ORDINAL NUMBERS AND TRANSFINITE INDUCTION

1. Motivation

There are many examples of inductive constructions in math, where we want to build some object achieving some conditions, and we simply march through all of the conditions, achieving them one at a time:

- In linear algebra, to build a basis for a vector space V, we can start with a linearly independent set of vectors, and enlarge it to a basis by going through a list of generating vectors and adding each which is not already in the span of the linearly independent set.
- In graph theory, to find a maximal independent (i.e., without any edges between them) set of vertices in a graph G, we can start with any independent set, and go through all the vertices, adding any which is not connected to any vertex in our independent set.
- In our proof of the completeness theorem for propositional logic, we enlarged an arbitrary consistent theory to a complete and consistent theory by going through all formulas ϕ and adding whichever of ϕ or $\neg\phi$ could be added without breaking consistency.

In the simplest cases (e.g., if V is known to be finite-dimensional, or G is a finite graph), such an inductive procedure terminates after finitely many steps. In slightly more complicated cases (where the structure we're working with is "countable" in some sense), the finite steps may not terminate, but nonetheless they "converge" to the answer we want. But in the most general cases, even countably many steps may not be enough; rather, we have to keep going after infinitely many steps, with "step $\infty, \infty+1, \infty+2, \dots$". Such a procedure that can continue after infinitely many steps is called **transfinite induction**, and these generalized indices like "$\infty, \infty+1, \dots$" are called **ordinal numbers**.

2. Ordinal numbers

Formally, **ordinal numbers** are mathematical objects equipped with a (total) ordering $<$ and constructed inductively according to a single rule:

- Given any set S of ordinal numbers, there is a least ordinal number strictly greater than all elements of S, denoted $\sup^+ S$.[1]

Note that there is no need for a "base case". To get started, we can take $S = \varnothing$, which is (vacuously) a set of ordinal numbers; its $\sup^+$ is then simply the least ordinal number, usually denoted

$$0 := \sup^+ \varnothing.$$

We can then take the next ordinal that comes after 0:

$$1 := \sup^+\{0\},$$

[1]If you've seen the notion of **supremum** or **least upper bound** (say, in real analysis), the only difference is that $\sup S$ is the least number *greater than or equal to* all elements of S. Thus, we always have $\sup^+ S \geq \sup S$, assuming both exist; equality may or may not hold. For example, the open interval $(0,1)$ has supremum (in $\mathbb{R}$ with the usual ordering) 1, which also happens to be the $\sup^+$ since it is strictly greater than every element of $(0,1)$. On the other hand, the set of integers $\{0,1,2\}$, in $\mathbb{Z}$ with the usual ordering, has $\sup = 2$ but $\sup^+ = 3$. (In $\mathbb{R}$, $\{0,1,2\}$ also has supremum 2, but does not have a $\sup^+$, since there is no "next" real number after 2.)

followed by

$$2 := \sup^+\{1\},$$
$$3 := \sup^+\{2\},$$
$$\vdots$$

After all the finite ordinals, we then have the first infinite ordinal, which is usually denoted

$$\omega := \sup^+\{0, 1, 2, \dots\} = \sup^+ \mathbb{N}$$

(rather than "∞" like in the motivational discussion we gave above, since there are many different notions of "infinity" in math, and since ω is only the *smallest* infinite ordinal). We can continue:

$$\omega + 1 := \sup^+\{\omega\},$$
$$\omega + 2 := \sup^+\{\omega + 1\},$$
$$\vdots$$

eventually leading to

$$\omega + \omega := \sup^+\{\omega, \omega + 1, \omega + 2, \dots\}.$$

Now we can repeat this entire procedure infinitely many times, yielding[2]

$$\begin{aligned} \omega + \omega + \omega &:= \sup^+\{\omega + \omega, \omega + \omega + 1, \omega + \omega + 2, \dots\}, \\ \omega + \omega + \omega + \omega &:= \sup^+\{\omega + \omega + \omega, \omega + \omega + \omega + 1, \omega + \omega + \omega + 2, \dots\}, \\ &\vdots \end{aligned}$$

and eventually,

$$\omega^2 := \omega \cdot \omega := \sup^+\{\omega, \omega + \omega, \omega + \omega + \omega, \dots\}.$$

We can keep going:

$$\omega^2 + 1, \omega^2 + \omega, \omega^2 + \omega^2, \omega^2 + \omega^2 + \omega^2, \dots,$$
$$\omega^3, \omega^4, \dots, \omega^\omega, \omega^\omega + \omega^\omega, \dots, \omega^\omega \cdot \omega = \omega^{\omega+1}, \omega^{\omega+2}, \dots, \omega^{\omega+\omega}, \omega^{\omega^2}, \dots, \omega^{\omega^\omega}, \dots$$

The above inductive rule for constructing ordinals can be read as saying:

> *You can never run out of ordinals: if you think you know what all the ordinals are, there's always a next one you haven't thought of.*[3]

3. Von Neumann representation of ordinals

As with all inductively constructed objects, we can formally represent ordinals as "expressions":

$$\sup^+ S := (\sup^+, S).$$

However, the fact that ordinals are intrinsically ordered (and that $\sup^+$ is defined as the *least* ordinal $>$ every element of S) complicates things a little, since we could have two such expressions representing the same ordinal, e.g.,

$$\omega + 1 = \sup^+\{\omega\} = \sup^+\{0, 1, 2, \dots, \omega\},$$

since clearly if an ordinal is $> \omega$, then it is also $>$ everything below ω.

[2] We carefully avoid writing $2\omega, 3\omega$, etc., since according to the official definition of ordinal multiplication (see Exercise 4.5 below), these should actually be denoted $\omega \cdot 2, \omega \cdot 3$, etc.

[3] But see the Burali-Forti paradox below.

We could resolve this issue by taking a quotient of all expressions $\sup^+ S$ to identify those which "should" denote the same ordinal. However, in this case (unlike with, say, formulas) there is an easier approach: among all the expressions $\sup^+ S$ which denote an ordinal α, we can designate one as the "canonical form" of α, namely the one with the *largest* possible S, consisting of *all* ordinals $< \alpha$. How can we tell if a given expression $\sup^+ S$ is in such "canonical form"? As suggested by the example of $\omega + 1$ above, the key is that S must be **downward-closed**, meaning that whenever it contains an ordinal, it contains all smaller ordinals:

$$\gamma < \beta \in S \implies \gamma \in S.$$

We can put any $\sup^+ S$ into its "canonical form" by replacing S with its **downward-closure**

$$\{\gamma \mid \exists \beta \in S\,(\gamma \le \beta)\},$$

which clearly has the same $\sup^+$ as S. For example, the downward-closure of $\{\omega\}$ is $\{0, 1, 2, \dots, \omega\}$, yielding the second representation of $\omega + 1$ above as its "canonical form".

Finally, we note that there is no need to include the symbol "$\sup^+$" in the formal representation of the expression $\sup^+ S$, since there is only one operation for inductively forming new ordinals from old ones (unlike with, say, formulas, where we need to record the $\wedge$ in $P \wedge Q = (\wedge, P, Q)$). In other words, *we may simply represent an ordinal as the set of ordinals less than it*:

$$\alpha = \{\beta \mid \beta < \alpha\}.$$

This is called the **von Neumann representation of ordinals**. For example,[4]

$$\begin{aligned}
0 &:= \varnothing, \\
1 &:= \{0\} = \{\varnothing\}, \\
2 &:= \{0, 1\} = \{\varnothing, \{\varnothing\}\}, \\
&\ \ \vdots \\
\omega &:= \mathbb{N} = \{0, 1, 2, \dots\} = \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, \dots\}, \\
&\ \ \vdots
\end{aligned}$$

One advantage of this representation is that many "numerical" notions become identified with set-theoretic ones. For example, note how the ordering is now simply set membership:

$$\beta < \alpha \iff \beta \in \alpha.$$

Thus, our original inductive definition of ordinals becomes, simply,

- Any downward-closed set α of ordinals (i.e., $\gamma \in \beta \in \alpha \implies \gamma \in \alpha$) is itself an ordinal.

The $\sup^+$ operation for an arbitrary set of ordinals S is now given by downward-closure, which under the von Neumann representation also has a simple formula:

$$\sup^+ S = S \cup \bigcup S$$

(i.e., $\gamma < \sup^+ S$ iff either $\gamma \in S$, or else $\gamma < \beta$ for some $\beta \in S$, i.e., $\gamma \in \beta$, whence $\gamma \in \bigcup S$).

The Burali-Forti paradox. This says that the collection of *all* ordinals is too big to form a set! Indeed, if there were a set of all ordinals, call it Ord, then $\sup^+ \mathrm{Ord}$ (which under the von Neumann representation is simply Ord) would be an ordinal strictly bigger than every ordinal, including itself, which would be a contradiction. Thus, Ord is only a "meta" collection or *proper class*, a collection of mathematical objects which cannot itself exist as an object in the mathematical universe, much like the collection of all sets (by Russell's paradox).

[4]Note how ω is simply identified with the set of natural numbers $\mathbb{N}$. In fact, many set theorists prefer to always write ω in place of $\mathbb{N}$, so that e.g., an infinite (countable) sequence is denoted $(x_i)_{i \in \omega}$ instead of $(x_i)_{i \in \mathbb{N}}$.

4. Transfinite induction

As with all inductively defined objects, we have a principle of induction for ordinals:

Principle of transfinite induction. To prove a statement $\Phi(\alpha)$ for all ordinals α:

- Prove that if $\Phi(\beta)$ for all $\beta < \alpha$ (i.e., $\beta \in \alpha$), then $\Phi(\alpha)$.

Principle of transfinite inductive definition. To define a thing $f(\alpha)$ for each ordinal α:

- Given $f(\beta)$ for all $\beta < \alpha$, define $f(\alpha)$.

Note how (as in the inductive definition of ordinals) there is no special "base case": when $\alpha = 0$, the requirement is simply that we must prove $\Phi(0)$, resp., define $f(0)$. (So this is more analogous to "strong induction" for $\mathbb{N}$, which is just the special case of transfinite induction where we stop at ω.)

Example 4.1. The sum $\alpha + \beta$ of ordinals is defined by induction on β as follows:

$$\alpha + \beta := \max(\alpha, \sup^+\{\alpha + \gamma \mid \gamma < \beta\}).$$

In other words, $\alpha + \beta$ is the least ordinal which is $\geq \alpha$, and also $> \alpha + \gamma$ for all $\gamma < \beta$. For example,

$$\begin{aligned}
\alpha + 0 &= \max(\alpha, 0) = \alpha && \text{(there are no } \gamma < \beta\text{)},\\
\alpha + 1 &= \max(\alpha, \sup^+\{\alpha + 0\})\\
&= \max(\alpha, \sup^+\{\alpha\})\\
&= \sup^+\{\alpha\} && \text{(since } \alpha < \sup^+\{\alpha\}\text{)},\\
\alpha + 2 &= \max(\alpha, \sup^+\{\alpha + 0, \alpha + 1\})\\
&= \max(\alpha, \sup^+\{\alpha, \sup^+\{\alpha\}\})\\
&= \sup^+\{\sup^+\{\alpha\}\},\\
&\ \vdots\\
\alpha + \omega &= \max(\alpha, \sup^+\{\alpha + 0, \alpha + 1, \alpha + 2, \dots\})\\
&= \sup^+\{\alpha, \sup^+\{\alpha\}, \sup^+\{\sup^+\{\alpha\}\}, \dots\}.
\end{aligned}$$

Note how this agrees with the names for some particular ordinals (e.g., $\omega + 1$) we used in Section 2. Note also that the max in the above definition of $+$ is only needed in the case $\beta = 0$. Note, finally,

$$\begin{aligned}
1 + \omega &= \sup^+\{1, \sup^+\{1\}, \sup^+\{\sup^+\{1\}\}, \dots\}\\
&= \sup^+\{1, 2, 3, \dots\}\\
&= \omega \neq \omega + 1.
\end{aligned}$$

Thus, *ordinal addition (as defined above) is not commutative!*

In contrast, we can prove some other familiar laws of arithmetic by transfinite induction:

Proposition 4.2. Ordinal addition has 0 as identity element: $\alpha + 0 = \alpha = 0 + \alpha$, for all α.

Proof. The first equation, $\alpha + 0 = \alpha$, follows directly from the definition (see example above). For the second, $\alpha = 0 + \alpha$, by induction on α, we may assume that $\beta = 0 + \beta$ for all $\beta < \alpha$. But then

$$\begin{aligned}
0 + \alpha &= \max(0, \sup^+\{0 + \beta \mid \beta < \alpha\}) && \text{by definition}\\
&= \max(0, \sup^+\{\beta \mid \beta < \alpha\}) && \text{by IH}\\
&= \max(0, \alpha) = \alpha.
\end{aligned}$$

□

Exercise 4.3. Prove that ordinal addition is monotone on both sides: if $\alpha \leq \beta$, then $\alpha + \gamma \leq \beta + \gamma$ and $\gamma + \alpha \leq \gamma + \beta$. Can you replace $\leq$ with $<$ throughout?

Proposition 4.4. Ordinal addition is associative: $(\alpha+\beta)+\gamma = \alpha+(\beta+\gamma)$.

Proof. By induction on γ, we may assume that for all $\delta < \gamma$, we have $(\alpha+\beta)+\delta = \alpha+(\beta+\delta)$. If $\gamma = 0$, then both sides of $(\alpha+\beta)+\gamma = \alpha+(\beta+\gamma)$ simplify to $\alpha+\beta$; so suppose $\gamma > 0$. Then

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

To see that the last two RHS's are the same: by definition, $\beta+\gamma = \sup^+\{\beta+\delta \mid \delta < \gamma\}$; thus each element of the set labelled B above is also in the set C, whence $\sup^+ B \le \sup^+ C$. On the other hand, every $\varepsilon < \beta+\gamma$ must be in the downward-closure of $\{\beta+\delta \mid \delta < \gamma\}$, i.e., $\varepsilon \le \beta+\delta$ for some $\delta < \gamma$, whence also $\alpha+\varepsilon \le \alpha+(\beta+\delta)$ by the preceding Exercise, which shows each element of C is $\le$ some element of B, whence $\sup^+ C \le \sup^+ B$. $\square$

Even though the principle of transfinite induction is stated with a single inductive case, in practice, it is often helpful to treat the case of 0 separately as a "base case"; we did so in the preceding proof. The non-0 cases are often also usefully split into two types:

- An ordinal α is a **successor** if $\alpha = \sup^+\{\beta\}$ $(= \beta+1$, according to the above definition of $+)$ for some β. For example, $1, 2, 3, \omega+1$ are successors.
- An ordinal α which is neither 0 nor a successor is called a **limit ordinal**, e.g., ω. This means that the set of ordinals $< \alpha$ (i.e., the set α itself) is nonempty but has no greatest element; hence, α can be "approached" from below. Note that for a limit ordinal α,
$$\alpha = \sup{}^{+}\{\beta \mid \beta < \alpha\} = \sup\{\beta \mid \beta < \alpha\};$$
indeed, α is clearly an upper bound for the set of $\beta < \alpha$, and there is no smaller upper bound, since for $\beta < \alpha$ we have $\beta+1 < \alpha$ so β is not an upper bound for the ordinals $< \alpha$.

Exercise 4.5. Define the product $\alpha \cdot \beta$ of ordinals by induction on β as follows:

$$\begin{aligned}\alpha \cdot 0 &:= 0, \\ \alpha \cdot (\beta+1) &:= \alpha \cdot \beta + \alpha, \\ \alpha \cdot \beta &:= \sup\{\alpha \cdot \gamma \mid \gamma < \beta\} \quad \text{if } \beta \text{ is a limit ordinal.}\end{aligned}$$

(a) Compute: $\omega \cdot 2, 2 \cdot \omega, \omega \cdot \omega$, and compare with the examples in Section 2.
(b) Show that $0 \cdot \alpha = 0$.
(c) Show that 1 is an identity element for $\cdot$ on both sides.
(d) Show that $\cdot$ is monotone on both sides.
(e) Show that $\cdot$ distributes over $+$ on one side only.

5. The well-ordering theorem

In order to fulfill the motivating examples we gave in Section 1, we need to know that the elements of an arbitrary set can be listed via a transfinite sequence indexed by the ordinals, much as the elements of a countable set can be listed via a sequence indexed by $\mathbb{N}$. This is provided by the

Well-ordering theorem. For any set X, there is an ordinal α and a bijection

$$\begin{aligned}\alpha &\longleftrightarrow X \\ \beta &\longmapsto x_\beta.\end{aligned}$$

In other words, we have a length α enumeration $(x_\beta)_{\beta<\alpha}$ of all the elements of X without repetition.

Proof. We define a sequence $(x_\alpha)_\alpha$ indexed by all ordinals by transfinite induction as follows:

- Given x_β for all $\beta < \alpha$, if $X \setminus \{x_\beta \mid \beta < \alpha\} \neq \varnothing$, then pick any element from it and call it x_α. Otherwise, stop the construction of the sequence (or if you prefer, let x_α be some junk element, call it ☹, which is not in X, to indicate that we've stopped).

By definition, all the x_α which are defined will be distinct from each other (since we picked each x_α to be distinct from all the previous ones x_β for $\beta < \alpha$); and if the construction stops at some α (i.e., $x_\alpha =$ ☹), then that means we've hit every element of X, whence we have a bijection.

So the only thing that could go wrong is if the construction never stops, i.e., for each ordinal α, we are able to pick a new element x_α of X. Let $Y := \{x \in X \mid x = x_\alpha \text{ for some } \alpha\}$ be the subset of all elements which are picked at some point. Then since (as noted above) the $x_\alpha \in Y$ are all distinct, each $y \in Y$ will be x_α for a unique α; call it α_y. But then $\{\alpha_y \mid y \in Y\}$ is the collection of all ordinals, which is too big to be a set by the Burali-Forti paradox, and yet is small enough to be a set because it's indexed by the set Y, a contradiction.[5] □

As an application, we now fulfill one of the examples we gave in Section 1. For present purposes, a **graph** $G = (V, E)$ consists of an arbitrary vertex set V together with an arbitrary binary edge relation $E \subseteq V^2$. A set of vertices $I \subseteq V$ is **independent** if no two $v, w \in I$ are connected by an edge. A **maximal** independent set is one to which no additional vertices may be added without breaking independence; by definition of independence, this means every $v \in V \setminus I$ must be connected to some $w \in I$ by an edge (in either direction, i.e., $(v, w) \in E$ or $(w, v) \in E$).

Theorem 5.1. Every graph $G = (V, E)$ has a maximal independent set of vertices.

Proof. By the well-ordering theorem, transfinitely enumerate $V = \{v_\beta\}_{\beta<\alpha}$ for some ordinal α. We now implement the greedy algorithm described in Section 1, by marching through the v_β's. We define a transfinite sequence of subsets $I_\beta \subseteq V$ for each $\beta \leq \alpha$ by induction as follows:

$$\begin{aligned}
I_0 &:= \varnothing, \\
I_{\beta+1} &:= \begin{cases} I_\beta \cup \{v_\beta\} & \text{if } I_\beta \cup \{v_\beta\} \text{ is independent,} \\ I_\beta & \text{otherwise,} \end{cases} \\
I_\beta &:= \textstyle\bigcup_{\gamma<\beta} I_\gamma \quad \text{for a limit ordinal } \beta.
\end{aligned}$$

It is obvious that this sequence is monotone: for $\beta \leq \gamma$, we have $I_\beta \subseteq I_\gamma$ (formally, this would be a rather trivial proof by induction on γ). We now show that each I_β is independent by induction on β:

- Clearly $I_0 = \varnothing$ is independent.
- If I_β is independent, then clearly the definition of $I_{\beta+1}$ ensures it is also independent.
- Suppose β is a limit ordinal, and I_γ is independent for every $\gamma < \beta$. Then $I_\beta = \bigcup_{\gamma<\beta} I_\gamma$ is also independent, since any two vertices in it must both occur in some I_γ for $\gamma < \beta$ (namely for the max of the two γ's in which they each occur, since that I_γ contains the other one, using the monotonicity of the I_γ's noted above), and I_γ is independent by the IH.

Finally, we claim that I_α is maximal independent. This is because every $v \in V \setminus I_\alpha$ is v_β for some $\beta < \alpha$; since $v \notin I_\alpha \supseteq I_{\beta+1}$, we did not add v_β in the βth stage of the construction, which means $I_\beta \cup \{v_\beta\}$ must not be independent, whence neither is $I_\alpha \cup \{v\} \supseteq I_\beta \cup \{v_\beta\}$, i.e., we cannot add v to I_α without breaking independence. □

[5] The precise explanation of what's going on in this proof requires some knowledge of the axioms of ZFC set theory. The set Y is allowed, since it's a subset of X, and we can pick it out using the Comprehension Axiom. The set $\{\alpha_y \mid y \in Y\}$ is allowed if Y is a set, by the Replacement Axiom. Finally, note that we blatantly used the Axiom of Choice (AC) in picking x_α arbitrarily from $X \setminus \{x_\beta \mid \beta < \alpha\}$ at each stage; hence this proof is potentially non-constructive, in that we may have proved that a transfinite enumeration $(x_\alpha)_\alpha$ exists, yet be unable to actually give an example. This is indeed the case, since the well-ordering theorem is actually equivalent to AC; see Section 6.

Many results of a similar kind can be either deduced as easy consequences, or else proved in an analogous manner:

Corollary 5.2 (Hausdorff maximality principle)**.** Every partially ordered set $(P, \leq)$ has a maximal totally ordered subset $Q \subseteq P$.

Proof. Let

$$E := \{(x, y) \in P \mid x \not\leq y \text{ and } y \not\leq x\}.$$

Then a subset $Q \subseteq P$ is independent with respect to the graph (P, E) iff it is totally ordered with respect to $\leq$, so we can directly apply the preceding theorem. $\square$

Corollary 5.3 (Zorn's lemma)**.** Let $(P, \leq)$ be a partially ordered set such that every totally ordered subset has an upper bound. Then P has a maximal element.

Proof. Let $Q \subseteq P$ be a maximal totally ordered subset, and let u be an upper bound for Q. Then u is maximal, since any $x \geq u$ is also an upper bound for Q, whence $Q \cup \{x\}$ is also totally ordered, whence $x \in Q$ by maximality, whence $u \geq x$ since u is an upper bound for Q. $\square$

Exercise 5.4. An **antichain** in a partially ordered set is a set of elements, no two of which are comparable. Show that every partially ordered set has a maximal antichain.

Exercise 5.5. A **hypergraph** is a set V together with a set E of *finite* subsets of V (called **hyperedges**). A subset $I \subseteq V$ is **independent** (with respect to E) if it doesn't contain any hyperedge. Show that every hypergraph has a maximal independent set.

Exercise 5.6. Show that every vector space has a basis.

Exercise 5.7. Show that Exercise 5.5 fails if we allow hyperedges to be infinite.
[Hint: first, consider $V := \mathbb{N}$ and $E = \{\mathbb{N}\}$; think about why this doesn't work. Then, add more hyperedges to E to make it work.]

6. Foundational remarks

As we remarked in Footnote 5, the above proof of the well-ordering theorem directly uses the

Axiom of Choice. For any indexed family $(X_i)_{i \in I}$ of nonempty sets, there is a family $(x_i)_{i \in I}$ of elements such that $x_i \in X_i$ for each $i \in I$.

In fact, the following results are all equivalent to each other:[6]

(a) the axiom of choice;
(b) the well-ordering theorem;
(c) every graph has a maximal independent set;
(d) the Hausdorff maximality principle;
(e) Zorn's lemma.

(All of the other Exercises in the preceding section also turn out to be equivalent to these.) Above we proved (a) $\Longrightarrow$ (b) $\Longrightarrow$ (c) $\Longrightarrow$ (d) $\Longrightarrow$ (e); to complete the circle, we now give the

Proof of axiom of choice from Zorn's lemma. Let P be the set of all partial families $(x_j)_{j \in J}$ where $J \subseteq I$ and $x_j \in X_j$ for each $j \in J$, partially ordered by $(x_j)_{j \in J} \leq (y_k)_{k \in K}$ iff $J \subseteq K$ and $x_j = y_j$ for all $j \in J$. For a totally ordered $Q \subseteq P$, the union of all the partial families in Q is easily seen to be still a partial family; the point is that all partial families in Q must agree wherever they are defined, since Q is totally ordered. Thus by Zorn's lemma, P has a maximal $(x_j)_{j \in J}$, which must have $J = I$ or else we could add $x_i \in X_i$ for some $i \in I \setminus J$ to it. $\square$

[6]A well-known joke, intended to capture many people's intuitions about these results: "The axiom of choice is obviously true, the well-ordering principle obviously false, and who can tell about Zorn's lemma?"

You may have heard before that the axiom of choice is "nonconstructive": it asserts that something exists (namely, the family of choices $(x_i)_{i\in I}$ from each of the nonempty sets), without giving any indication of how to come up with a particular example of one such. All of the other results in the preceding section are similarly nonconstructive, since they all imply the axiom of choice. The following examples are meant to give some intuition about the nonconstructivity of these results:

Example 6.1. Consider the graph with vertex set $V := \mathbb{R}$ and $E := \{(x, y) \in \mathbb{R}^2 \mid 0 \neq x - y \in \mathbb{Q}\}$. An independent set $I \subseteq \mathbb{R}$ must consist of numbers whose pairwise differences are all irrational; for example, it can contain at most one rational (since any two would have rational difference). If I is *maximal* independent, then it must contain exactly one rational (or else we could add one), as well as exactly one number differing from $\sqrt{2}$ by a rational, or from e, or π, etc.; how do you choose whether to include $\sqrt{2}$ or $1+\sqrt{2}$? [A maximal independent set for this graph is called a **Vitali set**.]

Exercise 6.2. Consider $\mathbb{R}$ as a vector space with field of scalars $\mathbb{Q}$. Note that $\mathbb{R}$ is highly infinite-dimensional as a $\mathbb{Q}$-vector space; e.g., $\sqrt{2}$ cannot be written as a scalar multiple of 1, and $\sqrt{3}$ cannot be written as $a1 + b\sqrt{2}$ for $a, b \in \mathbb{Q}$, etc. By Exercise 5.6 (which works just as well for vector spaces over any field), $\mathbb{R}$ has a $\mathbb{Q}$-basis. Using such a $\mathbb{Q}$-basis, define a function $f : \mathbb{R} \to \mathbb{R}$ such that

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

for all $x, y \in \mathbb{R}$, yet f is not equal to multiplication by a constant.

Can you think of any example of such a function? [Hint: no.]

On the other hand, there's nothing inherently "nonconstructive" about ordinals themselves, or about transfinite induction. We just need to formalize the inductive definition of ordinals, in the von Neumann form given in Section 3, into (ZF) set theory. For inductive definitions of "smaller" mathematical objects, such as natural numbers or formulas, we can simply say that the set of all such is the *smallest* set closed under the inductive construction operations; this immediately yields the principle of induction for such objects (from which the principle of inductive definitions can then be proved). For ordinals, we would similarly like to say:

> "The collection of ordinals, Ord, is smallest such that whenever α is a set of elements of Ord which is downward-closed in that $\gamma \in \beta \in \alpha \implies \gamma \in \alpha$, then $\alpha \in \mathrm{Ord}$."

The only added wrinkle is that because Ord is not a set (by the Burali-Forti paradox), only a "meta" collection, it doesn't make formal sense to say the "smallest" collection: we would need to say "for all other such 'meta' collections, Ord is contained in it". There are two ways around this issue:

- We can assert the principle of transfinite induction *locally* below each ordinal, rather than for Ord as a whole. That is, we define an ordinal to be a set α which is downward-closed in the above sense,[7] and such that for any subset $\alpha' \subseteq \alpha$, if every $\beta \in \alpha$ with $\gamma \in \beta \implies \gamma \in \alpha'$ is in α', then $\alpha' = \alpha$. This is enough, essentially because to know that induction works for a *particular* α, we can consider induction *below* any bigger ordinal, say $\alpha + 1$.
- Alternatively, we can forget about induction and focus on downward-closure: we define an ordinal to be a set α which is downward-closed in the above sense, and each of whose elements is downward-closed, as are each of *their* elements, etc.; that is,[8]

$$\alpha_n \in \cdots \in \alpha_2 \in \alpha_1 \in \alpha \implies \alpha_n \in \alpha_{n-2}.$$

These two approaches turn out to be equivalent: we can prove that each element of each element of ... an ordinal is downward-closed using induction. On the other hand, once we know that all elements of elements of ... ordinals are ordinals, transfinite induction follows from another basic axiom of ZF set theory called the **axiom of foundation** (which essentially says you can do induction on the $\in$ relation between arbitrary sets).

[7] The technical term is a **transitive** set.

[8] In fact it turns out to be enough to consider $n = 2, 3$; that is, an ordinal is a transitive set of transitive sets.