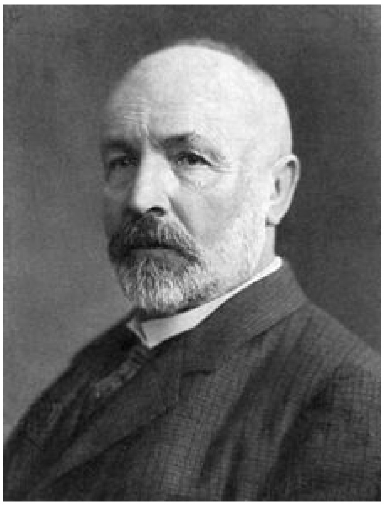
2.4 $\mathbb{R}$ is uncountable

Our goal in this section is to show that the set $\mathbb{R}$ of real numbers is *uncountable* or *non-denumerable*; this means that its elements cannot be *listed*, or cannot be put in one-to-one correspondence with the natural numbers. We saw at the end of Section 1.3 that $\mathbb{R}$ has the same cardinality as the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$, or the interval $(-1, 1)$, or the interval $(0, 1)$. We will show that the open interval $(0, 1)$ is uncountable. This assertion and its proof date back to the 1890's and to Georg Cantor. The proof is often referred to as "Cantor's diagonal argument" and applies in more general contexts than we will see in these notes.



Georg Cantor: born in St Petersburg (1845), died in Halle (1918)

Theorem 2.4.1. *The open interval* $(0, 1)$ *is not a countable set.*

Before embarking on a proof, we recall precisely what this set is. It consists of all real numbers that are greater than zero and less than 1, or equivalently of all the points on the number line that are to the right of 0 and to the left of 1. It consists of all numbers whose decimal representation have only 0 before the decimal point (except $0.000\ldots$ which is equal to 0, and $0.99999\ldots$ which is equal to 1). Note that the digits after the decimal point may terminate in an infinite string of zeros, or may have a repeating pattern to their digits, or may not have either of these properties. The interval $(0, 1)$ includes all these possibilities.

Our goal is to show that the interval $(0, 1)$ cannot be put in bijective correspondence with the set $\mathbb{N}$ of natural numbers. Our strategy is to show that no attempt at constructing a bijective correspondence between these two sets can ever be complete; it can never involve *all* the real numbers in the interval $(0, 1)$ no matter how it is devised. In order to achieve this, we will imagine that we had a listing of the elements of the interval $(0, 1)$; i.e. a bijective correspondence between this interval and $\mathbb{N}$. Such a correspondence would have to look something like the following.

$$
\begin{array}{ccl}
\mathbb{N} & & (0,1) \\
\hline
1 & \longleftrightarrow & 0.13567324\ldots \\
2 & \longleftrightarrow & 0.10000000\ldots \\
3 & \longleftrightarrow & 0.32323232\ldots \\
4 & \longleftrightarrow & 0.56834662\ldots \\
5 & \longleftrightarrow & 0.79993444\ldots \\
\vdots & & \vdots
\end{array}
$$

Note: The exact numbers that appear in the right-hand column above are not important, the point is that a bijective correspondence between $\mathbb{N}$ and $(0, 1)$ would have this general form. We

don't know whether any particular decimal number in the right hand side terminates in zeros (or repeats) or not, but we know that some do and some don't.

So the entries in the right hand column above are basically infinite sequences of digits from 0 to 9. The right hand column must then consist somehow of a list of *all* such sequences. Our problem is to show that this is impossible: that no matter how the right hand column is constructed, it can't contain *every* sequence of digits from 1 to 9. We can do this by exhibiting an example of a sequence that can't possibly be there.

Suppose our list starts as follows.

$\mathbb{N}$		$(0, 1)$
1	$\longleftrightarrow$	$0.\mathbf{1}3567324\ldots$
2	$\longleftrightarrow$	$0.1\mathbf{0}000000\ldots$
3	$\longleftrightarrow$	$0.32\mathbf{3}23232\ldots$
4	$\longleftrightarrow$	$0.568\mathbf{3}4662\ldots$
5	$\longleftrightarrow$	$0.7999\mathbf{3}444\ldots$
$\vdots$		$\vdots$

We will construct an element x of $(0, 1)$ that is not in the list. To do so :

1. Look at the *first* digit after the decimal point in Item 1 in the list. If this is 1, write 2 as the first digit after the decimal point in x. Otherwise, write 1 as the first digit after the decimal point in x. *So x differs in its first digit from Item 1 in the list.*

2. Look at the *second* digit after the decimal point in Item 2 in the list. If this is 1, write 2 as the second digit after the decimal point in x. Otherwise, write 1 as the second digit after the decimal point in x. *So x differs in its second digit from Item 2 in the list.*

3. Look at the *third* digit after the decimal point in Item 3 in the list. If this is 1, write 2 as the third digit after the decimal point in x. Otherwise, write 1 as the third digit after the decimal point in x. *So x differs in its third digit from Item 3 in the list.*

4. Continue to construct x digit by digit in this manner. At the nth stage, look at the nth digit after the decimal point in Item n in the list. If this is 1, write 2 as the nth digit after the decimal point in x. Otherwise, write 1 as the nth digit after the decimal point in x. *So x differs in its nth digit from Item n in the list.*

What this process constructs is an element x of the interval $(0, 1)$ that does not appear in the proposed list. The number x is not Item 1 in the list, because it differs from Item 1 in its 1st digit, it is not Item 2 in the list because it differs from Item 2 in its 2nd digit, it is not Item n in the list because it differs from Item n in its nth digit.

We conclude that the set of real numbers $\mathbb{R}$ is *not countable* (or *uncountable*).

Note:

1. In our example, the number x would start $0.21111\ldots$.

2. According to our construction, our x will always have all its digits equal to 1 or 2. So not only have we shown that the interval $(0, 1)$ is uncountable, we have even shown that the set of all numbers in this interval whose digits are all either 1 or 2 is uncountable.

3. A challenging exercise: why would the same proof not succeed in showing that the set of rational numbers in the interval $(0, 1)$ is uncountable?

Informally, Cantor's diagonal argument tells us that the "infinity" that is the cardinality of the real numbers is "bigger" than the "infinity" that is the cardinality of the natural numbers, or integers, or rational numbers. He was able to use the same argument to construct examples

of infinite sets of different (and bigger and bigger) cardinalities. So he actually established the notion of infinities of different magnitudes.

The work of Cantor was not an immediate hit within his own lifetime. It met some opposition from the finitist school which held that only mathematical objects that can be constructed in a finite number of steps from the natural numbers could be considered to exist. Foremost among the proponents of this viewpoint was Leopold Kronecker. From the book "The Honors Class" by Ben Yandell:

> Many mathematicians, Leopold Kronecker in Berlin, in particular, were bothered by this headlong leap into the infinite, accessible only by inference, not finite construction. Georg Cantor (1845-1918), teaching at Halle in 1888, had invented set theory in the 1870s and was writing about infinities of different sizes and even doing arithmetic with them. But Kronecker would admit only numbers or other mathematical objects that were finitely 'constructible'.

Leopold Kronecker (1823-1891)

> *"God made the integers, all else is the work of man."*
>
> *"What good your beautiful proof on π? Why investigate such problems, given that irrational numbers do not even exist?"*

The work of Cantor had influential admirers too, among them David Hilbert, who set the course of much of 20th Century mathematics in his address to the International Congress of Mathematicians in Paris in 1900.

David Hilbert (1862-1943)

> *"No one shall expel us from the paradise that Cantor has created for us."*
>
> *"What new methods and new facts in the wide and rich field of mathematical thought will the new centuries disclose?"*

Hilbert's address to the Paris Congress is one of the most famous mathematical lectures ever. In it he posed 23 unsolved problems, the first of which was Cantor's *Continuum Hypothesis*. The Continuum Hypothesis proposes that every subset of $\mathbb{R}$ is either countable (i.e. has the same cardinality as $\mathbb{N}$ or $\mathbb{Z}$ or $\mathbb{Q}$) or has the same cardinality as $\mathbb{R}$. This seems like a question to which the answer is either a straightforward yes or no, but it took the work of Kurt Gödel in the 1930s and Paul Cohen in the 1960s to reach the remarkable conclusion that the answer to the question is *undecidable*. This means essentially that the standard axioms of set theory do not provide enough structure to determine the answer to the question. Both the Continuum Hypothesis and its negation are consistent with the working rules of mathematics. People who work in set theory can legitimately assume that either the Continuum Hypothesis is satisfied or not. Fortunately most of us can get on with our mathematical work without having to worry about this very often.

Learning Outcomes for Section 2.4

After studying this section you should be able to

- Use Cantor's diagonal argument to prove that the interval $(0, 1)$ is uncountable.
- Make a few remarks about the history of this discovery.