

Cantor's Theorem

This DOCUMENT *was WRITTEN* in L^AT_EX 2_ε, so it **must** be L^AT_EX

intvl @subsheaf

August 2021

Cantor's diagonal argument illustrates how one may encounter infinite quantities that are larger than other infinite quantities. The usual presentation shows how one can define a set with strictly “more elements” than the set of natural numbers $\mathbb{N} = \{1, 2, 3, 4, \dots\}$ and it goes as follows.

We shall look at the collection of infinite sequences of zeros and ones, with an example of such sequence being represented by, for example,

00010111101010001010...

Suppose we attempted to list all of these sequences, from the first one, the second one, the third one and so on, allowing the list to grow infinitely, in a way that this enumeration of the sequences represents a correspondence between the natural numbers \mathbb{N} and each element of this collection. A possible attempt could be the one partially represented in Table 1.

Position in the list	Sequence						
1	0	0	1	0	1	1	...
2	0	1	0	0	0	1	...
3	1	1	1	0	1	0	...
4	0	0	1	0	0	0	...
5	1	0	1	0	0	1	...
⋮	⋮						
New Sequence	1	0	0	1	1	...	

Table 1: The diagonal argument.

It turns out that such enumeration cannot cover all the possible infinite sequences of zeros and ones. One reason for this is that, no matter how we construct the list, we may take the first number of the first sequence—and invert it, making it a one if the original is a zero and vice-versa—then take the second number of the second list and invert it again, then take the third number of the third list and invert it, and so on. The resulting sequence of numbers, which is represented in the final row of Table 1, would be guaranteed to be different from the first sequence at the first number, and different from the second sequence at the second number, from the third at the third number, and so on. In other words, the new sequence differs from the n -th element of the list at the n -th number, so that this new sequence is different from all the sequences present in the enumeration—it is therefore absent from there. This can be done no matter how we construct the list: we can always find an element that does not belong to it. The conclusion is that, in some sense, the collection of infinite sequences of zeros and ones is larger than the collection of natural numbers. There are infinities larger than the infinity of natural numbers.

This is the typical pop-sci presentation of Cantor's diagonal argument. However, the argument, with almost no changes, can show much stronger results. For this, however, we must define some set-theoretical machinery. We start with this basic classification of functions.

Definition 1. Let X and Y be sets and $f: X \rightarrow Y$ a function between them.

1. When, for all $a, b \in X$ with $a \neq b$ we have $f(a) \neq f(b)$, f is said to be *injective*.
2. When, for all $y \in Y$ there exists $x \in X$ with $f(x) = y$, f is said to be *surjective*.
3. When f is both injective and surjective, it is said to be *bijective*.

Cantor idea on how to quantify infinities is with the notion of equipotent sets, or, with more modern terminology, the concept of cardinality.

Definition 2. Let X and Y be two sets.

1. If *there exists* an injective function $f: X \rightarrow Y$, then we say that the cardinality of X is not greater than that of Y , denoted as $|X| \leq |Y|$.
2. If *there does not exist* a surjective function $f: X \rightarrow Y$, then we say that the cardinality of X is smaller than that of Y , denoted as $|X| < |Y|$.
3. If *there exists* a bijective function $f: X \rightarrow Y$, then we say that X and Y have the same cardinality, or $|X| = |Y|$.

As an example, let $\mathcal{P}(X)$ be the collection of all subsets of X . Let, also, Y be the collection of all functions $f: X \rightarrow \{0, 1\}$. We claim that $|\mathcal{P}(X)| = |Y|$. In order to prove this, it suffices to construct a bijection $\phi: \mathcal{P}(X) \rightarrow Y$.

We construct ϕ as follows. For each $A \subset X$, there is a function $f_A: X \rightarrow \{0, 1\}$ that is defined as

$$f(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases}.$$

For each A , this function is well-defined, since we know to compute $f_A(x)$ uniquely for all $x \in X$. So we construct ϕ by the rule $\phi(A) = f_A$.

1. ϕ is injective, for, if $A \subset X$ and $B \subset X$ are such that $A \neq B$, then there exists $x \in A$ such that $x \notin B$, therefore $f_A(x) = 1$ and $f_B(x) = 0$, thus $\phi(A) = f_A \neq f_B = \phi(B)$.
2. ϕ is surjective, for, let $f: X \rightarrow \{0, 1\}$. Let A be the set of $x \in X$ such that $f(x) = 1$. It follows, by construction, that $f(x) = 0$ for all $x \notin A$, therefore $f = f_A = \phi(A)$ —in other words, given arbitrary $f \in Y$, there exists $A \subset X$ such that $\phi(A) = f$.
3. Therefore, ϕ is a bijection and we conclude that $|\mathcal{P}(X)| = |Y|$.¹

This example is not random. These functions $f: X \rightarrow \{0, 1\}$ are called *indicator functions* and define every subset of X , via $f = f_A$, for each $A \subset X$. These functions are 1 when $x \in A$ and zero otherwise—hence the name indicator, as they indicate when an element belongs to A . Now, suppose $X = \mathbb{N}$. The functions $f: \mathbb{N} \rightarrow \{0, 1\}$ are functions that have the form, for example,

$$\begin{aligned} f(1) &= 0 \\ f(2) &= 0 \\ f(3) &= 1 \\ f(4) &= 0 \\ f(5) &= 1 \\ f(6) &= 1 \\ &\vdots \end{aligned}$$

that is, each f can be defined and represented by a specific infinite sequence of zeros and ones, that is, Table 1 tries to compare the size of \mathbb{N} to the size of the set of indicator functions Y , which,

¹Side note: since the von Neumann construction represents the set $\{0, 1\}$ via the symbol 2, and the collection of functions $f: X \rightarrow Y$ is also commonly represented as $f \in Y^X$, this example justifies the commonly used notation for $\mathcal{P}(X)$ as 2^X .

via the previous example, is the size of the collection of subsets of \mathbb{N} . This means that Cantor's diagonal argument proves that $\mathcal{P}(\mathbb{N})$ is “strictly larger” than \mathbb{N} itself. But not only this. The same argument given the next result.

Theorem 1 (Cantor's Theorem). *Let X be a set. Then it follows—always—that $|X| < |\mathcal{P}(X)|$.*

Proof. Let $\phi: X \rightarrow \mathcal{P}(X)$ be some function. Via the previous example, we can identify each element $A \in \mathcal{P}(X)$ to an indicator function f_A . Let us define, via ϕ , a function $g: X \rightarrow \{0, 1\}$ in the way that

$$g(x) = \begin{cases} 1, & \text{if } f_{\phi(x)} = 0 \\ 0, & \text{if } f_{\phi(x)} = 1 \end{cases}.$$

This function is also an indicator function, so $g = f_B$ for some $B \subset X$. However, for each $x \in X$, we have $f_{\phi(x)} \neq g(x) = f_B$, which means that $\phi(x) \neq B$, for all $x \in X$. In other words, we found B such that there exists no $x \in X$ such that $\phi(x) = B$, therefore, ϕ cannot be surjective. Since this argument applies for any $\phi: X \rightarrow \mathcal{P}(X)$, it follows that there exist no surjective function from X to $\mathcal{P}(X)$, hence $|X| < |\mathcal{P}(X)|$. \square

This is fundamentally the same argument as in Table 1. For example, ϕ corresponds to the attempt of listing all elements. The diagonal of the sequence is represented by the functions $f_{\phi(x)}$. Using the same simple argument, we have shown that, if X is any infinite set, we can find a set which represents a larger infinity—namely, the set of its subsets $\mathcal{P}(X)$. In particular, there is no bound on how large infinities get. For example, if we reach the infinity represented by $\mathcal{P}(X)$, then we can find a larger infinity via $\mathcal{P}(\mathcal{P}(X))$.