# Transfinite induction

## A brief introduction

Harald Hanche-Olsen

hanche@math.ntnu.no

**Abstract.**
Transfinite induction is like ordinary induction, only more so. The salient feature of transfinite induction is that it works by not only moving beyond the natural numbers, but even works in uncountable settings.

## Wellordering

A *partial order* on a set $S$ is a binary relation, typically written as $<$ or $\prec$ or some similar looking symbol (let us pick $\prec$ for this definition), which is *transitive* in the sense that, if $x \prec y$ and $y \prec z$, then $x \prec z$, and *antireflexive* in the sense that $x \prec x$ *never* holds. The order is called *total* if, for every $x, y \in S$, either $x \prec y$, $x = y$, or $y \prec x$. We write $x \succ y$ if $y \prec x$.

Furthermore, we write $x \preccurlyeq y$ if $x \prec y$ or $x = y$. When $\prec$ is a partial order then $\preccurlyeq$ is also a transitive relation. Furthermore, $\preccurlyeq$ is *reflexive*, i.e., $x \preccurlyeq x$ always holds, and if $x \preccurlyeq y$ and $y \preccurlyeq x$ both hold, then $x = y$. If a relation $\preccurlyeq$ satisfies these three conditions (transitivity, reflexivity, and the final condition) then we can define $\prec$ by saying $x \prec y$ if and only if $x \preccurlyeq y$ and $x \neq y$. This relation is then a partial order (exercise: prove this). We will call a relation of the form $\preccurlyeq$ a partial order, too. We rely on the shape of the symbol used (whether it includes something vaguely looking like an equals sign) to tell us which kind is meant.

Obvious examples of total orders are the usual orders on the real numbers, written $<$ or $\leq$.

A much less obvious example is the *lexicographical* order on the set $\mathbb{R}^{\mathbb{N}}$ of sequences $x = (x_1, x_2, \ldots)$ of real numbers: $x < y$ if and only if, for some $n$, $x_i = y_i$ when $i < n$ while $x_n < y_n$. Exercise: Show that this defines a total order.

An example of a partially ordered set is the set of all real functions on the real line, ordered by $f \leq g$ if and only if $f(x) \leq g(x)$ for all $x$. This set is not totally ordered. For example, the functions $x \mapsto x^2$ and $x \mapsto 1$ are not comparable in this order.

Another example is the set $\mathcal{P}(S)$ of subsets of a given set $S$, partially ordered by inclusion $\subset$. This order is not total if $S$ has at least two elements.

A *wellorder* on a set $S$ is a total order $\prec$ so that every nonempty subset $A \subseteq S$ has a smallest element. That is, there is some $m \in A$ so that $m \preccurlyeq a$ for every $a \in A$.





One example of a wellordered set is the set of natural numbers $\{1, 2, \ldots\}$ with the usual order.

Morover, every subset of a wellordered set is wellordered in the inherited order.

**1 Proposition. (Principle of induction)** *Let $S$ be a wellordered set, and $A \subseteq S$. Assume that, for every $x \in S$, if $y \in A$ for every $y \prec x$, then $x \in A$. Then $A = S$.*

**Proof:** Let $B = S \setminus A$. If $A \neq S$ then $B \neq \varnothing$. Let $x$ be the smallest element of $B$. But then, whenever $y \prec x$ then $y \in A$. It follows from the assumption that $x \in A$. This is a contradiction which completes the proof. ∎

An *initial segment* of a partially ordered set $S$ is a subset $A \subseteq S$ so that, if $a \in A$ and $x \prec a$, then $x \in A$. Two obvious examples are $\{x \in S : x \prec m\}$ and $\{x \in S : x \preccurlyeq m\}$ where $m \in S$. An initial segment is called *proper* if it is not all of $S$.

*Exercise*: Show that every initial segment of a wellordered set $S$ is either of the form $\{x \in S : x \prec m\}$, where $m \in S$, or all of $S$.

A map $f : S \to T$ between partially ordered sets $S$ and $T$ is called *order preserving* if $x \prec y$ implies $f(x) \prec f(y)$. It is called an *order isomorphism* if it has an inverse, and both $f$ and $f^{-1}$ are order preserving. Two partially ordered sets are called *order isomorphic* if there exists an order isomorphism between them.

Usually, there can be many order isomorphisms between order isomorphic sets. However, this is not so for wellordered sets:

**2 Proposition.** *There can be only one order isomorphism between two wellordered sets.*

**Proof:** Let $S$ and $T$ be wellordered, and $f$, $g$ two order isomorphisms from $S$ to $T$. We shall prove by induction that $f(x) = g(x)$ for all $x \in S$. We do this by applying Proposition 1 to the set of all $x \in S$ for which $f(x) = g(x)$.

Assume, therefore, that $x \in S$ and that $f(y) = g(y)$ for all $y \prec x$.

Let $t$ be the smallest element of $T$ which is greater than $f(y)$ for all $y \prec x$. It must exist, for $f(x)$ is one such $t$, and $T$ is wellordered. In particular, $f(x) \succcurlyeq t$.

We shall prove that $f(x) = t$. Since this must hold equally well for $g$, it will follow that $f(x) = g(x)$, and this will finish the proof.

But we cannot have $f(x) \succ t$, for then $t = f(y)$ for some $y \in S$ with $x \succ y$, because $f$ is an order isomorphism. But this would contradict the definition of $t$.

Thus $f(x) = t$ as claimed, and the proof is complete. ∎

**3 Corollary.** *If $S$ and $T$ are wellordered and $f: S \to T$ is an order isomorphism of $S$ to an initial segment of $T$, then for each $s \in S$, $f(s)$ is the smallest $t \in T$ greater than every $f(x)$ where $x \prec s$.*

**Proof:** Examine the proof of the previous proposition. ∎

**4 Proposition.** *Given two wellordered sets, one of them is order isomorphic to an initial segment of the other (which may be all of the other set).*

**Proof:** Let $S$ and $T$ be wellordered sets, and assume that $T$ is not order isomorphic to any initial segment of $S$. We shall prove that $S$ is order isomorphic to an initial segment of $T$.

Let $W$ be the set of $w \in S$ so that $\{y \in S : y \preccurlyeq w\}$ is order isomorphic to an initial segment of $T$.

Clearly, $W$ is an initial segment of $S$. In fact, if $w_1 \in S$ and $w_2 \in W$ with $w_1 \prec w_2$ and we restrict the order isomorphism of $\{y \in S : y \preccurlyeq w_2\}$ to the set $\{y \in S : y \preccurlyeq w_1\}$, we obtain an order isomorphism of the latter set to an initial segment of $T$. By using Corollary 3, we conclude that the union of all these mappings is an order isomorphism $f$ of $W$ to an initial segment of $T$. Since $T$ not order isomorphic to an initial segment of $S$, $f[W] \neq T$.

Assume that $W \neq S$. Let $m$ be the smallest element of $S \setminus W$. Extend $f$ by letting $f(m)$ be the smallest element of $T \setminus f[W]$. Then the extended map is an order isomorphism, so that $m \in W$. This is a contradiction

Hence $W = S$, and the proof is complete. ∎

It should be noted that if $S$ and $T$ are wellordered and each is order isomorphic to an initial segment of the other, then $S$ and $T$ are in fact order isomorphic. For otherwise, $S$ is order isomorphic to a proper initial segment of itself, and that is impossible (the isomorphism would have to be the identity mapping).

**5 Theorem. (The wellordering principle)** *Every set can be wellordered.*

**Proof:** The proof relies heavily on the axiom of choice. Let $S$ be any set, and pick a "complementary" choice function $c: \mathcal{P}(S) \setminus \{S\} \to S$.

More precisely, $\mathcal{P}(S)$ is the set of all subsets of $S$, and so $c$ is to be defined on all subsets of $S$ with *nonempty complement*. We require that $c(A) \in S \setminus A$ for each $A$. This is why we call it a complementary choice function: It chooses an element of each nonempty complement for subsets of $S$.

We consider subsets $G$ of $S$. If $G$ is provided with a wellorder, then $G$ (and the wellorder on it) is called *good* if

$$c(\{x \in G : x \prec g\}) = g$$

for all $g \in G$.





The idea is simple, if its execution is less so: In a good wellorder, the smallest element must be $x_0 = c(\varnothing)$. Then the next smallest must be $x_1 = c(\{x_0\})$, then comes $x_2 = c(\{x_0, x_1\})$, and so forth. We now turn to the formal proof.

If $G_1$ and $G_2$ are good subsets (with good wellorders $\prec_1$ and $\prec_2$) then one of these sets is order isomorphic to an initial segment of the other. It is easily proved by induction that the order isomorphism must be the identity map. Thus, one of the two sets is contained in the other, and in fact one of them *is* an initial segment of the other. Let $G$ be the union of *all* good subsets of $S$. Then $G$ is itself a good subset, with an order defined by extending the order on all good subsets. In other words, $G$ is the largest good subset of $S$.

Assume $G \neq S$. Then let $G' = G \cup \{c(G)\}$, and extend the order on $G$ to $G'$ by making $c(G)$ greater than all elements of $G$. This is a wellorder, and makes $G'$ good. This contradicts the construction of $G$ is the greatest good subset of $S$, and proves therefore that $G = S$. ∎

**Ordinal numbers.** We can define the natural numbers (including 0) in terms of sets, by picking one set of $n$ elements to stand for each natural number $n$. This implies of course

$$0 = \varnothing,$$

so that will be our starting point. But how to define 1? There is one obvious item to use as the element of 1, so we define

$$1 = \{0\}.$$

Now, the continuation becomes obvious:

$$2 = \{0, 1\}, \quad 3 = \{0, 1, 2\}, \ldots$$

In general, given a number $n$, we let its *successor* be

$$n^+ = n \cup \{n\}.$$

We define $n$ to be an *ordinal number* if every element of $n$ is in fact also a *subset* of $n$, and the relation $\in$ wellorders $n$.

Obviously, 0 is an ordinal number. Perhaps less obviously, if $n$ is an ordinal number then so is $n^+$. Any element of an ordinal number is itself an ordinal number, and each element is in fact the set of all smaller elements.

On the other hand, you may verify that, e.g., $\{0, 1, 2, 4\}$ is *not* an ordinal number, for though it is wellordered by $\in$, 4 is not a subset of the given set.

If $m$ and $n$ are ordinal numbers, then either $m = n$, $m \in n$, or $n \in m$. For one of them is order isomorphic to an initial segment of the other, and an induction proof shows that this order isomorphism must be the identity map.

For the proof of our next result, we are going to need the concept of *definition by induction*. This means to define a function $f$ on a well-ordered set $S$ by defining $f(x)$ in terms of the values $f(z)$ for $z \prec x$. This works by letting $A$ be the subset of $S$ consisting of those $a \in S$ for which there exists a unique function on $\{x \in S\colon x \preccurlyeq a\}$ satisfying the definition for all $x \preccurlyeq a$, and then using transfinite induction to show that $A = S$. In the end we have a collection of functions, each defined on an initial segment of $S$, all of which extend each other. The union of all these functions is the desired function. We skip the details here.

**6 Proposition.** *Every wellordered set is order isomorphic to a unique ordinal number.*

**Proof:** The uniqueness part follows from the previous paragraph. We show existence.

Let $S$ be wellordered. Define by induction

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

In particular, this means that $f(m) = \varnothing = 0$ where $m$ is the smallest element of $S$. The second smalles element of $S$ is mapped to $\{f(m)\} = \{0\} = 1$, the next one after that to $\{0, 1\} = 2$, etc.

Let

$$n = \{f(s)\colon s \in S\}.$$

Then every element of $n$ is a subset of $n$. Also $n$ is ordered by $\in$, and $f$ is an order isomorphism. Since $S$ is wellordered, then so is $n$, so $n$ is an ordinal number. ∎

An ordinal number which is not 0, and is not the successor $n^+$ of another ordinal number $n$, is called a *limit ordinal*.

We call an ordinal number *finite* if neither it nor any of its members is a limit ordinal. Clearly, 0 is finite, and the successor of any finite ordinal is finite. Let $\omega$ be the set of all finite ordinals. Then $\omega$ is itself an ordinal number. Intuitively, $\omega = \{0, 1, 2, 3, \ldots\}$. $\omega$ is a limit ordinal, and is in fact the smallest limit ordinal.

There exist uncountable ordinals too; just wellorder any uncountable set, and pick an order isomorphic ordinal number. There is a smallest uncountable ordinal, which is called $\Omega$. It is the set of all countable ordinals, and is a rich source of counterexamples in topology.

Arithmetic for ordinals can be tricky. If $m$ and $n$ are ordinals, let $A$ and $B$ be wellordered sets order isomorphic to $m$ and $n$, with $A \cap B = \varnothing$. Order $A \cup B$ by placing all elements of $B$ after those of $A$. Then $m + n$ is the ordinal number order





isomorphic to $A \cup B$ ordered in this way. You may verify that $0 + n = n + 0 = n$ and $n^+ = n + 1$. However, addition is not commutative on infinite ordinals: In fact $1 + n = n$ whenever $n$ is an infinite ordinal. (This is most easily verified for $n = \omega$.) You may also define $mn$ by ordering the cross product $m \times n$ lexicographically. Or rather, the convention calls for reverse lexicographic order, in which $(a, b) < (c, d)$ means either $b < d$ or $b = d$ and $a < c$. For example, $0n = n0 = 0$ and $1n = n1 = n$, but $\omega 2 = \omega + \omega$ while $2\omega = \omega$:

$$\omega \times 2 \text{ is ordered } (0,0), (1,0), (2,0), \ldots, (0,1), (1,1), (2,1), \ldots,$$
$$2 \times \omega \text{ is ordered } (0,0), (1,0), (0,1), (1,1), (0,2), (1,2), \ldots$$

## Zorn's lemma and the Hausdorff maximality principle

As powerful as the wellordering principle may be, perhaps the most useful method for doing transfinite induction is by Zorn's lemma. We need some definitions.

A *chain* in a partially ordered set is a subset which is totally ordered. A partially ordered set is called *inductively ordered* if, for every chain, there is an element which is greater than or equal to any element of the chain. An element of a partially ordered set is called *maximal* if there is no element of the set greater than the given element.

**7 Theorem. (Zorn's lemma)** *Every inductively ordered set contains a maximal element.*

**Proof:** Denote the given, inductive, order on the given set $S$ by $\prec$. Let $<$ be a wellorder of $S$.

We shall define a chain (whenever we say *chain* in this proof, we mean a chain with respect to $\prec$) on $S$ by using induction on $S$. First, we include the $<$-smallest element $s_0$ of $S$ in our chain. Next, we include the second smallest element $s_1$ if $s_1 \succ s_0$. We continue in the same way, always including the next element $s$ if $s \succ x$ for every $x$ included so far.

To put this on a firm footing, let us say that $s$ *dominates* a subset $X \subset S$ if $s \succcurlyeq x$ for each $x \in X$. We shall define $f(s)$ so that it becomes a chain built from a subset of $\{x\colon x \le s\}$ for each $s$, and so that $f(s) \subseteq f(t)$ whenever $s < t$. The induction becomes a bit complicated to account for the case when $s$ has no predecessor in the $<$ order; the auxiliary function $F$ will take care of that little problem.

More precisely, define $f\colon S \to \mathcal{P}(S)$ by induction as follows. When $s \in S$, let $F(s) = \bigcup_{x<s} f(x)$. Assuming the induction hypothesis that $f(x)$ is a chain for each $x < s$ and also $f(x) \subseteq f(y)$ whenever $x < y < s$, then $F(s)$ is also a chain. Let $f(s) = F(s) \cup \{s\}$ if $s$ dominates $F(s)$, $f(s) = F(s)$ otherwise. By induction, $f(x)$ is a chain for every $x$, and $f(x) \subseteq f(y)$ whenever $x < y$.

Let $C = \bigcup_{s \in S} f(s)$. Then $C$ is a chain with respect to $\prec$. Since $S$ is inductively ordered, there is an element $s \in S$ so that $s$ dominates $C$.

We claim that $s$ is maximal. For if not, then there is some $t \succ s$. Then, in particular, $t$ dominates $F(t)$. But then $t \in f(t)$ by the definition of $f$, so $t \in C$. This contradicts the fact that $t \succ s$ and $s \nsucceq x$ for every $x \in C$. ∎

The next theorem is very similar to Zorn's lemma. Some people seem to prefer one, some the other. They can usually be used interchangably.

**8 Theorem. (Hausdorff's maximality principle)** *Any partially ordered set contains a maximal chain.*

**Proof:** Let $S$ be a partially ordered set, and let $\mathcal{C}$ be the set of chains in $S$, ordered by inclusion. Then $\mathcal{C}$ is inductively ordered (exercise: prove this), so it has a maximal element by Zorn's lemma. ∎

*Exercise*: Use Hausdorff's maximality principle to prove Zorn's lemma.

The wellordering principle, Zorn's lemma, and Hausdorff's maximality lemma are all equivalent to the Axiom of choice. To see this, in light of all we have done so far, we only need to prove the Axiom of choice from Zorn's lemma.

To this end, let a set $S$ be given, and let a function $f$ be defined on $S$, so that $f(s)$ is a nonempty set for each $s \in S$. We define a partial choice function to be a function $c$ defined on a set $D_c \subseteq S$, so that $c(x) \in f(x)$ for each $x \in D_c$. We create a partial order on the set $\mathcal{C}$ of such choice function by saying $c \preccurlyeq c'$ if $D_c \subseteq D_{c'}$ and $c'$ extends $c$. It is not hard to show that $\mathcal{C}$ is inductively ordered. Thus it contains a maximal element $c$, by Zorn's lemma. If $c$ is not defined on all of $S$, we can extend $c$ by picking some $s \in S \setminus D_c$, some $t \in f(s)$, and letting $c'(s) = t$, $c'(x) = c(x)$ whenever $x \in D_c$. This contradicts the maximality of $c$. Hence $D_c = S$, and we have proved the Axiom of choice.

We end this note with an application of Zorn's lemma. A *filter* on a set $X$ is a set $\mathcal{F}$ of subsets of $X$ so that

- $\varnothing \notin \mathcal{F}$, $X \in \mathcal{F}$,
- $A \cap B \in \mathcal{F}$ whenever $A \in \mathcal{F}$ and $B \in \mathcal{F}$,
- $B \in \mathcal{F}$ whenever $A \in \mathcal{F}$ and $A \subseteq B \subseteq X$.

A filter $\mathcal{F}_1$ is called *finer* than another filter $\mathcal{F}_2$ if $\mathcal{F}_1 \supseteq \mathcal{F}_2$. An *ultrafilter* is a filter $\mathcal{U}$ so that no other filter is finer than $\mathcal{U}$.

**9 Proposition.** *For every filter there exists at least one finer ultrafilter.*





**Proof:** The whole point is to prove that the set of all filters on $X$ is inductively ordered by inclusion $\subseteq$. Take a chain $\mathcal{C}$ of filters, that is a set of filters totally ordered by inclusion. Let $\mathcal{F} = \bigcup \mathcal{C}$ be the union of all these filters. We show the second of the filter properties for $\mathcal{F}$, leaving the other two as an exercise.

So assume $A \in \mathcal{F}$ and $B \in \mathcal{F}$. By definition of the union, $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$ where $\mathcal{F}_1, \mathcal{F}_2 \in \mathcal{C}$. But since $\mathcal{C}$ is a chain, we either have $\mathcal{F}_1 \subseteq \mathcal{F}_2$ or vice versa. In the former case, both $A \in \mathcal{F}_2$ and $B \in \mathcal{F}_2$. Since $\mathcal{F}_2$ is a filter, $A \cap B \in \mathcal{F}_2$. Thus $A \cap B \in \mathcal{F}$. ∎

Ultrafilters can be quite strange. There are some obvious ones: If $x \in X$, $\{A \subseteq X : x \in A\}$ is an ultrafilter. Any ultrafilter that is not of this kind, is called *free*. It can be proved that no explicit example of a free ultrafilter can be given, since there are models for set theory without the axiom of choice in which no free ultrafilters exist. Yet, if the axiom of choice is taken for granted, there must exist free ultrafilters: On the set $\mathbb{N}$ of natural numbers, one can construct a filter $\mathcal{F}$ consisting of precisely the *cofinite* subsets of $\mathbb{N}$, i.e., the sets with a finite complement. Any ultrafilter finer than this must be free.

Let $\mathcal{U}$ be a free ultrafilter on $\mathbb{N}$. Then

$$\Bigl\{ \sum_{k \in A} 2^{-k} : A \in \mathcal{U} \Bigr\}$$

is a non-measurable subset of $[0, 1]$.

The existence of maximal ideals of a ring is proved in essentially the same way as the existence of ultrafilters. In fact, the existence of ultrafilters is a special case of the existence of maximal ideals: The set $\mathcal{P}(X)$ of subsets of $X$ is a ring with addition being symmetric difference and multiplication being intersection of subsets. If $\mathcal{F}$ is a filter, then $\{X \setminus A : A \in \mathcal{F}\}$ is an ideal, and similarly the set of complements of sets in an ideal form a filter.

Finally we should mention that the axiom of choice has many unexpected consequences, the most famous being the Banach–Tarski paradox: One can divide a sphere into a finite number of pieces, move the pieces around, and assemble them into two similar spheres.

## Further reading

A bit of *axiomatic set theory* is really needed to give these results a firm footing. A quite readable account can be found on the Wikepedia:
`http://en.wikipedia.org/wiki/Axiomatic_set_theory`

## Acknowledgement

Xavier Raynaud found a couple of mistakes in an earlier version of this note. His questions also helped me discover some points where the exposition was much too brief.