Math 171 WIM 1:
To Infinity... And Beyond!

Lucas Garron

December 18, 2011

Is there something bigger than infinity?

If you have ever thought about the concept of "infinity", the preceding question might seem natural to ask. Perhaps the next inquiry might be "Does that even make sense?" or "Can infinity even exist?" Although they might seem like childish questions, the math behind answering these questions is rather subtle, and it was not until about a century ago that mathematicians developed a good system to describe concepts related to infinity. A modern mathematician might answer these questions like this:

- In short: Yes, there is something bigger than "infinity".
- No, the question doesn't quite make sense to ask, because there isn't a single "infinity". However, we can define infinite things that exist about as much as the number 5 does.

The most important thing to note is there is not a single value of "infinity". Even though mathematicians use it as a shortcut to mean "more than anything you can count to", infinity is not really a number. Instead, it is more accurate to say that things *can be infinite*. (You might express this by saying that the concept of infinity is actually an adjective, not a noun.) A better way to answer our questions might be:

- We can define things that are infinite.
- Indeed, some infinite things are bigger than others. (In fact, if you name any infinite thing, there is *always* a bigger one.)

Once we have these things, it can be instructive to think of different infinities existing, in the sense that each infinity is the size of something that is infinite.

Infinite Sets

A set is simply a collection of things. You are probably familiar with the set containing the numbers you can use to count things: $\{0, 1, 2, 3, 4, ...\}$. This is the set of natural numbers, called $\mathbb{N}$. It seems clear that $\mathbb{N}$ should have an infinite number of elements; if we try to count the elements of $\mathbb{N}$ in order, we always have more to go, no matter how far we have counted.

First of all: if we can't write it all down, why should we be content in believing that $\mathbb{N}$ exists? There are a few reasons why $\mathbb{N}$ should be a set that exists, but it turns out simply to be useful to assume that it exists and that we can call it by a name (see Appendix A for a more thorough explanation of set theory). It turns out to cause no problems to think of $\mathbb{N}$ as existing, and once we've defined it and start using it, $\mathbb{N}$ is just as useful as the number 5 in helping us use mathematics to describe things. This answers one of our questions: Yes, there things that are infinite, such as the set $\mathbb{N}$.

If we tried to describe the number of elements in $\mathbb{N}$, it could not be any finite number. If we take this as a defining property of infinite sets, there are also other infinite sets. A simple example might be the even natural numbers: $2\mathbb{N} = \{0, 2, 4, 6, ...\}$. Although we could never finish, we can count up to any particular element of $\mathbb{N}$ (or of $2\mathbb{N}$), and for this reason sets like this are called *countable* infinite sets. There is also a special name for the number of elements in a countable set, and it is called "aleph-nought" and written as $\aleph_0$. Mathematicians consider $\aleph_0$ an infinite kind of number, and it is perhaps the closest number we have to the most common meaning of "infinity" (more on that later).

So, there are different infinite sets. Are they all countable? The surprising answer is that no, not all infinite sets are countable. A good example of this is a set you're probably familiar with: the set of real numbers, called $\mathbb{R}$. This set contains all the numbers on the real number line, such as 0 and -2.5 and $\frac{1}{3}$ and π and 10^{100}. Although we have to be careful to assume the real numbers can be collected into a set (see Appendix A), there are mathematically sound ways to do this, and $\mathbb{R}$ is a set.

If we try simple approaches to counting up the elements of $\mathbb{R}$, we run into issues: $\{0, 1, 2, ...\}$ skips all the numbers between 0 and 1, but if we spend infinitely long counting up the numbers from 0 to 1, we might not get anywhere else. If we try a few approaches, it seems that there is no way to count the elements of $\mathbb{R}$. But how could we know for sure that there is no way to do this, no matter how clever we are?

Cantor's Diagonalization Proof

The famous way to show that $\mathbb{R}$ is uncountable this is called a *diagonalization proof*, named for *Georg Cantor*, who introduced it exactly for exactly this reason. It proves that there is a set that is infinite but "even larger" than $\mathbb{N}$, and therefore that there are different "sizes of infinity".

Cantor's proof uses a technique called *proof by contradiction*: Suppose the real numbers are countable. If we can show that this leads to a contradiction (and it's generally a good idea to insist that contradictions should be impossible), our assumption must have been bad. This would mean that the real numbers are not countable.

Unique Decimal Representations of Real Numbers

In order to simplify the proof, let's just try to list *some* of the real numbers, the ones from 0 to 1, i.e. the numbers belonging to the set $C = [0, 1]$. For this proof, C is more convenient because we can write out every number uniquely as an infinite sequence of digits after the decimal point, and every such number is in C. Cantor's proof can easily be extended to all reals, but it is simpler to give the important ideas using C.

First note that any number x in $[0, 1)$ (i.e. $0 \leq x < 1$; 1 doesn't matter for now) has a representation as a sequence of digits $\{x_n\}$ such that:

$$X = 0.(x_1)(x_2)(x_3)... = \sum_{i=1}^{\infty} x_i$$

(See Appendix B for more details about what this means.)

There is a simple, familiar way to find out the digits for such a representation for a given number: Just pick the largest digit possible at any particular place, and keep going (if we assume that we have always used the largest digit possible at every place so far, we can show that it is always possible to pick such a digit that makes the sum sufficiently closer to X).

It should also be clear that any sequence digits defines some unique real number in $[0, 1)$. In order to prove the converse, we have to show that every real number has a unique representation using digits. This is a little troublesome if we consider the fact that some numbers *do* have two representations.

$$0.99999... = 1$$

(Again, see Appendix B for an explanation of what this means and why it is true.)

However, it turns out that numbers ending in all 9s / all 0s rare the *only* ones that can cause an issue. This means that a real number has a unique decimal representation if we insist that any representation that end in 99999... (e.g. after some point, each digit d_i is 9) is not a valid representation under our definition. This is proved in detail in a section of Appendix B.

C is Uncountable

Now, back to $C = [0, 1]$. Either C is countable or C is not countable, but only one of these can be true. What would it mean for C to be countable? It would mean that we could write the numbers in C down in a single list, one by one: a first element, a second element, and so on, with every element at a position we can count to. In order to arrive at a contradiction, we assume there must be at least one way to do this. Let's assume we have such an ordering call it $\{O_n\}$; it is a particular ordering that exists and cannot change.

Now, let's write out $\{O_n\}$ using the unique representation discussed above. It doesn't matter what the order is, but as an example, let's suppose it was the following order:

$$\begin{array}{cccccccccccc}
O_1 & = & 0 & . & \boxed{1} & 0 & 1 & 0 & 1 & 0 & 1 & ... \\
O_2 & = & 0 & . & 1 & \boxed{4} & 6 & 4 & 3 & 2 & 1 & ... \\
O_3 & = & 0 & . & 9 & 9 & \boxed{9} & 9 & 9 & 9 & 9 & ... \\
O_4 & = & 0 & . & 3 & 8 & 4 & \boxed{7} & 5 & 6 & 3 & ... \\
O_5 & = & 0 & . & 2 & 1 & 9 & 3 & \boxed{4} & 5 & 2 & ... \\
O_6 & = & 0 & . & 6 & 5 & 4 & 5 & 6 & \boxed{2} & 7 & ...
\end{array}$$

Now, we will take all the digits that are boxed, which lie on the "diagonal" of the list. We will use them to make a new number that differs from all the numbers that are already on the list. One way to do this would be to take the nth digit from the nth number of the list, and change the digit to another digit; let's say say we change d to 4 unless it's already 4, in which case we change it to 5. (Thought: What other rules can we use for such a change?) The process could be visualized like this:

$$\begin{array}{ccccccccccc} & & 0 & . & \boxed{1} & \boxed{4} & \boxed{9} & \boxed{7} & \boxed{4} & \boxed{2} & ... \\ & & & & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \\ D & = & 0 & . & 4 & 5 & 4 & 4 & 5 & 4 & ... \end{array}$$

Let us call our new number D $(= 0.454454...)$. D is clearly in $[0, 1]$ because every digit is still from 0 to 9. Furthermore, it does not end in all $9s$, so D must represent a real number in $[0, 1)$. However, D differs from the real number O_n in the nth digit for every value of n. This means that $D \neq O_1$ (because they have representations that differ in the first digit, and there fore must be different numbers), and that $D \neq O_n$ for any $n \in \mathbb{N}$. The implication is that D could not have been anywhere on our list, and in fact that this would be the case no matter what our list was. This is impossible if we believe our assumption that $\{O_n\}$ contains a list of all real numbers in $[0, 1]$, *which includes* D. Therefore, our assumption that such a list exists leads to a contradiction. It must be the case that such a list does not exist, which means that $[0, 1]$ is not countable.

Many people have tried to raise objections to this argument. Some of these miss the point that some representations have to be infinite. Others try to deny the conclusion that D was not in the list, because we assumed that the list *was* complete. And yet others try to insist that you can fix the problem by adding our new number D somewhere into the list. I will not try to explain why these objections are invalid; there are some good articles on the internet countering them. In any case, most mathematicians agree that Cantor's proof can be made rigorous enough to be a valid proof of the uncountability of $[0, 1]$. There are also other proofs of this fact, and a lot of mathematical results rely on the fact that such a proof holds.

Cardinal Numbers

Once we know that there are infinite sets with infinite sizes, it seems natural to wonder exactly what properties the sizes of infinite set have. Mathematicians have tools for dealing with this, and the name they use for the size of a set is *cardinality*. If S is a set, then $|S|$ is the cardinality of S.

Sets of equal Cardinality

We say that $|A| \leq |B|$ if there is an *injective mapping* from one to the other. An injective mapping is basically a function that converts every element into the first set into an element from the second set, in such a way that no two elements are converted to the same one.

For example, "$n \mapsto 4n$" is an injective mapping of the natural numbers into $\{0, 4, 8, 12, ...\}$, which is a subset of $2\mathbb{N} = \{0, 2, 4, 6, 8...\}$. Thus, $|\mathbb{N}| \leq |2\mathbb{N}|$. This might give you an idea for why our definition is "smaller than or equal to": The mapping $n \mapsto n$ shows that it is also the case that $2\mathbb{N} \leq \mathbb{N}$. Thus, $\mathbb{N}$ and $2\mathbb{N}$ have "small or equal" cardinality to each other. This is not a contradiction, but a very useful rule for equal cardinality: Two sets A and B have equal cardinality if $|A| \leq |B|$ and $|B| \leq |A|$.

However, a different definition is normally used to define sets as having equal cardinalities:

Definition 1 *The sets have equal cardinality if there is a bijective mapping between them.*

A *bijective mapping* is a mapping that is injective and also surjective, which means that every single element of the second set is mapped to. For example, $n \mapsto 4n$ is just an injective mapping from $\mathbb{N}$ to $2\mathbb{N}$, but $n \mapsto 2n$ is a bijective mapping.

The *Cantor-Bernstein Theorem* is an important fact of set theory that states that if there are injective mappings from two sets into each other, then there is a bijective mapping between them. In other words, this justifies our definition by stating that "$|A| = |B|$" precisely when $|A| \leq |B|$ and $|B| \leq |A|$. Such a mapping is simple in our case of $\mathbb{N}$ and $2\mathbb{N}$, but this is a useful fact for set theory in general.

This also leads to the nice fact that is that we can *always* compare the cardinality of two sets: one of them is always "smaller than or equal" to the other in cardinality. (If this holds both ways, they are equal. Else, one of them is smaller.) In addition, $|A| \leq |B|$ and $|B| \leq |C|$ means that $|A| \leq |C|$. Mathematicians would express these two properties by saying that cardinality provides us with a *total* and *transitive* order on all sets.

Cardinalities of $\mathbb{N}$ and $\mathbb{R}$

Let's say we have an infinite set S. This means that we should be able to keep counting elements of S and never be able to finish. It might be that S is countable and we can get to every element, like for $\mathbb{N}$. Otherwise, S is uncountable like $\mathbb{R}$ and we cannot (eventually) get to every element. However, the fact that we can always count some infinite subset of an infinite set demonstrates that $\mathbb{N}$ is smaller than or equal to (in cardinality) to every infinite set. Thus, $\mathbb{N}$ has the smallest possible size of an infinite set. In fact, it has a name:

Definition 2 $\aleph_0$ *(pronounced* aleph-naught*) is the smallest cardinal number, equal to* $|\mathbb{N}|$.

Since cardinal numbers can be infinite, this provides a final answer to our original question: There are infinite numbers that we might consider "infinities", and some are bigger than others.

Now, we know that $\mathbb{R}$ is bigger then $\mathbb{N}$ because $\mathbb{R}$ is not countable, and thus $\aleph_0 \leq |\mathbb{R}|$. So how large is $\mathbb{R}$? The next-biggest size that a set can have after $\aleph_0$ is considered the next cardinal number, denoted by $\aleph_1$. Clearly, $\aleph_1 \leq |R|$. But is $\aleph_1 = |\mathbb{R}|$?

The surprising conclusion is that there is no answer! The statement that $\aleph_1 = |\mathbb{R}|$ is known as the *continuum hypothesis*. In 1963, Paul Cohen (building on work by Kurt Gödel) proved that it is impossible to use the basic Zermelo-Fraenkel axioms to show whether the continuum hypothesis holds. In essence, the continuum hypothesis is undecidable using the tools mathematicians normally use. We could assume that either choice is correct, and it would be "a consistent way" to do mathematics. When this issue came up with infinite sets, it turned out to be fruitful to accept that infinite sets should exist, because this leads to a useful theory. However, in the case of the continuum hypothesis, it seems to be hard to come up with nice examples of the kinds of sets that might fit between $\mathbb{N}$ and $\mathbb{R}$ in size, but it is equally hard to come up with a non-arbitrary reason to show that they cannot exist. Thus, mathematicians have not come up with a good, intuitive reason for assuming whether the continuum hypothesis is true. Since 1963, it has has not been a conventional "unsolved" problem, but it is also not a "solved" problem in the sense that it has been proved either true or false. Thus, the continuum hypothesis is a good example of a problem that could be considered an *undecidable problem.*

Appendix A: Motivations for Set Theory

In mathematics, infinity is closely related to the idea of "sets". A set is basically a collection of various things, where the only thing that matters is whether something is in it or not. You can think of it as a list of things, except that it doesn't matter how often something is in the list, or in what order the list is. We could write $\{5, \pi, carrot\}$ to describe a set that contains 5, π, and "carrot", but $\{\pi, carrot, \pi, 5\}$ would mean the same set. Anything can be in a set; even sets can be contained in other sets! If something is contained in a set, we say that it is an *element* of the set, and once have a set where we can tell what is an element it (and what is not), we can give it a name. If we give our set with three elements the name X, we can say that

- $5 \in X$ (5 is *in* X, or 5 is *an element of* X), but
- $music \notin X$ (the item *music* is not in X).

We could try to describe every set from scratch, but many important sets have names that every mathematician uses: for example, the set of "natural numbers" is $\{0, 1, 2, 3, ...\}$ and it is written as $\mathbb{N}$.

Since a set is determined by what is in it, we could also describe a set using a sentence like "something is in our set (let's call it X) if it is a counting number between 0 and 100". A mathematician would use *set-builder notation* to write it like this:

$$Y = \{i : 0 < i < 100\}$$

In this case, it means "Y is the set of all elements i for which $0 < i < 100$ is true". We can tell that $5 \in Y$, but $1000 \notin Y$. What about *carrot*? We'd have to be able to tell whether $0 < carrot < 100$.

This is an issue, but our definitions of sets so far have a bigger issue: Russell's Paradox. Russell's Paradox comes from the following definition:

$$Z = \{S : S \notin S\}$$

Z is a set. What's in it? Well, any set (let's call it S) is in Z if S *does not contain itself*. How can a set contain itself? Well, $A = \{\{5\}\}$ (the set containing the set that contains 5) contains the set $\{5\}$ (the set that contains 5 itself), but A doesn't contain A itself. If we had a set that could be written as $X = \{X, 4, \pi\}$, then $X \in X$, so X would contain itself. The main issue arises when we ask: is Z in Z? If so, then it doesn't satisfy the definition it has to satisfy to be a member of itself (that $Z \notin Z$). If not, then it *does* satisfy the definition is has to satisfy to be a member of itself. Regardless of whether Z is in itself or not, the opposite must be true because of the definition of Z itself. Thus, we have a contradiction, and we cannot tell whether Z is in Z. If we don't want to get in deep trouble, Z cannot exist.

Although a lot of formal details have been left out, our exploration roughly parallels European mathematics in the late 1800's. Georg Cantor came up with formal ways to describe mathematical objects as sets, and used this to prove fundamental things that were previously vague or informal. Cantor's newly created "set theory" was wonderful, but in 1895, Bertrand Russell pointed out the paradox that is now named after him. "Naive" set theory was in trouble because it contained a fundamental contradiction that undermined all of it.

Mathematicians came up with a clever way around this. The most common approach, *Zermelo-Fraenkel set theory* (ZF) starts with just the empty set and has axioms that define what other sets exist. You can show that certain sets exist by building them from other sets, but you cannot define things like $Z = \{S : S \notin S\}$ anymore; instead, you would have to build it out of sets S that you already know to exist (e.g. if you have a set T, you could define $Z = \{S : (S \in T \text{ and } S \notin S)\}$). If we want ZF not to have any contradictions, then $Z = \{S : S \notin S\}$ is not a valid definition, and it is not a set. This also has other repercussions, like the fact that the "set of all sets" cannot exist. However, very many sets do exist, and mathematicians believe that this approach doesn't produce any contradictions. This is called the set of *natural numbers*, which is called $\mathbb{N}$. It seems obvious that $\mathbb{N}$ is infinite, but what does that mean?

Generally, finite can be taken to mean "bigger than anything finite". You might say that the length of a line is infinite because it is longer than any finite length. In our case, $\mathbb{N}$ is infinite because you can count more elements of $\mathbb{N}$ than elements of any finite set (which you can do by counting to 0 or 1 or 2... i.e. any element of $\mathbb{N}$ itself!). This concept is such a fundamental expression of being infinite that when mathematicians informally refer to "infinity", they generally mean $\mathbb{N}$ or "a set with everything that we could if we counted far enough".

A mathematical way to define infinite sets in ZF would be to start by building some finite sets. $\{0\}$ is finite, and so is $\{0, 1\}$, and so is $\{0, 1, 2\}$, and so on. In each case, we can count the number of elements in the set. But what if we take this process "to infinity"? Set theory provides a way for us to do this: there is an axiom (i.e. a rule) that says that there is a set that contains any element that is either in $\{0\}$, or in $\{0, 1\}$, or in $\{0, 1, 2\}$, and so on. In other words, this axiom lets us state that the set $\mathbb{N}$ exists. The axiom has a sensible name: the Axiom of Infinity.

This might seem a little silly. Isn't it cheating to say that something infinite like $\mathbb{N}$ exists just because we say it exists? Perhaps so. However, the careful approach of ZF requires us to build up all sets. It turns out that the only reasonable way to build up a set that contains an infinite number of elements... is to say that it can be done. This gives us a definition we can work with. As long as we can use it to produce a useful theory (ideally, one without contradictions), the Axiom of Infinity is useful, and modern mathematicians are generally content to assume that things are defined this way.

Appendix B: $0.99999... = 1$ and Real Numbers as Sums of Sequences

Grade school students learning mathematics are often surprised (or repulsed) to be told that $0.99999... = 1$. This statement tends to bring up heated arguments about what it means for a number to have a value, and what it means for two numerous to be the same. There are numbers proofs suggesting why this should be true, such as:

$$\begin{aligned} X &= 0.99999... \\ 10 \cdot X &= 9.99999... \\ 10 \cdot X - X &= 9.00000.... \\ 9 \cdot X &= 9 \\ X &= 1 \end{aligned}$$

From the perspective of analysis, the decimal expansion of real number in $[0, 1)$ can be defined as a sequence of digits $\{d_n\}$ that be written as $0.(a_1)(a_2)(a_3)....$

Definition 3 *A* (decimal) representation *of a real number is a sequence of digits (*0, 1*,* 2, 3*,* 4*,* 5, 6*,* 7*,* 8*, or* 9*)*

$$\{d_n\}$$

which may also be written out as $0.(d_1)(d_2)(d_3)...$ *without spaces between the digits.*

For example, $d_1 = 9, d_2 = 9, d_3 = 9, ...$ is the sequence that corresponds to the real number we mean with 0.99999.... Such a number can be calculated by adding the values of the tenths digit, hundredths digit, thousandths digit, etc.

$$\sum_{i=1}^{\infty} d_i \cdot 10^{-i} = \frac{d_1}{10} + \frac{d_2}{100} + \frac{d_3}{1000} + ...$$

More formally:

Definition 4 *The* value *of a real number* $\{d_n\}$ *is*

$$\sum_{i=1}^{\infty} d_i \cdot 10^{-i}$$

The value of an infinite sum like this is defined to be a number X if we can show that it gets arbitrarily close to X (and stays there). If someone gives us a number ϵ greater than 0, we have to be able to show that the sum eventually gets within ϵ of X.

Definition 5 *The value of a sum* $\left(\sum_{i=1}^{\infty} d_i\right)$ *is defined to be a* X *if we can do the following: Any time someone gives us a real number* ϵ *that is greater than* 0*, we can always find some natural number* N *such that*

$$\left(\sum_{i=1}^{N} d_i\right)$$

differs from X *by at most* ϵ*, and that the same thing holds if we replace* N *by a larger natural number.*

To see that 0.99999... = 1, let's say somebody gives us a number ϵ like 0.0000000434. Since $10^{-8} = 0.00000001 < 0.0000000434$, this means that

$$\sum_{i=1}^{8} 9 \cdot 10^{-i} = 0.99999999$$

differs from 1 by less than ϵ (and the difference only gets smaller if we replace 8 by a larger natural number). Since the negative powers of 10 get arbitrarily small, we can find such a natural number for any $\epsilon > 0$, and therefore we can say that the limit of the sum (and the value of 0.99999...) is defined to be 1. In real analysis, it has to be the case that two numbers are the same if there are no numbers between them, and our formal definition handles that.

Every real number in $[0, 1)$ corresponds to a Unique Decimal Representation (But There's a Catch!)

To see that this is the case, consider two different decimal representations for the same number x; let's call $\{a_n\}$ and $\{b_n\}$. Since are different, they must be different in *some* digit; let k be the first index of such a digit, i.e. $a_k \neq b_k$ but $a_i = b_i$ for $i < k$. For example, the real number0.14527 has two representations where k is 5.

$$\begin{array}{lccccccccccccc}
a_n\colon & x = & 0 & . & 1 & 4 & 5 & 2 & 6 & 9 & 9 & 9 & 9 & \ldots \\
 & & & & \parallel & \parallel & \parallel & \parallel & \nparallel & & & & & \\
b_n\colon & x = & 0 & . & 1 & 4 & 5 & 2 & 7 & 0 & 0 & 0 & 0 & \ldots
\end{array}$$

The sums of the representations up to digit k must differ by $|a_k - b_k| \cdot 10^{-k}$, which is at least 10^{-k}. However, two digits can only differ by at most 9, so the maximum difference the remaining digits can contribute to the sum is

$$\sum_{i=k+1}^{\infty} |a_i - b_i| \cdot 10^{-i} \sum_{i=k+1}^{\infty} 9 \cdot 10^{-i} = 10^{-k}$$

Thus, a change in the nth digit changes the sum by at least as much as the remaining digits can change back. The only way the sums of the two representations can be the same is if a_k and b_k differ by 1, and all the remaining digits change between a_n and b_n in the same way to produce the maximum sum, i.e. all from 0 to 9 or all from 9 to 0.

If there is any of the remaining digits that doesn't differ by 9, then the remaining sum of 10^{-k} cannot be reached, and $\{a_n\}$ and $\{b_n\}$ cannot have the same sum: they cannot represent the same real number.

Now, if a representation ends in all 0s or all 9s, we can change all repeating digits (and the one digit before) to produce a different representation for the same number. But if a representation of the number does *not* end in all 0s (or all 9s), this means that no matter how far you go, there will always be different digit from 0 (or 9). This means that there is not a point from which we can make this change, because that digit cannot be changed to 9 (or to 0) for a difference of 9. Thus:

Theorem 1 *Every real number in* $[0, 1)$ *has unique representation (that does not end in all* 9*s), and vice-versa.*