exactly one step*; the preexisting elements are not necessarily considered to be "trivially derivable". Show that this is also an inessential distinction:

(1) Verify that $T'(A) := A \cup T(A)$ is also a monotone set operator.
(2) Verify that T'-closed sets are the same as T-closed sets, hence $\overline{T'} = \overline{T}$.

In spite of this trick, it is useful to consider arbitrary monotone T in the general theory of induction. One advantage of not requiring $A \subseteq T(A)$ is that for the same monotone T, we also have

Exercise 3.20 (dual Knaster–Tarski)**.** Show, using Knaster–Tarski (without redoing its proof), that for every $A \subseteq X$, there is a greatest T**-open** $T^\circ(A) := B \subseteq A$, meaning $B \subseteq T(B)$, called the T**-interior** of A. Moreover, $T^\circ(X) = T(T^\circ(X))$; and we have the **principle of coinduction**: for any other T-open $B \subseteq T(B) \subseteq X$, we have $B \subseteq T^\circ(X)$.

3.B. **Examples of induction.** In this subsection, we assume we know about ordinary induction, and other basic facts, for $\mathbb{N}$, $\mathbb{R}$, etc. Our goal is to demonstrate the power of the general framework of induction, via some interesting examples from many different areas of math.

First, an amusing example of ordinary induction for $\mathbb{N}$:

Example 3.21 (blue-eyed islanders)**.** On an island live 500 inhabitants, 100 of whom have blue eyes while the other 400 have brown eyes. These islanders are extremely smart, able to immediately deduce any logically true statements. However, they have a very strict religion that forbids one from knowing one's own eye color; anyone who finds out their own eye color is required to commit ritual suicide the following day at noon in the village square, where all the other islanders can see. One day, a foreigner visits the island and casually remarks at a village gathering with everyone attending, "It's lovely to see another blue-eyed person like myself in this part of the world." What happens?

Solution. We claim that all of the blue-eyed people will simultaneously commit suicide 100 days after the foreigner makes the remark. More generally, we will prove by induction that if there are $n \geq 1$ blue-eyed people, they will all commit suicide n days after hearing the remark. If $n = 1$, the blue-eyed person finds out they have blue eyes, and so must commit suicide the next day. Now suppose the claim holds for n; we prove it for $n + 1$. Each blue-eyed person sees n other blue-eyed people, hence knows there are either $n + 1$ blue-eyed people in total (if they also have blue eyes) or n (if they don't). If there were n blue-eyed people, by the IH, they would commit suicide on the nth day. So on the nth day, since no one dies, every blue-eyed person figures out there are $n + 1$ blue-eyed people, hence that they have blue eyes, hence must commit suicide on day $n + 1$. □

Exercise 3.22. What happens to the brown-eyed people?

Exercise 3.23. What new information did the foreigner introduce that wasn't already known?

Remark 3.24. The philosophical/sociological/economic phenomenon this puzzle illustrates is known as *common knowledge*: everyone knows something, and everyone knows that everyone knows it, and everyone knows that everyone knows that everyone knows it, etc., which can be quite different than simply everyone knowing it. More complicated forms of induction can show up in common knowledge situations; see **TODO**.

We now turn to other forms of induction, i.e., other inductive set operators $T : \mathcal{P}(X) \to \mathcal{P}(X)$. In practice, one usually does not bother to explicitly write out the T; rather, one merely states the principle of induction, which recall is equivalent to the assertion that T is inductive, and from which it is usually easy to read off the definition of T.

Proposition 3.25 (principle of Cauchy induction)**.** Let $f : \mathbb{N} \to \mathbb{N}$ be a strictly increasing function, i.e., $f(n) < f(n+1)$. Suppose $A \subseteq \mathbb{N}$ such that

- $f(0) \in A$;
- $f(n) \in A \implies f(n+1) \in A$;
- $n+1 \in A \implies n \in A$.

Then $A = \mathbb{N}$.

Proof. By ordinary induction, $f(n) \in A$ for every $n \in \mathbb{N}$. Since $0 \le f(0) < f(1) < \cdots < f(n)$, $n \le f(n)$ for all $n \in \mathbb{N}$ (technically, this is again by ordinary induction on n). By ordinary induction on k and the third property above, $n + k \in A \implies n \in A$. Thus for every $n \in \mathbb{N}$, from $f(n) = n + (f(n) - n) \in A$, we get $n \in A$. □

Exercise 3.26. What is the $T : \mathcal{P}(\mathbb{N}) \to \mathcal{P}(\mathbb{N})$ for which the principle of induction yields the above?

Theorem 3.27 (AM–GM inequality)**.** For any $n \ge 1$ and $x_1, \dots, x_n \in [0, \infty)$, we have

$$\frac{x_1 + \cdots + x_n}{n} \ge \sqrt[n]{x_1 \cdots x_n}.$$

Proof. By Cauchy induction with the increasing function $f(n) = 2^n$.

- For $n = 1$, it's trivially true: $x_1 \le x_1$.
- For $n = 2$, expand $(\sqrt{x} - \sqrt{y})^2 \ge 0$ to get $x - 2\sqrt{xy} + y \ge 0$.
- Suppose it's true for 2^n; we prove it for 2^{n+1}.

$$\begin{aligned}
\frac{x_1 + \cdots + x_{2^{n+1}}}{2^{n+1}} &= \frac{\frac{x_1 + \cdots + x_{2^n}}{2^n} + \frac{x_{2^n+1} + \cdots + x_{2^{n+1}}}{2^n}}{2} \\
&\ge \frac{\sqrt[2^n]{x_1 \cdots x_{2^n}} + \sqrt[2^n]{x_{2^n+1} \cdots x_{2^{n+1}}}}{2} && \text{by IH} \\
&\ge \sqrt{\sqrt[2^n]{x_1 \cdots x_{2^n}}\ \sqrt[2^n]{x_{2^n+1} \cdots x_{2^{n+1}}}} && \text{by } n = 2 \text{ case} \\
&= \sqrt[2^{n+1}]{x_1 \cdots x_{2^{n+1}}}.
\end{aligned}$$

- Finally, suppose it's true for $n+1$; we prove it for n. WLOG not every $x_i \ne 0$. Then

$$\begin{aligned}
\frac{x_1 + \cdots + x_n}{n} &= \frac{x_1 + \cdots + x_n + \frac{x_1 + \cdots + x_n}{n}}{n+1} \\
&\ge \sqrt[n+1]{x_1 \cdots x_n \left(\frac{x_1 + \cdots + x_n}{n}\right)} && \text{by IH.}
\end{aligned}$$

Raise to the $(n+1)$th power, divide by $\frac{x_1 + \cdots + x_n}{n} > 0$, and take the nth root. □

Definition 3.28. The **lexicographical ordering** on $\mathbb{N}^2$ is the binary relation defined as follows:

$$(a,b) <_{\text{lex}} (c,d) \iff (a < c) \text{ or } (a = c \text{ and } b < d).$$

Proposition 3.29 (principle of lexicographical induction on $\mathbb{N}^2$)**.** Let $A \subseteq \mathbb{N}^2$ such that

- for every $(a,b) \in \mathbb{N}^2$, if every $(c,d) <_{\text{lex}} (a,b)$ is in A, then $(a,b) \in A$.

Then $A = \mathbb{N}^2$.

Proof. We prove by (strong) induction on a that for every $a \in \mathbb{N}$, for every $b \in \mathbb{N}$, $(a,b) \in A$.

- Assume (IH) that for every $c < a$, for every $d \in \mathbb{N}$, $(c,d) \in A$. We now induct on b.
  - Assume (IH2) that for every $d < b$, $(a,d) \in A$. Then for every $(c,d) <_{\text{lex}} (a,b)$, either
    - $c < a$ in which case $(c,d) \in A$ by (IH), or
    - $c = a$ and $d < b$ in which case $(c,d) \in A$ by (IH2).

    Thus every $(c,d) <_{\text{lex}} (a,b)$ is in A, and so $(a,b) \in A$ by our assumption on A. □

Example 3.30 (Ackermann function)**.** Define the following computation on finite nonempty sequences of natural numbers, that takes a sequence and replaces the last two terms as follows:

$$\begin{aligned}
(a_0, \ldots, a_{n-1}, 0, y) &\longrightarrow (a_0, \ldots, a_{n-1}, y+1),\\
(a_0, \ldots, a_{n-1}, x+1, 0) &\longrightarrow (a_0, \ldots, a_{n-1}, x, 1),\\
(a_0, \ldots, a_{n-1}, x+1, y+1) &\longrightarrow (a_0, \ldots, a_{n-1}, x, x+1, y).
\end{aligned}$$

For example:

$$\begin{aligned}
(1,2) &\longrightarrow (0,1,1)\\
&\longrightarrow (0,0,1,0)\\
&\longrightarrow (0,0,0,1)\\
&\longrightarrow (0,0,2)\\
&\longrightarrow (0,3)\\
&\longrightarrow (4).
\end{aligned}$$

Try starting with $(3,3)$ instead. [Hint: if you're very fast, it'll take you around 2 hours.]

Theorem 3.31. This computation always terminates with a single term.

Proof. First, we prove that starting from any sequence $(a_0, \ldots, a_n, a_{n+1})$ with at least two terms, the computation eventually reaches some $(a_0, \ldots, b)$, by lexicographical induction on (a_n, a_{n+1}). Assume (IH) that this happens for every $(b_0, \ldots, b_m, b_{m+1})$ with $(b_m, b_{m+1}) <_{\text{lex}} (a_n, a_{n+1})$.

- If $a_n = 0$, we immediately get $(a_0, \ldots, a_{n+1} + 1)$.
- If $a_n > 0$ but $a_{n+1} = 0$, we get $(a_0, \ldots, a_n - 1, 1)$, which has the same length; since $(a_n - 1, 1) <_{\text{lex}} (a_n, 0)$, by the IH, we eventually reach some $(a_0, \ldots, b)$.
- If $a_n, a_{n+1} > 0$, we get $(a_0, \ldots, a_n - 1, a_n, a_{n+1} - 1)$ which has one more term, and by the IH eventually reaches some $(a_0, \ldots, a_n - 1, b)$; now since $(a_n - 1, b) <_{\text{lex}} (a_n, a_{n+1})$, this eventually becomes some $(a_0, \ldots, c)$.

Now by induction on n, every sequence of length $n > 0$ eventually reaches a single term. □

Remark 3.32. The Ackermann function $A : \mathbb{N}^2 \to \mathbb{N}$, that computes the single term above resulting from a pair of terms, is historically important as the first example of a function which can be computed by a program, but cannot be computed in a programming language that has only `if...else` clauses and loops of the form `for i = 0,...,n`. Such programs are called **primitive recursive**, and include virtually all algorithms used in the real world. (They include way more than commonly considered classes of functions in computational complexity theory, e.g., NP, EXPSPACE.)

For a more general discussion of lexicographical induction, see Exercises 3.159 and 3.163.

Proposition 3.33 (real number induction)**.** Let $A \subseteq [0,\infty)$ (the nonnegative reals) such that

(i) $0 \in A$;
(ii) A is downward-closed, i.e., $y \le x \in A \implies y \in A$;
(iii) A is closed under increasing limits, i.e., if $x_0 < x_1 < \cdots \in A$ is bounded, then $\lim_{n\to\infty} x_n \in A$;
(iv) for every $x \in A$, there is some $\varepsilon(x) > 0$ such that $\varepsilon(x) \in A$.

Then $A = [0,\infty)$.

Proof. We need to use Dedekind-completeness of $\mathbb{R}$: any nonempty subset A of $\mathbb{R}$ with an upper bound has a least upper bound $\sup A$. (This is a defining property of $\mathbb{R}$ that distinguishes it from $\mathbb{Q}$, that you would see in a real analysis course.) Suppose $A \neq [0,\infty)$. Then since A is downward-closed, any element of $[0,\infty) \setminus A$ is an upper bound for A. We also know $0 \in A$, so $\sup A$ exists. We must have $\sup A \in A$: if not, then $\sup A > 0$ (since $0 \in A$), and we can find a sequence $0 \le x_0 < x_1 \cdots < \sup A$ converging to $\sup A$ from below; since each $x_i < \sup A$, x_i is not an upper bound for A, hence is below some element of A, hence in A by downward-closure; but this contradicts (iii). But then $\sup A \in A$ is the greatest element of A, contradicting (iv). □

Theorem 3.34. $\mathbb{R}$ is **connected**: if we partition $\mathbb{R} = A \sqcup B$ nontrivially so that both $A, B \neq \varnothing$, then some point in one of A, B is a limit of points in the other set.

This is false if we replace $\mathbb{R}$ with e.g., $[0,1] \cup [2,3]$, or $\mathbb{Q} = (\mathbb{Q} \cap (-\infty, \sqrt{2})) \sqcup (\mathbb{Q} \cap (\sqrt{2}, \infty))$.

Proof. Let $\mathbb{R} = A \sqcup B$, WLOG with $0 \in A$. Suppose that no point in either A or B may be obtained as a limit of points in the other set. We prove by induction that $[0,x) \subseteq A$ for all $x \in [0,\infty)$.

(i) Clearly $[0,0) = \varnothing \subseteq A$.
(ii) Clearly, if $[0,x) \subseteq A$ and $y \le x$, then $[0,y) \subseteq A$.
(iii) If $x_0 < x_1 < \cdots \nearrow x$ such that each $[0,x_n) \subseteq A$, then $[0,x) = \bigcup_n [0,x_n) \subseteq A$.
(iv) Finally, suppose $[0,x) \subseteq A$, but for every $\varepsilon > 0$, we have $[0,x+\varepsilon) \not\subseteq A$, i.e., $[0,x+\varepsilon) \cap B \neq \varnothing$. If $x \in A$, then we may write x as a limit of points $x_n \in [0, x+1/n) \cap B \subseteq (x, x+1/n)$. If $x \in B$, then $x > 0$ since $0 \in A$; then we may write x as a limit of points in $[0,x) \subseteq A$. Both of these contradict our assumption, so there must be some $\varepsilon > 0$ such that $[0,x+\varepsilon) \subseteq A$.

It follows that $[0,\infty) = \bigcup_{x\in[0,\infty)} [0,x) \subseteq A$; similarly $(-\infty, 0] \subseteq A$. □

Theorem 3.35 (Heine–Borel)**.** Let $\mathcal{A}$ be a set of open intervals $(a,b) \subseteq \mathbb{R}$ such that $[0,1] \subseteq \bigcup \mathcal{A}$. Then there is finite $\mathcal{F} \subseteq \mathcal{A}$ such that $[0,1] \subseteq \bigcup \mathcal{F}$.

Proof. We will prove by induction that for any $x \in [0,1]$, there is a finite $\mathcal{F} \subseteq \mathcal{A}$ with $[0,x] \subseteq \bigcup \mathcal{F}$.

(i) For $x = 0$, $[0,0] = \{0\}$ is contained a single interval in $\mathcal{A}$.
(ii) If finitely many intervals in $\mathcal{A}$ cover $[0,x]$, then they also cover $[0,y]$ for $y \le x$.
(iii) Let $x_0 < x_1 < \cdots \nearrow x$, and assume each $[0,x_n]$ is contained in a finite union of intervals in $\mathcal{A}$. Since $[0,1] \subseteq \bigcup \mathcal{A}$, x belongs to one interval $(a,b) \in \mathcal{A}$. Since $x_i \nearrow x$, there is n such that $x_n \in (a,b)$, whence $[x_n, x] \subseteq (a,b)$. By the induction hypothesis, there is finite $\mathcal{F} \subseteq \mathcal{A}$ covering $[0,x_n]$. Then $\mathcal{F} \cup \{(a,b)\}$ covers $[0,x]$.
(iv) Finally, assume that $[0,x]$ has a finite subcover. Then x belongs to one interval $(a,b) \in \mathcal{A}$. For sufficiently small $\varepsilon > 0$ (e.g., $\varepsilon := (b-x)/2$), also $x + \varepsilon \in (a,b)$, whence $[x, x+\varepsilon] \subseteq (a,b)$, whence a finite subcover $\mathcal{F} \subseteq \mathcal{A}$ of $[0,x]$ together with (a,b) covers $[0, x+\varepsilon]$. □

Corollary 3.36 (Extreme Value Theorem)**.** Every continuous $f : [0,1] \to \mathbb{R}$ achieves a maximum.

Proof. Suppose not. Then $[0,1] \subseteq f^{-1}[(-\infty, \sup f)] = \bigcup_n f^{-1}[(-\infty, \sup f - 1/n)]$; but $[0,1]$ is not contained in the union of any finitely many of these preimages, by definition of $\sup f$. Since f is continuous, each of these preimages is itself a union of open intervals $f^{-1}[(-\infty, \sup f - 1/n)] = \bigcup_i (a_{n,i}, b_{n,i})$; so $[0,1]$ is not contained in a union of finitely many of the $(a_{n,i}, b_{n,i})$ either. □

3.C. **Well-founded relations.** A general monotone set operator $T : \mathcal{P}(X) \to \mathcal{P}(X)$ provides a method of "generating" or "deriving" new elements from existing ones. However, it need not provide a meaningful notion of the "predecessors" of a given element $x \in X$, which are "required" to derive it: indeed, note that if $x \in T(A)$, then we may always enlarge A with some redundant elements, e.g., to all of X, and still have $x \in T(X)$. We next consider those special kinds of operators T which do come equipped with a meaningful notion of "predecessor", which admit a richer array of inductive techniques, as well as an elegant classification theory that allows us to say precisely when one notion of induction is "stronger" than another.

Definition 3.37. Let $\prec \subseteq X^2$ ("precedes", `\prec` in TEX) be an arbitrary binary relation on a set. The **induced monotone set operator** is

$$\begin{aligned} T = T_\prec : \mathcal{P}(X) &\longrightarrow \mathcal{P}(X) \\ A &\longmapsto \{x \in X \mid \forall y \prec x \, (y \in A)\} \\ &= \{x \in X \mid \downarrow x \subseteq A\} \end{aligned}$$

where

$$\downarrow x = \downarrow_\prec x := \{y \in X \mid y \prec x\}$$

is the set of $\prec$**-predecessors** of x. We call $A \subseteq X$ $\prec$**-closed** if it is T-closed, i.e., if every $x \in X$ with $\downarrow x \subseteq A$ is itself in A. Thus, we also call $\overline{T}(A)$ the $\prec$**-closure** of A.

This notion captures precisely the aforementioned intuition about a monotone $T : \mathcal{P}(X) \to \mathcal{P}(X)$ giving to each $x \in X$ a set of "required predecessors". The most conceptual way to see this is via some abstract nonsense involving general set algebra:

Exercise 3.38. Let X, Y be two sets.

(a) Recall (Exercise 2.56) that for a binary relation $R \subseteq X \times Y$ and $A \subseteq X$, we have the **image** $R[A] \subseteq Y$; and this operation preserves unions in A. Show that this yields a bijection

$$\begin{aligned} \mathcal{P}(X \times Y) &\cong \{F : \mathcal{P}(X) \to \mathcal{P}(Y) \mid F \text{ preserves arbitrary } \bigcup\} \\ R &\mapsto (A \mapsto R[A]). \end{aligned}$$

(The RHS may be thought of as a different encoding of the concept of "binary relation".)

(b) Conclude that we also have a bijection

$$\begin{aligned} \mathcal{P}(X \times Y) &\cong \{F : \mathcal{P}(X) \to \mathcal{P}(Y) \mid F \text{ preserves arbitrary } \bigcap\} \\ R &\mapsto (A \mapsto R\langle A\rangle), \end{aligned}$$

where $R\langle A\rangle$ is the **coimage**, the de Morgan dual of the image:

$$R\langle A\rangle := \neg R[\neg A] = \{y \in Y \mid \forall x \; R \, y \, (x \in A)\}.$$

A special case is $T_\prec(A)$ defined above (when $R = \prec$ lives on a single set).

(c) Show that moreover,

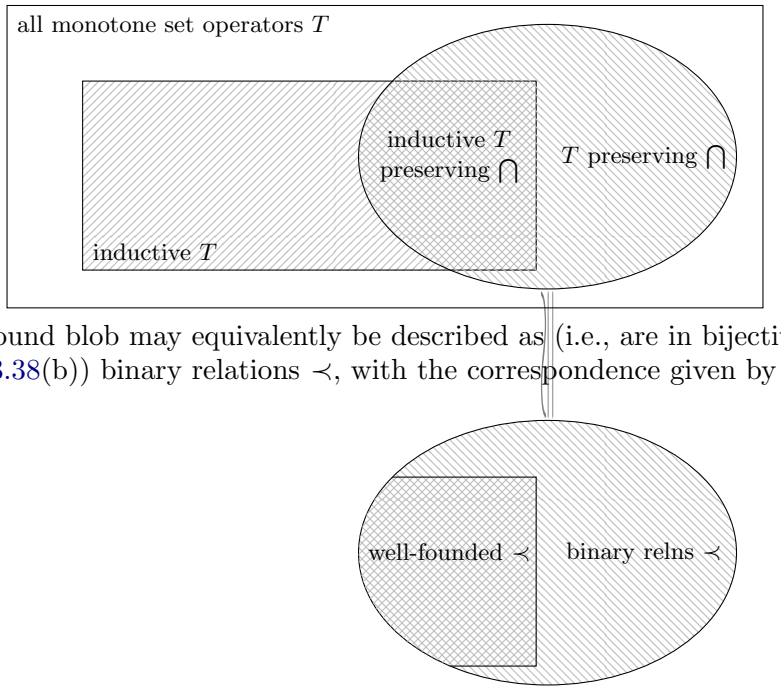
$$F : \mathcal{P}(X) \to \mathcal{P}(Y) \text{ preserves } \bigcap \iff \forall y \in Y \, \exists \text{ smallest } A \subseteq X \text{ s.t. } y \in F(A),$$

namely $A = R^{-1}[\{y\}]$ for the R corresponding to F via (b). In the case $R = \prec$, this is the set of predecessors $\downarrow x$ from above.

(d) Show that most of the non-standard examples of monotone T considered in the two preceding subsections are *not* induced by any binary relation $\prec$, including most algebraic "closure" operators (e.g., closing under vector operations in a vector space, Example 3.4), Cauchy induction (Proposition 3.25), and real number induction (Proposition 3.33).
[Find two disjoint sets which both generate the same element.]

To summarize, we have the following landscape of "notions of induction":

(3.39)



The T inside the round blob may equivalently be described as (i.e., are in bijective correspondence with, by Exercise 3.38(b)) binary relations $\prec$, with the correspondence given by $T = T_\prec$:

Definition 3.40. We say that a binary relation $\prec \subseteq X^2$ is **well-founded** if the induced monotone set operator $T_\prec$ is inductive, i.e.,

(a) We have the **principle of well-founded induction** for $\prec$: the only $\prec$-closed $A \subseteq X$ is X.
(b) Equivalently, to prove $\phi(x)$ for all x, it suffices to prove $\phi(x)$ assuming $\phi(y)$ for all $y \prec x$.
(c) Equivalently, every $\varnothing \neq B \subseteq X$ contains a $\prec$**-minimal** $x \in B$, i.e., $B \cap {\downarrow}x = \varnothing$.

More generally, the **well-founded part** of $\prec$ is $\mathrm{WF}(\prec) = \mathrm{WF}(X, \prec) := \overline{T_\prec}(\varnothing)$.

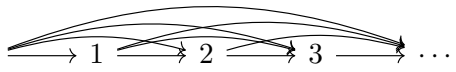
Example 3.41. On $\mathbb{N}$, the graph of the successor function

$$m \prec n \ :\Longleftrightarrow\ n = m + 1$$

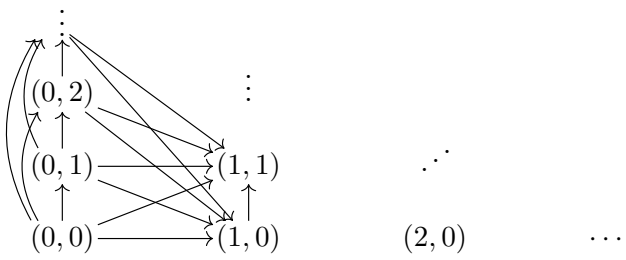
defines a relation whose induced T is precisely that from Example 3.2, thus whose principle of well-founded induction is ordinary induction on $\mathbb{N}$. Here is a picture of this $\prec$:

$$0 \longrightarrow 1 \longrightarrow 2 \longrightarrow 3 \longrightarrow \cdots$$

Example 3.42. Still on $\mathbb{N}$, $\prec := <$ defines a relation whose induced T is that from Example 3.3, thus whose principle of well-founded induction is strong induction on $\mathbb{N}$. Picture:



Example 3.43. The lexicographical order $<_{\mathrm{lex}}$ on $\mathbb{N}^2$ is well-founded, by Proposition 3.29:



As indicated by the above pictures, a common way to visualize an arbitrary binary relation R on a set X is as arrows or "directed edges" between the elements or "vertices". When thinking of R in this way, we also call it a **directed graph**, which is formally just a synonym for *binary relation*.

Definition 3.44. A binary relation $\prec \subseteq X^2$ is:

- **reflexive** if $x \prec x$ for all $x \in X$;
- **irreflexive** if $x \not\prec x$ for all $x \in X$;
- **symmetric** if $x \prec y \implies y \prec x$ for all $x, y \in X$;
- **antisymmetric** if $x \prec y \prec x \implies x = y$ for all $x, y \in X$;
- **transitive** if $x \prec y \prec z \implies x \prec z$ for all $x, y, z \in X$;
- **trichotomous** if $x \prec y$ or $x = y$ or $y \prec x$ for all $x, y \in X$.

Note that "irreflexive" is not the same as "not reflexive", and "antisymmetric" is not the same as "not symmetric". Note also that given irreflexivity, antisymmetry is equivalent to: $x \not\prec y$ or $y \not\prec x$ for all x, y; while given reflexivity, trichotomy is equivalent to **dichotomy**: $x \prec y$ or $y \prec x$ for all x, y.

- $\prec$ is a **preorder** if it is reflexive and transitive.
- $\prec$ is a **partial order** if it is an antisymmetric preorder.
- $\prec$ is a **linear order** (or **total order**) if it is a dichotomous partial order.
- $\prec$ is an **equivalence relation** if it is a symmetric preorder.

Proposition 3.45. A well-founded relation $\prec$ is irreflexive and antisymmetric. More generally, there are no **directed cycles** $x_0 \prec x_1 \prec \cdots \prec x_n = x_0$ of any length $n \geq 1$.

Proof. The directed cycle would be a subset with no minimal element. □

Proposition 3.46. More generally, a binary relation $\prec \subseteq X^2$ is well-founded iff there are no infinite descending sequences $x_0 \succ x_1 \succ x_2 \succ \cdots$ (where $\succ := \prec^{-1}$).

Proof. $\implies$: Such a sequence would form a subset with no minimal element.

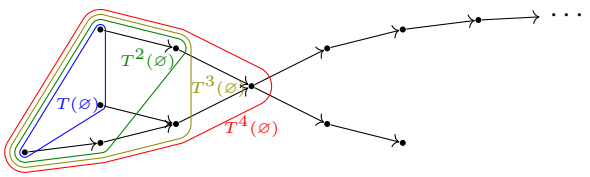
$\impliedby$: Suppose $\prec$ is not well-founded; let $C \subseteq X$ be nonempty with no minimal element. Pick $x_0 \in C$, then inductively, given $x_n \in C$ which cannot be minimal, pick $x_n \succ x_{n+1} \in C$. □

Remark 3.47. This might seem like a more intuitive definition of well-foundedness. However, from a foundational standpoint, the above proof is rather nontrivial: not only does it assume $\mathbb{N}$, i.e., the Axiom of Infinity 3.138, but it even uses the Axiom of Axiom 4.2 in order to pick x_{n+1} arbitrarily from among the potentially many predecessors of x_n at each stage. (See Exercise 4.7.) This characterization is thus best used for visual intuition only; the conceptual significance of well-foundedness is our official definition: that we can do induction on it.

Example 3.48. A **simple undirected graph**, usually called simply a **graph**, is an irreflexive *symmetric* binary relation. Given an irreflexive antisymmetric $\prec$, we may symmetrize it into $\sim := \prec \cup \succ$. We may visualize this as vertices connected by *unoriented* edges (no arrows); the original $\prec$ amounts to picking one of the two possible orientations for each edge.

A graph $\sim$ is **acyclic** if it has no cycles of ≥ 3 distinct vertices. (Of course, any edge yields a cycle of length 2.) An acyclic graph is also called a **forest**. A **tree** is a connected acyclic graph.

Note that if $\sim$ is a forest which is the symmetrization of an irreflexive antisymmetric $\prec$, then there are no non-vacuous instances of transitivity which hold for $\prec$, i.e., no x, y, z for which $x \prec y \prec z$ and also $x \prec z$, or else we would have a cycle of length 3. Example 3.41 is a tree (after symmetrizing); Examples 3.42 and 3.43 are not, being transitive. Here is another example:

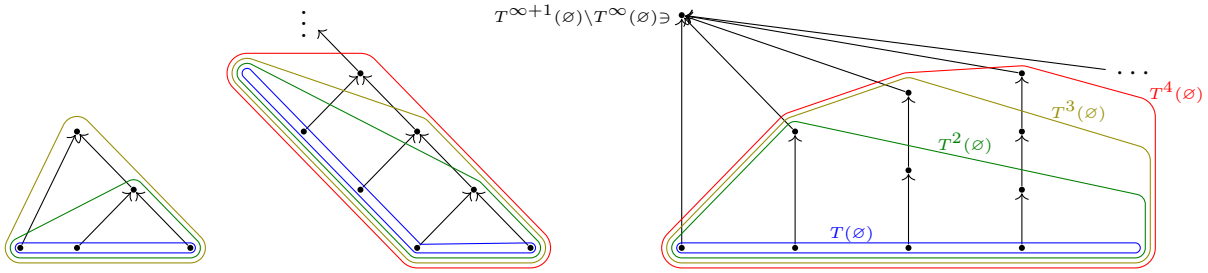


It is a bit silly to have "branches" of the tree pointing both backwards and forwards; usually, we would pick the orientations of the edges to point towards or away from a specified "root". For well-founded relations, pointing *towards* the root is the more natural condition, due to the following:

Proposition 3.49. Let $\prec$ be a well-founded relation. Suppose each vertex x has at most one $y \succ x$ (i.e., $\prec$ is the graph of a partial function). Then the symmetrization $\sim := \prec \cup \succ$ is a forest.

Proof. Suppose we had an undirected cycle $x_0 \sim x_1 \sim x_2 \sim \cdots \sim x_n = x_0$ of $n \geq 3$ distinct vertices. Then each $\sim$ is either $\prec$ or $\succ$. They cannot all be $\prec$ or $\succ$, since that would mean a directed cycle, contradicting Proposition 3.45. Suppose WLOG $x_0 \succ x_1$, and let $1 \leq i < n$ be least so that $x_i \prec x_{i+1}$; then $x_{i-1} \succ x_i \prec x_{i+1}$, so x_i has two distinct successors. □

Call $\prec$ a **well-founded forest** if the above conditions hold. Note that there is little risk of confusion with undirected forests, since well-founded implies antisymmetric by Proposition 3.45. If a vertex x has no successor, it is a **root vertex** (of its connected component; some components may have no root vertex, in which case they are instead "rooted at infinity"). For example, here is a well-founded forest with three connected components, one of them "rooted at infinity":



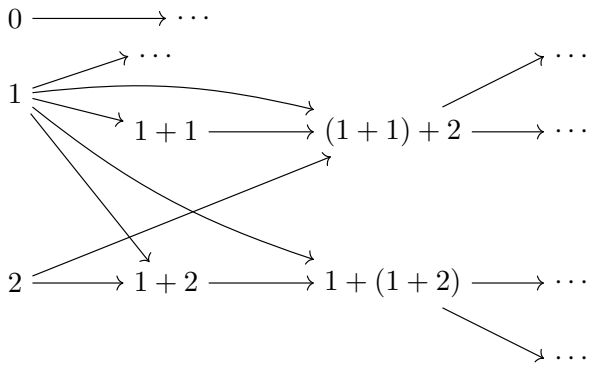
Well-founded forests are nice examples of induction, because we can draw pictures of them. But they are also quite general: we will see below (**TODO**) that every well-founded relation can be "simulated" by a forest.

Example 3.50. Closing under algebraic operations on a set X, as in Example 3.4, does not typically correspond to a relation $\prec$, by Exercise 3.38. For example, there is no unique "predecessor step" in generating $30 = 2 \cdot 3 \cdot 5$ from the prime numbers under binary multiplication.

However, if the set X consists of *formal symbolic expressions*[11] built from the operations f_i (in the notation of Example 3.4), then T does come from a well-founded $\prec$, namely

$$s \prec t \iff s \text{ is a immediate subexpression of } t = f_i(\dots, s, \dots) \text{ for some } f_i.$$

For example, if X consists of all expressions built from the single binary operation $+$, starting from the symbols $0, 1, 2$, then $\prec$ is given by a directed graph that looks like



[11] that is, terms in first-order logic

A key feature of notions of induction given by well-founded relations is that we may not only *prove* statements $\phi(x)$ by induction on x, but also *define* objects $f(x)$ inductively:[12]

Theorem 3.51 (principle of well-founded inductive definition)**.** Let $\prec$ be a well-founded relation on X, let $(Y_x)_{x\in X}$ be a family of sets, and let

$$\left(F_x : \prod_{z\prec x} Y_z \to Y_x\right)_{x\in X} \in \prod_{x\in X} Y_x^{\prod_{z\prec x} Y_z}.$$

Then there is a unique $f \in \prod_{x\in X} Y_x$ such that for each $x \in X$,

$$f(x) = F_x((f(z))_{z\prec x}).$$

In other words, "to define a family $(f(x) \in Y_x)_{x\in X}$, it suffices for each x to define $f(x) \in Y_x$ assuming given $f(z) \in Y_z$ for each $z \prec x$"; this definition of $f(x)$ given $(f(z))_{z\prec x}$ is specified by F_x.

Proof. Uniqueness follows easily from well-founded induction: if $f, g \in \prod_{x\in X} Y_x$ are two such functions, and $f(z) = g(z)$ for all $z \prec x$, then

$$f(x) = F_x((f(z))_{z\prec x}) = F_x((g(z))_{z\prec x}) = g(x).$$

We now prove existence. We identify f with its graph, which is to be a set of pairs

$$f \subseteq \textstyle\bigcup_{x\in X}(\{x\} \times Y_x)$$

The requirement on f says that for each $x \in X$ and $y \in Y_x$,

$$(x, y) \in f \iff \exists (y_z)_{z\prec x} \in \textstyle\prod_{z\prec x} Y_z \,(y = F_x((y_z)_{z\prec x}) \text{ and } \forall z \prec x \,((z, y_z) \in f)).$$

By the Knaster–Tarski Theorem 3.6, there is such a set of pairs f. We prove that f is a function, i.e., $\forall x \in X \,\exists! y \in Y_x \,((x, y) \in f)$, by $\prec$-induction on x. Assume $\forall z \prec x \,\exists! y_z \, Y_z \,((z, y_z) \in f)$. Then by the above $\iff$, the unique y such that $(x, y) \in f$ is $y = F_x((y_z)_{z\prec x})$. □

Example 3.52. $! : \mathbb{N} \to \mathbb{N}$ (factorial), i.e., $(n!)_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} \mathbb{N}$, is defined inductively as follows:

$$0! := 1,$$
$$(n + 1)! := (n + 1) \cdot n!.$$

In the above formalism: take $\prec$ to be the successor graph from Example 3.41, $X := Y_n := \mathbb{N}$, and

$$\begin{aligned} F_0 : \textstyle\prod_{m\in\downarrow 0} \mathbb{N} = \mathbb{N}^{\varnothing} &\longrightarrow \mathbb{N} \\ \varnothing &\longmapsto 1, \\ F_{n+1} : \textstyle\prod_{m\in\downarrow(n+1)} \mathbb{N} = \mathbb{N}^{\{n\}} &\longrightarrow \mathbb{N} \\ (y) &\longmapsto (n + 1) \cdot y. \end{aligned}$$

Example 3.53. The Fibonacci sequence $(f(n))_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} \mathbb{N}$ is defined via *strong* induction:

$$\begin{aligned} f(0) &:= 0, \\ f(1) &:= 1, \\ f(n) &:= f(n - 1) + f(n - 2) \quad \text{for } n \geq 2. \end{aligned}$$

In the above formalism: take $\prec := <$, $X := Y_n := \mathbb{N}$, and

$$F_n : \textstyle\prod_{m<n} \mathbb{N} = \mathbb{N}^n \longrightarrow \mathbb{N}$$
$$(y_0, \ldots, y_{n-1}) \longmapsto \begin{cases} 0 & \text{if } n = 0, \\ 1 & \text{if } n = 1, \\ y_{n-1} + y_{n-2} & \text{if } n \geq 2. \end{cases}$$

[12] Often, set theorists will insist that the correct terminology for "inductive definition" is **recursion**.

Example 3.54. Recall the Ackermann function $A : \mathbb{N}^2 \to \mathbb{N}$ defined via an algorithm as in Example 3.30 and Remark 3.32. We may also define it directly, via lexicographical induction:

$$\begin{aligned} A(0, y) &:= y + 1, \\ A(x + 1, 0) &:= A(x, 1), \\ A(x + 1, y + 1) &:= A(x, A(x + 1, y)). \end{aligned}$$

Note that Example 3.30 is the algorithm for expanding this definition:

$$\begin{aligned} A(1, 2) &= A(0, A(1, 1)) \\ &= A(0, A(0, A(1, 0))) \\ &= A(0, A(0, A(0, 1))) \\ &= A(0, A(0, 2)) \\ &= A(0, 3) \\ &= 4; \end{aligned}$$

now erase the A's and nested parentheses to recover the computation from Example 3.30.

In the formalism of Theorem 3.51: we use the lexicographical ordering $<_{\text{lex}}$ on $X := \mathbb{N}^2$ from Proposition 3.29, which says precisely that $<_{\text{lex}}$ is well-founded; $Y_{(x,y)} := \mathbb{N}$; and

$$\begin{aligned} F_{(0,y)} : \textstyle\prod_{(u,v)<_{\text{lex}}(0,y)} \mathbb{N} = \mathbb{N}^{\{(0,0),\ldots,(0,y-1)\}} &\longrightarrow \mathbb{N} \\ \vec{a} = (a_{(0,0)}, \ldots, a_{(0,y-1)}) &\longmapsto y + 1, \\ F_{(x+1,0)} : \textstyle\prod_{(u,v)<_{\text{lex}}(x+1,0)} \mathbb{N} = \mathbb{N}^{(x+1)\times\mathbb{N}} &\longrightarrow \mathbb{N} \\ \vec{a} = (a_{(u,v)})_{u\le x, v\in\mathbb{N}} &\longmapsto a_{(x,1)}, \\ F_{(x+1,y+1)} : \textstyle\prod_{(u,v)<_{\text{lex}}(x+1,y+1)} \mathbb{N} = \mathbb{N}^{((n+1)\times\mathbb{N})\cup\{(x+1,0),\ldots,(x+1,y)\}} &\longrightarrow \mathbb{N} \\ \vec{a} &\longmapsto a_{x,a_{x+1,y}}. \end{aligned}$$

Remark 3.55. In many examples, such as those above, the sets Y_x in Theorem 3.51 are the same. In fact, the general case where the Y_x's vary can be reduced to this simpler case, since we may take $Y := \bigcup_{x\in X} Y_x$, define $f : X \to Y$ inductively, and then prove by induction that in fact, each $f(x) \in Y_x$. On the other hand, the general form of Theorem 3.51 has the advantage that we may

Exercise 3.56. Deduce the principle of induction (i.e., that $\prec$ is well-founded) from the principle of inductive definition (Theorem 3.51).

Example 3.57. To see why the principle of inductive definition, unlike the principle of induction, only works for a well-founded relation $\prec$ rather than a general monotone $T : \mathcal{P}(X) \to \mathcal{P}(X)$: take

$$\begin{aligned} T : \mathcal{P}(\mathbb{N}) &\longrightarrow \mathcal{P}(\mathbb{N}) \\ A &\longmapsto \{0, 1, 2\} \cup \{x + y \mid x, y \in A\}. \end{aligned}$$

This is clearly inductive. If we try to define a function $f : \mathbb{N} \to \mathbb{N}$ T-inductively via

$$\begin{aligned} f(0) := f(1) := f(2) &:= 2, \\ f(x + y) &:= f(x)f(y), \end{aligned}$$

we get a contradiction $f(4) = f(2 + 2) = f(2)f(2) = 4$ and $f(4) = f(1 + 2 + 1) = f(1)f(2)f(1) = 8$.

On the other hand, if X consists of *formal symbolic expressions* built from $0, 1, 2, +$ as in Example 3.50, then such a well-founded inductive definition would be allowed: we would get $f(2 + 2) = 4$ and $f((1 + 2) + 1) = 8$, but these two expressions $2 + 2$ and $(1 + 2) + 1$ are different. We can use this for example to define the *evaluation* of each such expression e.

3.D. **Simulations.** Our goal for the next few subsections is, roughly speaking, to "compare" and "classify" different notions of induction. We will focus on well-founded relations, although some things can be generalized to monotone operators T, as indicated in Exercises.

Example 3.58. Ordinary induction on $\mathbb{N}$ can clearly be "reduced" to strong induction.

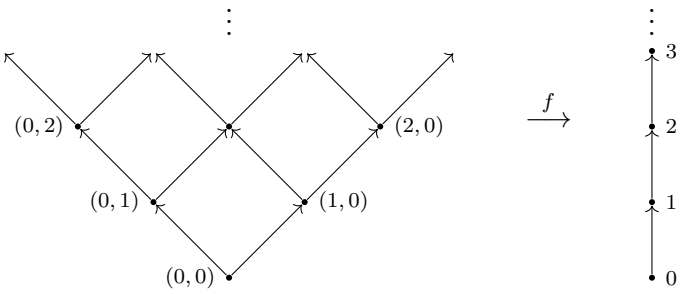
What does this mean, precisely? The successor relation $\prec$ on $\mathbb{N}$ (Example 3.41) is a subgraph of the $<$ relation (Example 3.42), i.e., the predecessors $\downarrow_{\prec} x$ of the former are a subset of the precedessors $\downarrow_{<} x$ of the latter, for any $x \in \mathbb{N}$. Thus, if we know strong induction, then we can easily deduce ordinary induction, whose induction hypotheses are a subset of those for strong induction.

Exercise 3.59. Show that in general, if $\prec_1 \subseteq \prec_2 \subseteq X^2$ and $\prec_2$ is well-founded, then so is $\prec_1$.

Example 3.60. Consider the following induction principle for $\mathbb{N}^2$: for $A \subseteq \mathbb{N}^2$, if

- $(0,0) \in A$,
- $(0,y) \in A \implies (0,y+1) \in A$,
- $(x,0) \in A \implies (x+1,0) \in A$,
- $(x,y+1), (x+1,y) \in A \implies (x+1,y+1) \in A$,

then $A = \mathbb{N}^2$. This is induction for the following well-founded graph on the left:



We may prove this induction principle by proving that $(x,y) \in A$ for all $(x,y) \in \mathbb{N}^2$, by ordinary induction on $x+y$. In other words, we have the addition function $f : \mathbb{N}^2 \to \mathbb{N}$ which "reduces" this induction principle to ordinary induction for $\mathbb{N}$.

Definition 3.61. For two sets equipped with binary relations $(X, \prec_X)$ and $(Y, \prec_Y)$, a function $f : X \to Y$ is a **homomorphism** (between the relations) if for all $x, x' \in X$,

$$x' \prec_X x \implies f(x') \prec_Y f(x).$$

Proposition 3.62. For a homomorphism f as above, if $\prec_Y$ is well-founded, then so is $\prec_X$.

Proof idea. If $\prec_X$ is not well-founded, then (by Proposition 3.46) there is a descending sequence $x_0 \succ_X x_1 \succ_X \cdots$, whence $f(x_0) \succ_Y f(x_1) \succ \cdots$, whence $\prec_Y$ is not well-founded. □

However, we don't like this "proof" very much, for it uses not only $\mathbb{N}$ but also the Axiom of Choice (cf. Remark 3.47). The following proof is similar in spirit, being based on the "contrapositive" form of the principle of induction. Below, we give another "forward" proof, generalizing the usual way that principles of induction may be used to prove each other (such as in Example 3.60).

Proof 1. If $\prec_X$ is not well-founded, then there is $\varnothing \neq B \subseteq X$ with no $\prec_X$-minimal element, whence $\varnothing \neq f[B] \subseteq Y$ has no $\prec_Y$-minimal element, since for every $f(x) \in f[B]$ where $x \in B$, there is $x' \prec_X x$, whence $f(x') \prec_Y f(x)$. □

Proof 2. Let $A \subseteq X$ be $\prec_X$-closed; we show that $\forall x \in X\, (x \in A)$, by $\prec_Y$-induction on $f(x)$. That is, we show that $\forall y \in Y\, \forall x \in f^{-1}(y)\, (x \in A)$, by $\prec_Y$-induction on y. Suppose (IH) this holds for all $y' \prec_Y y$. Then for all $x \in f^{-1}(y)$, for all $x' \prec_X x$, we have $f(x') \prec_Y f(x) = y$, whence by the IH, $x' \in A$. Thus since A is $\prec_X$-closed, $x \in A$. □

Example 3.63. If $\prec$ is a well-founded relation on X, then any subrelation of $\prec$ is also well-founded (because id_X is a homomorphism). This covers Exercise 3.59.

Example 3.64. If $\prec$ is a well-founded relation on X, then for any $Y \subseteq X$, the restriction $\prec|Y := (\prec) \cap Y^2$ is well-founded on Y (because the inclusion $Y \hookrightarrow X$ is a homomorphism).

Example 3.65. For sets with relations $X = (X, \prec_X)$ and $Y = (Y, \prec_Y)$, an **isomorphism** $f : X \cong Y$ is a bijection such that both f, f^{-1} are homomorphisms, i.e.,

$$x' \prec_X x \iff f(x') \prec_Y f(x).$$

As usual in mathematics, when two structures are called "isomorphic" (in whatever sense appropriate to that context), then they should be considered "the same structure", possibly with elements labeled differently, but sharing all of the same "structural" properties. For example,

$$(\mathbb{N}, <) \cong (-\mathbb{N}, >) \quad \text{where } -\mathbb{N} := \{-n \mid n \in \mathbb{N}\}.$$

If $f : X \cong Y$ is an isomorphism, then in particular, $\prec_X$ is well-founded iff $\prec_Y$ is.

In spite of these examples, however, homomorphisms are not the most natural or general way of comparing notions of induction:

Example 3.66. Strong induction on $\mathbb{N}$ can also be "reduced" to ordinary induction, as follows. Suppose $A \subseteq \mathbb{N}$ satisfies the hypothesis of the principle of strong induction:

$(*)$ $\forall x \in \mathbb{N}\,((\forall y < x,\, y \in A) \implies x \in A)$.

We prove that $A = \mathbb{N}$, by proving by *ordinary* induction that for each $x \in \mathbb{N}$, every $y \leq x$ is in A.

- For $x = 0$, the only $y \leq x$ is $x = 0$, which is in A by $(*)$ applied to $x = 0$.
- Suppose (IH) every $y \leq x$ is in A; we prove that every $y \leq x + 1$ is in A. The only new y to consider is $y = x + 1$. By the IH, every $y < x + 1$ is in A. Thus by $(*)$, $x + 1 \in A$.

Definition 3.67. For two sets equipped with binary relations $(X, \prec_X)$ and $(Y, \prec_Y)$, a relation $R \subseteq X \times Y$ is a **simulation** (of $\prec_X$ in $\prec_Y$) if for all $x, x' \in X$ and $y \in Y$,

$$x' \prec_X x \; R \; y \implies \exists y' \in Y\,(x' \; R \; y' \prec_Y y),$$

i.e.,

$$x \; R \; y \implies \forall x' \prec_X x,\, \exists y' \prec_Y y\,(x' \; R \; y') \iff:\; x \; T_{\mathrm{sim}}(R) \; y,$$
$$R \circ (\prec_X) \subseteq (\prec_Y) \circ R.$$

We use wavy arrows to denote R:

$$\begin{array}{ccc} x' & \xrightarrow{\prec_X} & x \\ {\scriptstyle R}\wr & & \wr{\scriptstyle R} \\ y' & \dashrightarrow_{\prec_Y} & y \end{array}$$

Proposition 3.68. For a simulation R with $\mathrm{dom}(R) = X$, if $\prec_Y$ is well-founded, then so is $\prec_X$. More generally, $R^{-1}[\mathrm{WF}(\prec_Y)] \subseteq \mathrm{WF}(\prec_X)$.

The following proof idea, using $\mathbb{N}$ and the Axiom of Choice, explains the name "simulation": we are "simulating" the "history" of an element $x \in X$ in Y.

$$\begin{array}{ccccccc} \cdots & \longrightarrow & x_2 & \longrightarrow & x_1 & \longrightarrow & x_0 \\ & & \wr & & \wr & & \wr \\ \cdots & \dashrightarrow & y_2 & \dashrightarrow & y_1 & \dashrightarrow & y_0 \end{array}$$

Exercise 3.69. Give proofs using (a) every nonempty set has a minimal element; (b) induction.

Example 3.70. When R above is the graph of a function f, then the definition of simulation says precisely that f is a homomorphism (since $y = f(x)$ and $y' = f(x')$).

Exercise 3.71.

(a) Show that a binary relation $R \subseteq X \times Y$ is a simulation iff its domain is a $\prec_X$-downward-closed subset of X, i.e., $x' \prec x \in \mathrm{dom}(R) \implies x' \in \mathrm{dom}(R)$, and $R \subseteq \mathrm{dom}(R) \times Y$ is a simulation (of the restriction of $\prec_X$ to $\mathrm{dom}(R)$ in $\prec_Y$).

(b) Conclude that a *partial* function f is a simulation iff _____.

Example 3.72. $\leq \subseteq \mathbb{N} \times \mathbb{N}$ is a simulation of $<$ in $\prec$:

$$x' < x \leq y \implies x' \leq y' := y - 1 \prec y.$$

Proof (b) of Exercise 3.69 should recover the proof of strong induction in Example 3.66 in this case.

Example 3.73. The addition map $f : \mathbb{N}^2 \to \mathbb{N}$ from Example 3.60 is a simulation from the "grid" graph $\prec_{\mathbb{N}^2}$ on the left to the usual successor graph $\prec_{\mathbb{N}}$ (since it is a homomorphism), and its inverse is also a simulation, despite not being a function. Indeed, if $m \prec_{\mathbb{N}} n \; f^{-1}\,(x, y)$, i.e., $m+1 = n = x+y$, WLOG with $x > 0$, then $m \; f^{-1}\,(x-1, y) \prec_{\mathbb{N}^2} (x, y)$.

Definition 3.74. R is a **cosimulation** if R^{-1} is a simulation, and a **bisimulation** if it is both a simulation and a cosimulation, i.e.,

$$x \; R \; y \implies \forall x' \prec x \, \exists y' \prec y \, (x' \; R \; y') \text{ and } \forall y' \prec y \, \exists x' \prec x \, (x' \; R \; y') \iff: x \; T_{\mathrm{bisim}}(R) \; y.$$

This intuitively means that x and y have "histories which look the same".

Corollary 3.75 (of Proposition 3.68)**.** If $f : X \twoheadrightarrow Y$ is a surjective cosimulation, and $\prec_X$ is well-founded, then so is $\prec_Y$. More generally, $f[\mathrm{WF}(\prec_X)] \subseteq \mathrm{WF}(\prec_Y)$. □

Example 3.76 (cf. Example 3.70)**.** A bijection is a bisimulation iff it is an isomorphism.

Proposition 3.77. Let $(X, \prec_X), (Y, \prec_Y), (Z, \prec_Z)$ be three sets with binary relations.

(a) $\mathrm{id}_X : X \to X$ is a (bi)simulation.

(b) If $R \subseteq X \times Y$ and $S \subseteq Y \times Z$ are (bi)simulations, then so is $S \circ R \subseteq X \times Z$.

(c) If $R \subseteq X \times Y$ is a bisimulation, then so is $R^{-1} \subseteq Y \times X$.

Proof. (a) because it is an isomorphism; (c) is obvious.

(b) $S \circ R \circ (\prec_X) \subseteq S \circ (\prec_Y) \circ R \subseteq (\prec_Z) \circ S \circ R$; similarly for their inverses. □

Exercise 3.78. Note that proof 2 of Proposition 3.62 uses the coimage (Exercise 3.38)

$$f\langle A \rangle = \{y \in Y \mid \forall x \in f^{-1}(y) \, (x \in A)\};$$

similarly for Exercise 3.69(b). Based on this observation, generalize the concept of simulation to monotone set operators as follows.

Let X, Y be sets with monotone set operators T_X, T_Y respectively, and let $R \subseteq X \times Y$.

(a) Show that the following are equivalent:

(i) For all $B \subseteq Y$, we have $R^{-1}[T_Y(B)] \subseteq T_X(R^{-1}[B])$.

(ii) For all $A \subseteq X$, we have $T_Y(R\langle A \rangle) \subseteq R\langle T_X(A) \rangle$.

(iii) For all $A \subseteq X$, we have $R[S_X(A)] \subseteq S_Y(R[A])$, where

$$S_X(A) := X \setminus T_X(X \setminus A)$$

is the de Morgan dual of T_X (we can think of $S_X(A)$ as those x which "depend on A", i.e., for which A is necessary, rather than sufficient, to derive x); similarly for S_Y.

(b) Verify that these conditions are equivalent to R being a simulation, when $T_X = T_{\prec_X}$ and $T_Y = T_{\prec_Y}$ for relations $\prec_X \subseteq X^2$ and $\prec_Y \subseteq Y^2$.

(c) Show that if these hold, then they continue to hold when T_X, T_Y are replaced with $\overline{T_X}, \overline{T_Y}$.

(d) Conclude that $R^{-1}[\overline{T_Y}(\varnothing)] \subseteq \overline{T_X}(\varnothing)$. Thus, if T_Y is inductive, then so is T_X.

3.E. **Simulability and bisimilarity.** Note that Definitions 3.67 of *simulation* and 3.74 of *bisimulation* are both of the form "$R \subseteq T(R)$", for suitable monotone set operators $T = T_{\mathrm{sim}}, T_{\mathrm{bisim}} : \mathcal{P}(X \times Y) \to \mathcal{P}(X \times Y)$ as indicated in those definitions. In other words, they amount to saying that R is "T-open", in the sense of the dual Knaster–Tarski Theorem 3.20. Consequently,

Definition 3.79. For sets with binary relations $(X, \prec_X), (Y, \prec_Y)$, by dual Knaster–Tarski, there is a *largest* simulation and bisimulation between them, denoted respectively

$$(\lesssim) = (\lesssim_{X,Y}) := T^{\circ}_{\mathrm{sim}}(X \times Y) \subseteq X \times Y,$$
$$(\equiv) = (\equiv_{X,Y}) := T^{\circ}_{\mathrm{bisim}}(X \times Y) \subseteq X \times Y.$$

(These depend on $\prec_X, \prec_Y$, even though we don't write $\prec_X, \prec_Y$ in the notation.) If $x \lesssim y$, we say x is **simulable** by y or that y can **simulate** x; if $x \equiv y$, we say x, y are **bisimilar**. We also write

$$x \eqsim y \ :\Longleftrightarrow\ x \lesssim y \lesssim x,$$

and call x, y **mutually simulable** if this holds.

By virtue of the definitions of $\lesssim, \equiv$ as the largest T-open sets for some T, we have

$$x \lesssim y \iff x\, T_{\mathrm{sim}}(\lesssim)\, y \iff \forall x' \prec x\, \exists y' \prec y\, (x' \lesssim y') \tag{3.80}$$
$$\iff x\, R\, y \quad \text{for } \textit{some} \text{ simulation } R, \tag{3.81}$$
$$x \equiv y \iff x\, T_{\mathrm{bisim}}(\equiv)\, y \iff \forall x' \prec x\, \exists y' \prec y\, (x' \equiv y') \text{ and } \forall y' \prec y\, \exists x' \prec x\, (x' \equiv y') \tag{3.82}$$
$$\iff x\, R\, y \quad \text{for } \textit{some} \text{ bisimulation } R. \tag{3.83}$$

Corollary 3.84 (of Proposition 3.77)**.** Let $(X, \prec_X), (Y, \prec_Y), (Z, \prec_Z)$ be sets with binary relations.

(a) $x \equiv x$ for any $x \in X$.
(b) $x \equiv y \equiv z \implies x \equiv z$, and $x \lesssim y \lesssim z \implies x \lesssim z$, for $x \in X$, $y \in Y$, $z \in Z$.
(c) $x \equiv y \implies y \equiv x$ for $x \in X$, $y \in Y$.

In other words, $\lesssim$ is a preorder, while $\equiv$ is an equivalence relation, *between all elements of all sets* $(X, \prec)$ *equipped with a relation* (formally, on the disjoint union $\bigsqcup_{(X,\prec)} X$).

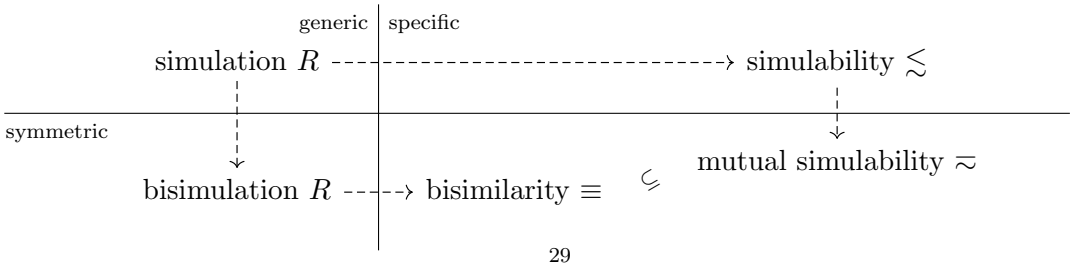
Proof. For (b): $(\equiv_{Y,Z}) \circ (\equiv_{X,Y}) \subseteq X \times Z$ is a bisimulation by Proposition 3.77(b), hence contained in $(\equiv_{X,Z})$; similarly for $\lesssim$. The other two parts follow similarly. □

Corollary 3.85. If $x \equiv y$, then $x \eqsim y$.

Proof. $\equiv$ is a simulation, hence contained in $\lesssim$; it follows that also $(\equiv) = (\equiv^{-1}) \subseteq (\lesssim^{-1})$. □

Remark 3.86. *Simulation* is a generic class of relations: there are many examples of simulations. Among them, there is a specific one, called *simulability* $\lesssim$.

Starting from the generic concept of *simulation*, we may symmetrize it to obtain the generic concept of *bisimulation*, of which there are again many examples; among these, we have the specific example of *bisimilarity* $\equiv$. Or, we may symmetrize the specific simulation of *simulability* $\lesssim$, to obtain *mutual simulability* $\eqsim$. *These two equivalence relations are not the same!* The preceding result shows that $\equiv$ implies $\eqsim$, but the converse may fail; see Examples 3.89 and 3.90. In the following diagram, the arrows denote informal derivation of one concept from another:

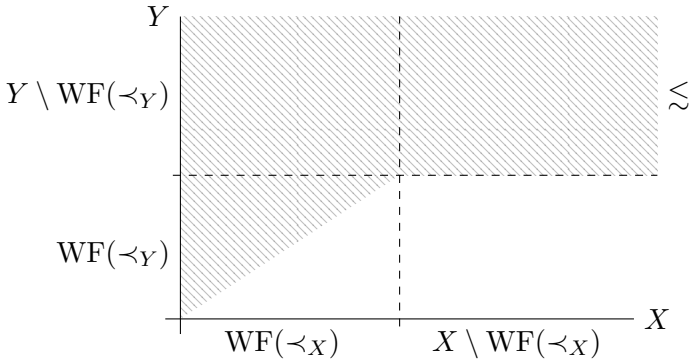


Example 3.87. If $x \in X$ has no predecessors, then it is simulable by all $y \in Y$, while it can only simulate $y \in Y$ which also has no predecessors, immediately by (3.80).

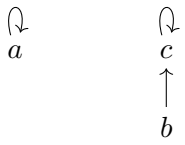
Thus, if $x \eqsim y$, then x has no predecessors iff y does. Conversely, if x, y both have no predecessors, then clearly $x \equiv y$ (however, if both have predecessors, this may or may not hold).

Example 3.88. If $y \in Y \setminus \mathrm{WF}(\prec_Y)$, then $x \lesssim y$ for all $x \in X$. Indeed, since $\mathrm{WF}(\prec_Y)$ is $\prec_Y$-closed by definition, (for any $x' \prec x$) there is some $y' \prec y$ also in $Y \setminus \mathrm{WF}(\prec_Y)$; thus $X \times (Y \setminus \mathrm{WF}(\prec_Y))$ is a simulation, hence contained in $\lesssim$.

Thus, elements not in the well-founded part are equivalent according to $\eqsim$ (they may or may not be equivalent according to $\equiv$; see the following example), and "above" all other elements according to $\lesssim$. The generic picture of $\lesssim \subseteq X \times Y$, between two sets X, Y with relations $\prec_X, \prec_Y$ (drawn here as if they were the $<$ relation on the $\mathbb{R}$ or $\mathbb{N}$ line), looks like:



Example 3.89. Let $X = \{a\}$ and $Y = \{b, c\}$ be equipped with the following ill-founded graphs:



We may determine the $\lesssim$ (namely $\lesssim_{X,X}, \lesssim_{X,Y}, \lesssim_{Y,X}, \lesssim_{Y,Y}$), $\eqsim$, and $\equiv$ relations between all pairs of elements of either of these sets (a ✓ means row $\lesssim$ column, etc.):

$\lesssim$	a	b	c
a	✓	✗	✓
b	✓	✓	✓
c	✓	✗	✓

$\eqsim$	a	b	c
a	✓	✗	✓
b	✗	✓	✗
c	✓	✗	✓

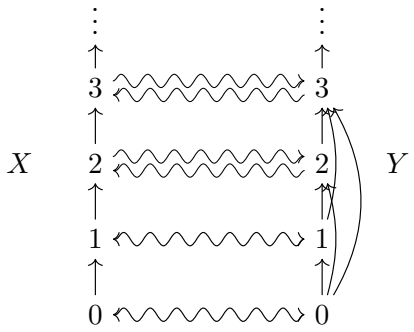
$\equiv$	a	b	c
a	✓	✗	✗
b	✗	✓	✗
c	✗	✗	✓

For $\lesssim$, we know the first and third columns are ✓ while the second column is ✗ except for the diagonal entry, by the preceding example, since b is in the well-founded part while a, c are not. Intersecting this matrix with its transpose yields the matrix for $\eqsim$. For $\equiv$, we know it must be contained in $\eqsim$ (Corollary 3.85), and it is still reflexive, so the only possibility is that a, c are not bisimilar (even though they are mutually simulable); this is indeed the case, since the predecessor $b \prec c$ is not bisimilar to any predecessor of a (the only such predecessor being a itself which we've already determined is not $\eqsim b$).

Example 3.90. Let $X = \mathbb{N}$ equipped with the successor relation $x \prec_X y \iff x + 1 = y$, and $Y = \mathbb{N}$ equipped with the usual order relation $\prec_Y := <$. We've already seen that $=$ is a homomorphism from X to Y (Example 3.58), while $\leq$ is a simulation from Y to X (Example 3.72). In fact it is also a simulation from X to Y, or from each of these sets to itself. Indeed, as with any relation, $\prec_X$ is trivially a simulation from itself to itself; thus $\lesssim_{X,X} \subseteq X \times X$ is a preorder

containing $\prec_X$, hence contains the preorder generated by $\prec_X$, which is easily seen to be $\leq$. It follows that $\lesssim_{X,Y} \subseteq X \times Y$ contains $=$ and is closed under composition with $\lesssim_{X,X} = \leq$; said composition is easily seen to be $\leq$. And similarly, $\lesssim_{Y,Y} \supseteq \lesssim_{X,Y} \circ \lesssim_{Y,X} \supseteq \leq \circ \leq = \leq$.

We claim that in fact, all four of these $\lesssim$ relations are equal to $\leq$. To see this for $\lesssim_{X,Y}$, we can show by induction that $x \lesssim_{X,Y} y \implies x \leq y$, by induction on x. For $x = 0$ the conclusion is trivially true. Now let x be such that for all y, we have $x \lesssim_{X,Y} y \implies x \leq y$, and suppose $x+1 \lesssim_{X,Y} y$. Then $x \prec_X x+1 \lesssim_{X,Y} y$, so there exists $y' \prec_Y y$, i.e., $y' < y$, such that $x \lesssim_{X,Y} y'$, which by the IH means $x \leq y' < y$, whence $x+1 \leq y$. This proves it for $\lesssim_{X,Y}$; the other three relations $\lesssim_{X,X}, \lesssim_{Y,X}, \lesssim_{Y,Y}$ are easier, since we still have $x \prec x+1$ according to the weaker $\prec$ on the domain, while the stronger $\prec$ on the codomain still yields $y' < y$.



In particular, we see that the $\approx$ relation between X, Y (or between X, X, etc.) is equality. The $\equiv$ relation between X and itself, or between Y and itself, is thus also equality, since it must be reflexive. But $\equiv_{X,Y}$ is smaller: at the beginning, we do have $0 \equiv_{X,Y} 0$, since 0 is minimal in both X, Y; and thus $1 \equiv_{X,Y} 1$, since every $x' \prec_X 1$ (namely $x' = 0$) is bisimilar to some $y' \prec_Y 1$ (namely $y' = 0$), and similarly vice-versa. But $2 \not\equiv_{X,Y} 2$, since $0 \prec_Y 2$ is not bisimilar to the unique $x' = 1 \prec_X 2$.

These examples demonstrate that *in the well-founded part*,[13] we may inductively determine whether $\lesssim$ or $\equiv$ holds by "checking all predecessors". This is made precise by the following:

Proposition 3.91. Let X, Y be sets equipped with binary relations $\prec_X, \prec_Y$, at least one of which is well-founded. Then $\lesssim_{X,Y}, \equiv_{X,Y}$ are the *unique* fixed points of $T_{\text{sim}}, T_{\text{bisim}}$ respectively, i.e., the unique binary relations between X, Y obeying (3.80), (3.82). Equivalently, they are also the *smallest* $T_{\text{sim}}, T_{\text{bisim}}$*-closed* relations.

Proof. If say $\prec_X$ is well-founded, then note that (3.80) defines the truth value of $x \lesssim y$ over all y, by well-founded induction on x.

More formally, via the canonical bijection $\mathcal{P}(X \times Y) \cong 2^{X \times Y} \cong (2^Y)^X \cong \mathcal{P}(Y)^X$ (Exercise 2.82), $\lesssim$ corresponds to a function $f : X \to \mathcal{P}(Y)$, taking $X \ni x \mapsto \{y \in Y \mid x \lesssim y\}$; now (3.80) says

$$f(x) = \{y \in Y \mid \forall x' \prec x \, \exists y' \prec y \, (y' \in f(x'))\},$$

which is a well-founded inductive definition of f (Theorem 3.51) and hence $\lesssim$. Similarly for $\equiv$, or if $\prec_Y$ is well-founded.

The assertion that $\overline{T_{\text{sim}}}(\varnothing) = \lesssim_{X,Y} := T^\circ_{\text{sim}}(X \times Y)$ is equivalent to $\lesssim_{X,Y}$ being the unique T_{sim}-fixed point, because $\overline{T_{\text{sim}}}(\varnothing)$ is also a T_{sim}-fixed point, by the Knaster–Tarski Theorem 3.6; similarly for $\overline{T_{\text{bisim}}}(\varnothing)$. □

Exercise 3.92. Show that in general, without the assumption that either $\prec_X$ or $\prec_Y$ is well-founded, the least fixed points $\overline{T_{\text{sim}}}(\varnothing), \overline{T_{\text{bisim}}}(\varnothing) \subseteq X \times Y$ are the same as $\lesssim_{X,Y}, \equiv_{X,Y}$ but with all of $(X \setminus \mathrm{WF}(\prec_X)) \times (Y \setminus \mathrm{WF}(\prec_Y))$ excluded (recall Example 3.88).

[13] In the ill-founded part, the behavior of $\lesssim$ is completely determined by Example 3.88, while $\equiv$ may be more complicated; see Example 3.89 and Remark 3.102.

3.F. **The Mostowski collapse and Axiom of Foundation.** For the bisimilarity relation $\equiv$, there is an even better way to check when it holds between two well-founded relations. Instead of checking inductively whether two elements x, y "have the same history", according to the definition of $\equiv$, we may compute a set $\xi(x)$ which records "the history" of a single element x.

In order to motivate this, note that (by Corollary 3.84) $\equiv$ is an equivalence relation (on $\bigsqcup_{(X,\prec)} X$). Abstractly, *any* equivalence relation is given by equality of *some* value $\xi(x)$ assigned to each x; this is the point of the quotient set construction. The usual way of constructing quotients is to take equivalence classes $\xi(x) := [x]$, which for $\equiv$ would be proper classes (since $\equiv$ relates elements of *all* sets equipped with a binary relation). But there is another way: we want

$$\begin{aligned}
(3.93) \qquad \xi(x) = \xi(y) &\overset{?}{\iff} x \equiv y \\
&\iff \forall x' \prec x\, \exists y' \prec y\, (x' \equiv y') \text{ and } \forall y' \prec y\, \exists x' \prec x\, (x' \equiv y') \\
\text{(by IH)} \quad &\iff \forall x' \prec x\, \exists y' \prec y\, (\xi(x') = \xi(y')) \text{ and } \forall y' \prec y\, \exists x' \prec x\, (\xi(x') \equiv \xi(y')) \\
&\iff \{\xi(x') \mid x' \prec x\} = \{\xi(y') \mid y' \prec y\} \quad \text{by Extensionality 2.1,}
\end{aligned}$$

which will hold if ξ satisfies

$$\xi(x) = \{\xi(x') \mid x' \prec x\}.$$

Definition 3.94. Let $\prec \subseteq X^2$ be a well-founded relation. The **Mostowski collapse** (or **extensional collapse**) of $(X, \prec)$ is the function ξ on X defined inductively by the above equation.

Remark 3.95. There is a subtle technical problem with this definition: it is not valid according the principle of well-founded inductive definition as we have proved in Theorem 3.51! There, we assumed given codomain sets Y_x for the function we're defining (possibly varying with x); whereas ξ builds sets of arbitrary complexity from scratch in the universe, with no connection to the set X we start with. The codomains were essential in the proof of Theorem 3.51 using the impredicative Knaster–Tarski Theorem 3.6, which had to intersect over arbitrary subsets of the given sets.

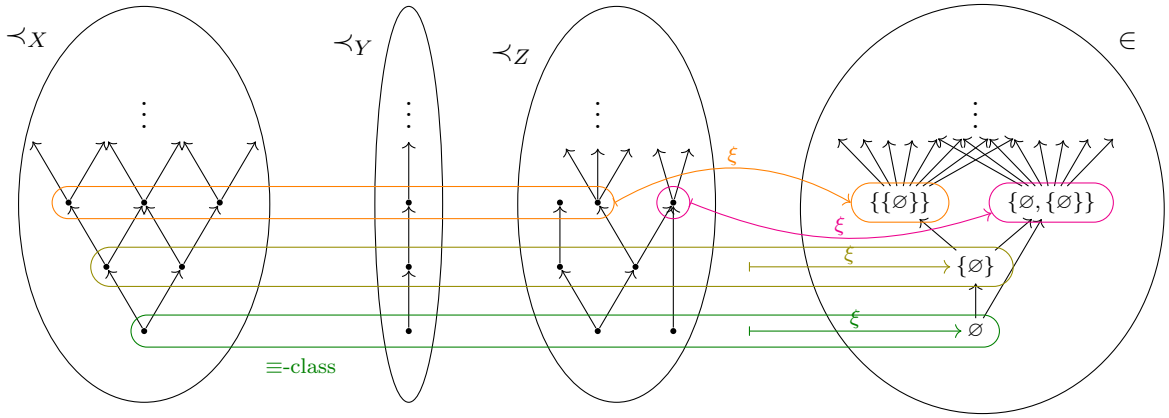
Later in Theorem 3.167, we will give a more powerful version of Knaster–Tarski that allows precisely this kind of "unbounded" construction from scratch in the universe. Until then, statements that mention ξ should be understood as "provided the ξ function exists for the given $(X, \prec)$".

Proposition 3.96. $\xi : (X, \prec) \to (V, \in)$ is a bisimulation.

Proof. This is a restatement of the definition of ξ:

- If $x' \prec x$ ξ y, i.e., $\xi(x) = y$, then $\xi(x') =: y' \in y$.
- If $\xi(x) = y \ni y'$, then $y' = \xi(x')$ for some $x' \prec x$. □

We thus have a single "ultimate" notion of induction, namely the universe V equipped with $\in$, which contains exactly one set in each bisimilarity class, and ξ picks out this set:



Do *all* sets represent notions of induction, i.e., is the universe itself well-founded?

Axiom of Foundation/Regularity 3.97. For any set X, $\in_X \subseteq X^2$ is well-founded.
That is, for any $A \subseteq X$, if every $x \in X$ with $x \subseteq A$ is in A, then $A = X$.
Equivalently, every $\varnothing \neq B$ $(\subseteq X)$ contains a $\in$-minimal x, i.e., $x \cap B = \varnothing$.

Example 3.98. Foundation rules out the existence of a set $x \in x$. To see this, we cannot simply take B above to be x, since x could contain a different element disjoint from x, e.g., if $x = \{\varnothing, x\}$. Instead, we can take $B := \{x\}$.

Example 3.99. Similarly, Foundation rules out $x \in y \in x$ (take $B = \{x, y\}$).

Example 3.100. Similarly, Foundation rules out a sequence $(x_0, x_1, \dots)$ such that $x_0 \ni x_1 \ni \cdots$; and in fact, assuming the Axiom of Choice, this is equivalent to Foundation (cf. Proposition 3.46).

Remark 3.101. The above means only what it says: Foundation rules out such a sequence $(x_0, x_1, \dots)$ *existing in the universe*. There *could* still be a set x_0, and a set x_1, and a set x_2, etc., such that $x_0 \ni x_1 \ni \cdots$; *but the entire sequence* $(x_0, x_1, \dots)$ *cannot exist as a set in the universe*. That is, Foundation in some sense only says that "the universe thinks it's well-founded", not that "the universe is actually well-founded". It turns out that the latter property is not expressible via any axiom of set theory that is "finitely expressible" (hence usable by humans)![14]

Remark 3.102. People have occasionally considered alternatives to the Axiom of Foundation. One example is *Aczel's Anti-Foundation Axiom*, which says "*every* binary relation $\prec \subseteq X^2$ has a unique Mostowski collapse function ξ". For instance, this implies "there is a unique x such that $x = \{x\}$" (called a *Quine atom*), given by Mostowski collapsing the loop on a single vertex 🔄!

3.G. **Transitivity and rank.** We begin this subsection by tying up a loose end from the last. Under the Axiom of Foundation, arbitrary sets are precisely the Mostowski collapses $\xi(x)$ of *elements* of sets with well-founded relations $(X, \prec)$. The entire set X is thus bisimilar to the image $\xi[X]$ (with $\in$); but this image is no longer an arbitrary set: it must be downward-closed (Exercise 3.71).

Definition 3.103. Let $\prec \subseteq X^2$ be a binary relation. We say that $x \in X$ is **$\prec$-transitive** if

$$\downarrow x \text{ is } \prec\text{-downward-closed},$$
$$\text{i.e., } \forall z \prec y \prec x \, (z \prec x),$$
$$\text{i.e., } \forall y \prec x \, (\downarrow y \subseteq \downarrow x).$$

Thus, $\prec$ is a transitive *relation* iff every $x \in X$ is a $\prec$-transitive *element*.

We say that a set X is **transitive** if it is $\in$-transitive, i.e., the following equivalent conditions:

$$X \text{ is } \in\text{-downward-closed},$$
$$\forall y \in x \in X \, (y \in X),$$
$$\forall x \in X \, (x \subseteq X),$$
$$X \subseteq \mathcal{P}(X).$$

Lemma 3.104. Let $(X, \prec_X), (Y, \prec_Y)$ be sets with binary relations, $f : X \to Y$ a function which is a bisimulation. If $x \in X$ is transitive, then so is $f(x) \in Y$.

Proof. Let $y'' \prec y' \prec f(x)$. Since f^{-1} is a simulation, there exist $x'' \prec x' \prec x$ such that $f(x') = y'$ and $f(x'') = y''$. Since x is transitive, $x'' \prec x$. Since f is a homomorphism, $y'' = f(x'') \prec f(x)$. □

[14] Using the compactness theorem in first-order logic, given any "actually well-founded" universe V of set theory, we can produce a bigger universe V' with a sequence $x_0 \ni x_1 \ni \cdots$ which obeys exactly the same first-order axioms as V. Such "non-standard" models of set theory have interesting applications in analysis, topology, and combinatorics, where they can be used to make rigorous sense of "infinities/infinitesimals" like in calculus. See Exercise 4.70.

Proposition 3.105. For a set X, the following are equivalent:

(i) X is a transitive set and $\in_X \subseteq X^2$ is well-founded (i.e., Foundation holds for X).
(ii) $\in_X$ is well-founded and $\xi_{\in_X} = \mathrm{id}_X$.
(iii) $\in_X$ is well-founded and $X = \xi_{\in}[X]$.
(iv) $X = \xi_{\prec}[Y]$ for some well-founded relation $\prec \subseteq Y^2$ on some set Y.
(v) $X = \xi_{\prec}(y)$ for some well-founded relation $\prec \subseteq Y^2$ and $\prec$-transitive element $y \in Y$.

Proof. (i) $\implies$ (ii): Note that id_X obeys the inductive 3.94 of ξ.
(ii) $\implies$ (iii) $\implies$ (iv) is obvious.
(iv) $\implies$ (v): Let y be a new element not in Y,[15] and put $z \prec y$ for all $z \in Y$; then $\xi[Y] = \xi(y)$.
(v) $\implies$ (i) by the preceding lemma. □

We now turn to classifying well-founded relations up to simulability $\lesssim$, rather than bisimilarity. It turns out that transitivity plays a key role here. A paradigmatic example is ordinary versus strong induction on $\mathbb{N}$, which are mutually simulable by Examples 3.58 and 3.72; note that the latter relation $(<)$ is the transitive closure of the former $(\prec)$. In general, we will show that

$$\text{“simulation = bisimulation + transitive closure”}.$$

Definition 3.106. A binary relation is a **strict partial order** if it is irreflexive, transitive, and antisymmetric, and a **(reflexive) partial order** if it is reflexive, transitive and antisymmetric (Definition 3.44). It is easily seen that for any set X, we have a bijection

$$\begin{aligned}\{\text{strict partial orders } < \subseteq X^2\} &\cong \{\text{(reflexive) partial orders } \leq \subseteq X^2\}\\ < &\mapsto < \cup (=_X)\\ \leq \setminus (=_X) &\mapsfrom \leq.\end{aligned}$$

When we have a $<$ or $\leq$, by default we always use the other symbol to denote its (pre)image under this bijection. We will simetimes also loosely refer to $<$ as a “partial order”.

A **partial well-order** is a well-founded *strict* partial order, or equivalently by Proposition 3.45, just a well-founded transitive relation, typically denoted $<$.

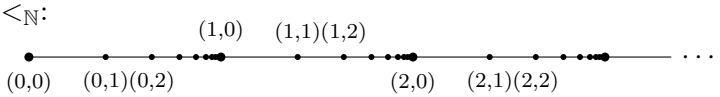
A **well-order** is a well-founded strict *linear* order, i.e., a partial well-order which furthermore obeys trichotomy (Definition 3.44).

Exercise 3.107. Show that a well-founded relation obeying trichotomy is automatically transitive.

Definition 3.108. The **transitive closure** of an arbitrary binary relation $\prec \subseteq X^2$ is the smallest transitive relation containing it (which exists by Knaster–Tarski; cf. Example 3.5).

Example 3.109. $<$ is a well-order on $\mathbb{N}$, and is the transitive closure of $x \prec y \iff x + 1 = y$.

Example 3.110. $<_{\mathrm{lex}}$ is a well-order on $\mathbb{N}^2$ (well-foundedness is by Proposition 3.29), which is much “longer” than $<_{\mathbb{N}}$:



Example 3.111. The transitive closure of the “grid” in Example 3.60 is the partial well-order

$$(a, b) < (c, d) \iff a \leq c \text{ and } b \leq d \text{ and } (a < c \text{ or } b < d)$$

(which is well-founded again because addition $h : \mathbb{N}^2 \to \mathbb{N}$ is a homomorphism, now to $<$).

Exercise 3.112. Assuming ordinary induction on $\mathbb{N}$, show that the transitive closure of $\prec$ is

$$x < y \iff \exists n \geq 1\, \exists x = x_0 \prec x_1 \prec \cdots \prec x_{n-1} \prec x_n = y.$$

[15] Under Foundation, we may take $y := Y$; regardless, Russell's paradox says that $y := \{z \in Y \mid z \notin z\}$ works.

Proposition 3.113. Let $\prec \subseteq X^2$ be a binary relation, $<$ be its transitive closure. Then for each $z \in X$, $\downarrow_< z$ is the $\prec$-downward-closure of $\downarrow_\prec z$. That is, $<$ is the smallest relation $\supseteq \prec$ such that

$$(*) \qquad x < y \prec z \implies x < z,$$
$$\text{i.e., } \prec \circ < \subseteq <.$$

Note that by definition, $<$ is the smallest relation $\supseteq \prec$ such that $< \circ < \subseteq <$; this gives us an induction principle for showing that $x < y$ implies some other binary relation $\phi(x, y)$ (by showing that ϕ is also transitive and contains $\prec$). In contrast, the above yields an induction principle for showing that $x < y$ *for fixed* x implies some other *unary* relation $\phi(y)$.

The proof is easy using the preceding Exercise: it is easily seen that the same "finite path" definition also works for the smallest relation containing $\prec$ and obeying $(*)$. But as usual, we want to give a proof avoiding the use of $\mathbb{N}$.

Proof. It is clear that $<$ obeys $(*)$, hence contains the smallest relation $<'$ containing $\prec$ and obeying $(*)$. It remains to show that $<'$ is also transitive, hence contains $<$. That is, we must show

$$\forall y <' z \underbrace{\forall x \in X \, (x <' y \implies x <' z)}_{\phi(y,z)}.$$

We induct on $y <' z$, i.e., we show that the set of (y, z) satisfying ϕ also contains $\prec$ and obeys $(*)$. Indeed, $y \prec z \implies \phi(y, z)$ since $<'$ satisfies $(*)$ by definition. And we have

$$\begin{aligned} \phi(x, y) \text{ and } y \prec z &\iff \forall w \in X \, (w <' x \implies w <' y) \text{ and } y \prec z \\ &\implies \forall w \in X \, (w <' x \implies w <' z) \iff \phi(x, z) \end{aligned}$$

again since $<'$ satisfies $(*)$. □

Corollary 3.114. Let $\prec \subseteq X^2$, $<$ be its transitive closure, and $\leq := < \cup (=)$. Then

$$< = \prec \cup (\prec \circ <) = \prec \circ \leq.$$

Proof. By Exercise 3.18(c) (applied to $T(R) := \prec \circ R$). □

Corollary 3.115. Let $\prec \subseteq X^2$ be a binary relation, $<$ be its transitive closure, and $\leq = < \cup (=)$. Then $\leq$ is a simulation between each pair among $(X, \prec), (X, <)$. In particular, for all $x \in X$,

$$x \mathrel{\overline{\sim}}_{(X,\prec),(X,<)} x;$$

and $\mathrm{WF}(\prec) = \mathrm{WF}(<)$, so $\prec$ is well-founded iff $<$ is.

This generalizes Example 3.90 in the case of $\mathbb{N}$.

Proof. To say that $\leq$ is a simulation of $<$ in $\prec$ means (Definition 3.67)

$$(\leq \circ <) \subseteq (\prec \circ \leq).$$

The LHS is $(< \cup =) \circ < = (< \circ <) \cup < = <$, which is equal to the RHS by the preceding result.

For the other three simulations, we either decrease $<$ on the LHS to $\prec$, or increase $\prec$ on the RHS to $<$, so the $\subseteq$ is even more true. The last claim follows from Proposition 3.68. □

Example 3.116. In contrast to the case of $\mathbb{N}$ (Example 3.90), even between a partial well-order $<$ and itself, $\lesssim$ could be strictly more than $\leq$. Take $a < b, c$, where b, c are incomparable; then $b \equiv c$.

On the other hand:

Lemma 3.117. For any well-founded $\prec \subseteq X^2$, we have $\lesssim_{X,X} \cap \succ = \varnothing$.

Proof. If $y \prec x \lesssim y$, then there exists $z \prec y$ such that $y \lesssim z$; thus the set of such x (or such y) has no minimal element, hence must be empty. □

Corollary 3.118. For a well-order $(X, <)$, we have $\lesssim_{X,X} = \leq$; thus $(\eqsim) = (\equiv) = (=)$.

Proof. By the two preceding results, $\leq \subseteq \lesssim \subseteq X^2 \setminus > = \leq$. Thus $(\eqsim) = (\lesssim \cap \gtrsim) = (\leq \cap \geq) = (=)$. □

Corollary 3.119. For any binary relation $(X, \prec)$ and well-order $(Y, <)$, $\eqsim_{X,Y}$ is a partial function.

Proof. $(\eqsim_{X,Y}) \circ (\eqsim_{X,Y})^{-1} = (\eqsim_{X,Y}) \circ (\eqsim_{Y,X}) \subseteq (\eqsim_{Y,Y}) = (=_Y)$ (cf. Exercise 2.55). □

We now show that the global structure of $\lesssim$ is (essentially) a proper-class-sized well-order:

Lemma 3.120. Let $(X, \prec_X), (Y, \prec_Y)$ be sets with binary relations, $x \in X$, and $y \in Y$. Then

$$x \lesssim y \text{ or } \exists x' \prec x \, (y \lesssim x');$$

in particular,

$$x \lesssim y \text{ or } y \lesssim x.$$

Proof. The second line follows from the first, since $x' \prec x \implies x' \lesssim x$ (Corollary 3.115).

To prove the first line: if $y \notin \mathrm{WF}(\prec_Y)$, then $x \lesssim y$ by Example 3.88. Now suppose $y \in \mathrm{WF}(\prec_Y)$. We prove by induction that $\forall x \not\lesssim y \, \exists x' \prec x \, (y \lesssim x')$. Assume this holds for all $y' \prec y$. If $x \not\lesssim y$, then

$$\exists x' \prec x \, \forall y' \prec y \, (x' \not\lesssim y'),$$

which by the IH means

$$\exists x' \prec x \, \forall y' \prec y \, \exists x'' \prec x' \, (y' \lesssim x''),$$

or

$$\exists x' \prec x \, (y \lesssim x').$$

□

Corollary 3.121. Let $(X, \prec_X), (Y, <_Y)$ be sets with binary relations, with the latter a partial well-order. Then for $x \in X$ and $y \in Y$,

$$x \lesssim y \iff \exists y' \leq y \, (x \eqsim y'),$$

i.e.,

$$\lesssim_{X,Y} = \leq_Y \circ \eqsim_{X,Y}.$$

Thus

$$x \lesssim y \iff \forall x' \prec x \, \exists y' < y \, (x' \eqsim y').$$

Proof. $\Longleftarrow$ follows from Corollary 3.115. Conversely, suppose $x \lesssim y$. Let $y' \leq y$ be minimal such that $x \lesssim y'$. Then for every $y'' < y'$, we have $x \not\lesssim y''$ (here using that $y'' \leq y$ by transitivity), so by the preceding lemma, $y'' \lesssim x'$ for some $x' \prec x$. Thus $y' \lesssim x$. For the last statement: we have

$$\begin{aligned} x \lesssim y &\iff \forall x' \prec x \, \exists y' < y \, (x' \lesssim y') \\ &\iff \forall x' \prec x \, \exists y'' \leq y' < y \, (x' \eqsim y'') \\ &\iff \forall x' \prec x \, \exists y'' < y \, (x' \eqsim y''). \end{aligned}$$

□

Corollary 3.122. Let $(X, <_X), (Y, <_Y)$ be sets with partial well-orders. Then $(\eqsim_{X,Y}) = (\equiv_{X,Y})$.

Proof. $\supseteq$ is by Corollary 3.85. To show $\subseteq$: by the preceding result,

$$\begin{aligned} \eqsim_{X,Y} \circ <_X &\subseteq \lesssim_{X,Y} \circ <_X \\ &\subseteq <_Y \circ \lesssim_{X,Y} \\ &= <_Y \circ \leq_Y \circ \eqsim_{X,Y} \\ &= <_Y \circ \eqsim_{X,Y}, \end{aligned}$$

and similarly $\eqsim_{Y,X} \circ <_Y \subseteq <_X \circ \eqsim_{Y,X}$, thus $\eqsim_{X,Y}$ is a bisimulation. □

Corollary 3.115 tells us that up to simulability, every (well-founded) relation may be replaced with a partial (well-)order; while the two preceding results tell us that simulability between partial well-orders may essentially be understood in terms of bisimilarity, for which we already have the complete classification via Mostowski collapse ξ. Putting everything together, we get: for two well-founded relations $(X, \prec_X), (Y, \prec_Y)$, with transitive closures $<_X, <_Y$, for $x \in X$ and $y \in Y$,

$$\begin{aligned}
x \mathrel{\overline{\sim}}_{\prec_X,\prec_Y} y &\iff x \mathrel{\overline{\sim}}_{<_X,<_Y} y && \text{by Corollary 3.115}\\
&\iff x \equiv_{<_X,<_Y} y && \text{by Corollary 3.122}\\
&\iff \xi_{<_X}(x) = \xi_{<_Y}(y) && \text{by (3.93)},\\
x \mathrel{\lesssim}_{\prec_X,\prec_Y} y &\iff x \mathrel{\lesssim}_{<_X,<_Y} y && \text{by Corollary 3.115}\\
&\iff \exists y' \le y\, (x \equiv_{<_X,<_Y} y') && \text{by Corollaries 3.121 and 3.122}\\
&\iff \xi_{<_X}(x) \in \xi_{<_Y}(y) \text{ or } \xi_{<_X}(x) = \xi_{<_Y}(y) && \text{by (3.93)}\\
&\iff \forall x' < x\, \exists y' < y\, (x' \equiv_{<_X,<_Y} y') && \text{by Corollaries 3.121 and 3.122}\\
&\iff \xi_{<_X}(x) \subseteq \xi_{<_Y}(y).
\end{aligned}$$

Definition 3.123. Let $\prec \subseteq X^2$ be a well-founded relation, $<$ be its transitive closure. The **$\prec$-rank** of $x \in X$ is defined inductively by[16]

$$\begin{aligned}
\rho(x) = \rho_\prec(x) := \xi_<(x) &= \{\xi_<(x') \mid x' < x\}\\
&= \{\rho_\prec(x) \mid x' < x\}\\
&= \{\rho(x') \mid x' \prec x\} \cup \{\rho(x'') \mid x'' < x' \prec x\} && \text{by Corollary 3.114}\\
&= \{\rho(x') \mid x' \prec x\} \cup \textstyle\bigcup_{x' \prec x} \rho(x')\\
&= \textstyle\bigcup_{x' \prec x} (\rho(x) \cup \{\rho(x)\}).
\end{aligned}$$

Thus, for two sets with well-founded relations $(X, \prec_X), (Y, \prec_Y)$, for $x \in X$ and $y \in Y$, by the above,

$$\begin{aligned}
x \mathrel{\overline{\sim}} y &\iff \rho(x) = \rho(y),\\
x \mathrel{\lesssim} y &\iff \rho(x) \in \rho(y) \text{ or } \rho(x) = \rho(y) \iff \rho(x) \subseteq \rho(y).
\end{aligned} \tag{3.124}$$

Similarly to Proposition 3.96, $\rho(x)$ in fact picks out a representative in the $\simeq$-class of x:

$$(X, \prec) \xrightarrow[\subseteq \mathrel{\overline{\sim}} \text{ (by 3.115)}]{\text{id}} (X, <) \xrightarrow[\subseteq \equiv \text{ (by 3.96)}]{\xi} (V, \in), \qquad (X, \prec) \xrightarrow{\ \rho\ } (V, \in)$$

Example 3.125. In $\mathbb{N}$ with the successor relation $x \prec y :\iff x + 1 = y$:

$$\begin{aligned}
\rho(0) &= \xi(0) = \varnothing,\\
\rho(1) &= \rho(0) \cup \{\rho(0)\} = \varnothing \cup \{\varnothing\} = \{\varnothing\},\\
\rho(2) &= \rho(1) \cup \{\rho(1)\} = \{\varnothing\} \cup \{\{\varnothing\}\} = \{\varnothing, \{\varnothing\}\},\\
\rho(3) &= \rho(2) \cup \{\rho(2)\} = \{\varnothing, \{\varnothing\}\} \cup \{\{\varnothing, \{\varnothing\}\}\} = \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}.
\end{aligned}$$

(We will soon define $\mathbb{N}$ so that this is the identity function; see Axiom 3.138.)

Exercise 3.126. Let $T : \mathcal{P}(X) \to \mathcal{P}(X)$ be a monotone set operator. Call T **transitive** if $T \subseteq T \circ T$.

(a) Show that for $T = T_\prec$, this is equivalent to transitivity of $\prec$.

(b) Show that $\breve{T} := T^\circ \circ T$ is the pointwise largest operator pointwise below T, i.e., $\breve{T}(A) \subseteq T(A)$.

(c) Show that $\overline{T}(\varnothing) = \overline{\breve{T}}(\varnothing)$. [Recall 3.78. Note that id_X is a simulation of T in $\breve{T}$, while T°, which corresponds to $T_\le$ when $T = T_\prec$, is a "simulation" in the reverse direction.]

[16] As in Remark 3.95, we do not yet know that this inductive definition is justified.

3.H. **Ordinal numbers.**

Proposition 3.127 (cf. Proposition 3.105)**.** For a set α, the following are equivalent:

(i) α is transitive and $\in_\alpha \subseteq \alpha \times \alpha$ is a well-order.
(ii) α is transitive and $\in_\alpha$ is a partial well-order, i.e., transitive and well-founded.
(iii) α is transitive, each $\beta \in \alpha$ is transitive, and $\in_\alpha$ is well-founded (i.e., Foundation for α).
(iv) $\in_\alpha$ is a partial well-order and $\xi_{\in_\alpha} = \mathrm{id}_\alpha$.
(v) $\in_\alpha$ is a partial well-order and $\alpha = \xi_\in[\alpha]$.
(vi) $\in_\alpha$ is well-founded and $\rho_{\in_\alpha} = \mathrm{id}_\alpha$.
(vii) $\in_\alpha$ is well-founded and $\alpha = \rho_\in[\alpha]$.
(viii) $\alpha = \xi_<[X]$ for some partial well-order $< \subseteq X^2$.
(ix) $\alpha = \xi_<(x)$ for some partial well-order $< \subseteq X^2$ and $x \in X$.
(x) $\alpha = \rho_\prec[X]$ for some well-founded relation $\prec \subseteq X^2$.
(xi) $\alpha = \rho_\prec(x)$ for some well-founded relation $\prec \subseteq X^2$ and $x \in X$.

Proof. (i) $\implies$ (ii) $\iff$ (iii) is obvious.
(ii) $\iff$ (iv) by Proposition 3.105.
(iv) $\implies$ (v) $\implies$ (viii) and (vi) $\implies$ (vii) $\implies$ (x) are obvious.
(iv) $\implies$ (vi), (v) $\implies$ (vii), (viii) $\implies$ (x), and (ix) $\implies$ (xi) since $\rho = \xi$ for transitive relations.
(viii) $\impliedby$ (x) and (ix) $\impliedby$ (xi) since $\rho_\prec = \xi_<$ by definition.
(viii) $\implies$ (ix) and (x) $\implies$ (xi) by adding a new greatest element to X, as in Proposition 3.105.
(ix), (xi) $\implies$ (i): $\alpha = \xi_<(x)$ is transitive and $\in_\alpha$ is well-founded by Proposition 3.105; and linearity holds since $\rho(x) \le \rho(y) \iff x \lesssim y$ and $\lesssim$ is linear by Lemma 3.120. □

Definition 3.128. An **ordinal number** is a set obeying the above equivalent conditions.

The class of all ordinal numbers is denoted $\mathbb{ON}$ (also known as Ord or ∞[17]), and ordered via

$$\alpha < \beta \ :\iff\ \alpha \in \beta,$$
$$\alpha \le \beta \ :\iff\ \alpha \in \beta \text{ or } \alpha = \beta \iff \alpha \subseteq \beta.$$

Proposition 3.129. $\mathbb{ON}$ is a transitive class, i.e., every element of an ordinal is an ordinal; and $\in$ is a well-order on the class $\mathbb{ON}$, where well-foundedness means:

(a) The **principle of transfinite induction**: if $A \subseteq \mathbb{ON}$ is a *subclass*, and every $\alpha \in \mathbb{ON}$ with $\alpha \subseteq A$ is in A, then $A = \mathbb{ON}$.
(b) Equivalently, every *class* $\varnothing \ne B \subseteq \mathbb{ON}$ has a $\in$-minimal element.
(c) In particular, this holds for nonempty sets B of ordinals.

Proof. Linearity again follows from linearity of $\lesssim$ (Lemma 3.120). The only thing left to check is that well-foundedness of each individual $\alpha \in \mathbb{ON}$ implies well-foundedness of the entirety of $\mathbb{ON}$. Indeed, let $\varnothing \ne B \subseteq \mathbb{ON}$ be a class; then there is some $\alpha \in B$. If α is minimal in B, we're done. Otherwise, $\alpha \cap B \subseteq \alpha$ is nonempty, hence has a minimal element, which is also minimal in B. □

Corollary 3.130 (Burali-Forti paradox)**.** $\mathbb{ON}$ is a proper class.

Proof. Otherwise, it is an ordinal, whence $\mathbb{ON} \in \mathbb{ON}$, contradicting well-foundedness. □

Corollary 3.131. For a set A, there is a **minimum excluded** ordinal $\operatorname{mex} A \notin A$. □

We can explicitly describe $\operatorname{mex} A$ as a set: an ordinal α will be in $\operatorname{mex} A$ iff it is $< \operatorname{mex} A$, iff it and all its predecessors $\beta \in \alpha$ are in A, iff $\alpha \subseteq A$. Thus

$$\operatorname{mex} A = \{\alpha \in A \cap \mathbb{ON} \mid \alpha \subseteq A\}.$$

[17] because $\mathbb{ON}$ obeys the same properties as the ordinals, except for being a proper class, hence can be thought of as an "absolute infinity" bigger than all ordinals; see Corollary 3.130

Example 3.132. If $A \cap \mathbb{ON}$ is already downward-closed (i.e., transitive), then $\operatorname{mex} A = A \cap \mathbb{ON}$.

Example 3.133. There is a least ordinal $0 := \operatorname{mex} \varnothing = \varnothing =$ the rank $\rho(x)$ of any minimal element x with respect to any binary relation.

Example 3.134. There is a next least $1 := \operatorname{mex}\{0\} = \{0\} =$ the rank of any non-minimal element all of whose predecessors are minimal.

Example 3.135. There is a next least $2 := \operatorname{mex}\{0, 1\} = \{0, 1\}$. (Note however that $\operatorname{mex}\{1\} = \{0\}$.)

Definition 3.136. More generally, any ordinal α has a **successor**[18]

$$\begin{aligned} \alpha^+ &:= \operatorname{mex}\{\beta \in \mathbb{ON} \mid \beta \leq \alpha\} \\ &= \operatorname{mex}(\alpha \cup \{\alpha\}) = \alpha \cup \{\alpha\}. \end{aligned}$$

Definition 3.137. An ordinal α is a **limit ordinal** if it is neither 0 nor a successor, i.e., $0 < \alpha$, and for every $\beta < \alpha$, also $\beta^+ < \alpha$, i.e., α contains 0 and is closed under the successor operation.

Axiom of Infinity 3.138. There exists a limit ordinal, hence a least one, called $\mathbb{N}$ or ω.

Note that by Knaster–Tarski, $\mathbb{N}$ (if it exists) is equivalently the intersection of all limit ordinals, i.e., $n \in \mathbb{N}$ iff n is less than (i.e., belongs to) every limit ordinal. Regardless of whether the Axiom of Infinity holds, we call such ordinals n **natural numbers**, and denote the class of them by $\mathbb{N}$. Thus,

$$\mathsf{Infinity} \iff \mathbb{N} \text{ is a set} \iff \mathbb{N} \neq \mathbb{ON}.$$

Exercise 3.139. Suppose there exists a set X which contains $\varnothing$ and is closed under the operation $x \mapsto x \cup \{x\}$. Then by Knaster–Tarski, there is a smallest such X. Prove (without using Infinity or Foundation) that X is then transitive and $\in_X$ is transitive and well-founded, hence $X = \mathbb{N}$.

Thus, the Axiom of Infinity may also be stated as follows, without mentioning transitivity: there is a set containing $\varnothing$ and closed under $x \mapsto x \cup \{x\}$ (whence there is a smallest such set $\mathbb{N}$).

Definition 3.140. Zermelo–Fraenkel set theory ZF consists of the axioms of $\mathsf{ZF}^- - \mathsf{Infinity}$ (2.26), plus the Axioms of Foundation 3.97 and Infinity 3.138. ZF^- is the same but without Foundation.[19]

Under ZF^-, the ordinals look like:

$$0 < 1 < 2 < \cdots < \omega < \underbrace{\omega^+}_{=:\omega+1} < \underbrace{\omega^{++}}_{=:\omega+2} < \cdots < \sup{}^{(+)}\{\omega, \underbrace{\omega^+, \omega^{++}}_{=:\omega+\omega}, \dots\} < \cdots$$

We see that while ω is an "infinite" number, it is actually the smallest "infinity"; thus we use the more precise symbol ω, rather than ∞. (As noted above, when ∞ is used in a context involving ordinals, it usually denotes the "absolute infinity" $\mathbb{ON}$.)

Definition 3.141. For any set $A \subseteq \mathbb{ON}$, its **least upper bound** or **supremum** is

$$\sup A := \bigcup A = \{\beta \in \mathbb{ON} \mid \exists \alpha \in A\, (\beta < \alpha)\}$$

(since $\beta \geq \sup A \iff \forall \alpha \in A\, (\beta \geq \alpha)$), while its **least strict upper bound** is

$$\begin{aligned} \sup{}^+ A &:= A \cup \bigcup A = \{\beta \in \mathbb{ON} \mid \exists \alpha \in A\, (\beta \leq \alpha)\} = \sup\nolimits_{\alpha \in A} \alpha^+ \\ &= \begin{cases} \sup A & \text{if } A \text{ does not have a maximum element,} \\ \alpha^+ = \alpha \cup \{\alpha\} & \text{if } A \text{ has maximum } \alpha. \end{cases} \end{aligned}$$

[18] Not to be confused with the **successor cardinal**; see Definition 5.20. This notation α^+ is almost never used for successor ordinal, which is usually denoted as the soon-to-be-introduced $\alpha + 1$; see Example 3.152.

[19] Zermelo set theory Z^- is missing Foundation and Replacement, but includes Infinity; see (2.27). In fact, Zermelo originally introduced the "wrong" version of Infinity, where n is encoded as $\{\{\cdots\{\varnothing\}\cdots\}\}$; in other words, as the Mostowski collapse with respect to the successor graph, rather than the $<$ relation. It turns out that this version of Infinity is insufficient to prove the nowadays standard version, i.e., on some foundational level, *strong induction* really is stronger than *ordinary induction*! The modern "strong" encoding of naturals is called the *von Neumann encoding*.

Remark 3.142. Definition 3.123 of rank of well-founded $\prec \subseteq X^2$ now says

$$\rho(x) = \sup^+_{y \prec x} \rho(y).$$

Using suprema and Infinity, we may produce many bigger examples of limit ordinals:

Example 3.143. The countable sequence $\omega < \omega^+ < \omega^{++} < \cdots$ (technically defined by induction on ω; see below) has a $\sup^{(+)}$, which is usually called $\omega + \omega$ (see Definition 3.148).

Taking successor ω many more times yields an ordinal called $\omega + \omega + \omega$.

Repeating this process ω times, we get an even bigger ordinal called ω^2 (see Exercise 3.159).

Similarly, we may produce $\omega^3 < \omega^4 < \cdots$, with supremum ω^ω, and then $\omega^\omega \cdot \omega = \omega^{\omega+1}$, ..., $\omega^{\omega+\omega}$, ..., ω^{ω^2}, ..., ω^{ω^ω}, ..., $\omega^{\omega^{\omega^\omega}}$, $\omega^{\omega^{\omega^{\omega^\omega}}}$, ...; the supremum of *this* sequence is called ε_0.

Note that all of these increasingly huge ordinals are all still countable (i.e., have countably many predecessors, by repeated use of the fact that a countable union of countable is countable).

Theorem 3.144 (Hartogs)**.** For any set X, there is an ordinal that does not inject into X, hence a least such ordinal $\eta(X)$, called the **Hartogs number** of X.

Proof. By definition, an ordinal $\alpha < \eta(X)$ iff α does inject into X; we must show that the set of all such α forms a set. Indeed, if $f : \alpha \hookrightarrow X$ is an injection, then $f : \alpha \cong \mathrm{im}(f)$ is a bijection, hence we may transfer the ordering $<_\alpha = \in_\alpha$ on α to a well-ordering $<_{\mathrm{im}(f)} \subseteq \mathrm{im}(f)^2$ such that f becomes an isomorphism, whence $f^{-1} = \xi_{<_{\mathrm{im}(f)}} : \mathrm{im}(f) \cong \alpha$. Thus

$$\eta(X) = \{\alpha \mid \exists A \subseteq X,\ <_A \subseteq A^2\ (<_A \text{ is a well-order on } X \text{ and } \alpha = \xi_{<_A}[A])\}$$

is a set by Replacement. □

Example 3.145. There is a least uncountable ordinal, called $\omega_1 := \eta(\omega)$, much bigger than all of the ordinals built from ω above. Then there is an even bigger $\omega_2 := \eta(\omega_1)$, etc. (See Definition 5.20.)

As usual for a well-founded relation, we have not only a principle of transfinite induction for $\mathbb{ON}$, but also a principle of inductive definitions as in 3.51. We will in fact give a new proof of this, more powerful than that in Theorem 3.51 because it does not rely on the impredicative Knaster–Tarski Theorem 3.6, hence does not require either the domain or codomain of the function to be sets:

Theorem 3.146 (principle of transfinite inductive definition)**.** Let $(Y_\alpha)_{\alpha \in \mathbb{ON}}$ be a family of classes,

$$\left(F_\alpha : \textstyle\prod_{\beta<\alpha} Y_\beta \to Y_\alpha\right)_{\alpha \in \mathbb{ON}} \in \textstyle\prod_{\alpha \in \mathbb{ON}} Y_\alpha^{\prod_{\beta<\alpha} Y_\beta}.$$

Formally, the family $(Y_\alpha)_\alpha$ is (not a function from ordinals to classes but) a single proper class Y of pairs (α, y) such that $y \in Y_\alpha$; likewise, the F_α's can be represented as a single proper class function $F(\alpha, \vec{y})$. Then there is a unique (proper class) function $f \in \prod_{\alpha \in \mathbb{ON}} Y_\alpha$ such that for each α,

$$(*) \qquad f(\alpha) = F_\alpha((f(\beta))_{\beta<\alpha}).$$

Proof. We will first prove that for each $\alpha \in \mathbb{ON}$, it is possible to define a unique initial segment $f_\alpha \in \prod_{\beta \le \alpha} Y_\beta$ of our desired f obeying $(*)$, by induction on α. Assume that for all $\alpha' < \alpha$, it is possible to define a unique $f_{\alpha'}$ on all $\beta \le \alpha'$. Then for $\alpha'' < \alpha' < \alpha$, by uniqueness, $f_{\alpha''}$ must be the restriction of $f_{\alpha'}$ to $\beta \le \alpha''$, since said restriction also obeys $(*)$. It follows that $g_\alpha := \bigcup_{\alpha' < \alpha} f_{\alpha'}$ is still a function (on domain α), since any two $f_{\alpha'}, f_{\alpha''}$ for $\alpha', \alpha'' < \alpha$ agree on their common domain. The desired f_α is then g_α extended with the single value $f_\alpha(\alpha) := F_\alpha(g_\alpha)$, by $(*)$.

Now given the unique f_α defined on $\beta \le \alpha$ for all α, by the same uniqueness argument as above, for $\alpha' < \alpha$, $f_{\alpha'}$ must be the restriction of f_α. Then the desired f is $\bigcup_{\alpha \in \mathbb{ON}} f_\alpha$, i.e., $f(\alpha) := f_\alpha(\alpha)$. □

Exercise 3.147. Along similar lines, give a direct proof of well-founded inductive definition (3.51) for any partial well-order $< \subseteq X^2$ on a set, with the codomains $(Y_x)_{x \in X}$ allowed to be classes. Using transitive closure, extend this to arbitrary well-founded relations $\prec \subseteq X^2$.

3.I. **Ordinal arithmetic.**

Definition 3.148. The **sum** of two ordinals α, β is defined by induction on β as follows:

$$\begin{aligned}\alpha + \beta &:= \alpha \cup \{\alpha + \gamma \mid \gamma < \beta\} \\ &= \begin{cases} \alpha & \text{if } \beta = 0, \\ \sup^+_{\gamma<\beta}(\alpha+\gamma) & \text{if } \beta > 0, \end{cases} \\ &= \begin{cases} \alpha & \text{if } \beta = 0, \\ (\alpha+\gamma)^+ & \text{if } \beta = \gamma^+, \\ \sup_{\gamma<\beta}(\alpha+\gamma) & \text{if } \beta \text{ is a limit ordinal.} \end{cases}\end{aligned}$$

Exercise 3.149. How do these definitions fit into the formalism of Theorem 3.146?

Proposition 3.150. The three definitions above really are equivalent.

These three definitions illustrate a common pattern in many definitions by transfinite induction: we may either separate into a "base case" for zero, an "inductive case" for successor, and a limit case which typically just takes supremum; or we may combine successor and limit (and sometimes also zero) by applying the inductive case to all predecessors and then taking their supremum.

Proof. First, we take the first definition as given, and check that the second recursive equation then also holds. If $\beta = 0$, then clearly $\alpha + \beta = \alpha$ as desired. Otherwise, we have

$$\begin{aligned}\sup^+_{\gamma<\beta}(\alpha+\gamma) &= \{\alpha+\gamma \mid \gamma<\beta\} \cup \textstyle\bigcup_{\gamma<\beta}(\alpha+\gamma) \quad \text{by Definition 3.141} \\ &= \{\alpha+\gamma \mid \gamma<\beta\} \cup \textstyle\bigcup_{\gamma<\beta}(\alpha \cup \{\alpha+\delta \mid \delta<\gamma\}) \\ &= \{\alpha+\gamma \mid \gamma<\beta\} \cup \alpha \quad \text{since } \delta<\gamma<\beta \implies \delta<\beta.\end{aligned}$$

Next, note that by either the first or second definition,

Proposition 3.151. $+$ is strictly monotone in the second argument: $\gamma < \beta \implies \alpha+\gamma < \alpha+\beta$. □

Now we check that the second and third definitions are equivalent. If $\beta = \gamma^+$, then by monotonicity, the largest $\alpha+\delta$ among $\delta<\beta$ is $\alpha+\gamma$, hence the $\sup^+$ in the second definition reduces to $(\alpha+\gamma)^+$. And if β is a limit ordinal, then by strict monotonicity, the set of $\alpha+\delta$ for $\delta<\beta$ has no upper bound either, hence the $\sup^+$ reduces to a sup. □

Example 3.152. For any ordinal α, we have

$$\begin{aligned}\alpha + 0 &= \alpha, \\ \alpha + 1 &= (\alpha+0)^+ = \alpha^+.\end{aligned}$$

(So from now on, we will rarely write α^+.) For instance,

$$1 + 1 = 1^+ = 2.$$

Note that the second clause of the third definition above becomes

$$\alpha + (\gamma+1) = (\alpha+\gamma) + 1.$$

Proposition 3.153. $+$ on ordinals is associative: $(\alpha+\beta)+\gamma = \alpha+(\beta+\gamma)$.

Proof. By induction on γ. If $\gamma = 0$, both sides simplify to $\alpha+\beta$. Otherwise,

$$\begin{aligned}(\alpha+\beta)+\gamma = \sup^+_{\delta<\gamma}((\alpha+\beta)+\delta) &= \sup^+_{\delta<\gamma}(\alpha+(\beta+\delta)) \quad \text{by IH} \\ &= \sup^+_{\varepsilon<\beta+\gamma}(\alpha+\varepsilon) = \alpha+(\beta+\gamma),\end{aligned}$$

using in the second-last step that every $\varepsilon < \beta+\gamma = \beta \cup \{\beta+\delta \mid \delta<\gamma\}$ is either some $\beta+\delta$, or $< \beta = \beta+0$ whence $\alpha+\varepsilon < \alpha+(\beta+0)$. □

Exercise 3.154. Prove that $0 + \alpha = \alpha$ for every ordinal α.

Example 3.155. We have $1 + \omega = \sup^+\{1+0, 1+1, 1+2, \dots\} = \omega \neq \omega + 1$. It follows by induction that $1 + \alpha = \alpha \neq \alpha + 1$ for any $\alpha \geq \omega$.

Proposition 3.156. $+$ on *naturals* is commutative: $m + n = n + m$ for all $m, n \in \mathbb{N}$.

Proof. First, we prove $1+n = n+1$ by induction. We have $1+0 = 1 = 0+1$. Assuming $1+n = n+1$,

$$\begin{aligned} 1 + (n+1) &= (1+n) + 1 && \text{by definition of } + \\ &= (n+1) + 1 && \text{by IH.} \end{aligned}$$

We now prove $m + n = n + m$ by induction on n. We have

$$\begin{aligned} m + 0 &= m && \text{by definition of } + \\ &= 0 + m && \text{by Exercise 3.154.} \end{aligned}$$

Now suppose $m + n = n + m$. Then

$$\begin{aligned} m + (n+1) &= (m+n) + 1 && \text{by definition of } + \\ &= (n+m) + 1 && \text{by IH} \\ &= n + (m+1) && \text{by definition of } + \\ &= n + (1+m) && \text{by previous case} \\ &= (n+1) + m && \text{by associativity.} \end{aligned}$$

□

Exercise 3.157. Prove that $+$ is weakly monotone in the first argument: $\alpha \leq \beta \implies \alpha + \gamma \leq \beta + \gamma$.

Exercise 3.158. Prove that $\alpha \leq \beta$ iff $\alpha + \gamma = \beta$ for some γ, and in that case, γ is unique.

Exercise 3.159. Let $(X, <)$ be a well-ordered set (e.g., an ordinal), and $(\alpha_x)_{x \in X}$ be a family of ordinals. Define the **indexed sum** $\sum_{x \in X} \alpha_\gamma$ by induction on $\rho[X]$ as follows:

$$\textstyle\sum_{x \in X} \alpha_x := \sup_{x \in X}(\sum_{y < x} \alpha_y + \alpha_x).$$

(a) Give an equivalent definition, split into "zero", "successor", and "limit" cases, depending on whether X is empty or has a greatest element.
(b) Verify that $\sum_{x \in 1} \alpha_x = \alpha_0$ and $\sum_{x \in 2} \alpha_x = \alpha_0 + \alpha_1$.
(c) Prove that $\sum_{x < \beta} 1 = \beta$.

As a special case, define the **product** of ordinals α, β by

$$\alpha \cdot \beta := \textstyle\sum_{\gamma < \beta} \alpha.$$

(d) Conclude that $\alpha \cdot 1 = \alpha = 1 \cdot \alpha$.
(e) Prove that $\alpha \cdot 0 = 0 = 0 \cdot \alpha$.
(f) What are $\omega \cdot 2$, $2 \cdot \omega$?

We now give another perspective on ordinal sums and products. Let $(X, <_X)$ be a well-ordered set, and for each $x \in X$, let $(Y_x, <_{Y_X})$ be a well-ordered set. Recall from Definition 2.76

$$\textstyle\bigsqcup_{x \in X} Y_x := \{(x, y) \mid x \in X \text{ and } y \in Y_x\}.$$

The **lexicographical order** $<_{\text{lex}}$ on $\bigsqcup_{x \in X} Y_x$ is defined as in 3.28, and is a well-order as in 3.29.

(g) Prove that $\rho_{<_{\text{lex}}}[\bigsqcup_{x \in X} Y_x] = \sum_{x \in X} \rho[Y_x]$. In particular, $\rho_{<_{\text{lex}}}[\bigsqcup_{x \in X} \alpha_x] = \sum_{x \in X} \alpha_x$.
(h) Prove the *indexed associative law*: for any well-ordered sets X, $(Y_x)_{x \in X}$ and ordinals $(\alpha_{x,y})_{x \in X, y \in Y_x}$,

$$\textstyle\sum_{x \in X} \sum_{y \in Y_x} \alpha_{x,y} = \sum_{(x,y) \in \bigsqcup_{x \in X} Y_x} \alpha_{x,y}.$$

(i) Conclude that $\cdot$ on ordinals is associative and distributes over $\sum$ on one side [see (f)].

Exercise 3.160. Prove that $\cdot$ on *naturals* is commutative.

Exercise 3.161. What can you say about monotonicity of $\cdot$?

Exercise 3.162. Prove that for any ordinals α, β, there are unique γ (quotient) and δ (remainder) such that $\alpha = \beta \cdot \gamma + \delta$ and $\delta < \beta$.

Exercise 3.163. Let $(X, <)$ be a well-ordered set, and $(\alpha_x)_{x \in X}$ be a family of ordinals. Define the **indexed product** $\prod_{\gamma<\beta} \alpha_\gamma$ by induction on $\rho[X]$ as follows:[20]

$$\prod_{x \in X} \alpha_x := \begin{cases} 1 & \text{if } X = \varnothing, \\ \prod_{y<x} \alpha_y \cdot \alpha_x & \text{if } X \text{ has a maximum } x, \\ \sup_{z<x \in X} \prod_{y<x} \alpha_y & \text{if } X \text{ is neither empty nor has a maximum.} \end{cases}$$

In particular, for two ordinals α, β, define **ordinal exponentiation**

$$\alpha^\beta := \prod_{\gamma<\beta} \alpha.$$

(a) What is 0^α? More generally, what is $\prod_{\gamma<\beta} \alpha_\gamma$ if some $\alpha_\gamma = 0$?
(b) What is 1^α?
(c) What is 2^ω? [See Footnote 20.]
(d) Prove that $(m \cdot n)^k = m^k \cdot n^k$ for *naturals* m, n, k. Give a counterexample for ordinals.

For a linearly ordered set $(X, <_X)$ and partially ordered sets $(Y_x, <_{Y_x})_{x \in X}$, the **lexicographical order** on $\prod_{x \in X} Y_x$ is given by

$$\vec{y} <_{\text{lex}} \vec{z} :\iff \exists x \in X \, (y_x < z_x \text{ and } \forall x' < x \, (y_x = z_x)).$$

(e) Verify that this is a partial order.
(f) Verify that if $<_X$ is a well-order and each $<_{Y_x}$ is a linear order, then $<_{\text{lex}}$ is linear.
(g) Show that even if $<_X$ and each $<_{Y_x}$ are well-orders, $<_{\text{lex}}$ might not be.

Now suppose $>_X$ is a well-order, and each $<_{Y_x}$ is a well-order, with least element $0_x \in Y_x$. Let

$$\textstyle\bigoplus_{x \in X} Y_x := \left\{ \vec{y} \in \prod_{x \in X} Y_x \;\middle|\; \{x \in X \mid y_x \neq 0_x\} \text{ is finite} \right\}.$$

(*Finite* means there is a bijection with some $n \in \mathbb{N}$; see Definition 4.4.)

(h) Prove that $<_{\text{lex}}$ restricted to $\bigoplus_{x \in X} Y_x$ is a well-order.
(i) Prove that $\rho_{<_{\text{lex}}}[\bigoplus_{x \in X} Y_x] = \prod_{x \in X} \rho[Y_x]$, where $\prod$ uses the well-order $>_X$ on X. In particular, $\rho_{<_{\text{lex}}}[\bigoplus_{x \in X} \alpha_x] = \prod_{x \in X} \alpha_x$ if every $\alpha_x \neq 0$ (otherwise apply (a)).
(j) Use this to prove that $\alpha^{\sum_{x \in X} \beta_x} = \prod_{x \in X} \alpha^{\beta_x}$. In particular, $\alpha^{\beta+\gamma} = \alpha^\beta \cdot \alpha^\gamma$ and $\alpha^{\beta \cdot \gamma} = (\alpha^\beta)^\gamma$.

Exercise 3.164. Let α, β ("base") be ordinals.

(a) Prove that if $\alpha > 0$, then there are unique γ (exponent), $0 < \delta < \beta$ ("digit"), and $\varepsilon < \beta^\gamma$ (remainder) such that $\alpha = \beta^\gamma \cdot \delta + \varepsilon$.
(b) Prove that there is a unique finite sequence $\gamma_{k-1} > \gamma_{k-1} > \cdots > \gamma_0$ (possibly with $k = 0$) and "digits" $0 < \delta_0, \ldots, \delta_{k-1} < \beta$ such that

$$\alpha = \beta^{\gamma_{k-1}} \cdot \delta_{k-1} + \cdots + \beta^{\gamma_0} \cdot \delta_0.$$

When $\beta = \omega$, this "base ω expansion" is known as the **Cantor normal form** of α. [Warning: it is possible for the largest exponent γ_{k-1} to be equal to α.]
(c) What are the base 2, respectively base ω, expansions of: $\omega + 3$, $\omega^3 + \omega \cdot 3$, $\omega^{\omega^3} \cdot \omega \cdot 3$, ω_1?

[20] Warning: this is distinct from cardinal product; see Remark 5.30. Unlike the indexed sum, the indexed product of ordinals doesn't even have the same cardinality as the cardinal product.

3.J. **The revenge of Knaster–Tarski.** Recall that the proof of Knaster–Tarski given in Theorem 3.6 was "top-down" or *impredicative* (Remark 3.14). Using transfinite induction, we now give a "bottom-up" proof, that also has the benefit of generalizing to proper classes (to a certain extent).

Definition 3.165. For a class X, the **powerclass** $\mathcal{P}(X) = \{A \mid A \subseteq X\}$ is defined as before (Example 2.6). Note that this is now the *class* of all sub*sets* of X, since the elements of a comprehension must be sets. In particular, if X is a proper class, $\mathcal{P}(X)$ has no greatest element.

Definition 3.166. Let X be a class. A **monotone set operator** $T : \mathcal{P}(X) \to \mathcal{P}(X)$ means the same thing as in Definition 3.1, i.e., a monotone *class* function mapping each sub*set* to another.

Given such T, if we instead have a sub*class* $A \subseteq X$, we define

$$\begin{aligned} T(A) &:= \bigcup\{T(B) \mid B \subseteq A\} \\ &= \{x \mid \exists B \subseteq A\, (x \in T(B))\}. \end{aligned}$$

Again, here B ranges over all sub*sets*. Note that if A happens to be a subset, then there is a largest such $B \subseteq A$, namely A; hence this agrees with the original value of T on A (using monotonicity).

We say a subclass $A \subseteq X$ is **T-closed** if $T(A) \subseteq A$, i.e., $T(B) \subseteq A$ for every sub*set* $B \subseteq A$.

Theorem 3.167 (Knaster–Tarski II)**.** Let X be a class, $T : \mathcal{P}(X) \to \mathcal{P}(X)$ be a monotone set operator, $A \subseteq X$ be a sub*set*. Define sets $A_\alpha \subseteq X$ for each $\alpha \in \mathbb{ON}$ by induction as follows:

$$\begin{aligned} A_0 &:= A, \\ A_{\alpha+1} &:= A_\alpha \cup T(A_\alpha), \\ A_\alpha &:= \textstyle\bigcup_{\beta<\alpha} A_\beta \quad \text{for a limit ordinal } \alpha. \end{aligned}$$

This is equivalent to

$$\begin{aligned} A_\alpha &:= \textstyle\bigcup_{\beta<\alpha}(A_\beta \cup T(A_\beta)) \quad \text{for } \alpha > 0, \\ A_\alpha &:= A \cup \textstyle\bigcup_{\beta<\alpha} T(A_\beta) \quad \text{for all } \alpha. \end{aligned}$$

Then $(A_\alpha)_\alpha$ is a monotone transfinite sequence, i.e., $\beta \le \alpha \implies A_\beta \subseteq A_\alpha$; and

$$\overline{T}(A) := \textstyle\bigcup_{\alpha\in\mathbb{ON}} A_\alpha$$

is the smallest T-closed subclass of X containing A. Moreover, if $A \subseteq T(A)$, then we have

$$\begin{aligned} A_{\alpha+1} &:= T(A_\alpha), \\ A_\alpha &:= \textstyle\bigcup_{\beta<\alpha} T(A_\beta) \quad \text{for } \alpha > 0, \end{aligned}$$

hence $T(\overline{T}(A)) = \overline{T}(A)$. In particular, this holds for $A = \varnothing$. (See the picture (3.15).)

Proof. First, we assume that $A \subseteq T(A)$. We take $A_0 := A$ and the last equation as the definition of A_α for $\alpha > 0$. We then have $A_\alpha \subseteq T(A_\alpha)$ for all α: for $\alpha > 0$, assuming $A_\beta \subseteq T(A_\beta)$ for all $\beta < \alpha$, we have $T(A_\beta) \subseteq T(T(A_\beta)) \subseteq T(A_\alpha)$ for all $\beta < \alpha$, whence $A_\alpha = \bigcup_{\beta<\alpha} T(A_\beta) \subseteq T(A_\alpha)$. Then $\beta < \alpha \implies A_\beta \subseteq T(A_\beta) \subseteq A_\alpha$, i.e., the sequence is monotone. The rest of the recursions defining A_α now immediately follow from the last, except for $A_\alpha = \bigcup_{\beta<\alpha} A_\beta$ for limit α, which follows from $\beta < \alpha \implies \beta + 1 < \alpha$ and so $\bigcup_{\beta<\alpha} A_\beta = \bigcup_{\beta<\alpha} A_{\beta+1} = \bigcup_{\beta<\alpha} T(A_\beta)$ (cf. Proposition 3.150).

We now verify that $\overline{T}(A) \subseteq X$ is the smallest T-closed subclass containing A. If $B \subseteq X$ is any T-closed subclass containing A, then we prove $\overline{T}(A) = \bigcup_\alpha A_\alpha \subseteq B$, i.e., $A_\alpha \subseteq B$ for all α, by induction on α: we have $A_0 = A \subseteq B$, and for $\alpha > 0$, assuming $A_\beta \subseteq B$ for all $\beta < \alpha$, then $T(A_\beta) \subseteq T(B) \subseteq B$ for all $\beta < \alpha$, whence $A_\alpha = \bigcup_{\beta<\alpha} T(A_\beta) \subseteq B$. Clearly $A = A_0 \subseteq \overline{T}(A)$. To show $\overline{T}(A)$ is T-closed, let $B \subseteq \overline{T}(A)$ be a sub*set*, and for each $x \in B$, let α_x be least such that $x \in A_{\alpha_x}$; then letting $\alpha := \sup_{x\in B} \alpha_x$, we have $B \subseteq A_\alpha$, whence $T(B) \subseteq T(A_\alpha) = A_{\alpha+1} \subseteq \overline{T}(A)$.

This concludes the proof assuming $A \subseteq T(A)$. To deduce the general case, apply the special case to $T'(A) := A \cup T(A)$, or to $T_A(B) := A \cup T(B)$ from Exercise 3.18. □

In the case where $A \subseteq T(A)$, it is convenient to think of A_α as the "αth iterate of T of A":

$$T^\alpha(A) := \begin{cases} A & \text{if } \alpha = 0, \\ \bigcup_{\beta<\alpha} T(T^\beta(A)) & \text{if } \alpha > 0. \end{cases}$$

(It is literally $(T \circ \cdots \circ T)(A)$ when $\alpha \in \mathbb{N}$.) Then thinking of $\mathbb{ON}$ as "∞" (Footnote 17), we write

$$T^\infty(A) := \overline{T}(A) = \bigcup_{\alpha<\infty} T^\alpha(A).$$

In particular, we may always write these for $A = \varnothing$.

Definition 3.168. For a monotone $T : \mathcal{P}(X) \to \mathcal{P}(X)$ and $x \in \overline{T}(\varnothing) = T^\infty(\varnothing)$, the T**-rank** of x is

$$\rho_T(x) := \min\{\alpha \in \mathbb{ON} \mid x \in T^{\alpha+1}(\varnothing)\}.$$

(Note that the least β with $x \in T^\beta(\varnothing)$ must be a successor, since $T^\beta(\varnothing) = \bigcup_{\alpha<\beta} T^\alpha(\varnothing)$ for limit β.)

By definition,

$$\rho_T(x) = \alpha \iff x \in T^{\alpha+1}(\varnothing) \setminus T^\alpha(\varnothing).$$

It follows that

$$\rho_T[X] := \{\rho_T(x) \mid x \in X\} = \{\alpha \in \mathbb{ON} \mid T^\alpha(\varnothing) \subsetneq T^{\alpha+1}(\varnothing)\}.$$

This is an initial segment of $\mathbb{ON}$, since if $T^\alpha(\varnothing)$ is already T-closed, then it must be $\overline{T}(\varnothing)$; in other words, it is either an ordinal, or the entirety of $\mathbb{ON} = \infty$. If it is an ordinal, then the transfinite sequence $\varnothing \subseteq T(\varnothing) \subseteq T^2(\varnothing) \subseteq \cdots$ increases up to $T^{\rho_T[X]}(X) = \overline{T}(\varnothing)$ and then stops.

Note that in the above proof of Theorem 3.167, to show that $\overline{T}(\varnothing)$ is T-closed, we essentially used that for a sub*set* $B \subseteq \overline{T}(\varnothing)$, necessarily $\alpha := \sup^+ \rho_T[B] < \infty$, whence $B \subseteq T^\alpha(\varnothing)$. By replacing ∞ here with a smaller cardinality, we get

Proposition 3.169. Suppose a monotone set operator $T : \mathcal{P}(X) \to \mathcal{P}(X)$ is **finitary**, meaning

$$T(A) = \bigcup_{\text{finite } B \subseteq A} T(B),$$

i.e., whenever $x \in T(A)$, then $x \in T(B)$ for some finite $B \subseteq A$.[21] Then for any $A \subseteq X$, the transfinite sequence $(A_\alpha)_{\alpha \in \mathbb{ON}}$ from Theorem 3.167 stabilizes at $A_\omega = A_{\omega+1}$; thus $\overline{T}(A) = A_\omega$.

In particular, $\overline{T}(A)$ is a set if A is.

Proof. For any finite $B \subseteq A_\omega$, each $x \in B$ is in some (least) $\alpha_x < \omega$; then letting $\alpha := \sup_{x \in B} \alpha_x$, we have $B \subseteq A_\alpha$, whence $T(B) \subseteq T(A_\alpha) = A_{\alpha+1} \subseteq A_\omega$. Thus $T(A_\omega) = \bigcup_{\text{finite } B \subseteq A_\omega} T(B) \subseteq A_\omega$, so A_ω is already T-closed and hence equal to $\overline{T}(A)$. □

Example 3.170. To generate a subgroup of a group (Example 3.4), or an equivalence relation from an arbitrary binary relation (Example 3.5), only takes ω many steps, since each newly derived group element or pair depends on ≤ 2 existing elements.

Remark 3.171. Analogous bounds hold for higher cardinalities; see Corollary 5.49 and Exercise 5.69.

Exercise 3.172. Show that for a monotone set operator $T : \mathcal{P}(X) \to \mathcal{P}(X)$ on a class X, the T-interior $T^\circ(A)$ (cf. Exercise 3.20) of every subclass $A \subseteq X$ exists, and is a set if A is.

Exercise 3.173. Recall the theory of rank for well-founded relations (Definition 3.123).

(a) For a well-founded $\prec \subseteq X^2$, we have $\rho_\prec = \rho_{T_\prec}$.
(b) For an arbitrary inductive monotone set operator $T : \mathcal{P}(X) \to \mathcal{P}(X)$, $\rho_T : X \to \mathbb{ON}$ is a simulation from (X, T) to $(\mathbb{ON}, T_\in)$, in the sense of Exercise 3.78.
(c) If T is transitive in the sense of Exercise 3.126, then ρ_T^{-1} is also a simulation.
(d) Thus, for arbitrary inductive monotone T, each $x \in X$ is "mutually simulable" with its rank. [Recall 3.126(c).]

[21] Other synonymous terminology include **of finite character**, **Scott-continuous**.

3.K. **Transitive closure, $\in$-induction, and the cumulative hierarchy.** The axioms of set theory yield two basic monotone set operators on the universe V, namely $\bigcup$ and $\mathcal{P}$. Note that[22]

$$\bigcup A \subseteq B \iff A \subseteq \mathcal{P}(B) \tag{3.174}$$

for any classes $A, B \subseteq V$. Thus in particular, A is $\bigcup$-closed iff it is $\mathcal{P}$-open, iff it is transitive.

Definition 3.175. The **transitive closure** of a class A is the smallest transitive class $\overline{\bigcup} A \supseteq A$. Since $\bigcup$ is clearly a finitary (indeed unary) monotone set operator, by Proposition 3.169,

$$\overline{\bigcup} A = A \cup \bigcup A \cup \bigcup\bigcup A \cup \cdots = \bigcup_{n\in\mathbb{N}} \bigcup^n A;$$

in particular, the transitive closure of a set is still a set.[23]

Example 3.176. $\overline{\bigcup}\{\{\varnothing\}\} = \{\{\varnothing\}\} \cup \{\varnothing\} \cup \varnothing = \{\{\varnothing\}, \varnothing\} = 2$.

Corollary 3.177. For every set x, there is a set X containing x such that $x = \xi_{\in_X} x$.

Proof. Let $X = \overline{\bigcup}\{x\}$; then for every $y \in X$, by transitivity, $\xi_{\in_X} y = \{z \in X \mid z \in y\} = y$. □

Corollary 3.178 (Axiom (Schema) of ($\in$-)Induction)**.** Assume ZF, and let $\phi(x)$ be a property. If

- for every set x, if every $y \in x$ satisfies $\phi(y)$, then $\phi(x)$,

then for every set x, $\phi(x)$. In other words, the global $\in$ relation on V is well-founded.

Proof. To show $\phi(x)$, by Foundation, we may do $\in_X$-induction on $X = \overline{\bigcup}\{x\}$. □

Remark 3.179. Conversely, clearly well-foundedness of the global $\in$ implies well-foundedness of $\in_X$ on each set X, i.e., Induction implies Foundation.[24] This is analogous to the passage between well-foundedness of each ordinal α and well-foundedness of the entire class $\mathbb{ON}$ in Proposition 3.129.

Exercise 3.180. Every class A also has a **transitive interior** $\mathcal{P}^\circ(A)$ by Exercise 3.172. Show that

$$\mathcal{P}^\circ(A) = \{x \in A \mid \overline{\bigcup} x = x \cup \bigcup x \cup \bigcup\bigcup x \cup \cdots \subseteq A\}.$$

Such x are called **hereditarily in** A, i.e., x, its elements, its elements' elements, etc. are all in A. For example, $\{\varnothing, \{\varnothing\}, \{\{\varnothing\}\}\}$ is a **hereditarily finite set**, i.e., in $\mathcal{P}^\circ(A)$ for $A = \{x \mid x \text{ finite}\}$.

This covers two possible combinations of closure/interior and $\bigcup/\mathcal{P}$, that are related via (3.174). What about the other two combinations? One of them is again related to Foundation: note that

$$\mathcal{P}(A) = \{x \mid \forall y \in x\, (y \in A)\} = T_\in(A) \tag{3.181}$$

is the monotone set operator induced (as in Definition 3.37) by the relation $\in \subseteq V^2$.

Definition 3.182. The **von Neumann cumulative hierarchy** is the transfinite sequence used to build $\overline{\mathcal{P}}(\varnothing)$ as in the Knaster–Tarski Theorem 3.167:

$$\begin{aligned}
V_0 &:= \mathcal{P}^0(\varnothing) = \varnothing,\\
V_1 &:= \mathcal{P}^1(\varnothing) = \{\varnothing\},\\
V_2 &:= \mathcal{P}^2(\varnothing) = \{\varnothing, \{\varnothing\}\},\\
V_3 &:= \mathcal{P}^3(\varnothing) = \{\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \{\varnothing, \{\varnothing\}\}\},\\
V_\alpha &:= \mathcal{P}^\alpha(\varnothing) = \textstyle\bigcup_{\beta<\alpha} \mathcal{P}(V_\beta).
\end{aligned}$$

Thus, $V_\infty := \overline{\mathcal{P}}(\varnothing)$ is the well-founded part of the universe. The **rank** $\rho(x)$ of $x \in V_\infty$ is its $\in$-rank

$$\rho(x) := \rho_\in(x) = \min\{\alpha \in \mathbb{ON} \mid x \in V_{\alpha+1} = \mathcal{P}(V_\alpha)\}.$$

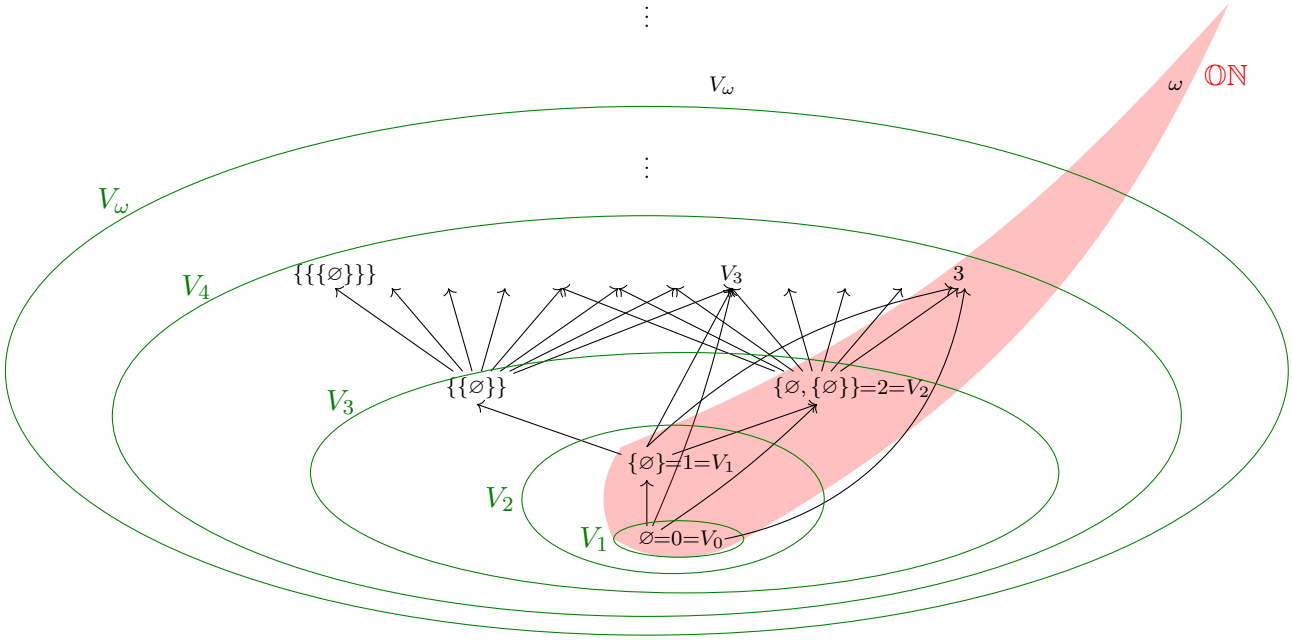
[22] In other words, $\bigcup$ and $\mathcal{P}$ form a (monotone) Galois adjunction $\mathcal{P}(V) \rightleftarrows \mathcal{P}(V)$.

[23] It turns out that this fact cannot be proved without the Axiom of Infinity!

[24] Once again, in the absence of Infinity, Induction is strictly stronger than Foundation.

Corollary 3.183. Assuming only $\mathsf{ZF}^- - \mathsf{Infinity}$, Induction (hence Foundation, assuming Infinity) is equivalent to $V = V_\infty$, i.e., every set x is in V_α for some ordinal α. □

The universe under Foundation (or Induction, in the absence of Infinity) looks like:



Each level V_α is a set; while the union is the entirety of V. Note that this is a much more precise version of the picture (2.12): the "boundary" of the universe is the upper fringe; a class $A \subseteq V$ is proper iff it keeps going up, i.e., is not contained in any V_α. For a set A, the least α such that $A \subseteq V_\alpha$, i.e., $A \in V_{\alpha+1}$, is its rank $\rho(A)$. The ordinals $\mathbb{ON}$ form a linearly ordered "spine" that contains exactly one representative of each rank α (namely α).

Corollary 3.184. Assume Foundation (or Induction, in the absence of Infinity). Then any other class $A \subseteq V$ can also be written as a transfinite increasing union of sets, namely $A = \bigcup_{\alpha\in\mathbb{ON}}(A \cap V_\alpha)$. □

An important application is a way of building "quotients" by proper class-sized equivalence relations. Given an equivalence relation $\sim \subseteq X^2$, the point of the quotient set construction is to represent each $x \in X$ by some "invariant" such that $\sim$ between x's becomes equality between invariants. The usual equivalence class $[x]$ takes all elements equivalent to x; if $\sim$ (hence X) is a proper class, then $[x]$ may also be, hence need not exist in the universe even though x does. But we may instead take a sub*set* of $[x]$ as the "invariant". For example, this gives one possible representation of "cardinal numbers" within the universe; see Definition 5.4.

Corollary 3.185 (Scott's trick)**.** Assume Foundation. Let X be a class, $\sim \subseteq X^2$ be an equivalence relation (also a class). Then there is a (class) function $\pi : X \to V$ such that $\pi(x) = \pi(y) \iff x \sim y$.

Proof. Let $\alpha(x) := \min\{\alpha \in \mathbb{ON} \mid [x] \cap V_\alpha \neq \varnothing\}$, which is always defined since $x \in V_\alpha$ for some α by Foundation, and $\pi(x) := [x] \cap V_{\alpha(x)}$. If $x \sim y$, then $[x] = [y]$, whence $\alpha(x) = \alpha(y)$ and $\pi(x) = \pi(y)$. Conversely, if $\pi(x) = \pi(y)$, then there is some $z \in [x] \cap V_{\alpha(x)} = \pi(x)$ by definition of $\alpha(x)$, whence also $z \in \pi(y) \subseteq [y]$, whence $x \sim z \sim y \implies x \sim y$. □

Exercise 3.186. Show (assuming ZF) that V_ω consists precisely of the hereditarily finite sets (Exercise 3.180).

Exercise 3.187. Show (assuming ZF) that if A is a $\bigcup$-open set, then $\rho(A)$ is either 0 or a limit ordinal. (I don't know of any particular conceptual significance of such sets, unfortunately.)

4. Choice

4.A. **The Axiom of Choice.** Given a well-ordered set, we may prove statements and construct things for each element one by one. (Whereas for a general well-founded relation, we sort of have to handle incomparable elements simultaneously, since they are not allowed to depend on each other.) It is thus of interest to know: which sets can be well-ordered?

Proposition 4.1. For a set X, the following are equivalent:

(i) There exists a well-order $<$ on X (we say X is **well-orderable**).
(ii) There exists a bijection between X and an ordinal.
(iii) There exists an injection from X into a well-ordered set.
(iv) There exists a surjection from a well-ordered set onto X.
(v) There exists a **choice function** $c \in \prod_{A \in \mathcal{P}(X)\setminus\{\varnothing\}} A$, i.e., $c : \mathcal{P}(X) \setminus \{\varnothing\} \to X$ such that $c(A) \in A$ for each $\varnothing \neq A \subseteq X$.
(vi) For any set I and family of nonempty subsets $(A_i)_{i\in I} \in (\mathcal{P}(X) \setminus \{\varnothing\})^I$, $\prod_{i\in I} A_i \neq \varnothing$.
(vii) For any set I and relation $R \subseteq I \times X$ with $\mathrm{dom}(R) = I$, there is a function $R \supseteq f : I \to X$.

Proof. (i) $\implies$ (ii): The rank function $\rho_< : X \cong \rho_<[X]$ is a bijection.
(ii) $\implies$ (iv) is obvious.
(iv) $\implies$ (iii): Let $(Y, <)$ be well-ordered, $f : Y \twoheadrightarrow X$; then $g : X \hookrightarrow Y$ where $g(x) := \min f^{-1}(x)$.
(iii) $\implies$ (i): Let $(Y, <)$ be well-ordered, $f : X \hookrightarrow Y$; then $x < x' :\iff f(x) < f(x')$ well-orders X.
(v) $\implies$ (vi): $(c(A_i))_{i\in I} \in \prod_{i\in I} A_i$.
(vi) $\iff$ (vii) follows from the canonical bijection $\mathcal{P}(I \times X) \cong \mathcal{P}(X)^I$ (Exercise 2.82).
(vii) $\implies$ (v): Take $R := \ni \subseteq (\mathcal{P}(X) \setminus \{\varnothing\}) \times X$.
(i) $\implies$ (v): $c(A) := \min A$.
(v) $\implies$ (ii): Define $A_\alpha \subseteq X$, which is either empty or a singleton, for each $\alpha \in \mathbb{ON}$ inductively:

$$A_\alpha := \begin{cases} \varnothing & \text{if } \bigcup_{\beta<\alpha} A_\beta = X, \\ \{c(X \setminus \bigcup_{\beta<\alpha} A_\beta)\} & \text{else.} \end{cases}$$

Note that if $A_\alpha = \varnothing$, then clearly $A_\beta = \varnothing$ for all $\beta > \alpha$. Also, if $A_\alpha, A_\beta \neq \varnothing$ and $\alpha \neq \beta$, WLOG with $\beta < \alpha$, then (the unique elements of) A_α, A_β differ by definition of A_α. We thus have an injection from an initial segment of $\mathbb{ON}$, namely all those α such that $A_\alpha \neq \varnothing$, mapping such α to the unique element of $A_\alpha \subseteq X$. Since X is a set, the initial segment of those α's must also form a set, hence an ordinal $\gamma \subseteq \mathbb{ON}$, such that $A_\gamma = \varnothing$ since $\gamma \notin \gamma$; thus $\alpha \mapsto$ unique element of A_α is a bijection $\gamma \cong X$. □

Axiom 4.2 (Choice (AC))**.** Every set is well-orderable, i.e., the equivalent conditions above hold.
(Conditions (i), (ii), (iii) and (iv) are known as the **well-ordering theorem**.)

Definition 4.3. Zermelo–Fraenkel set theory with **Choice** ZFC consists of the axioms of ZF (Definition 3.140) together with the Axiom of Choice. (See the diagram (2.27).)

The Axiom of Choice is unique among the axioms of ZFC in asserting the existence of a mathematical object, namely a choice function or a well-order, which obeys certain properties that by no means uniquely characterize it. By contrast, most of the other axioms (2.27) are special cases of unrestricted Comprehension, which assert the existence of a set which is necessarily unique by Extensionality (while Foundation asserts that certain kinds of sets *don't* exist). For this reason, Choice is often regarded as a "non-constructive" axiom; see Theorem 4.20.

Nonetheless, the Axiom of Choice is extremely useful, to the point that one often uses it without thinking, and doing math without it can feel quite bizarre, as the following applications illustrate:

Definition 4.4. A set X is **finite** if there exists a bijection between X and a natural.

Proposition 4.5. If a set X is infinite (i.e., not finite), then there is an injection $f : \mathbb{N} \hookrightarrow X$.

Proof. Choose inductively $f(n) \in X \setminus f[n]$; this is always possible, or else we would have $f : n \cong X$.

More precisely, we first fix a choice function $c \in \prod_{A \in \mathcal{P}(X)\setminus\{\varnothing\}} A$, and then define $A_n \subseteq X \setminus \bigcup_{m<n} A_m$ which is either empty or a singleton $\{f(n)\}$ inductively for each $n \in \mathbb{N}$ as in Proposition 4.1. Having done so, we may show that $A_n \neq \varnothing$ for all n, or else for the least n such that $A_n = \varnothing$, we would have a bijection $f : n \cong X$ taking each $m < n$ to the unique element of A_m.

Or more concisely, we've already done the work in Proposition 4.1 of showing that there exists a bijection $f : \alpha \cong X$ from an *ordinal*; now by assumption, α is not a natural, hence $\omega \leq \alpha$, and so the restriction of f to ω is the desired injection. □

Exercise 4.6. Show that if $A \subseteq \mathbb{R}$ has no upper bound (i.e., no $b \in \mathbb{R}$ such that $a \leq b$ for all $a \in A$), then there is a sequence $\mathbb{N} \to A$ converging to ∞. Be explicit about uses of Choice.

The preceding two applications of Choice both follow from the following weakening:

Exercise 4.7 (Countable Dependent Choice (DC))**.** Show (over ZF) that the following statements are equivalent:

(i) For every set $X \neq \varnothing$ and relation $R \subseteq X^2$ with $\mathrm{dom}(R) = X$, there exists a sequence $f : \mathbb{N} \to X$ such that $f(n)\ R\ f(n+1)$ for all $n \in \mathbb{N}$.

(ii) For every sequences of sets $(X_n)_{n\in\mathbb{N}}$ with $X_0 \neq \varnothing$ and relations $(R_n \subseteq X_n \times X_{n+1})_{n\in\mathbb{N}}$ with $\mathrm{dom}(R_n) = X_n$, there exists a sequence $f \in \prod_{n\in\mathbb{N}} X_n$ such that $f(n)\ R_n\ f(n+1)$ for all n.

Show that these statements are implied by the full Axiom of Choice, and explain how the above two results follow from ZF + DC.

Exercise 4.8. Show using DC that for a function $f : \mathbb{R} \to \mathbb{R}$, the following are equivalent:

(i) For every convergent sequence $(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, $\lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n)$.

(ii) For every $\varepsilon > 0$ and $x \in \mathbb{R}$, there exists $\delta > 0$ such that for every $y \in \mathbb{R}$ with $|x - y| < \delta$, we have $|f(x) - f(y)| < \varepsilon$.

Exercise 4.9. Let V be a vector space (over $\mathbb{R}$, say). Show that V is finite-dimensional iff there is no strictly increasing sequence of vector subspaces $W_0 \subsetneq W_1 \subsetneq \cdots \subseteq V$.

Many other familiar "basic" results in analysis and algebra also fail in the absence of at least DC.

Example 4.10. The following common situation does *not* require Choice, despite appearances. Given a function $f : X \to Y$ and an equivalence relation $\sim \subseteq X^2$, if $\forall x_1 \sim x_2\,(f(x_1) = f(x_2))$, then f descends to a function on the quotient set $\tilde{f} : X/\!\sim \to Y$, defined by $\tilde{f}([x]) := f(x)$.

It may appear that we are "choosing an arbitrary representative x from the equivalence class"; however, because the choice doesn't matter, we can simply define $\tilde{f}$ via comprehension:

$$\begin{aligned}\tilde{f} &:= \{(C, y) \in X/\!\sim \times\, Y \mid \exists x \in C\,(f(x) = y)\} \\ &= \{(C, y) \in X/\!\sim \times\, Y \mid \forall x \in C\,(f(x) = y)\}.\end{aligned}$$

Exercise 4.11. Show that full Choice is equivalent to: every surjection $f : X \twoheadrightarrow Y$ has a section (right inverse) $g : Y \hookrightarrow X$.

Exercise 4.12. Show that the following statement is equivalent to full Choice. For any set X, set I, sets $(J_i)_{i\in I}$, and sets $(A_{i,j} \subseteq X)_{i\in I, j\in J_i}$,

$$\bigcap_{i\in I}\bigcup_{j\in J_i} A_{i,j} = \bigcup_{(j_i)_{i\in I}\in\prod_{i\in I} J_i}\ \bigcap_{i\in I} A_{i,j_i}.$$

[It suffices to consider $X = 1$.] Similarly, for a set I, sets $(J_i)_{i\in I}$, and reals $(x_{i,j} \in [0,1])_{i\in I, j\in J_i}$,

$$\inf_{i\in I}\sup_{j\in J_i} x_{i,j} = \sup_{(j_i)_{i\in I}\in\prod_{i\in I} J_i}\ \inf_{i\in I} x_{i,j_i}.$$

4.B. **Basic combinatorial constructions.** Returning to our original motivation for the Axiom of Choice: a large class of applications consists of inductively constructing objects satisfying various constraints. Roughly speaking, the general format of such constructions is: if all the constraints are *finitary* in nature, and *finitely consistent* with each other, then they can be satisfied.

Theorem 4.13. Let X be a set, $R \subseteq X^2$ be a directed graph (i.e., binary relation). Then there is a maximal R**-clique** $A \subseteq X$, i.e., $A^2 \subseteq R$.

Proof. Fix a well-order $< \subseteq X^2$. Define $<$-inductively

$$A := \{x \in X \mid ((A \cap {\downarrow}x) \cup \{x\})^2 \subseteq R\}.$$

(That is, define inductively the indicator function $f : X \to 2$ of A by

$$f(x) = 1 \ :\iff \ ((f|{\downarrow}x)^{-1}(1) \cup \{x\})^2 \subseteq R,$$

and then put $A := f^{-1}(1)$.) Then $A^2 \subseteq R$, since for any $(x_0, x_1) \in A^2$, letting x_i be the maximum coordinate, we have $(x_0, x_1) \in ((A \cap {\downarrow}x_i) \cup \{x_i\})^2 \subseteq R$. And A is maximal as such, since for $x \notin A$, we have $((A \cap {\downarrow}x) \cup \{x\})^2 \not\subseteq R$, hence also $(A \cup \{x\})^2 \not\subseteq R$, so x cannot be added to A. □

Corollary 4.14 (Hausdorff maximality principle)**.** Every poset $(X, \leq)$ has a maximal linearly ordered subset.

Proof. Take $R := \leq \cup \geq$ above. □

Corollary 4.15 (Zorn's lemma)**.** Let $(X, \leq)$ be a poset such that every linearly ordered subset has an upper bound. Then X has a maximal element.

Proof. Let $C \subseteq X$ be maximal linearly ordered, u be an upper bound of it. Then u is maximal, since if $v \geq u$, then v is also an upper bound of C, whence $v \in C$ by maximality, whence $v \leq u$. □

Exercise 4.16. Show (over ZF) that Zorn's lemma implies Choice. Thus the preceding three results are all equivalent to Choice.

Zorn's lemma is frequently used in "ordinary" math outside of logic, often in proofs that have the flavor of transfinite/well-ordered induction (but do not require knowing what these words mean). The following generalization of Theorem 4.13 is a typical example; we give two proofs.

Theorem 4.17. Let X be a set, $R_n \subseteq X^n$ for each $n \in \mathbb{N}$ be a family of finitary relations (a "directed hypergraph"). Then there is a maximal $A \subseteq X$ such that $A^n \subseteq R_n$ for each n.

Proof 0. Copy Theorem 4.13 (replacing A^2 with A^n). □

Proof 1. Consider the poset of all such A, ordered by inclusion. If $C \subseteq \mathcal{P}(X)$ is a linearly ordered set of such A, then $\bigcup C$ is also such an A, i.e., $(\bigcup C)^n \subseteq R_n$ for each n, since for any $\vec{x} = (x_0, \ldots, x_{n-1}) \in (\bigcup C)^n$, each $x_i \in A_i$ for some $A_i \in C$, whence the largest A_i contains all of $x_0, \ldots, x_{n-1}$, whence $\vec{x} \in A_i^n \subseteq R_n$. Now apply Zorn's lemma. □

Corollary 4.18. Every vector space V over every field K has a basis.[25] □

Corollary 4.19. There exists a function $f : \mathbb{R} \to \mathbb{R}$ such that

$$f(x + y) = f(x) + f(y) \quad \forall x, y \in \mathbb{R},$$

yet f is not given by $f(x) = cx$ for any constant $c \in \mathbb{R}$.

Proof. Pick a $\mathbb{Q}$-basis $B \subseteq \mathbb{R}$ and define the $\mathbb{Q}$-linear transformation f by scaling basis elements by different amounts. □

[25] A theorem of Andreas Blass shows that this is equivalent over ZF to Choice!

Can you think of such a function? The following theorem is well beyond the scope of this course:

Theorem 4.20. It is not possible to explicitly define any such function $f : \mathbb{R} \to \mathbb{R}$. More precisely:

(a) It is not possible[26] to prove from $\mathsf{ZF} + \mathsf{DC}$ that any such function exists.

(b) It *is* possible to explicitly define such a function $f : \mathbb{R}^L \to \mathbb{R}$ on a $\mathbb{Q}$-linear subspace $\mathbb{R}^L \subseteq \mathbb{R}$ (called the **constructible reals**). Thus by (a), it is not possible to prove in $\mathsf{ZF} + \mathsf{DC}$ that $\mathbb{R}^L = \mathbb{R}$. Moreover, it is also not possible to prove, even in ZFC, that $\mathbb{R}^L \neq \mathbb{R}$.

In other words, not only can you not define such an f, but you cannot even prove that you cannot define such an f! (Exercise: do the two uses of "define" here mean the same thing?)

4.C. **Prime ideals and ultrafilters.** The following is an important application of Zorn's lemma:

Theorem 4.21 (prime ideal theorem)**.** Let $(R, +, 0, \cdot, 1)$ be a **commutative rig**, i.e., a set equipped with two commutative, associative, and unital operations, such that $\cdot$ distributes over $+, 0$. Let $F \subseteq R$ be a **multiplicative submonoid**, i.e., closed under $\cdot, 1$, and $I \subseteq R$ be an **ideal**, i.e., closed under $+, 0$, and $r \cdot (-)$ for each $r \in R$. Suppose $F \cap I = \varnothing$. Then there is an ideal $I \subseteq J \subseteq R$ which is **prime**, meaning $R \setminus J$ is a multiplicative submonoid, such that $F \cap J = \varnothing$.

Proof. Consider the poset of all ideals $I \subseteq J \subseteq R$ such that $F \cap J = \varnothing$. For a linearly ordered set C of such J, $I \cup \bigcup C$ is still such an ideal: it clearly contains I and is disjoint from F; and it is still an ideal, because $\{I\} \cup C$ is linearly ordered and the definition of "ideal" involves being closed under some finitary conditions. Thus, by Zorn's lemma there is a maximal such J. It remains to show that $R \setminus J$ is a multiplicative submonoid. We have $1 \in R \setminus J$, since $1 \in F$ and $F \cap J = \varnothing$. Now let $a, b \in R \setminus J$. It is easily verified that

$$J + Ra := \{c + ra \mid c \in J \text{ and } r \in R\}$$

is an ideal, which contains J (take $r = 0$) and a (take $c = 0$ and $r = 1$), hence by maximality of J, $c + ra \in F$ for some $c \in J$, $r \in R$. Similarly, $d + sb \in F$ for some $d \in J$, $s \in R$. Then

$$\begin{aligned} F &\ni (c + ra)(d + sb) \\ &= cd + csb + rad + rasb \end{aligned}$$

where the first three terms are (multiples of c or d, hence) in J, whence $ab \in R \setminus J$ or else the last term is also in J, contradicting $F \cap J = \varnothing$. □

The set of prime ideals of a commutative rig R is called its **spectrum**, denoted $\mathrm{Spec}(R)$.

Example 4.22. For any rig R, the smallest ideal in R is clearly $\{0\}$. This ideal is prime iff $1 \neq 0$ (equivalently by the multiplicative identity law, $R \neq \{0\}$), and the product of any two nonzero elements of R is nonzero. Recall that a *ring* R (rig with additive inverses) with this property is called an **integral domain**. Any field or subring thereof is clearly an integral domain, e.g., $\mathbb{C}, \mathbb{R}, \mathbb{Q}$.

Examples of rings which are not integral domains include $\mathbb{Z}/6\mathbb{Z}$ (in which $2 \cdot 3 = 0$), or the ring of polynomials $\mathbb{R}[\varepsilon]$ quotiented by the ideal generated by ε^2, i.e., the ring of expressions of the form $x + y\varepsilon$, defined similarly to $\mathbb{C}$, but subject to $\varepsilon^2 = 0$ (rather than $i^2 = -1$).

Exercise 4.23. If R is a commutative ring, and $I \subseteq R$ is an ideal, then I is prime iff R/I is an integral domain. (R/I is a field iff I is a *maximal* proper ideal, i.e., maximal disjoint from the multiplicative submonoid $\{1\}$, which implies prime by the proof of Theorem 4.21.)

Example 4.24. For any set X and rig R, we have the rig R^X of R-valued functions on X, with pointwise $+, \cdot$. If R is such that $\{0\} \subseteq R$ is prime, then for each $x \in X$, we have a prime ideal

$$I_x := \{f : X \to R \mid f(x) = 0\}.$$

More generally, if we have a rig R_x for each $x \in X$, then we may consider the product rig $\prod_{x \in X} R_x$.

[26] assuming that ZF is consistent to begin with; otherwise everything is provable

One intuitive interpretation of the Prime Ideal Theorem 4.21 is that it allows an abstract rig R to be concretely represented as such a rig of functions, with domain $X = \mathrm{Spec}(R)$ and codomain some "nice" rig (possibly varying with the point $x \in X$). For example:

Corollary 4.25. A commutative ring R embeds into a product of fields iff it has no nonzero **nilpotent** elements $a \in R$, meaning elements with $a^n = 0$ for some $n \geq 1$.

Proof. If $f : R \to \prod_i K_i$ is a ring homomorphism to a product of fields, and $a \in R$ is nilpotent with $a^n = 0$, then $f(a)^n = f(a^n) = 0$, whence each coordinate of $f(a)$ is 0 (since fields have no nonzero nilpotent elements); thus if f is injective, then $a = 0$. Conversely, if R has no nonzero nilpotent elements, then for any $0 \neq a \in R$, we may find a prime ideal $I \in \mathrm{Spec}(R)$ disjoint from the multiplicative submonoid $\{1, a, a^2, \dots\}$ (which is disjoint from the ideal $\{0\}$ by non-nilpotence of a); then the quotient ring R/I is an integral domain, hence embeds into its field of fractions K_I, such that a remains nonzero in K_I. Thus $f : R \to \prod_{I \in \mathrm{Spec}(R)} (R/I) \hookrightarrow \prod_{I \in \mathrm{Spec}(R)} K_I$ given coordinatewise by the quotient map $R \twoheadrightarrow R/I$ works. □

Exercise 4.26. Let X be a set, K_x be a field for each $x \in X$, and consider $R = \prod_{x \in X} K_x$.

(a) If X is finite, then we have a bijection

$$X \cong \mathrm{Spec}(R)$$
$$x \mapsto I_x.$$

Thus for example, $\mathbb{R}^3$ has precisely 3 prime ideals.

(b) If X is infinite, then the above map is not surjective. Thus for example, $\mathbb{R}^{\mathbb{N}}$ has prime ideals other than the I_n. [Consider the ideal I_∞ of all sequences $f : \mathbb{N} \to \mathbb{R}$ converging to 0.]

Exercise 4.27. Let $n \in \mathbb{N}$, let A be an $n \times n$ matrix with complex entries, and consider the commutative subring R of the ring of all $n \times n$ matrices (which is not itself commutative) generated by A and scalar multiples of the identity I_n, i.e., $R = \{c_0 I_n + c_1 A + \cdots + c_m A^m \mid c_0, \dots, c_m \in \mathbb{C}\}$.

(a) If A is diagonalizable, then $\mathrm{Spec}(R)$ is in bijection with the set of eigenvalues of A, where each eigenvalue λ corresponds to the ideal J_λ generated by $A - \lambda I_n$, each quotient ring R/J_λ is isomorphic to $\mathbb{C}$, and $R \cong \mathbb{C}^{|\mathrm{Spec}(R)|}$.

(b) In general, each prime ideal $J \in \mathrm{Spec}(R)$ is contained in J_λ for a unique eigenvalue λ; and the ideals J_λ are themselves maximal, hence prime. Each J_λ contains a finite number k_λ of prime ideals, which are linearly ordered, hence there is a least prime ideal $K_\lambda \subseteq J_\lambda$, such that $R/K_\lambda \cong \mathbb{C}[\varepsilon]/\varepsilon^{k_\lambda}$ (Example 4.22); and R is isomorphic to the product of these rings.

[Consider the Jordan normal form of A. The detailed study of eigenvalues in linear algebra is known as **spectral theory**; this was the original example of a *spectrum* of a ri(n)g.]

What about rigs that are not rings? The most important example is

Exercise 4.28. Consider $2 = \{0, 1\}$ with $+ = \vee = \max$ and $\cdot = \wedge = \min$. For a set X, the product rig 2^X is canonically isomorphic via Example 2.80 to the powerset $\mathcal{P}(X)$, with the rig operations

$$+ = \cup, \qquad \cdot = \cap, \qquad 0 = \varnothing, \qquad 1 = X.$$

(For example, the pointwise max of indicator functions $\chi_A \vee \chi_B$ is $\chi_{A \cup B}$.)

(a) A subset $I \subseteq 2^X$ is closed under "scalar multiplication" $r \cdot (-)$ iff it is downward-closed under pointwise $\leq$, i.e., the corresponding subset of $\mathcal{P}(X)$ is downward-closed under $\subseteq$.

(b) Thus, $I \subseteq \mathcal{P}(X)$ is an ideal iff it is downward-closed and closed under finite unions.

(c) For $I \subseteq \mathcal{P}(X)$, the following are equivalent:

(i) I is a prime ideal, i.e., downward-closed, closed under $\cup$ and contains $\varnothing$, and $\mathcal{P}(X) \setminus I$ is closed under $\cap$ and contains X.

(ii) $I \subseteq \mathcal{P}(X) \setminus \{X\}$ is a maximal subset closed under $\cup, \varnothing$.
(iii) $\chi_{\mathcal{P}(X)\setminus I} : \mathcal{P}(X) \to 2$ is a rig homomorphism, i.e., preserves $\wedge, \vee, 0, 1$.
(iv) $\chi_{\mathcal{P}(X)\setminus I} : \mathcal{P}(X) \to 2$ preserves $\wedge, 1, \neg$ (where $\neg$ denotes Boolean complement).
(v) $\mathcal{P}(X) \setminus I$ is closed under $\cap, X$ and is equal to $\{X \setminus A \mid A \in I\}$, i.e., for every $A \subseteq X$, exactly one of $A, X \setminus A$ is in I.
(vi) $\chi_{\mathcal{P}(X)\setminus I} : \mathcal{P}(X) \to 2$ preserves $0, 1$ and maps finite *disjoint* unions to sums:

$$\chi_{\mathcal{P}(X)\setminus I}(A \sqcup B) = \chi_{\mathcal{P}(X)\setminus I}(A) + \chi_{\mathcal{P}(X)\setminus I}(B).$$

(vii) $\chi_{\mathcal{P}(X)\setminus I} : \mathcal{P}(X) \to 2$ preserves *all* finite Boolean operations $2^n \to 2$ ($\vee, \neg$, XOR, etc.).

For such a prime ideal $I \subseteq \mathcal{P}(X)$, the "dual" set $\mathcal{P}(X) \setminus I$ is called an **ultrafilter**.

(d) Thus, the Prime Ideal Theorem 4.21, in the case of the rig $R = \mathcal{P}(X)$, is equivalent to:
(i) Every proper ideal $I \subseteq \mathcal{P}(X) \setminus \{X\}$ is contained in a prime (or maximal) ideal.
(ii) Every $F \subseteq \mathcal{P}(X) \setminus \{\varnothing\}$ closed under finite intersections is contained in an ultrafilter.
(iii) Every $F \subseteq \mathcal{P}(X)$ with the **finite intersection property**, meaning that the intersection of any finitely many sets in F is nonempty, is contained in an ultrafilter.

This special case of the PIT is called the **Boolean prime ideal theorem**.

Example 4.29. For any set X and $x \in X$, as in Example 4.24, we have an ideal $I_x \subseteq 2^X$ of all functions $f : X \to 2$ vanishing at x, i.e., the indicator functions of all $A \subseteq X$ not containing x.

The corresponding **principal ultrafilter** $U_x := \mathcal{P}(X) \setminus I_x$ consists of all subsets $A \subseteq X$ containing x. Its indicator function $u_x := \chi_{U_x} = \chi_{\mathcal{P}(X)\setminus I_x} : \mathcal{P}(X) \to 2$ takes a subset $A \subseteq X$ or "property" $\phi : X \to 2$, and returns the truth value of $\phi(x)$. Exercise 4.28(c)(vii) says that

$$u_x(\phi \vee \psi) = (\phi \vee \psi)(x) = \phi(x) \vee \psi(x), \qquad u_x(\phi \wedge \psi) = (\phi \wedge \psi)(x) = \phi(x) \wedge \psi(x),$$

and more generally the truth value of any finite Boolean combination of properties evaluated at x is the Boolean combination of the truth values.

Exercise 4.30. Show that conversely, if $u : \mathcal{P}(X) \to 2$ preserves *arbitrary* $\bigwedge, \bigvee$, then u must be u_x for a unique $x \in X$. [Note that $2 \cong \mathcal{P}(1)$; cf. Exercise 3.38.]

Thus, general ultrafilters $U \subseteq \mathcal{P}(X)$ (or their indicator functions $u = \chi_U : \mathcal{P}(X) \to 2$) may be seen as "fictitious elements", which are determined by specifying which subsets $A \subseteq X$ they "belong to", in a *finitely consistent* way, but which need not be realized by any actual elements of X:

Example 4.31. Let X be an infinite set. Then the ideal $I_{\mathsf{fin}} \subseteq \mathcal{P}(X)$ of finite subsets (called the **Fréchet ideal**) forms a proper ideal, hence is contained in a prime ideal I, whose dual ultrafilter $U = \mathcal{P}(X) \setminus I$ contains the complement of every finite set. In other words, U is a "fictitious element" of X which is outside every finite set; we may think of it as a "point at infinity".

Exercise 4.32. Let X be a set.

(a) A prime ideal $I \subseteq \mathcal{P}(X)$ is nonprincipal (i.e., not equal to I_x for any $x \in X$) iff $I \supseteq I_{\mathsf{fin}}$.
(b) There are infinitely many nonprincipal ultrafilters in $\mathcal{P}(\mathbb{N})$, i.e., infinitely many "points at infinity". [For example, in $\mathbb{Z}$ which is in bijection with $\mathbb{N}$, we may go to "$+\infty$" or "$-\infty$".]
(c) In fact, there are uncountably many ultrafilters in $\mathcal{P}(\mathbb{N})$.

(In fact, it is known that there are precisely $2^{2^{\aleph_0}}$ ultrafilters in $\mathcal{P}(\mathbb{N})$.)

Remark 4.33. As in Theorem 4.20, it is not possible to give any explicit examples of nonprincipal ultrafilters $U \subseteq \mathcal{P}(X)$, nor to prove that this is impossible. Indeed, there is a common generalization underlying these results. Recall that (the indicator function of) an ultrafilter $u : 2^X \to 2$ must be a group homomorphism with respect to XOR. It turns out that between any "reasonable" groups, such as $\mathbb{R}$ with addition or 2^X with XOR, any "definable" group homomorphism must be continuous![27]

[27] Precisely, this uses the Steinhaus–Weil–Pettis theorem, which shows that measurable group homomorphisms must be continuous, and Solovay's theorem, which shows that it is consistent with ZF that every set is measurable.

4.D. **Compactness.** Despite being somewhat nebulous objects, ultrafilters turn out to be an extremely versatile tool, that can serve as a fundamental "bridge to infinity", from which many other notions of "infinity" and "limit" in mathematics (e.g., topology, compactness) can be derived. We will spend the rest of this section on applications of Choice via this route.

Recall (see Theorem 4.13) that a general intuition behind many applications of Choice is: "finitary, finitely consistent" constraints may be simultaneously satisfied. The following makes this precise:

Definition 4.34. Let X be a set, $\mathcal{A} \subseteq \mathcal{P}(X)$ be an arbitrary family of subsets or "properties". We call X with the family $\mathcal{A}$ **compact** if every $\mathcal{C} \subseteq \mathcal{A}$ with the finite intersection property (Exercise 4.28) must itself have nonempty intersection. In other words, if $\mathcal{C}$ is a family of "finitely consistent" constraints from $\mathcal{A}$, then we may find $x \in \bigcap \mathcal{C}$ simultaneously satisfying all of them.[28]

Example 4.35. For an arbitrary set X, consider the constraints $C_x := X \setminus \{x\}$ of being $\neq x$ for each $x \in X$. If X is infinite, then any finitely many $C_{x_0}, C_{x_1}, \dots, C_{x_{n-1}}$ have nonempty intersection; but clearly $\bigcap_{x \in X} C_x = \varnothing$. Thus, $\mathcal{A} = \{C_x \mid x \in X\}$ is not compact.[29]

Example 4.36. Let I, Y be sets and $X := Y^I$. For each $i \in I$ and $y \in Y$, we may impose the constraint $C_{i,y} := \{f : I \to Y \mid f(i) = y\}$. If $\mathcal{A}$ is the set of all such $C_{i,y}$, and $\mathcal{C} \subseteq \mathcal{A}$ is a subset with the finite intersection property, then that means $\mathcal{C}$ cannot contain two $C_{i,y}, C_{i,y'}$ with $y \neq y'$ (since a function f cannot map i to both y and y'). We may then find $f \in \bigcap \mathcal{C}$ as follows: fix any $g \in X$ (which is possible, or else the finite intersection of $\varnothing \subseteq \mathcal{C}$ would be empty); now for any $i \in I$ such that some $C_{i,y}$ is in $\mathcal{C}$, put $f(i) := y$ for the unique such y; else put $f(i) := g(i)$. Thus $\mathcal{A}$ is compact.

This also works more generally for a family of sets $(Y_i)_{i \in I}$ and $X := \prod_{i \in I} Y_i$.[30]

Example 4.37. Take $X := \prod_{i \in I} Y_i$ as above, but $C_{i,y} := \{f : I \to Y \mid f(i) \neq y\}$ for $i \in I$ and $y \in Y_i$. If each Y_i is finite, then $\mathcal{A} := \{C_{i,j} \mid i \in I \text{ and } y \in Y_i\}$ is compact, since if $\mathcal{C} \subseteq \mathcal{A}$ has the finite intersection property, then for each $i \in I$, there must be at least one $y_i \in Y_i$ such that $C_{i,y} \notin \mathcal{C}$ (or else $\bigcap_{y \in Y_i} C_{i,y} = \varnothing$); we may then find $f \in \bigcap \mathcal{C}$ taking these values $f(i) = y_i$. But if some Y_i is infinite, then as in Example 4.35, the constraints $C_{i,y}$ for each $y \in Y_i$ are finitely consistent (assuming $X \neq \varnothing$) but cannot all be satisfied. So $\mathcal{A}$ is compact iff ($X = \varnothing$ or) each Y_i is finite.

Theorem 4.38 (Alexander subbasis lemma)**.** Let X be a set, $\mathcal{A} \subseteq \mathcal{P}(X)$ be a family of subsets. If $\mathcal{A}$ is compact, then so is the set $\mathcal{B}$ of finite unions of sets in $\mathcal{A}$.[31]

Proof. Let $\mathcal{C} \subseteq \mathcal{B}$ have the finite intersection property. Then there is an ultrafilter $\mathcal{C} \subseteq \mathcal{U} \subseteq \mathcal{P}(X)$, i.e., $\chi_{\mathcal{U}} : \mathcal{P}(X) \to 2$ preserves finite $\cup, \cap$ and maps every $C \in \mathcal{C}$ to 1. Since every $C \in \mathcal{C}$ is a finite union of sets in $\mathcal{A}$, it follows that there is $C \supseteq A \in \mathcal{A}$ such that $\chi_{\mathcal{U}}(A) = 1$, i.e., $A \in \mathcal{U}$. Thus $\bigcap \mathcal{C} \supseteq \bigcap(\mathcal{A} \cap \mathcal{U}) \neq \varnothing$, since $\mathcal{A} \cap \mathcal{U} \subseteq \mathcal{A}$ has the finite intersection property (since $\mathcal{U}$ does). □

Corollary 4.39 (de Bruijn–Erdős)**.** Let $G \subseteq X^2$ be a graph, $k \in \mathbb{N}$. Suppose every finite subgraph $F \subseteq G$ has a k**-coloring**, i.e., $f : X \to k$ such that $\forall (x, y) \in F\,(f(x) \neq f(y))$. Then so does G.

Proof. Consider k^X as in Example 4.37, and note that the set of k-colorings of G is

$$\{f : X \to k \mid \forall i \in k,\ (x, y) \in G\,(f(x) \neq i \text{ or } f(y) \neq i)\} = \textstyle\bigcap_{i \in k, (x,y) \in G} (C_{x,i} \cup C_{y,i});$$

the family of these conditions $\mathcal{C} := \{C_{x,i} \cup C_{y,i} \mid i \in k,\ (x, y) \in G\}$ has the finite intersection property, since for any finitely many $C_{x_0,i_0} \cup C_{y_0,i_0}, \dots, C_{x_{n-1},i_{n-1}} \cup C_{y_{n-1},i_{n-1}} \in \mathcal{C}$, the finite subgraph $F := \{(x_0, y_0), \dots, (x_{n-1}, y_{n-1})\} \subseteq G$ has a k-coloring. □

[28] If $\mathcal{A}$ is closed under arbitrary intersections and finite unions, then this means $\mathcal{A}$ forms the closed sets of a compact topology on X. Note that clearly, $\mathcal{A}$ is compact iff its closure under arbitrary intersections is (contrapositive: to find a finite subcover of an open cover, we may assume the open sets come from a given basis for the topology).

[29] In other words, the discrete topology on an infinite set is not compact.

[30] In other words, the product of the cofinite topologies on each Y_i is compact.

[31] In other words, to check compactness of a topology, it suffices to consider open covers from a given *sub*basis.

Exercise 4.40. Hall's marriage theorem states that for two finite sets X, Y and $R \subseteq X \times Y$, if

($*$) for every $n \in \mathbb{N}$ and $A \subseteq X$ with n elements, $R[A] \subseteq Y$ has at least n elements,

then there is an injection $f : X \hookrightarrow Y$ with graph contained in R.

(a) Taking Hall's marriage theorem as a black box, prove that the same statement holds without requiring X, Y to be finite, but still requiring that for each $x \in X$, $R[\{x\}] \subseteq Y$ is finite.
(b) Give a counterexample when this last assumption is also dropped.

Exercise 4.41. Let X be a set.

(a) Prove, using the Alexander subbasis lemma and no other applications of Choice, that there is a linear order $< \subseteq X^2$.

(Recall from Proposition 4.1 that full Choice is equivalent to *well*-orderability. Since the Boolean prime ideal theorem is strictly weaker than full Choice, it follows that linear orderability is strictly weaker than well-orderability. Nonetheless, it is also not possible to explicitly define a linear ordering of e.g., $\mathcal{P}(\mathbb{R})$.)

(b) Prove that more generally, every partial order on X extends to a linear order.

Exercise 4.42. Let K be a finite field, V be a K-vector space. Prove (without using full Choice) that for every vector subspace $W \subseteq V$ and linear transformation $f : W \to K$, there is an extension of f to a linear transformation $V \to K$. Conclude that V embeds into a power of K.

Exercise 4.43. For a set of functions $X = \prod_{i \in I} Y_i$ and the sets $C_{i,y}$ as in Example 4.37:

(a) Every finite Boolean combination of the sets $C_{i,y}$ can be written as an intersection of finite unions of them.
(b) Thus, every family $\mathcal{C}$ of such finite Boolean combinations with the finite intersection property has nonempty intersection, assuming each Y_i is finite.

This should give a slightly more convenient way of expressing the last few examples: we may simply write down the conditions we want using arbitrary (finite) Boolean connectives, without worrying about making them into finite unions. (Statement (b) is essentially the **compactness theorem for finitary propositional logic**, or **Tychonoff's theorem** for finite sets.)

The following generalizes Example 4.37 by allowing arbitrary constraints to be imposed on each coordinate, as long as those constraints are themselves compact:

Theorem 4.44 (Tychonoff)**.** Let I be a set, and for each $i \in I$, let Y_i be a set equipped with a family of subsets $\mathcal{A}_i \subseteq \mathcal{P}(Y_i)$ which is compact. Then $X := \prod_{i \in I} Y_i$ equipped with the family of all

$$\pi_i^{-1}[C] = \{\vec{y} = (y_j)_{j \in I} \in X \mid y_i \in C\}, \quad i \in I, C \in \mathcal{A}_i,$$

where $\pi_i : X \to Y_i$ is the coordinate projection, is compact. Thus so is the family of all finite unions of such $\pi_i^{-1}[C]$ (or more generally, arbitrary intersections of such finite unions).[32]

Proof. Let $\mathcal{C} \subseteq \mathcal{P}(X)$ be a set of such $\pi_i^{-1}[C]$ with the finite intersection property. For each i, let

$$\mathcal{C}_i := \{C \in \mathcal{A}_i \mid \pi_i^{-1}[C] \in \mathcal{C}\};$$

then $\mathcal{C} = \bigcup_{i \in I} \{\pi_i^{-1}[C] \mid C \in \mathcal{C}_i\}$, and so

$$\bigcap \mathcal{C} = \bigcap_{i \in I} \bigcap_{C \in \mathcal{C}_i} \pi_i^{-1}[C] = \bigcap_{i \in I} \pi_i^{-1}[\bigcap \mathcal{C}_i].$$

Each $\mathcal{C}_i \subseteq \mathcal{A}_i$ has the finite intersection property, since for finite $\mathcal{F} \subseteq \mathcal{C}_i$, we have $\pi_i^{-1}[\bigcap \mathcal{F}] = \bigcap_{C \in \mathcal{F}} \pi_i^{-1}[C] \neq \varnothing$ since each $\pi_i^{-1}[C] \in \mathcal{C}$ which has the finite intersection property. Thus choosing $y_i \in \bigcap \mathcal{C}_i$ for each i (using Choice again), we have $\vec{y} \in \bigcap_{i \in I} \pi_i^{-1}[\bigcap \mathcal{C}_i] = \bigcap \mathcal{C}$. □

[32] In other words, a product of compact topological spaces is compact.

Example 4.45. For finite sets Y_i, the full powerset $\mathcal{A}_i = \mathcal{P}(Y_i)$ (or just the complements of singletons) is compact; Theorem 4.44 recovers Example 4.37 in this case.

Exercise 4.46. Prove the original Prime Ideal Theorem 4.21, for an arbitrary rig R, using Tychonoff's theorem for 2^R and no other applications of Choice. Thus, the following are equivalent (over ZF): the PIT for arbitrary rigs; the Boolean PIT (for powersets); the Alexander subbasis lemma; Tychonoff's theorem for products of finite sets; Tychonoff's theorem for powers of 2.

Exercise 4.47. On the other hand, using the full Tychonoff's theorem, prove the full Axiom of Choice. [Given nonempty X_i, let $Y_i := X_i \sqcup \{\infty_i\}$ where $\infty_i \notin X_i$, equipped with the condition X_i.]

Example 4.48. Consider $Y = [0,1]$, equipped with $\mathcal{A} = \{[0,r] \mid r \in [0,1]\} \cup \{[s,1] \mid s \in [0,1]\}$. This is easily seen to be compact: given $\mathcal{C} \subseteq \mathcal{A}$ with the finite intersection property, for each $[0,r], [s,1] \in \mathcal{C}$, we must have $s \leq r$ (or else $[0,r] \cap [s,1] = \varnothing$); thus $\sup_{[s,1]\in\mathcal{C}} s \in \bigcap \mathcal{C}$. Since a union of two such intervals $[0,r] \cup [s,1]$ is the complement of the open interval (r,s), applying the Alexander subbasis lemma to $\mathcal{A}$ reproves the Heine–Borel Theorem 3.35 (in contrapositive form).

Now for any set I, applying Tychonoff's Theorem 4.44 yields that $[0,1]^I$, with the conditions

$$\{\vec{y} \in [0,1]^I \mid y_i \leq r\}, \qquad \{\vec{y} \in [0,1]^I \mid y_i \geq s\}$$

(or finite unions thereof), is compact.

Similarly to how 4.39–4.42 (secretly) used Tychonoff's theorem for finite sets to derive existence results for "objects made of finitary data" (e.g., graph colorings using finitely many colors), the compactness of $[0,1]^I$ has myriad applications concerning "objects made of real-valued data", e.g., solutions to differential equations, or various kinds of geometric objects. Many such applications can be found in functional analysis (e.g., the Hahn–Banach and Banach–Alaoglu theorems), but would take us too far from the focus of this course. We will consider a particular family of such applications with a more set-theoretical flavor in the following subsection.

4.E. **Measures and geometrical paradoxes.**

Theorem 4.49 (Banach)**.** Let $G = (G, +, 0, -)$ be an abelian group. There exists a function $\mu : \mathcal{P}(G) \to [0,1]$ such that

(i) $$\mu(G) = 1,$$

(ii) $$\mu(A \sqcup B) = \mu(A) + \mu(B) \quad \text{for disjoint } A, B \subseteq G,$$

(iii) $$\mu(A + g) = \mu(A) \quad \text{for } A \subseteq G,\ g \in G,$$

called an **invariant finitely additive probability measure** on G. (Here, $A+g := \{a+g \mid a \in A\}$.)

Proof. The set of such μ is the subset of $[0,1]^{\mathcal{P}(G)}$ defined by the conditions

(i) $$\mu(G) \geq 1,$$

(ii) $$\begin{cases} \mu(A) \leq r \text{ or } \mu(B) \leq s \text{ or } \mu(A \sqcup B) \geq r+s & \forall A, B \subseteq G,\ A \cap B = \varnothing,\ r, s \in [0,1], \\ \mu(A) \geq r \text{ or } \mu(B) \geq s \text{ or } \mu(A \sqcup B) \leq r+s & \forall A, B \subseteq G,\ A \cap B = \varnothing,\ r, s \in [0,1], \end{cases}$$

(iii) $$\begin{cases} \mu(A+g) \leq r+\varepsilon \text{ or } \mu(A) \geq r & \forall A \subseteq G,\ g \in G,\ r \in [0,1],\ \varepsilon > 0, \\ \mu(A+g) \geq r \text{ or } \mu(A) \leq r+\varepsilon & \forall A \subseteq G,\ g \in G,\ r \in [0,1],\ \varepsilon > 0. \end{cases}$$

Indeed, e.g., the first line of (ii) says "$\mu(A) > r$ and $\mu(B) > s \implies \mu(A \sqcup B) \geq r+s$"; imposing this for all $r, s \in [0,1]$ is easily seen to imply $\mu(A \sqcup B) \geq \mu(A) + \mu(B)$. Similarly for the other conditions. Since each of these conditions is a finite disjunction (= union) of closed inequalities involving a single "coordinate" of the "tuple" $\mu \in [0,1]^{\mathcal{P}(G)}$, by compactness it suffices to verify that any finitely many of these conditions may be satisfied at once.

To do so, we will show that for any finite $F \subseteq G$ and $\varepsilon > 0$, we may find a $\mu : \mathcal{P}(G) \to [0,1]$ obeying (i) and all of (ii) (meaning $\mu(A \sqcup B) = \mu(A) + \mu(B)$), and obeying (iii) for all $A \subseteq G$, all $r \in [0,1]$, all $g \in F$, and the given ε. Note that (iii) for a given A, g, ε (and all r) is equivalent to

$$|\mu(A+g) - \mu(A)| \leq \varepsilon. \tag{iii$'$}$$

Thus considering a single ε is enough, since given finitely many, we may take their minimum.

Now to find μ: let $1/\varepsilon \leq N \in \mathbb{N}$. We define the measure $\mu(A)$ of $A \subseteq G$ by sampling the first few integer linear combinations of the group elements in F, and counting what proportion lands in A:

$$\mu(A) := \frac{\left|\left\{(a_f)_{f\in F} \in N^F \;\middle|\; \sum_{f\in F} a_f f \in A\right\}\right|}{N^{|F|}}.$$

It is rather obvious that $\mu(G) = 1$ and $\mu(A \sqcup B) = \mu(A) + \mu(B)$, so (i) and (ii) hold. To check (iii$'$): note that for $g \in F$,

$$\begin{aligned}
\mu(A+g) &= \frac{\left|\left\{(a_f)_{f\in F} \in N^F \;\middle|\; \sum_{f\in F} a_f f \in A+g\right\}\right|}{N^{|F|}} \\
&= \frac{\left|\left\{(a_f)_{f\in F} \in N^F \;\middle|\; \sum_{f\in F} a_f f - g \in A\right\}\right|}{N^{|F|}} \\
&= \frac{\left|\left\{(b_f)_{f\in F} \in N^F - \vec{e}_g \;\middle|\; \sum_{f\in F} b_f f \in A\right\}\right|}{N^{|F|}},
\end{aligned}$$

where $\vec{b} = \vec{a} - \vec{e}_g$ is $\vec{a}$ with the gth coordinate decreased by 1 ($\vec{e}_g \in N^F$ is the "gth basis vector"). Since the former set counted by $\mu(A)$ and this set counted by $\mu(A+g)$ agree on all $N^{|F|-1}(N-1)$ elements of $N^F \cap (N^F - \vec{e}_g)$, their cardinalities can differ by at most $N^{|F|-1}$, and so

$$|\mu(A+g) - \mu(A)| \leq N^{|F|-1}/N^{|F|} = 1/N \leq \varepsilon. \qquad \square$$

Remark 4.50. Note that condition (ii) implies

$$\mu(\varnothing) = \mu(\varnothing) + \mu(\varnothing) \implies \mu(\varnothing) = 0, \tag{iv}$$

$$A \subseteq B \implies \mu(B) = \mu(A \sqcup (B \setminus A)) = \mu(A) + \mu(B \setminus A) \geq \mu(A). \tag{v}$$

Example 4.51. If G is a finite group (not necessarily abelian), then condition (iii) above implies that each singleton $\{g\} \subseteq G$ (is a translate of $\{0\}$, hence) has the same measure $\mu(\{g\})$; while conditions (i), (ii) imply that these measures have to sum to 1. Thus $\mu(\{g\}) = 1/|G|$, and so by (ii),

$$\mu(A) = |A|/|G|$$

is the unique finitely additive probability measure on G. Thus for infinite groups G, we may think of μ as a way of defining a notion of "proportion" for each subset of G.

Example 4.52. Consider $G = \mathbb{Z}$ (with the usual addition). Again, every singleton $\{g\}$ must have the same measure; but since for any $N \in \mathbb{N}$, we may find N disjoint singletons, it follows that $\mu(\{g\}) \leq 1/N$, and thus $\mu(\{g\}) = 0$. By finite additivity, any finite $A \subseteq \mathbb{Z}$ must have measure 0.

What about infinite $A \subseteq \mathbb{Z}$? For example, if $A = 2\mathbb{Z}$ is the even integers, then since $A \sqcup (A+1) = \mathbb{Z}$, both A and $A+1$ must have measure $1/2$, which perhaps agrees with our intuition that "half of the integers are even, the other half are odd". But for e.g.,

$$A = \bigcup_{n\in\mathbb{N}} [10^{2n}, 10^{2n+1}) = [1,10) \cup [100,1000) \cup [10000,100000) \cup \cdots,$$

A appears to include a lot of the first 10 naturals, then very few of the first 100, then a lot of the first 1000, etc. There is no constraint that uniquely determines the "proportion" of this A; the measure μ returned by Theorem 4.49 merely chooses *some* arbitrary number, for this A and all other sets at once, so that these numbers are all consistent with each other.

Exercise 4.53. Show that in fact, there is an invariant finitely additive probability measure $\mu : \mathcal{P}(\mathbb{Z}) \to [0,1]$ such that $\mu(\mathbb{N}) = 0$. [Take an existing measure ν, and try $\mu(A) := \nu(A \setminus \mathbb{N})$.]

Exercise 4.54. By Exercise 4.28(c)(vi), an ultrafilter $\mathcal{P}(\mathbb{Z}) \to 2 \subseteq [0,1]$ obeys all the axioms of μ above except for translation-invariance. Show that this is unavoidable.

Corollary 4.55. For any abelian group G, it is not possible to find a finite partition

$$G = A_0 \sqcup \cdots \sqcup A_{m-1} \sqcup B_0 \sqcup \cdots \sqcup B_{n-1}$$

such that $A_0, \ldots, A_{m-1}$ can be slid around and reassembled into all of G, as can $B_0, \ldots, B_{n-1}$:

$$G = (A_0 + g_0) \sqcup \cdots \sqcup (A_{m-1} + g_{m-1}) = (B_0 + h_0) \sqcup \cdots \sqcup (B_{n-1} + h_{n-1}).$$

Proof. Let μ be an invariant finitely additive probability measure on G. Then the above give

$$\begin{aligned} 1 = \mu(G) &= \mu(A_0) + \cdots + \mu(A_{m-1}) + \mu(B_0) + \cdots + \mu(B_{n-1}) \\ &= \mu(A_0 + g_0) + \cdots + \mu(A_{m-1} + g_{m-1}) + \mu(B_0 + h_0) + \cdots + \mu(B_{n-1} + h_{n-1}) \\ &= \mu(G) + \mu(G) = 2. \end{aligned}$$

□

Example 4.56. It is not possible to partition the plane $\mathbb{R}^2$ into finitely many pieces, and slide them around using translations, in order to make two copies of the plane.

Example 4.57. It is not possible to take the unit circle $S^1 = \{(x,y) \mid x^2 + y^2 = 1\} \subseteq \mathbb{R}^2$, partition it into finitely many pieces, and slide the pieces around using translations and rotations in order to make two disjoint unit circles. To see this: the translations are not really relevant; it is enough to consider rotations (to make two overlapping unit circles), i.e., we treat S^1 as a group under rotation (as the complex numbers $e^{i\theta}$ under $\cdot$, or isomorphically as the quotient group $\mathbb{R}/\mathbb{Z}$ under $+$).

Exercise 4.58.

(a) Show that it is also not possible to take the right half of the unit circle S^1, partition it into finitely many pieces, and rotate them to cover the full S^1.
(b) Show that it is not possible to take the unit square $[0,1]^2 \subseteq \mathbb{R}^2$, partition it into finitely many pieces, and translate them to form two disjoint unit squares. [Consider a torus.]

Exercise 4.59. Which groups G admit an invariant finitely additive probability measure?

(a) Show that if $N \subseteq G$ is a normal subgroup, and G has an invariant finitely additive probability measure, then so does G/N.
(b) Show that if $H \subseteq G$ is a subgroup, and G has an invariant finitely additive probability measure, then so does H. [Given $A \subseteq H$, measure a union of copies in the cosets of H.]
(c) Show that if $H \subseteq G$ is a subgroup of finite index (i.e., H has only finitely many cosets), and H has an invariant finitely additive probability measure, then so does G.
(d) Conclude that Example 4.57 remains valid if we allow reflections as well as rotations.
(e) (Følner) Show that G has an invariant finitely additive probability measure if for every finite $F \subseteq G$ and $\varepsilon > 0$, there is a finite $\varnothing \neq D \subseteq G$ such that for every $g \in F$, gD agrees with D except on $\varepsilon|D|$ elements. [Imitate the proof of Theorem 4.49.]

Such groups G are known as the **amenable** groups (pronounced a-*mean*-able, a bad pun), and are known to be closed under many other group-theoretic constructions, such as direct product, extension, and increasing union.

Why would anyone doubt Example 4.57, say? Recall one of the first "paradoxes" of cardinality:

Example 4.60. There are bijections $\mathbb{N} \cong \mathbb{Z} \cong 2\mathbb{Z} \cong 2\mathbb{Z} + 1 \cong \mathbb{Q} \cong \cdots$.

A trivial rephrasing is that

Example 4.61. There is a *countable* partition

$$\mathbb{Q} = A_0 \sqcup A_1 \sqcup \cdots \sqcup B_0 \sqcup B_1 \sqcup \cdots$$

so that $A_0, A_1, \ldots$ can be slid around to form another copy of $\mathbb{Q}$, as can $B_0, B_1, \cdots$. Just enumerate $\mathbb{Q}$, and take the even and odd singletons. Of course $\mathbb{Q}$ here can be any countably infinite group.

Now $\mathbb{Q}$ is countable, so perhaps it's unsurprising that countable partitions can "ruin" its size; but this in turn easily yields

Theorem 4.62 (Vitali)**.** There is a countable partition of $\mathbb{R}$ (or of S^1) into countably many pieces, so that they may be slid around to form two disjoint copies of $\mathbb{R}$ (or S^1).

Proof. Consider the coset equivalence relation of the subgroup $\mathbb{Q} \subseteq \mathbb{R}$:

$$x \sim y \iff y - x \in \mathbb{Q}.$$

Let $C \subseteq \mathbb{R}$ choose exactly one element from each equivalence class, i.e., each coset of $\mathbb{Q}$. Then

$$+ : \mathbb{Q} \times C \longrightarrow \mathbb{R}$$

is a bijection, such that the cosets of $\mathbb{Q}$ on the right correspond to the "cross-sections" $\mathbb{Q} \times \{c\}$ on the left; in other words, we have decomposed $\mathbb{R}$ into C-many copies of $\mathbb{Q}$, such that in each copy, translation by $\mathbb{Q}$ acts independently. Now the partition from the last example yields a partition

$$\mathbb{Q} \times C = (A_0 \times C) \sqcup (A_1 \times C) \sqcup \cdots \sqcup (B_0 \times C) \sqcup (B_1 \times C) \sqcup \cdots$$

such that these pieces may be translated by elements of $\mathbb{Q}$ in the first coordinate to form two copies of $\mathbb{Q} \times C$. The bijection $+$ thus takes this to a partition

$$\mathbb{R} = (A_0 + C) \sqcup (A_1 + C) \sqcup \cdots \sqcup (B_0 + C) \sqcup (B_1 + C) \sqcup \cdots$$

which can be moved around to form two copies of $\mathbb{R}$.

The proof for S^1 is the same, using any countably infinite subgroup, e.g., $\mathbb{Q}/\mathbb{Z} \subseteq \mathbb{R}/\mathbb{Z} \cong S^1$ (the rotations by rational multiples of 2π). □

Exercise 4.63. In fact, S^1 may be rearranged into countably many copies of S^1.

It follows that the proof of Corollary 4.55 must break if we try to apply it to these countable partitions. Namely, there cannot be an invariant *countably* additive probability measure $\mu : \mathcal{P}(\mathbb{R}) \to [0, 1]$ or $\mu : \mathcal{P}(S^1) \to [0, 1]$ (which was the original application of Vitali). For $\mathbb{R}$, this is not surprising, since one usually understands its "length" to be infinite; the (non-constructive) existence of Banach's 4.49 finitely additive notion of finite "length" is perhaps the real surprise. But for S^1, we have the "usual" notion of finite arc length (namely the Lebesgue measure, which can be normalized by dividing by 2π), which makes sense for "nice" subsets of S^1 such as intervals; Vitali's theorem says that there is *no way to consistently extend it to arbitrary subsets of* S^1.

Exercise 4.64. Show that the unit disk $B^2 = \{(x, y) \mid x^2 + y^2 \leq 1\} \subseteq \mathbb{R}^2$ may be partitioned into countably many pieces which can be slid around (using translations and rotations) into two disjoint unit disks. [Take pie slices, with some ad-hockery to deal with the center point.]

It turns out that in dimensions > 2, there cannot even be a *finitely* consistent notion of "volume":

Theorem 4.65 (Banach–Tarski)**.** Let $S^2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$ be the unit sphere. There is a way to partition S^2 into finitely many subsets, slide those subsets around using translations and (3D) rotations, and rearrange them into two disjoint unit spheres.

(As in the previous exercises, the same then holds for the ball $B^3 = \{(x, y, z) \mid x^2 + y^2 + z^3 \leq 1\}$; and in fact we may make infinitely (indeed *uncountably*) many copies of B^3 from one.)

The main reason behind the stark difference with S^1 is that the group SO(3) of 3D rotations is *highly nonabelian*. In fact, in some sense, it is as complicated a group as possible:

Definition 4.66. The **free group on two generators** $\mathbb{F}_2 = \langle a, b \rangle$ consists of all finite strings (including empty) of the four symbols a, b, a^{-1}, b^{-1}, such that no letter occurs consecutively with its inverse. Two such strings are multiplied by concatenation followed by cancelling inverses: e.g.,

$$(aba^{-1}bb)(\cancel{b^{-1}}ab) = aba^{-1}bab.$$

In other words, $\mathbb{F}_2$ is obtained by "declaring there to be two elements a, b, and taking all elements built from those, with no relations between them except those implied by the group axioms".

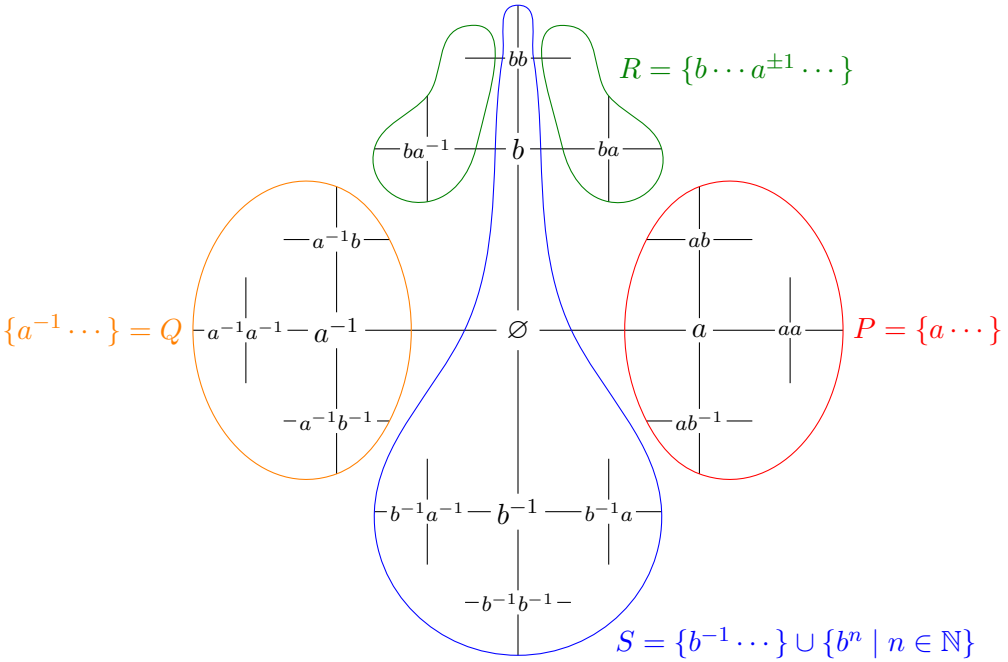
Theorem 4.67 (Hausdorff)**.** There are two orthogonal matrices $A, B \in \mathrm{SO}(3)$ obeying no relations not implied by the group axioms, i.e., such that we have an injective group homomorphism

$$\mathbb{F}_2 \longrightarrow \mathrm{SO}(3)$$
$$a^{n_0}b^{n_1}a^{n_2}\cdots \longmapsto A^{n_0}B^{n_1}A^{n_2}\cdots.$$

This is the key technical ingredient underlying the Banach–Tarski paradox, and its proof involves numerical computations with matrices: for example, one could take the rotations about the x- and y-axes, both by the angle $\arcsin(\frac{3}{5})$ in a 3-4-5 right triangle, yielding matrices with rational entries. Given this theorem, which we will not prove, the proof of Theorem 4.65 follows the same strategy as Theorem 4.62: first we do the finite partitioning and rearranging in the group $\mathbb{F}_2$; then we transfer it to the orbits of the action on S^2 via the Axiom of Choice.

Theorem 4.68. There is a finite partition $\mathbb{F}_2 = P \sqcup Q \sqcup R \sqcup S$ such that $\mathbb{F}_2 = P \sqcup aQ = R \sqcup bS$.

Proof.



□

Proof sketch of Theorem 4.65. We identify $\mathbb{F}_2$ with a subgroup of $\mathrm{SO}(3)$ via Theorem 4.67. Each non-identity rotation $I_3 \neq T \in \mathrm{SO}(3)$ fixes only two points on S^2. Thus, there are countably many points $F \subseteq S^2$ such that every $1 \neq T \in \mathbb{F}_2$ moves every point in $S^2 \setminus F$; we will prove the weaker statement that $S^2 \setminus F$ can be countably partitioned and then rearranged into two copies of itself. As in the proof of Theorem 4.62, let $C \subseteq S^2 \setminus F$ choose exactly one point from each $\mathbb{F}_2$-orbit. Then

$$S^2 \setminus F = PC \sqcup QC \sqcup RC \sqcup SC$$

is the desired partition; indeed, $S^2 \setminus F = PC \sqcup aQC = RC \sqcup bSC$. □

For much more on Banach–Tarski and related "geometrical paradoxes", including the details we omitted above, see [S. Wagon, *The Banach–Tarski Paradox*, 1985].

4.F. **Ultralimits and ultraproducts.** We close by sketching some more sophisticated applications of ultrafilters. The reader is forewarned that some comfort with abstraction is assumed here.

Exercise 4.69 (assuming you know topology)**.** We will show that (as promised in Section 4.D) ultrafilters can be used as a foundation for developing topology.

The key idea is that, as in Example 4.31, we may think of an ultrafilter $\mathcal{U} \subseteq \mathcal{P}(I)$ on a set I as a "point of I at infinity", which is "declared to be in A" for each $A \in \mathcal{U}$. For example, if $I = \mathbb{Z}$, then an ultrafilter containing $\{\mathbb{N}, \mathbb{N}+1, \mathbb{N}+2, \dots\}$ (which clearly has the finite intersection property) is a "point at $+\infty$", while an ultrafilter containing $\{-\mathbb{N}, -\mathbb{N}-1, -\mathbb{N}-2, \dots\}$ is a "point at $-\infty$".

Thus, given a topological space X, an I-sequence $\vec{x} \in X^I$, and a point $y \in X$, we write

$$\lim_{\mathcal{U}} \vec{x} = \lim_{i \to \mathcal{U}} x_i = x \iff \forall \text{ open } V \ni y,\ \exists A \in \mathcal{U}\, \forall i \in A\, (x_i \in V) \qquad (*)$$
$$\iff \forall \text{ open } V \ni y,\ \vec{x}^{-1}(V) \in \mathcal{U}.$$

(a) Verify that the two conditions above are indeed equivalent.

(b) Show that $V \subseteq X$ is open iff for every set I, ultrafilter $\mathcal{U} \subseteq \mathcal{P}(I)$, I-sequence $\vec{x} \in X^I$, and $y \in V$, if $\lim_{\mathcal{U}} \vec{x} = y$, then there is $A \in \mathcal{U}$ such that for all $i \in A$, $x_i \in V$.

(c) Thus, $F \subseteq X$ is closed iff it is closed under limits, i.e., for every set I, ultrafilter $\mathcal{U} \subseteq \mathcal{P}(I)$, I-sequence $\vec{x} \in F^I$, and $y \in X$, if $\lim_{\mathcal{U}} \vec{x} = y$, then $y \in F$.

(d) Show that a map $f : X \to Y$ to another topological space is continuous iff it preserves limits at every ultrafilter: if $\lim_{\mathcal{U}} \vec{x} = y$ in X, then $\lim_{\mathcal{U}} f(\vec{x}) = f(y)$ in Y.

(e) Show that X is compact iff for every set I, ultrafilter $\mathcal{U} \subseteq \mathcal{P}(I)$, and I-sequence $\vec{x} \in X^I$, there is at least one limit $\lim_{\mathcal{U}} \vec{x}$.

(f) Show that X is Hausdorff iff for every set I, ultrafilter $\mathcal{U} \subseteq \mathcal{P}(I)$, and I-sequence $\vec{x} \in X^I$, there is at most one limit $\lim_{\mathcal{U}} \vec{x}$.

(g) Give a direct proof of Tychonoff's Theorem 4.44 using ultrafilter limits (and not using the Alexander subbasis lemma), thereby showing that products of *compact Hausdorff* spaces are still compact (and Hausdorff), using only the PIT and no other applications of Choice.

(h) Verify that for $\vec{x} \in [0,1]^I$ and an ultrafilter $\mathcal{U} \in \mathcal{P}(I)$, $\lim_{\mathcal{U}} \vec{x}$ may be computed as either of

$$\liminf_{i \to \mathcal{U}} x_i := \sup_{A \in \mathcal{U}} \inf_{i \in A} x_i, \qquad \limsup_{i \to \mathcal{U}} x_i := \inf_{A \in \mathcal{U}} \sup_{i \in A} x_i.$$

(i) Verify that for $\vec{x} \in 2^I$ and (the indicator function of) an ultrafilter $u : 2^I \to 2$, equipping 2 with the discrete topology, $\lim_u \vec{x}$ is none other than $u(\vec{x})$.

(j) Verify that for $\vec{x} \in X^I$ and a principal ultrafilter $\mathcal{U}_i = \{A \subseteq I \mid i \in A\}$ (recall Example 4.29), we have $\lim_{\mathcal{U}_i} \vec{x} = x_i$ ("reflexivity").

(k) Verify that for $\vec{x} \in X^I$, $\vec{y} \in X^J$, and ultrafilters $\mathcal{V} \subseteq \mathcal{P}(J)$ and $\mathcal{U}_j \subseteq \mathcal{P}(I)$ for each $j \in J$,

$$\forall j \in J \left(\lim_{i \to \mathcal{U}_j} x_i = y_j \right) \text{ and } \lim_{j \to \mathcal{V}} y_j = z \implies \lim_{i \to \mathcal{W}} x_i = z$$

where $\mathcal{W} \subseteq \mathcal{P}(I)$ is the ultrafilter whose indicator function $2^I \to 2$ is the pointwise $\lim_{\mathcal{V}}$ of the indicator functions of the $\mathcal{U}_j$ ("transitivity").

(l) In particular, by taking $I = X$ and $\mathcal{U}_x$ above to be principal ultrafilters, show that the behavior of arbitrary ultrafilter limits $\lim_{j \to \mathcal{V}} y_j$ is completely determined by limits of the form $\lim_{x \to \mathcal{W}} x$ where $\mathcal{W} \subseteq \mathcal{P}(X)$.

(m) Show that for any set X and ternary relation "$\lim_{\mathcal{U}} \vec{x} = y$" between arbitrary ultrafilters $\mathcal{U}$ on arbitrary index sets I, I-sequences $\vec{x} \in X^I$, and points $y \in X$, such that (j), (k), and (l) hold, then the collection of sets $V \subseteq X$ obeying (b) forms the open sets of a topology such that $(*)$ recovers the given $\lim_{\mathcal{U}}$ relations. Thus in principle, it is possible to develop all of topology by taking ultrafilter limit as the primitive notion (but in practice quite painful).

[Hint: you should be making liberal use of the Boolean PIT throughout.]

Exercise 4.70. Let again $\mathcal{U} \subseteq \mathcal{P}(I)$ be an ultrafilter on an index set I. We can take not only $\mathcal{U}$-limits of real numbers or points in a topological space, as in the preceding exercise; we can even take "limits" of graphs or other (first-order) structures!

Let X_i be a set for each $i \in I$. The **ultraproduct** of these sets at $\mathcal{U}$ is their "lim inf":

$$\prod_{i \to \mathcal{U}} X_i := \varinjlim_{A \in \mathcal{U}} \prod_{i \in A} X_i.$$

The notation $\varinjlim$ (known as a **direct limit**) means: take the disjoint union of these products (over all $A \in \mathcal{U}$; note that there is no need to disjointify), and quotient by the equivalence relation

$$\textstyle\prod_{i \in A} X_i \ni \vec{x} \sim_{\mathcal{U}} \vec{y} \in \prod_{j \in B} X_j \iff \exists \mathcal{U} \ni C \subseteq A \cap B \left(\vec{x}|C = \vec{y}|C \in \prod_{k \in C} X_k\right).$$

Thus, concretely, $\prod_{i \to \mathcal{U}} X_i$ consists of equivalence classes of functions with domain in $\mathcal{U}$, two such functions being equivalent iff they agree on some common subset of their domains which is still in $\mathcal{U}$.

(a) Verify that this is indeed an equivalence relation.
(b) When $\mathcal{U}$ is the principal ultrafilter at some $i \in I$, $\prod_{j \to \mathcal{U}} X_j$ is in canonical bijection with X_i.
(c) When $X_i = X$ is constant, the ultraproduct $\prod_{i \to \mathcal{U}} X$ is called an **ultrapower**, usually denoted $X^{\mathcal{U}}$. We then have a canonical map $X \to X^{\mathcal{U}}$, taking x to the equivalence class of the constant tuple $(x)_{i \in I} \in X^I$, which is injective, called the **diagonal embedding**.
(d) When $I = X_i = \mathbb{N}$, and $\mathcal{U} \subseteq \mathcal{P}(\mathbb{N})$ is a nonprincipal ultrafilter, $\mathbb{N}^{\mathcal{U}}$ is uncountable. [For example, the equivalence class of $\mathrm{id} : \mathbb{N} \to \mathbb{N}$ is outside the image of the diagonal embedding.]

Now suppose that each X_i is also equipped with a graph (i.e., binary relation) $R_i \subseteq X_i^2$. We will abuse notation by freely identifying R_i with its indicator function $X_i^2 \to 2$, and similarly $\mathcal{U}$ with its indicator function $2^I \cong \mathcal{P}(I) \to 2$. Define the **ultraproduct graph** $\prod_{i \to \mathcal{U}} R_i$ on $\prod_{i \to \mathcal{U}} X_i$ by

$$\begin{aligned} \textstyle\prod_{i \to \mathcal{U}} R_i : (\prod_{i \to \mathcal{U}} X_i)^2 &\longrightarrow 2 \\ ([\vec{x}], [\vec{y}]) &\longmapsto \mathcal{U}((R_i(x_i, y_i))_{i \in I}). \end{aligned}$$

(e) When $\mathcal{U} \subseteq \mathcal{P}(\mathbb{N})$ is a nonprincipal ultrafilter, and $R_i = (<) \subseteq \mathbb{N}^2$ for each i, the **ultrapower graph** $(<^{\mathcal{U}}) := \prod_{i \to \mathcal{U}} (<) \subseteq (\mathbb{N}^{\mathcal{U}})^2$ is still a linear order with a least element; and the diagonal embedding $\mathbb{N} \to \mathbb{N}^{\mathcal{U}}$ is an order-isomorphism with its image, which is an initial segment of $(\mathbb{N}^{\mathcal{U}}, <^{\mathcal{U}})$. Thus, $[\mathrm{id}] \in \mathbb{N}^{\mathcal{U}}$ is strictly above this initial segment.
(f) More generally, if each $(X_i, <_i)$ is a linear order, then so is $(\prod_{i \to \mathcal{U}} X_i, \prod_{i \to \mathcal{U}} <_i)$.

This remarkable fact is not at all specific to "linear order":

(g) (Łoś's[33] theorem) Let $\phi(x^0, \ldots, x^{n-1})$ be a property of n points in a graph (X, R), which is expressed using the symbols $=, R$, connectives $\wedge, \vee, \neg$, and quantifiers $\forall, \exists$. For example,

$$\phi(x^0, x^1) := \text{“}R(x^0, x^1) \wedge \neg \exists x^2 \left(R(x^0, x^2) \wedge R(x^2, x^1)\right)\text{”}$$

says that x^1 is an immediate successor of x^0. Then for any graphs $((X_i, R_i))_{i \in I}$, the truth value of ϕ in the ultraproduct graph $(\prod_{i \to \mathcal{U}} X_i, \prod_{i \to \mathcal{U}} R_i)$ is given by

$$\begin{aligned} \phi : (\textstyle\prod_{i \to \mathcal{U}} X_i)^n &\longrightarrow 2 \\ ([\vec{x}^0], \ldots, [\vec{x}^{n-1}]) &\longmapsto \mathcal{U}((\phi(x_i^0, \ldots, x_i^{n-1}))_{i \in I}). \end{aligned}$$

[Induct on the syntactic expression ϕ, in the manner of Example 3.50.]
(h) In particular, for $n = 0$, if ϕ is a property of a graph not depending on any free variables, then $\prod_{i \to \mathcal{U}} X_i$ satisfies ϕ iff the set of i such that X_i satisfies ϕ is in $\mathcal{U}$.
(i) If each (X_i, R_i) satisfies the axioms of ZFC (with R_i as the membership relation), then so does $(\prod_{i \to \mathcal{U}} X_i, \prod_{i \to \mathcal{U}} R_i)$. Show that if (X, R) is such a model of ZFC, we may take an ultrapower to get a model which (still satisfies Foundation but) is ill-founded.

[33] pronounced "wosh"

More generally still, there is no reason to restrict to a single binary relation. Suppose each X_i is equipped with some family of relations $R_i, S_i, \dots$ of specified finite arities (the same across all i). Then Łoś's theorem holds by exactly the same proof, where now the property ϕ may of course talk about all of the relations. We may also handle operations of finite arities, e.g., each X_i might be an abelian group (with operations $+$ of arity 2, 0 of arity 0, and $-$ of arity 1); the simplest way, given we already know how to deal with relations, is to represent operations via their graphs, which amounts to defining e.g., a binary operation $+$ on the ultraproduct via

$$[\vec{x}^0] + [\vec{x}^1] := [(x_i^0 + x_i^1)_{i \in I}].$$

We will not bother with the details, and will freely use operations and/or relations as needed.

(j) Suppose each X_i is a ring or a field, respectively. Verify that so will be $\prod_{i \to \mathcal{U}} X_i$.

(k) Suppose each X_i is a field of characteristic $\geq p$ or 0, or more generally, this is true for a set of i in $\mathcal{U}$. Verify that so will be $\prod_{i \to \mathcal{U}} X_i$.

(l) Conclude that if $\mathcal{U} \subseteq \mathcal{P}(\mathbb{N})$ is a nonprincipal ultrafilter, and the X_i are finite fields of characteristics $\to \infty$, then $\prod_{i \to \mathcal{U}} X_i$ is a field of characteristic 0.

(m) Verify that if each X_i is algebraically closed, then so is $\prod_{i \to \mathcal{U}} X_i$.

(n) By well-known results of field theory, (i) any two uncountable algebraically closed fields of the same characteristic and cardinality are isomorphic; (ii) there is an algebraically closed field of each characteristic and uncountable cardinality.
Using these results, show that if a property ϕ of fields (finitely expressible as in Łoś's theorem) is true in every algebraically closed field of finite characteristic, then it is true in $\mathbb{C}$. [Some cardinal arithmetic is needed; see Section 5.E.]

Now, we focus on the case $I = \mathbb{N}$ and an ultrapower $X^{\mathcal{U}}$ for a fixed nonprincipal ultrafilter $\mathcal{U} \subseteq \mathcal{P}(\mathbb{N})$. Then for any n-ary relation $R \subseteq X^n$, we have a corresponding ultrapower relation $R^{\mathcal{U}} \subseteq (X^{\mathcal{U}})^n$; similarly for n-ary functions. In particular, for $n = 0$, a single element $x \in X$ regarded as a 0-ary function has ultrapower $x^{\mathcal{U}}$ given by the equivalence class of the constant function $[x]$; in other words, we may conveniently write the diagonal embedding $X \to X^{\mathcal{U}}$ as $x \mapsto x^{\mathcal{U}}$.

(o) Show that for any sequence of nonempty sets $X \supseteq C_0 \supseteq C_1 \supseteq \cdots$, we have $\varnothing \neq \bigcap_n C_n^{\mathcal{U}} \subseteq X^{\mathcal{U}}$.

(p) For example, $\varnothing \neq \bigcap_n [n, \infty)^{\mathcal{U}} \subseteq \mathbb{R}^{\mathcal{U}}$ (cf. (e)); elements of this set are called **positive infinite**.

(q) Similarly, elements of $\bigcap_n (-\infty, -n]^{\mathcal{U}} \subseteq \mathbb{R}^{\mathcal{U}}$ are **negative infinite**. Elements of $\mathbb{R}^{\mathcal{U}}$ which are not positive or negative infinite are **finite**; these are precisely elements of $\bigcup_n [-n, n]^{\mathcal{U}}$.

(r) Similarly, elements of $\bigcap_n (-1/n, 1/n)^{\mathcal{U}} \subseteq \mathbb{R}^{\mathcal{U}}$ are **infinitesimal**; these are finite.

(s) By Łoś's theorem, $\mathbb{R}^{\mathcal{U}}$ is an ordered field, with the operations $+^{\mathcal{U}}, -^{\mathcal{U}}, \cdot^{\mathcal{U}}, /^{\mathcal{U}}$. Verify that

$$\text{infinitesimal} +^{\mathcal{U}} \text{infinitesimal} = \text{infinitesimal},$$
$$\text{finite} +^{\mathcal{U}} \text{finite} = \text{finite},$$
$$\text{finite} \cdot^{\mathcal{U}} \text{infinitesimal} = \text{infinitesimal},$$
$$\text{finite} \cdot^{\mathcal{U}} \text{finite} = \text{finite}.$$

Thus, the finite elements of $\mathbb{R}^{\mathcal{U}}$ form a subring, in which the infinitesimals form an ideal. Let $a \approx b \iff (a -^{\mathcal{U}} b \text{ infinitesimal})$ denote the associated coset equivalence relation.

(t) Let $f : \mathbb{R} \to \mathbb{R}$ and $a, b \in \mathbb{R}$. Show that

$$\lim_{x \to a} f(x) = b \iff \forall a^{\mathcal{U}} \neq x \in \mathbb{R}^{\mathcal{U}} \, (x \approx a^{\mathcal{U}} \implies f^{\mathcal{U}}(x) \approx b^{\mathcal{U}}),$$
$$f \text{ continuous} \iff \forall x \in \mathbb{R} \, \forall y \in \mathbb{R}^{\mathcal{U}} \, (x^{\mathcal{U}} \approx y \implies f(x)^{\mathcal{U}} \approx f^{\mathcal{U}}(y)),$$
$$f \text{ uniformly continuous} \iff \forall x \in \mathbb{R}^{\mathcal{U}} \, \forall y \in \mathbb{R}^{\mathcal{U}} \, (x \approx y \implies f^{\mathcal{U}}(x) \approx f^{\mathcal{U}}(y)).$$

The application of ultrapowers to reason about "infinities" is part of *nonstandard analysis*.

5. Cardinality

5.A. **Equinumerosity and cardinality.**

Definition 5.1. Two sets A, B are **equinumerous** or **have the same cardinality**, denoted

$$A \cong B,$$

if there exists a bijection $f : A \cong B$.

Remark 5.2. This is a (proper class) equivalence relation on V, since bijections are closed under composition, inversion, and identity.

The concept of equinumerosity is more fundamental than that of "the cardinality" $|A|$ of a set. We would like to define the latter to mean any object we can assign to A, in such a way that equinumerosity indeed means "having the same cardinality":

$$|A| = |B| \iff A \cong B. \tag{5.3}$$

In other words, "cardinality" should be the quotient $V/\cong$. For an equivalence relation on a set, we can construct the quotient as the set of equivalence classes; but the $\cong$-equivalence classes will generally be proper classes. This means that "cardinality", like "class", is an informal meta-notion that does not exist in the universe. So we cannot have sets of cardinals, functions on cardinals, etc.

These issues cannot be avoided in the basic set theory ZF^-, but with a little more, we can:

Definition 5.4 (assuming Foundation)**.** The **cardinality** $|A|$ of a set A is defined using Scott's trick 3.185, i.e., as the set of all sets equinumerous with A with minimal rank among all such sets.

Thus, a **cardinal** κ is a maximal nonempty set of equinumerous sets of equal rank.

Definition 5.5 (assuming Choice)**.** Then in each $\cong$-equivalence class, there is at least one ordinal; and among those, we may canonically choose the least one. The **cardinality** $|A|$ of a set A is the least ordinal equinumerous with A.

Thus, a **cardinal** κ is an **initial ordinal**: one not equinumerous with any strictly smaller ordinal.

Either of these encodings allows to treat cardinals as objects existing in the universe. As with other coding choices (Sections 2.D and 2.E), the choice of which of these we call "cardinals" is largely irrelevant for ordinary mathematical practice; all we need to know is that (5.3) holds. Nonetheless, in the presence of Choice, the initial ordinals representation is somehow much more canonical and convenient, in the same way that the standard encoding of naturals is well-justified as the canonical representatives (i.e., Mostowski collapse) of ordinary induction. We therefore adopt the following

Convention 5.6. By default, we assume the cardinality of any well-orderable set to be encoded via Definition 5.5, as an initial ordinal. (Thus, assuming Choice, cardinal = initial ordinal.)

5.B. **Cardinal comparison.**

Definition 5.7. A set A **injects** into another set B, denoted

$$A \hookrightarrow B,$$

if there exists an injection $f : A \hookrightarrow B$.

Remark 5.8. This is a preorder on V, since injections are closed under composition and identity.

Remark 5.9. $A \cong B \implies A \hookrightarrow B \hookrightarrow A$, since bijections are injections.

It follows from both of these properties that $A' \cong A \hookrightarrow B \cong B' \implies A' \hookrightarrow B'$, justifying

Definition 5.10. For two cardinals $|A|, |B|$,

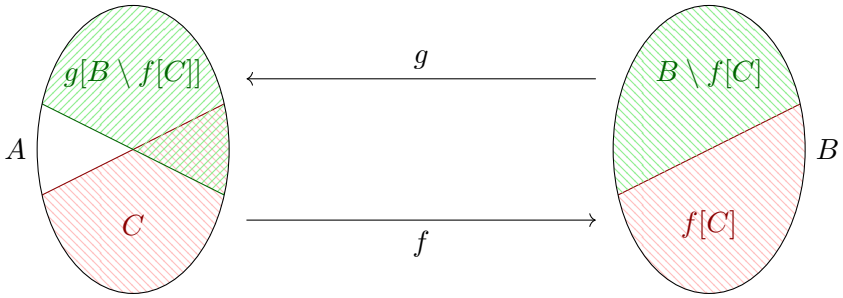
$$|A| \leq |B| \ :\iff A \hookrightarrow B.$$

This is a preorder on the class of all cardinals (however we choose to encode them).

Theorem 5.11 (Schröder–Bernstein)**.** For any two sets A, B, if $A \hookrightarrow B \hookrightarrow A$, then $A \cong B$.

Proof. Let $f : A \hookrightarrow B$ and $g : B \hookrightarrow A$; our goal is to construct a bijection $h : A \cong B$ which is equal to f on some elements and g^{-1} on others (or possibly both). Let $C \subseteq A$ be the elements on which h is given by f. Then part of h will be given by the bijection $f : C \cong f[C] \subseteq B$. The rest must be given by the inverse of $g : B \setminus f[C] \cong g[B \setminus f[C]] \subseteq A$; thus we must find $C \subseteq A$ such that

$$C = A \setminus g[B \setminus f[C]].$$



Since $C \mapsto A \setminus g[B \setminus f[C]]$ is monotone, by Knaster–Tarski 3.6 such a fixed point C exists. □

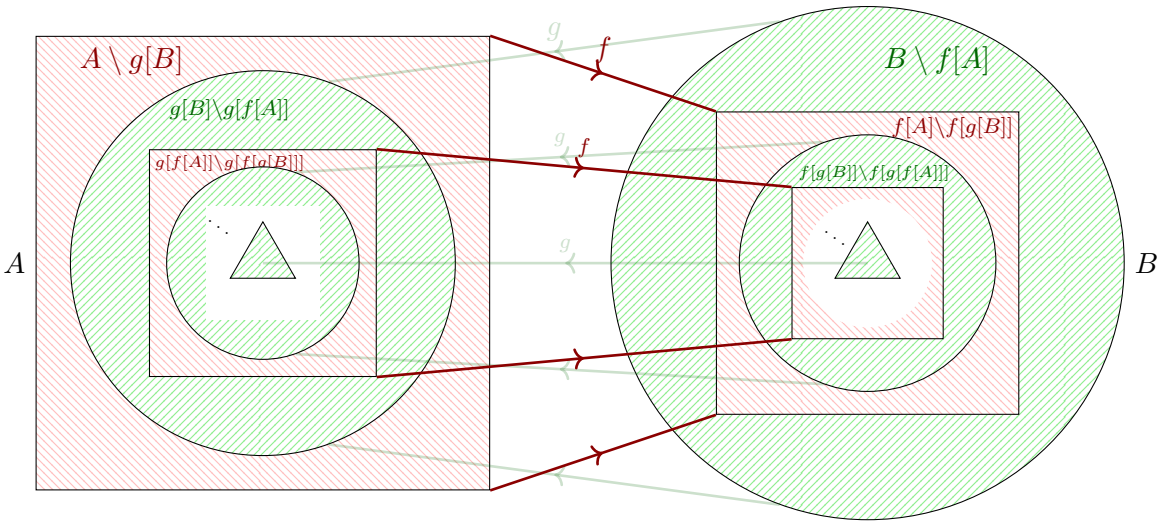
Remark 5.12. For a more informative proof, we may use the transfinite version of Knaster–Tarski 3.167. Note that since g is injective, we may write the above operator more intelligibly as

$$\begin{aligned} A \setminus g[B \setminus f[C]] &= A \setminus (g[B] \setminus g[f[C]]) \\ &= (A \setminus g[B]) \sqcup g[f[C]], \end{aligned}$$

which is clearly finitary; thus by Proposition 3.169, the least fixed point C is the ω-union of

$$\begin{aligned} \varnothing &\subseteq A \setminus g[B] \\ &\subseteq (A \setminus g[B]) \sqcup (g[f[A]] \setminus g[f[g[B]]]) \\ &\subseteq (A \setminus g[B]) \sqcup (g[f[A]] \setminus g[f[g[B]]]) \sqcup (g[f[g[f[A]]]] \setminus g[f[g[f[g[B]]]]]) \\ &\subseteq \cdots . \end{aligned}$$

This union consists of the red rings on the left below; the resulting bijection h simply switches each with the green ring inside it, while applying g^{-1} on the remaining "center".



Exercise 5.13. Verify that if we instead take the greatest fixed point C of $C \mapsto A \setminus g[B \setminus f[C]]$, the bijection h is defined the same way as above except being given by f instead of g^{-1} on the "center".

5.C. **Well-orderable cardinals.**

Proposition 5.14. For an ordinal α and initial ordinal κ, we have $\kappa \subseteq \alpha \iff \kappa \hookrightarrow \alpha$.

In other words, an initial ordinal (not only $\ncong$ but) does not inject into any smaller ordinal.

In particular, ordinal and cardinal comparison agree on initial ordinals.

Proof. Clearly, $\kappa \subseteq \alpha \implies \kappa \hookrightarrow \alpha$. Conversely, if $\kappa \not\subseteq \alpha$, then since ordinals are linearly ordered, $\alpha \subsetneq \kappa$; since κ is initial, this means $\alpha \ncong \kappa$, whence $\kappa \not\hookrightarrow \alpha$ by Schröder–Bernstein. □

Theorem 5.15 (pigeonhole principle)**.** For $m, n \in \mathbb{N}$, we have $m \leq n \iff m \hookrightarrow n$.

In other words, naturals are initial ordinals.

Proof. $\implies$ is obvious; we prove $\impliedby$ by induction on n. If $n = 0$, then clearly $m \hookrightarrow n = \varnothing$ implies $m = \varnothing$. Now suppose every $m \hookrightarrow n$ is $\leq n$, and let $f : m \hookrightarrow n + 1$. If $m = 0$, then clearly $m \leq n$. Now suppose $m > 0$, whence $m = m' + 1$ for some $m' \in \mathbb{N}$. If $f[m'] \subseteq n$, then by the IH, $m' \leq n$, whence $m = m' + 1 \leq n + 1$. Otherwise, there is some $k < m'$ such that $f(k) = n$, whence since f is injective, $f(m') < n$; modify f by swapping $f(k), f(m')$, and apply the previous case. □

Proposition 5.16. If K is a set of initial ordinals, then $\sup K \in \mathbb{ON}$ is initial.

Proof. If $\sup K \hookrightarrow \alpha < \sup K$, then $\alpha < \kappa$ for some $\kappa \in K$, and $\kappa \leq \sup K \hookrightarrow \alpha$, contradicting that κ is initial. □

Corollary 5.17. ω is an initial ordinal: $\omega \not\hookrightarrow n$ for all $n \in \omega$. □

Recall also Hartogs' Theorem 3.144, which we may now restate as

Theorem 5.18 (Hartogs)**.** For every set A, there is a least ordinal $\eta(A)$ (which is therefore initial), called the **Hartogs number** of A, such that $\eta(A) \not\lesssim |A|$.

Corollary 5.19. For every initial ordinal κ, there is a least **successor cardinal** $\kappa^+ = \eta(\kappa) > \kappa$.

Note that this is not to be confused with the successor ordinal (Definition 3.136). For naturals they agree, but otherwise κ^+ is much bigger.

Definition 5.20. $\aleph_\alpha = \omega_\alpha$ is the αth infinite initial ordinal. Thus

$$\begin{aligned}
\aleph_0 &= \omega_0 = \omega,\\
\aleph_1 &= \omega_1 = \omega^+,\\
\aleph_2 &= \omega_2 = \omega^{++},\\
&\vdots\\
\aleph_\omega &= \sup_{n<\omega} \aleph_n \quad \text{by Proposition 5.16,}\\
\aleph_{\omega+1} &= \aleph_\omega^+;
\end{aligned}$$

more generally,

$$\begin{aligned}
\aleph_{\alpha+1} &= \aleph_\alpha^+,\\
\aleph_\alpha &= \sup_{\beta<\alpha} \aleph_\beta \quad \text{for limit ordinals } \alpha\text{, by Proposition 5.16.}
\end{aligned}$$

(Usually, people write $\aleph_\alpha$ when thinking of it as a cardinal, versus ω_α as an ordinal.)

Remark 5.21. By Proposition 4.1(iii), every cardinal $\leq$ an initial ordinal is itself well-orderable.

Corollary 5.22 (over ZF)**.** The Axiom of Choice is equivalent to: cardinals are linearly ordered.

Proof. If Choice holds, then every cardinal is an initial ordinal, which are well-ordered.

Conversely, if a set A is comparable in cardinality with $\eta(A)$, then since $\eta(A) \not\lesssim A$, we have $A < \eta(A)$, whence A is well-orderable. □

Exercise 5.23. By essentially Proposition 4.1(iv), for a nonempty set A and well-ordered set B, we have $A \hookrightarrow B$ iff $B \twoheadrightarrow A$.

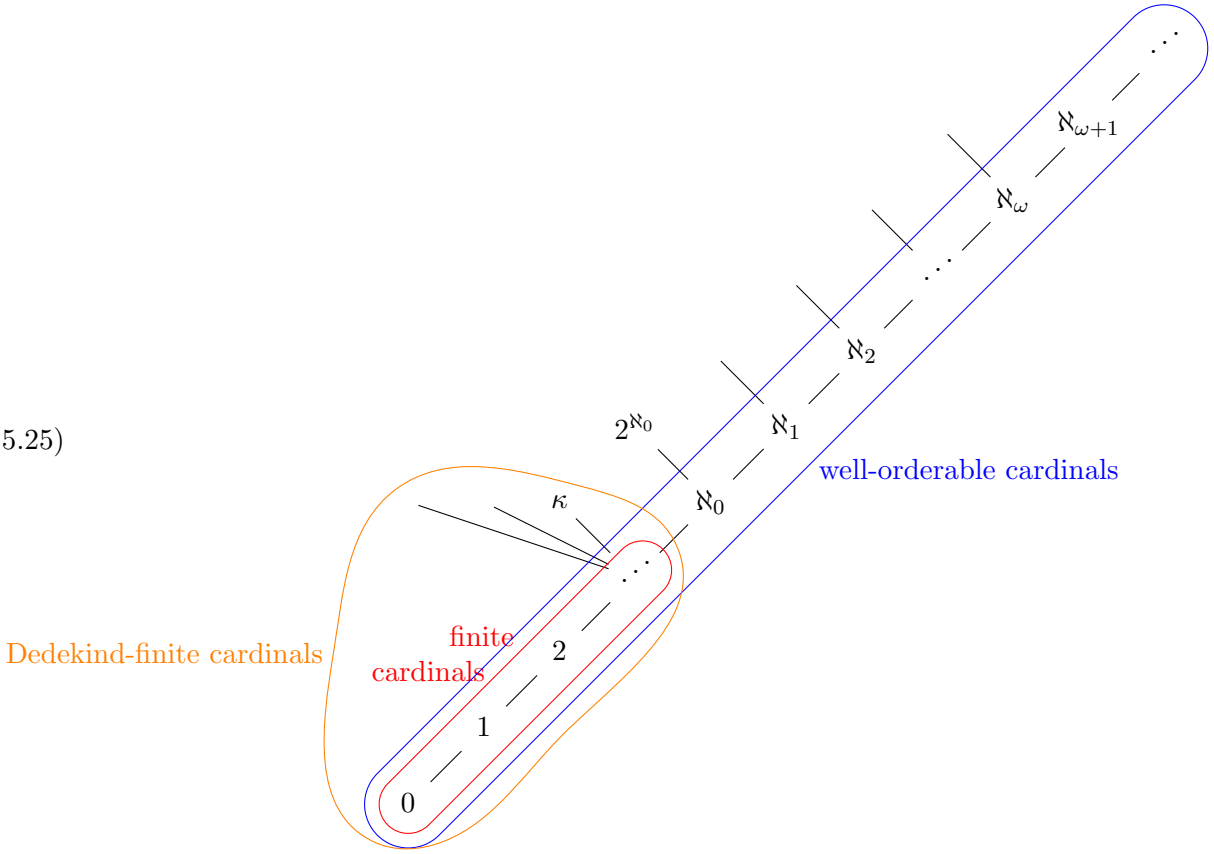
Thus for example, a nonempty set A is countable ($A \hookrightarrow \mathbb{N}$) iff it admits a surjection $\mathbb{N} \twoheadrightarrow A$.

Proposition 5.24. Every cardinal is comparable with every natural $n \in \mathbb{N}$.

Proof. Suppose $|A|$ is a cardinality not in $\mathbb{N}$; we prove by induction that $n \leq |A|$ for every $n \in \mathbb{N}$. Clearly $0 = \varnothing \hookrightarrow A$. Now suppose $n \leq |A|$, i.e., there is $f : n \hookrightarrow A$. Then $f[n] \neq A$, hence we may extend f to an injection $n + 1 \hookrightarrow A$ by taking any $f(n) \in A \setminus f[n]$. □

We thus get the following picture of cardinals in the absence of Choice:

(5.25)



The well-orderable cardinals, i.e., initial ordinals, form a well-ordered (because $\mathbb{ON}$ is well-ordered), downward-closed (by Remark 5.21) "spine" without an upper bound (by Hartogs). Among them, the finite cardinals $n \in \omega$ form an "initial segment" that are actually below everything else (by Proposition 5.24). Above them, there could be infinite cardinals that are not $\geq \aleph_0$, i.e., cardinalities of infinite sets without an infinite sequence, since Dependent Choice is required to construct such an infinite sequence (Proposition 4.5). Such sets without an infinite sequence are called **Dedekind-finite**; the infinite Dedekind-finite sets have Hartogs number $\aleph_0$. Similarly, even if DC holds, there could be uncountable cardinalities (not $\leq \aleph_0$, which means $> \aleph_0$ assuming DC) that are incomparable with $\aleph_1$, which will have Hartogs number $\aleph_1$.

Remark 5.26. In particular, $2^{\aleph_0}$, the cardinality of $\mathbb{R}$ (see Example 5.34), is usually considered to be "definably" incomparable with $\aleph_1$. Here "definable" has the same meaning as in Theorem 4.20: it is not possible to write down an "explicit" injection either way (nor to prove that this is impossible). Indeed, there is even a theorem saying that $\aleph_1$ and $2^{\aleph_0}$ are the only two minimal "definable" uncountable cardinalities!

5.D. **Cardinal arithmetic.** In general, given any (say binary) operation $*$ on sets which is *functorial* in the sense of Definition 2.68, functoriality implies that $*$ respects the equivalence relation $\cong$ on V, hence descends to an operation on the quotient class of cardinals defined via

$$|A| * |B| := |A * B|.$$

Definition 5.27. The **sum** of cardinals is induced by disjoint union (Definition 2.76):

$$|A| + |B| := |A \sqcup B| = \big|\{(i, x) \in 2 \times (A \cup B) \mid (i = 0 \text{ and } x \in A) \text{ or } (i = 1 \text{ and } x \in B)\}\big|.$$

Definition 5.28. The **product** of cardinals is induced by Cartesian product (Definition 2.32):

$$|A| \cdot |B| := |A \times B|.$$

Definition 5.29. Exponentiation of cardinals is induced by function sets (Definition 2.48):

$$|B|^{|A|} := \big|B^A\big|.$$

(Functoriality is by Example 2.71.)

Remark 5.30. For cardinals represented as initial ordinals κ, λ, these notions must not be confused with the ordinal arithmetic operations from Section 3.I with the same name!

For $+$ and $\cdot$, we at least have that the ordinal operation yields a (typically non-initial) ordinal whose cardinality is the cardinal operation:

$$|\text{ordinal } \kappa + \lambda| = \text{cardinal } \kappa + \lambda,$$
$$|\text{ordinal } \kappa \cdot \lambda| = \text{cardinal } \kappa \cdot \lambda.$$

These follow from the rank-based definitions of ordinal $+, \cdot$ from Exercise 3.159. It follows that

$$\text{ordinal } \kappa + \lambda \geq \text{cardinal } \kappa + \lambda,$$
$$\text{ordinal } \kappa \cdot \lambda \geq \text{cardinal } \kappa \cdot \lambda.$$

For exponentiation however, the ordinal power from Exercise 3.163 will usually be much smaller than the cardinal power! For example, the ordinal power $2^\omega = \sup\{1, 2, 4, 8, \dots\} = \omega$ is countable, whereas the cardinal power 2^ω is not, by Cantor's Theorem 2.9 (see also 5.33).

(Note that the cardinal operations *do* agree with ordinal ones on naturals, since these are all initial by Theorem 5.15.)

Just as any functorial set operation descends to an operation on cardinals, so does every natural bijection between two such operations (see Section 2.E) yield an algebraic identity:

Proposition 5.31. The following all refer to cardinal operations:

(a) $+$ and $\cdot$ are commutative and associative, with respective identity elements $0, 1$.
(b) $\cdot$ distributes over $+$ and $\kappa \cdot 0 = 0$. In particular, $\kappa \cdot n = \underbrace{\kappa + \cdots + \kappa}_{n}$ for $n \in \mathbb{N}$.
(c) $\kappa^{\lambda+\mu} = \kappa^\lambda \cdot \kappa^\mu$ and $\kappa^0 = 1$. In particular, $\kappa^n = \kappa \cdots \kappa$ for $n \in \mathbb{N}$.
(d) $(\kappa \cdot \lambda)^\mu = \kappa^\mu \cdot \lambda^\mu$ and $1^\mu = 1$.
(e) $(\kappa^\lambda)^\mu = \kappa^{\lambda \cdot \mu}$ and $\kappa^1 = \kappa$.
(f) $|\mathcal{P}(X)| = 2^{|X|}$.

Proof. By various canonical bijections. For example, (e) follows from Example 2.81. □

We also get various *in*equalities, derived from natural *in*jections:

Proposition 5.32. Again referring only to cardinal operations:

(a) $+, \cdot$ are monotone (in both arguments).
(b) $\kappa \leq \lambda \implies \kappa^\mu \leq \lambda^\mu$, and $\kappa \leq \lambda \implies \mu^\kappa \leq \mu^\lambda$, unless $\mu = \kappa = 0 < \lambda$.

Proof. Let $f : A \hookrightarrow B$; then for any C,

$$\begin{aligned} A \sqcup C &\longhookrightarrow B \sqcup C \\ (0,a) &\longmapsto (0, f(a)) \\ (1,c) &\longmapsto (1,c), \end{aligned} \qquad \begin{aligned} A \times C &\longhookrightarrow B \times C \\ (a,c) &\longmapsto (f(a),c), \end{aligned} \qquad \begin{aligned} A^C &\longhookrightarrow B^C \\ g &\longmapsto f \circ g, \end{aligned}$$

which shows $|A| + |C| \le |B| + |C|$, $|A| \cdot |C| \le |B| \cdot |C|$, and $|A|^{|C|} \le |B|^{|C|}$. To show $|C|^{|A|} \le |C|^{|B|}$:

- If $|C| > 0$, then pick any $c \in C$; then we have an injection
$$\begin{aligned} C^A &\longhookrightarrow C^B \\ h &\longmapsto \left(x \mapsto \begin{cases} h(f^{-1}(x)) & \text{if } x \in \operatorname{im}(f), \\ c & \text{else} \end{cases} \right). \end{aligned}$$
- If $|C| = 0 < |A|$, then $|C|^{|A|} = 0 \le |C|^{|B|}$.
- If $|B| = 0$, then $|A| = 0$ (since $A \hookrightarrow B$), so $|C|^{|A|} = 1 = |C|^{|B|}$. □

Theorem 5.33 (Cantor)**.** For any cardinal κ, $\kappa < 2^\kappa$.

Proof. Letting $\kappa = |A|$, we have $A \hookrightarrow \mathcal{P}(A)$ by $a \mapsto \{a\}$, but $A \ncong \mathcal{P}(A)$ by Theorem 2.9. □

Example 5.34. We have

$$\begin{aligned} \mathbb{R} &\longhookrightarrow \mathcal{P}(\mathbb{Q}) \\ r &\longmapsto \{q \in \mathbb{Q} \mid q < r\}, \end{aligned} \qquad \begin{aligned} \mathcal{P}(\mathbb{N}) &\longhookrightarrow \mathbb{R} \\ A &\longmapsto \textstyle\sum_{n \in A} 10^{-n}, \end{aligned}$$

whence

$$|\mathbb{R}| \le |\mathcal{P}(\mathbb{Q})| = 2^{|\mathbb{Q}|} = 2^{\aleph_0} = 2^{|\mathbb{N}|} = |\mathcal{P}(\mathbb{N})| \le |\mathbb{R}|,$$

and so by Schröder–Bernstein,

$$|\mathbb{R}| = |\mathcal{P}(\mathbb{N})| = 2^{\aleph_0} > \aleph_0.$$

Example 5.35. We have

$$2^{\aleph_0} \le 3^{\aleph_0} \le \cdots \le \aleph_0^{\aleph_0} \le (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0 \cdot \aleph_0} = 2^{\aleph_0}$$

(the last step using that $\mathbb{N}^2$ is countable, e.g., by the injections $n \mapsto (n,0)$ and $(m,n) \mapsto 2^m 3^n$). Thus by Schröder–Bernstein, these cardinals are all equal. For example, $|\mathbb{R}^{\mathbb{N}}| = (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0} = |\mathbb{R}|$.

We may also define indexed versions of cardinal operations:

Definition 5.36. For a set I and family of sets $(A_i)_{i \in I}$, the **indexed sum and product** of cardinals are defined via

$$\begin{aligned} \textstyle\sum_{i \in I} |A_i| &:= \left| \textstyle\bigsqcup_{i \in I} A_i \right|, \\ \textstyle\prod_{i \in I} |A_i| &:= \left| \textstyle\prod_{i \in I} A_i \right|. \end{aligned}$$

Exercise 5.37. Check that these are well-defined, i.e., depend only on the cardinalities of the A_i,[34] assuming the Axiom of Choice (why?).

Remark 5.38. Again, for sum only, this is related to the indexed ordinal sum from Exercise 3.159:

$$|\text{ordinal } \textstyle\sum_{\gamma < \beta} \kappa_\gamma| = \text{cardinal } \textstyle\sum_{\gamma < \beta} \kappa_\gamma \le \text{ordinal } \textstyle\sum_{\gamma < \beta} \kappa_\gamma.$$

Exercise 5.39. Prove indexed analogs of Proposition 5.31. [See Exercises 3.159 and 3.163.]

Exercise 5.40. Prove indexed analogs of Proposition 5.32, assuming the Axiom of Choice (why?).

We will say more about these indexed operations in Section 5.F below.

[34] See Exercise 2.79. Indeed, they are even invariant under replacing I with an equinumerous copy, provided we also reindex the A_i's. More precisely, $\sum$ and $\prod$ are functorial on a category called $\int_{I \in \mathsf{Set}} \mathsf{Set}^I$.

5.E. **Well-ordered cardinal arithmetic.** While the above laws of cardinal arithmetic had fairly concrete "structural" proofs, under Choice things become much more trivial:

Proposition 5.41. For every infinite well-orderable cardinal κ, we have $\kappa + 1 = \kappa$.

(Here + refers to cardinal sum.)

Proof. We have $\aleph_0 + 1 = \aleph_0$, i.e., there is a bijection

$$\begin{aligned} \omega \sqcup \{0\} &\cong \omega \\ (0, n) &\mapsto n + 1 \\ (1, 0) &\mapsto 0. \end{aligned}$$

Now since $\omega \leq \kappa$, we get a bijection $\kappa \sqcup \{0\} \cong \kappa$ which is the above together with the identity function on $\kappa \setminus \omega$. □

Corollary 5.42. Every infinite initial ordinal $\kappa = \aleph_\alpha$ is a limit ordinal.

Proof. For $\alpha < \kappa$, either $\alpha < \omega$ in which case $\alpha + 1 < \omega$ by definition of ω, or α is infinite in which case $|\alpha + 1| = |\alpha| + 1 = |\alpha| \leq \alpha < \kappa$ whence $\alpha + 1 < \kappa$ since κ is initial. (Here + means ordinal.) □

Exercise 5.43. Show that in general (without Choice), for a cardinal κ, the following are equivalent:

(i) $\kappa + 1 = \kappa$.
(ii) κ is Dedekind-infinite, i.e., $\aleph_0 \leq \kappa$ (see (5.25)).
(iii) For any set A with $|A| = \kappa$, there is a non-surjective injection $A \hookrightarrow A$.

Proposition 5.44. Let I be a well-orderable set, $(A_i)_{i \in I}$ be a family of sets. Then

$$\left|\bigcup_{i \in I} A_i\right| \leq \sum_{i \in I} |A_i|.$$

Proof. Map each $a \in \bigcup_{i \in I} A_i$ to $(i, a) \in \bigsqcup_{i \in I} A_i$ for the smallest i such that $a \in A_i$. □

Theorem 5.45. For every infinite well-orderable cardinal κ, we have $\kappa^2 = \kappa$.[35]

Proof. By induction. Suppose for all infinite cardinals $\lambda < \kappa$, we have $\lambda^2 = \lambda$. Then

$$\begin{aligned} \kappa^2 &= |\kappa \times \kappa| \\ &= \left|\bigcup_{\alpha < \kappa} (\alpha \times \alpha)\right| && \text{by Corollary 5.42} \\ &\leq \sum_{\alpha < \kappa} |\alpha|^2 && \text{(cardinal sum, by Proposition 5.44)} \\ &\leq \sum_{\alpha < \kappa} |\alpha|^2 && \text{(ordinal sum, by Remark 5.30)} \\ &\leq \sup_{\alpha < \kappa} \sum_{\beta < \alpha} |\beta|^2 && \text{(ordinal sum and sup, by Corollary 5.42 and Exercise 3.159);} \end{aligned}$$

but for each $\alpha < \kappa$, we have $\left|\sum_{\beta < \alpha} |\beta|^2\right| \leq |\alpha|^3 < \kappa$ by the IH (for $\alpha \geq \omega$) or definition of ω (for $\alpha < \omega$), whence $\sum_{\beta < \alpha} |\beta|^2 < \kappa$ since κ is initial. □

Corollary 5.46. For two well-orderable cardinals κ, λ at least one of which is infinite, we have

$$\begin{aligned} \kappa + \lambda &= \max(\kappa, \lambda), \\ \kappa \cdot \lambda &= \max(\kappa, \lambda) \quad \text{if } \kappa, \lambda \neq 0. \end{aligned}$$

Proof. WLOG $\kappa \leq \lambda$, hence $\lambda \geq \aleph_0$. Then $\kappa + \lambda \leq \lambda + \lambda = 2 \cdot \lambda \leq \lambda^2 = \lambda$, and clearly also $\lambda \leq \kappa + \lambda$ and $\kappa \cdot \lambda \leq \lambda \cdot \lambda = \lambda$. If $\kappa > 0$, then also $\lambda = 1 \cdot \lambda \leq \kappa \cdot \lambda$. □

Corollary 5.47. For an infinite well-orderable cardinal κ, we have $\kappa^n = \kappa$ for every $1 \leq n \in \mathbb{N}$. □

[35] More precisely, the proof yields an inductive way of defining an explicit injection $\kappa^2 \hookrightarrow \kappa$ for each κ; hence why this statement for well-orderable κ does not need Choice.

The following is a weird application of well-orderability of $\mathbb{R}$, similar in spirit to the "pathological" constructions using Choice from Section 4, but additionally using a bit of cardinal arithmetic:

Theorem 5.48 (Mazurkiewicz)**.** There exists $A \subseteq \mathbb{R}^2$ containing exactly two points on every line.[36]

Proof. Note that the set $\mathcal{L}$ of lines in $\mathbb{R}^2$ has cardinality $2^{\aleph_0} = |\mathbb{R}|$: for example, we have injections

$$\begin{aligned} \mathbb{R} &\hookrightarrow \mathcal{L} & \mathcal{L} &\hookrightarrow \mathbb{R}^3 \\ b &\longmapsto \{x = b\}, & L &\longmapsto \begin{cases} (1, m, b) & \text{if } L \text{ is a nonvertical line } y = mx + b, \\ (0, 1, b) & \text{if } L \text{ is a vertical line } x = b. \end{cases} \end{aligned}$$

Thus, let $(L_\alpha)_{\alpha<2^{\aleph_0}}$ be a transfinite enumeration of $\mathcal{L}$, here assuming that $2^{\aleph_0}$ is an initial ordinal. Define a sequence $(A_\alpha \subseteq \mathbb{R}^2)_{\alpha<2^{\aleph_0}}$, where each $|A_\alpha| \leq 2$, inductively as follows:

- If L_α already contains two points in $\bigcup_{\beta<\alpha} A_\beta$, then $A_\alpha := \varnothing$.
- Otherwise, note that $|\bigcup_{\beta<\alpha} A_\beta| \leq 2 \cdot |\alpha| < 2^{\aleph_0}$, thus also $|(\bigcup_{\beta<\alpha} A_\beta)^2| < 2^{\aleph_0} = |L_\alpha|$. For each pair of distinct points in $\bigcup_{\beta<\alpha} A_\beta$, the unique line through them is not L_α by assumption, hence intersects L_α in at most one point. Pick one or two (depending on whether $|L_\alpha \cap \bigcup_{\beta<\alpha} A_\beta| = 1$ or 0) points on L_α not on any such line and not in $\bigcup_{\beta<\alpha} A_\beta$, and let A_α be those points.

We claim that $A := \bigcup_{\alpha<2^{\aleph_0}} A_\alpha$ works. Indeed, to check that L_α contains exactly two points in A: by definition of A_α, $L_\alpha \cap \bigcup_{\beta\leq\alpha} A_\beta$ contains at least two points. If $L_\alpha \cap A$ contained at least three points, then there is a least β such that $L_\alpha \cap \bigcup_{\gamma\leq\beta} A_\gamma$ contains at least three points, which means $L_\alpha \cap \bigcup_{\gamma<\beta} A_\gamma$ still contains at most two points. But by definition of A_β, we would not have either added a new point to a line that already passes through two existing points, or added two new points to a line (namely $L_\alpha = L_\beta$) that already had a point. □

Another useful consequence of Theorem 5.45 is in the context of finitary monotone set operations $T : \mathcal{P}(X) \to \mathcal{P}(X)$, in the sense of Proposition 3.169, which recall meant that each $x \in T(A)$ depends on only finitely many elements of A, i.e., $T(A) = \bigcup_{b_0,\dots,b_{n-1}\in A} T(\{b_0, \dots, b_{n-1}\})$.

Corollary 5.49 (assuming Choice)**.** Let $T : \mathcal{P}(X) \to \mathcal{P}(X)$ be a finitary monotone set operator, and κ be an infinite cardinal such that T maps finite sets to sets of size $\leq \kappa$. Then for any $A \subseteq X$, we have $|\overline{T}(A)| \leq \max(|A|, \kappa)$.

Proof. Since T is finitary, for any $A \subseteq X$, we have

$$\begin{aligned} |A \cup T(A)| &= |A \cup \textstyle\bigcup_{n\in\mathbb{N}} \bigcup_{\vec{b}\in A^n} T(\{b_0, \dots, b_{n-1}\})| \\ &\leq |A| + \textstyle\sum_{n\in\mathbb{N}} \sum_{\vec{b}\in A^n} |T(\{b_0, \dots, b_{n-1}\})| \\ &\leq |A| + \textstyle\sum_{n\in\mathbb{N}} |A|^n \cdot \kappa \\ &= \max(|A|, \kappa). \end{aligned}$$

Now by induction, each A_n in the Knaster–Tarski sequence (3.167) obeys the same bound for each $n \in \mathbb{N}$, thus by Proposition 3.169, so does $\overline{T}(A) = A_\omega = \bigcup_{n<\omega} A_n$. □

Example 5.50. A $\mathbb{Q}$-vector space X of infinite (well-orderable) dimension κ, or even just with a generating set of cardinality κ, has cardinality κ. Indeed, $T(A) = \{\vec{0}\} \cup \{a\vec{x} + b\vec{y} \mid a, b \in \mathbb{Q}, \vec{x}, \vec{y} \in A\}$ is clearly finitary and maps finite A to $T(A)$ of size $\leq \aleph_0$. Similarly for κ-generated group, ring, etc.

Example 5.51. An $\mathbb{R}$-vector space of dimension $\kappa \geq 2^{\aleph_0}$ has cardinality κ.

For a generalization replacing "finitary" with larger arities, see **TODO**.

[36] It appears to be an open problem whether such a set can be constructed without Choice! (It is known that the existence of such a set does not imply that $\mathbb{R}$ is well-orderable; see [A. Miller, *Infinite Combinatorics and Definability*].)

5.F. **Regular cardinals.** We've shown that finite sums and products of well-orderable cardinals are trivial. What can we say about the indexed sums and products (Definition 5.36)? We henceforth work under full Choice, which is needed even in order for the indexed operations to be well-defined.

Proposition 5.52. Let $(\kappa_i)_{i\in I}$ be a family of cardinals, such that $\sup_{i\in I}\kappa_i \geq \max(|I|,\aleph_0)$. Then

$$\textstyle\sum_{i\in I}\kappa_i = \sup_{i\in I}\kappa_i.$$

Proof. $\geq$ is straightforward. For $\leq$, we have

$$\begin{aligned}\textstyle\sum_{i\in I}\kappa_i &\leq \textstyle\sum_{i\in I}\sup_{j\in I}\kappa_j\\ &= |I|\cdot\sup_{j\in I}\kappa_j\\ &\leq (\sup_{j\in I}\kappa_j)^2\\ &= \sup_{j\in I}\kappa_j.\end{aligned}$$

□

Example 5.53. $\sum_{n\in\mathbb{N}}\aleph_n = \sup_{n\in\mathbb{N}}\aleph_n = \aleph_\omega$.

Example 5.54. Clearly any $\kappa = \sum_{\alpha<\kappa}1$, hence we need the assumption that $\sup_i\kappa_i \geq \max(|I|,\aleph_0)$.

Exercise 5.55. Show that more generally,

$$\textstyle\sum_{i\in I}\kappa_i = \max(\sup_{i\in I}\kappa_i, |\{i\in I \mid \kappa_i > 0\}|),$$

provided the RHS is infinite. (A version of this was already used in the proof of Theorem 5.45.)

Corollary 5.56. For an infinite cardinal κ and a family of cardinals $(\lambda_i)_{i\in I}$ with $|I|,\lambda_i<\kappa$,

$$\textstyle\sum_{i\in I}\lambda_i < \kappa \iff \sup_{i\in I}\lambda_i < \kappa.$$

Proof. If the RHS above is finite, then clearly so is the LHS; thus one is $<\kappa$ iff the other is. □

Definition 5.57. A cardinal κ is **regular** if for any family of cardinals $(\lambda_i)_{i\in I}$ with $|I|,\lambda_i<\kappa$, we have $\sum_{i\in I}\lambda_i<\kappa$, or equivalently if κ is infinite, $\sup_{i\in I}\lambda_i<\kappa$.

Exercise 5.58. Show that κ is regular iff for any family of sets $(A_i)_{i\in I}$ with $|I|,|A_i|<\kappa$, we have $|\bigcup_{i\in I}A_i|<\kappa$.

Example 5.59. 0 is (vacuously) regular.

Example 5.60. 1 is regular: any family of cardinals $(\lambda_i)_{i\in I}$ with $|I|<1$ must be empty.

Example 5.61. 2 is regular: any family of cardinals $(\lambda_i)_{i\in I}$ with $|I|,\lambda_i<2$ has sum either 0 or $\lambda_i<2$ for the unique $i\in I$.

Example 5.62. No $3\leq n<\omega$ is regular, since $n=(n-1)+1$.

Remark 5.63. Not everyone agrees on which finite cardinals (if any) are considered regular. (Note that if we instead take $\sup_{i\in I}\lambda_i<\kappa$ as the primary definition, then every $n<\omega$ *except* 0 would be regular.) The definition we have given is the most useful from the perspective of Remark 5.68.

Example 5.64. $\aleph_0$ is regular, being closed under (binary) $+$. By Exercise 5.58, this means that a finite union of finite sets is finite.

Example 5.65. Any infinite successor cardinal κ^+ is regular: if $|I|,\lambda_i<\kappa^+$, then $|I|,\lambda_i\leq\kappa$, whence $\sum_{i\in I}\lambda_i\leq\kappa^2=\kappa<\kappa^+$. In other words, a union of $\leq\kappa$ sets, each of size $\leq\kappa$, is of size $\leq\kappa$.

Example 5.66. $\aleph_\omega=\sum_{n\in\mathbb{N}}\aleph_n$ is *not* regular.

Remark 5.67. The existence of infinite regular cardinals which are not successors, called **weakly inaccessible cardinals**, other than $\aleph_0$, is not provable in ZFC. The reason is related to (but slightly subtler than) Exercise 5.93: for such κ, a "definable subuniverse" of V_κ called L_κ, in which GCH holds (thus "weakly inaccessible" becomes "strongly inaccessible"), would be a model of ZFC. See Theorem 5.73.

Remark 5.68. In ordinary mathematics, regular cardinals are precisely the meaningful "arity bounds" on types of operations we can equip a set with. For example:

- A group, ring, vector space, etc., has only *finitary* operations, meaning arities $< \aleph_0$.
- For a fixed group G, a G-set (set equipped with group action) has only *unary* operations, meaning arities < 2. More things are true about such structures: for example, the substructures (subsets closed under the operations) are closed under unions, as well as intersections.
- The **Borel sets** in $\mathbb{R}$ are generated from the open sets via countable Boolean operations, which have arities $< \aleph_1$.
- It does not really make sense to consider only operations of arities < 3, say, because we can compose a binary operation $*$ to get a quaternary operation $(a * b) * (c * d)$.
- Similarly, it does not make sense to consider operations of arities $< \aleph_\omega$, because we can compose a countable operation with ones of arities $\aleph_0, \aleph_1, \aleph_2, \ldots$ to get an $\aleph_\omega$-ary operation.
- For any regular cardinal κ, there are good notions of **κ-Borel set** (ones built from intervals via Boolean operations of size $< \kappa$), **κ-ary first-order logic** $\mathcal{L}_{\kappa\omega}$ (with conjunctions $\bigwedge$ and disjunctions $\bigvee$ of size $< \kappa$), etc.

We may formalize this as follows. Call a monotone set operator $T : \mathcal{P}(X) \to \mathcal{P}(X)$ **$< \kappa$-ary**[37] if

$$T(A) = \bigcup_{\substack{B \subseteq A \\ |B| < \kappa}} T(B),$$

i.e., whenever $x \in T(A)$, then $x \in T(B)$ for some $B \subseteq A$ with $|B| < \kappa$.

Exercise 5.69. Let $\kappa \geq 2$ be a cardinal, $T : \mathcal{P}(X) \to \mathcal{P}(X)$ be a $< \kappa$-ary monotone set operator.

(a) Show that if κ is regular, then for any $A \subseteq X$, the transfinite sequence from Theorem 3.167 stabilizes at $\overline{T}(A) = A_\kappa$ (cf. Proposition 3.169).
(b) Show that this fails if κ is not regular. [Consider a suitable T on $X = \kappa + 1$.]

Exercise 5.70. How badly can regularity fail? For a limit ordinal α, the **cofinality** of α is

$$\mathrm{cf}(\alpha) := \min\{\rho_{\in_A}[A] \mid A \subseteq \alpha, \sup A = \alpha\}.$$

(a) Compute $\mathrm{cf}(\omega), \mathrm{cf}(\omega + \omega), \mathrm{cf}(\omega^2), \mathrm{cf}(\omega^\omega), \mathrm{cf}(\omega_1^2), \mathrm{cf}(\omega_1^\omega)$ (ordinal exponentiation).
(b) Show that

$$\mathrm{cf}(\alpha) = \min\{|A| \mid A \subseteq \alpha, \sup A = \alpha\}.$$

In particular, $\mathrm{cf}(\alpha)$ is always an initial ordinal such that $\omega \leq \mathrm{cf}(\alpha) \leq |\alpha|$.
(c) Thus, $\mathrm{cf}(\alpha) = \alpha$ iff α is a regular cardinal.
(d) Show that $\mathrm{cf}(\mathrm{cf}(\alpha)) = \mathrm{cf}(\alpha)$, i.e., $\mathrm{cf}(\alpha)$ is always a regular cardinal.
(e) Show that for an infinite cardinal κ,

$$\mathrm{cf}(\kappa) = \min\{|I| \mid |I| \leq \kappa, \forall i \in I\, (\lambda_i < \kappa), \textstyle\sum_{i \in I} \lambda_i = \kappa\}$$

where the $\lambda_i < \kappa$ are cardinals.
(f) Show that if α is a limit ordinal, then $\mathrm{cf}(\aleph_\alpha) = \mathrm{cf}(\alpha)$.
(g) Thus if κ is an uncountable weakly inaccessible cardinal, then $\kappa = \aleph_\kappa$. The converse is false.

[37] This usually gets abbreviated to "κ-ary", leading to potential confusion since $\kappa \not< \kappa$. For example, "2-ary", i.e., < 2-ary, means nullary or unary, but *not* binary.

5.G. **Powers, products, and inaccessibility.** Regularity is also highly relevant to powers and indexed products. Recall that by Cantor's Theorem 5.33, $\kappa < 2^\kappa$ for every κ; thus $2^\kappa \geq \kappa^+$.

Continuum Hypothesis 5.71 (CH)**.** $|\mathbb{R}| = 2^{\aleph_0} = \aleph_1$.

Generalized Continuum Hypothesis 5.72 (GCH)**.** For every infinite cardinal κ, $2^\kappa = \kappa^+$.

Theorem 5.73 (Gödel)**.** If ZFC is consistent, then so is ZFC + GCH, i.e., GCH cannot be disproved.

Proof idea. Take the ZFC universe V. Roughly speaking, GCH might fail because for a set X, Comprehension says it must have certain subsets; but it may have many more than just those. Gödel constructed a subuniverse $L \subseteq V$, the **constructible universe**, via an inductive procedure analogous to the cumulative hierarchy (Section 3.K) but adding at each level only those subsets of the previous level demanded by Comprehension, i.e., which are definable by a formula $\phi(x)$. Since there are only countably many formulas, each $X \in L$ will have the least possible $2^{|X|} = |X|^+$. □

Theorem 5.74 (Easton)**.** Assume ZFC is consistent. Let F be any (proper class) function from infinite regular cardinals to infinite regular cardinals such that

(i) $\kappa \leq \lambda \implies F(\kappa) \leq F(\lambda)$;
(ii) $\kappa < F(\kappa)$.

Then it is consistent with ZFC that $2^\kappa = F(\kappa)$ for all infinite regular κ.

Example 5.75 (Cohen)**.** In particular, CH cannot be proved from ZFC: we could declare $2^{\aleph_0} := \aleph_2$.

Remark 5.76. The assumption of regularity in Easton's Theorem 5.74 may seem to come from nowhere, but it is needed. For instance, we cannot have $2^{\aleph_0} = \aleph_\omega$; see Exercise 5.82(e). The possible behaviors of cardinal exponentiation at *singular* (i.e., non-regular) cardinals is quite subtle, and is a focus of modern set theory research (see e.g., Shelah's *PCF theory*).

Easton's theorem tells us that in ZFC, expressions of the form 2^κ essentially cannot be "evaluated" into any simpler form; they behave like "indeterminates", whose values may vary from universe to universe. But given these "indeterminates", we can evaluate many other powers and products.

Proposition 5.77. If λ is an infinite cardinal and $\mu \leq \kappa \leq \mu^\lambda$, then $\kappa^\lambda = \mu^\lambda$.

Proof. $\mu^\lambda \leq \kappa^\lambda \leq (\mu^\lambda)^\lambda = \mu^{\lambda^2} = \mu^\lambda$. □

This allows us to "evaluate" κ^λ as long as κ is not too big relative to λ. Namely, if $2 \leq \kappa \leq 2^\lambda$, then $\kappa^\lambda = 2^\lambda$. More generally, we may inductively "evaluate" κ^λ in terms of μ^λ for $\mu < \kappa$, *unless* the cardinals $< \kappa$ are closed under $(-)^\lambda$. In that exceptional case, we in particular have $\kappa > 2^\lambda > \lambda$.

Proposition 5.78. If κ is regular, λ is infinite, $\kappa > \lambda$, and $\mu^\lambda \leq \kappa$ for all $\mu < \kappa$, then $\kappa^\lambda = \kappa$.

Proof. Since $\kappa > \lambda$ is regular, every function $f : \lambda \to \kappa$ must land in $\sup^+_{\alpha\in\lambda} f(\alpha) < \kappa$. Thus

$$\begin{aligned} \text{(cardinal power) } \kappa^\lambda &= \left|\textstyle\bigcup_{\alpha<\kappa} \alpha^\lambda\right| && \text{(set of functions)} \\ &\leq \textstyle\sum_{\alpha<\kappa} |\alpha|^\lambda && \text{(cardinal power)} \\ &\leq \kappa^2 = \kappa. \end{aligned}$$

□

Example 5.79. For $2 \leq \kappa \leq 2^{\aleph_0}$, we have $\kappa^{\aleph_0} = 2^{\aleph_0}$ (cf. Example 5.35). In particular, $\aleph_1^{\aleph_0} = 2^{\aleph_0}$.

If CH holds, then this is $= \aleph_1 < \aleph_2$, whence the cardinals $< \aleph_2$ are closed under $(-)^{\aleph_0}$, whence $\aleph_2^{\aleph_0} = \aleph_2$, whence similarly $\aleph_3^{\aleph_0} = \aleph_3$, etc., up to $\aleph_\omega^{\aleph_0}$ which we can't compute since $\aleph_\omega$ isn't regular.

If CH fails, then $2^{\aleph_0} \geq \aleph_2$, whence $\aleph_2^{\aleph_0} = 2^{\aleph_0}$. Similarly to the case of CH, if now $2^{\aleph_0} = \aleph_2$, then $\aleph_3^{\aleph_0} = \aleph_3$, $\aleph_4^{\aleph_0} = \aleph_4$, etc.

Exercise 5.80 (Hausdorff formula)**.** For $\kappa \geq 2$ and $\lambda \geq \aleph_0$, we have $(\kappa^+)^\lambda = \max(\kappa^\lambda, \kappa^+)$.

Exercise 5.81. The above calculations may be generalized to singular cardinals κ, taking their cofinality (Exercise 5.70) into account; however, we can no longer reduce all powers to powers of 2.

(a) Generalize Proposition 5.78 to: if $\mathrm{cf}(\kappa) > \lambda \geq \aleph_0$ and $\mu^\lambda \leq \kappa$ for all $\mu < \kappa$, then $\kappa^\lambda = \kappa$.

(b) Show that if $\mathrm{cf}(\kappa) \leq \lambda$ and $\mu^\lambda \leq \kappa$ for all $\mu < \kappa$, then $\kappa^\lambda = \kappa^{\mathrm{cf}(\kappa)} =: \beth(\kappa)$ (pronounced "**gimel** κ", the third letter of the Hebrew alphabet).

(c) Conclude that κ^λ for $\kappa \geq 2$ and $\lambda \geq \aleph_0$ may be "evaluated" by induction on κ, as follows:

$$\kappa^\lambda = \begin{cases} 2^\lambda & \text{if } \kappa \leq 2^\lambda, \\ \beth(\mu) & \text{if } 2^\lambda < \kappa \text{ and there is a least } \aleph_0 \leq \mu < \kappa \text{ such that } \kappa \leq \mu^\lambda, \\ \beth(\kappa) & \text{if } \mu^\lambda < \kappa \text{ for all } \mu < \kappa, \text{ and } \mathrm{cf}(\kappa) \leq \lambda, \\ \kappa & \text{if } \mu^\lambda < \kappa \text{ for all } \mu < \kappa, \text{ and } \mathrm{cf}(\kappa) > \lambda. \end{cases}$$

This allows us to "compute" κ^λ recursively, provided we know the values of 2^λ and $\beth(\mu)$ (and can complete the transfinite process of computing μ^λ for all $\mu < \kappa$, as well as the more "elementary" transfinite computation of $\mathrm{cf}(\kappa)$).

(d) Show that 2^λ may in turn be reduced to $\beth$ by induction on λ as follows:

$$2^\lambda = \begin{cases} \beth(\lambda) & \text{if } \lambda \text{ is regular}, \\ \sup_{\nu<\lambda} 2^\nu & \text{if } \lambda \text{ is singular and this sup is achieved}, \\ \beth(\sup_{\nu<\lambda} 2^\nu) & \text{if } \lambda \text{ is singular and this sup is not achieved}. \end{cases}$$

Thus all cardinal exponentiation may be reduced to $\beth$.

The gimel function $\beth$ may appear somewhat strange: unlike the continuum function $\lambda \mapsto 2^\lambda$, it does not correspond to any natural set operation (like powerset). Much of its study is based on the following generalization of Cantor's Theorem 2.9:

Exercise 5.82 (König's theorem)**.**

(a) Let I be a set, $(X_i)_{i\in I}, (Y_i)_{i\in I}$ be families of sets. Suppose that for each i, there is no surjection $X_i \twoheadrightarrow Y_i$. Then there is no surjection

$$\bigsqcup_{i\in I} X_i \longrightarrow\!\!\!\!\!\rightarrow \prod_{i\in I} Y_i.$$

[Copy the proof of Theorem 2.9.]

(b) Thus for families of cardinals $(\kappa_i)_{i\in I}, (\lambda_i)_{i\in I}$, if $\kappa_i < \lambda_i$ for each i, then

$$\sum_{i\in I} \kappa_i < \prod_{i\in I} \lambda_i.$$

(c) Deduce Cantor's Theorem 2.9.

(d) Deduce that $\lambda < \mathrm{cf}(\kappa^\lambda)$ for all $\kappa \geq 2$ and $\lambda \geq \aleph_0$.

(e) Conclude that $2^{\aleph_0} \neq \aleph_\omega$.

(f) Conclude that $\kappa < \beth(\kappa)$ for all $\kappa \geq \aleph_0$.

(g) Conclude that under GCH, cardinal exponentiation is much simpler:

$$\kappa^\lambda = \begin{cases} \lambda^+ & \text{if } \kappa \leq \lambda, \\ \kappa^+ & \text{if } \mathrm{cf}(\kappa) \leq \lambda < \kappa, \\ \kappa & \text{if } \mathrm{cf}(\kappa) > \lambda. \end{cases}$$

(h) Conclude that GCH is equivalent to: $\beth(\kappa) = \kappa^+$ for all $\kappa \geq \aleph_0$.

For more information on cofinality, the gimel function, GCH, etc., see [T. Jech, *Set Theory*, Ch. 5].

Concerning indexed products $\prod_{i\in I} \kappa_i$ in general, we may reduce them to powers:

Proposition 5.83. Let $(\kappa_i)_{i\in I}$ be a family of nonzero cardinals, such that for each i, we have $|\{j \in I \mid \kappa_j \geq \kappa_i\}| = |I|$. Then

$$\prod_{i\in I} \kappa_i = (\sup_{i\in I} \kappa_i)^{|I|}.$$

Proof. $\leq$ is straightforward; we show $\geq$. WLOG $I = |I|$ is an initial ordinal. If I is finite, both sides are 1 (if $I = 0$) or $\max_i \kappa_i$; so assume I is infinite. Let $p : I \times I \cong I$ be a bijection. Define $q : I \times I \hookrightarrow I$ by induction on $p(i, j)$ as follows: $q(i, j) \in I$ is least so that

$$\kappa_{q(i,j)} \geq \kappa_j, \qquad q(i, j) \neq q(i', j') \quad \forall p(i', j') < p(i, j);$$

this is possible because by assumption, there are I-many κ's which are $\geq \kappa_j$, while there are only $|p(i, j)| < I$-many (i', j')'s with $p(i', j') < p(i, j)$. Then because each $\kappa \neq 0$,

$$\begin{aligned}\prod_{i\in I} \kappa_i &\geq \prod_{i\in \mathrm{im}(q)} \kappa_i \\ &= \prod_{(i,j)\in I\times I} \kappa_{q(i,j)} \\ &\geq \prod_{i\in I} \prod_{j\in I} \kappa_j \\ &\geq \prod_{i\in I} \sup_{j\in I} \kappa_j\end{aligned}$$

(using various laws for indexed products from Exercises 5.39 and 5.40). □

Example 5.84. $\prod_{n<\omega} \aleph_n = \aleph_0 \cdot \aleph_1 \cdot \aleph_2 \cdot \dots = \aleph_\omega^{\aleph_0}$.

Example 5.85. $\overbrace{2 \cdot 2 \cdot \dots}^{\aleph_0} \cdot \aleph_\omega \cdot \aleph_\omega = 2^{\aleph_0} \cdot \aleph_\omega$ is a product of $\aleph_0$ many cardinals $\leq \aleph_\omega$, which may however be strictly less than $\aleph_\omega^{\aleph_0}$, e.g., if CH holds (see Example 5.79 and Exercise 5.82(f)).

In this example, we have a product $\prod_{i\in I} \kappa_i$ where most of the κ_i's are bounded below $\sup_i \kappa_i$; we may then reduce inductively to computing a product of smaller cardinals $\prod_{j\in J} \kappa_j$ where $\sup^+_{j\in J} \kappa_j < \sup^+_{i\in I} \kappa_i$, as well as a smaller product $\prod_{i\in I\setminus J} \kappa_i$ where $J \subseteq I$ with $|I \setminus J| < |I|$, and then multiplying these two together.

Exercise 5.86. Verify that this procedure works in general: we may reduce the computation of an arbitrary indexed product $\prod_{i\in I} \kappa_i$ to powers and binary products (i.e., max), by lexicographical induction on $(|I|, \sup^+_{i\in I} \kappa_i)$.

We now consider cardinality bounds closed under powers and/or indexed products, analogously to how regularity amounts to closure under sums/unions:

Definition 5.87. A cardinal κ is a **strong limit** if it is infinite[38] and:

- for any $\lambda < \kappa$, we have $\lambda^\lambda < \kappa$;
- equivalently, for any $\lambda, \mu < \kappa$, we have $\lambda^\mu < \kappa$;
- equivalently, for any sets A, B of size $< \kappa$, we have $|A^B| < \kappa$;
- equivalently, for any $\lambda < \kappa$, we have $2^\lambda < \kappa$;
- equivalently, for any set A of size $< \kappa$, we have $|\mathcal{P}(A)| < \kappa$.

By Cantor's theorem, these imply that κ is a limit cardinal, i.e., $0 < \kappa$ and $\lambda < \kappa \implies \lambda^+ < \kappa$.

Example 5.88. $\aleph_0$ is a strong limit.

Example 5.89. The next strong limit is $\sup\{\aleph_0, 2^{\aleph_0}, 2^{2^{\aleph_0}}, 2^{2^{2^{\aleph_0}}}, \dots\}$.

Remark 5.90. The **beth cardinals** are $\beth_0 := \omega$, $\beth_1 := 2^\omega$, $\beth_2 := 2^{2^\omega}$, $\dots$, $\beth_\alpha := \sup_{\beta<\alpha} 2^{\beth_\beta}$. Thus, for any limit ordinal α, $\beth_\alpha$ is a strong limit cardinal (and $\aleph_\alpha$ is a weak limit cardinal). The GCH can be stated as: $\aleph_\alpha = \beth_\alpha$.

[38] I suppose that in some contexts, it may also be useful to regard 2 as a strong limit.

While regularity means that sets of size $< \kappa$ are closed under $\bigcup$, strong limit means they are closed under $\mathcal{P}$; these are the two fundamental set-theoretic operations. Combining them yields

Definition 5.91. A cardinal κ is **strongly inaccessible** if it is regular and a strong limit.

Exercise 5.92. Show that κ is strongly inaccessible iff it is infinite and for any family of cardinals $(\lambda_i)_{i\in I}$ with $|I|, \lambda_i < \kappa$, we have $\prod_{i\in I} \lambda_i < \kappa$.

As the name suggests, strongly inaccessible cardinals are called such because they are so large that other than $\aleph_0$, none of them can be constructed in ZFC. This is because the "things below them" are closed under everything that ZFC requires; thus if we had a strongly inaccessible κ, we may simply truncate the universe at κ to get a smaller universe without any strongly inaccessibles. (Similarly, $\aleph_0$ could not be constructed either; we had to declare its existence via the Axiom of Infinity, without which V_ω could as well be the entire universe.)

Exercise 5.93. A set A is **hereditarily of size** $< \kappa$ if it, its elements, its elements' elements, etc., are all of size $< \kappa$. In other words, every set in the transitive closure $\overline{\bigcup}\{A\}$ (Definition 3.175) is of size $< \kappa$. (Recall Exercise 3.186.) Let H_κ be the class of all sets hereditarily of size $< \kappa$.

(a) For which $n \in \mathbb{N}$ is $H_n \subseteq V_n$?
(b) Prove that for $\kappa \geq \aleph_0$, we have $H_\kappa \subseteq V_\kappa$ iff κ is regular. In particular, H_κ is a set for all κ.
(c) Prove that for $\kappa \geq \aleph_0$, we have $H_\kappa = V_\kappa$ iff κ is strongly inaccessible.
(d) Verify that $H_\kappa = V_\kappa$ for strongly inaccessible κ has the following properties:
  (i) H_κ is transitive.
  (ii) $\varnothing \in H_\kappa$, and $a, b \in H_\kappa \implies \{a, b\} \in H_\kappa$.
  (iii) If $I \in H_\kappa$ and $(A_i)_{i\in I} \in H_\kappa^I$, then $\bigcup_{i\in I} A_i \in H_\kappa$.
  (iv) $A \in H_\kappa \implies \mathcal{P}(A) \in H_\kappa$.
  (v) $A \subseteq B \in H_\kappa \implies A \in H_\kappa$.
  (vi) For any $A \in H_\kappa$ and function $f : A \to H_\kappa$, we have $f[A] \in H_\kappa$.
  (vii) $A \in H_\kappa \implies \bigcup A \in H_\kappa$.
  (viii) $A, B \in H_\kappa \implies A \times B, B^A \in H_\kappa$.
  (ix) If $I \in H_\kappa$ and $(A_i)_{i\in I} \in H_\kappa^I$, then $\prod_{i\in I} A_i, \bigsqcup_{i\in I} A_i \in H_\kappa$.
  (x) If κ is uncountable, then $\mathbb{N} \in H_\kappa$.
(e) Conclude that H_κ satisfies $\mathsf{ZFC} - \mathsf{Infinity}$ (assuming the real universe does), and ZFC if κ is uncountable.
(f) Note that for $A \in H_\kappa$, not only do we have $\mathcal{P}(A) \in H_\kappa$, but also that H_κ thinks that $\mathcal{P}(A)$ is the powerset of A. Explain why this might not be the case if, say, we did not have (i).
(g) Verify that for any $A \in H_\kappa$, H_κ thinks that A is an ordinal iff A is indeed an ordinal.
(h) Verify that for any $A \in H_\kappa$, H_κ thinks that A is an initial ordinal iff A is indeed such.
(i) Verify that for any $A \in H_\kappa$, H_κ thinks that A is a strongly inaccessible cardinal iff A is indeed such.
(j) Conclude that $\mathsf{Infinity}$ cannot be proved from $\mathsf{ZFC} - \mathsf{Infinity}$, and that the existence of a strongly inaccessible cardinal cannot be proved from ZFC.
(k) Prove that sets U with the above properties (i)–(iv) (called **Grothendieck universes**) are precisely all H_κ for strongly inaccessible κ.

Grothendieck universes are used in areas of math that need to work with "mathematical universes" as mathematical objects, e.g., category theory (the category of all groups, etc.).