Principles of Analysis: Fall 2002
Cardinality and Infinite Sets

Acknowledgements: Much of this material is borrowed with permission from the lecture notes *Countable and Uncountable Sets* written by John Anderson, Dept. of Mathematics and Computer Science, College of the Holy Cross. Another source is *Analysis with an Introduction to Proof* 3rd edition, by Steven Lay, Prentice Hall.

1 Sizes of Sets

Given two **finite** sets A, B, we can compare their size simply by counting the number of elements in each. How can we compare the sizes of two infinite sets? One method using functions was developed by the mathematician Georg Cantor in the 19th century, although the idea goes back to Galileo. Cantor's work was quite controversial in his time and was the subject of intense criticism. Today his theories are considered one of the cornerstones of mathematics. Using Cantor's ideas, we will be able to show that in a sense, most real numbers are irrational - in fact, most are transcendental.

Consider first two finite sets, say

$$A = \{a, b, c\}, \quad B = \{1, 2\}$$

We can easily find a function $f : A \to B$ for which range$(f) = B$, say

$$f(a) = 1, f(b) = 1, f(c) = 2$$

However, this function is not one-to-one, because $f(a) = f(b)$. You should be able to convince yourself that we can't find $f : A \to B$ that is one-to-one. The problem is that A has more elements than B. On the other hand, we can find a function $g : B \to A$ such that g is one-to-one, say

$$g(1) = a, g(2) = c$$

but it's impossible to find such a g with range$(g) = A$, because B has fewer elements than A. No matter how hard you try, there is no function from A to B or from B to A which is **both** one-to-one and onto. This suggests the following definition for the size of a finite set.

Definition 1.1 *We say that the set A has* **cardinality** *n, denoted $|A| = n$ iff there exists a bijection (one-to-one and onto function) between A and the set of natural numbers $\{1, 2, 3, \ldots, n\}$. (Cardinality is a synonym for size.) If such a bijection exists for some natural number n, then A is said to be* **finite**. *If A is not finite, A is said to be* **infinite**.

This definition is useful not only for precisely describing the size of a finite set, but also for comparing sets of infinite size. Two sets will have the same size (whether they are finite or infinite) if we can find a bijection between them. In other words, if there is a **one-to-one correspondence** between the sets, they have the same size. If every element of your set can be paired with every element of my set, with nobody left out and nobody paired up more than once, we have sets of the same size. It may sound simple and obvious, but the definition below has remarkable consequences.

Definition 1.2 *The sets A and B have the same cardinality ($|A| = |B|$) iff there exists a bijection between A and B.*

The most common infinite set, the one we use as our "yardstick" for other infinite sets, is the set of natural numbers

$$\mathbf{N} = \{1, 2, 3, \ldots\}$$

and the size of $\mathbf{N}$ is given the special symbol "aleph null" named for the first letter in the Hebrew alphabet.

$$|\mathbf{N}| = \aleph_0$$

Definition 1.3 *An infinite set A is called* **countable** *iff there exists a bijection between A and the natural numbers $\mathbf{N}$. Finite sets are also called countable. If a set is not countable, it is called* **uncountable**.

Let's begin by comparing the natural numbers with the set of integers. The integers is a bigger set, so we should expect their "size" to be bigger, right?

Theorem 1.4 *The set of integers $\mathbf{Z}$ is a countable set. That is, there exists a bijection between the sets $\mathbf{N}$ and $\mathbf{Z}$.*

Proof: We have to construct a function $f : \mathbf{N} \to \mathbf{Z}$. There are many ways to do this. Here is one possibility: let $f(1) = 0$, $f(2) = -1$, $f(3) = +1$, $f(4) = -2$, $f(5) = 2$, and so on. You can easily find a formula for this function. It is also not hard to see that f is both one-to-one and onto. $\square$

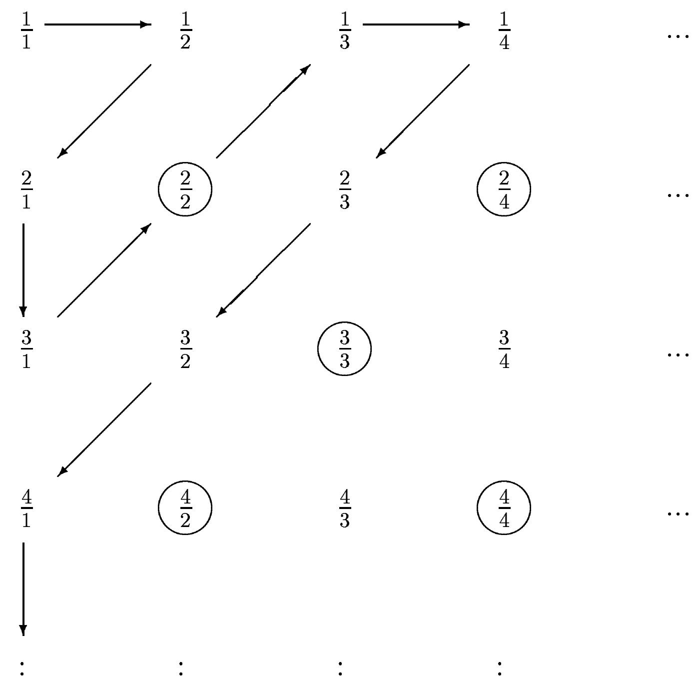
Thus from a certain point of view $\mathbf{N}$ and $\mathbf{Z}$ are the same size, even though there are "more" integers than natural numbers. Note that our proof essentially gives a way to count the integers, making sure we count every one and count each one only once. You can think of this as making a list:

$$0, 1, -1, 2, -2, 3, -3, \ldots$$

It is clear how the list continues and that every integer is eventually counted somewhere in the list. In general, you can think of countable sets as being those which it is possible to "count" making sure you obtain all of the set. In principle, if a set is countable, it can be listed as a sequence $x_1, x_2, \ldots$ of distinct values: if $f : \mathbf{N} \to A$ is both one-to-one and onto, then set $x_1 = f(1), x_2 = f(2)$, etc. The next theorem may surprise you.

Theorem 1.5 *The set of rationals $\mathbf{Q}$ is a countable set. That is, there exists a bijection between the sets $\mathbf{N}$ and $\mathbf{Q}$.*

Proof: This is a little trickier. We will describe pictorially a one-to-one correspondence $f : \mathbf{N} \to \mathbf{Q}$. First we make an array of all rationals p/q as shown below.



All fractions p/q with numerator $p = 1$ are in the first row, all those with numerator $p = 2$ are in the second row, etc. All those with denominator $q = 1$ are in the first column, all those with denominator $q = 2$ are in the second column and so on. The circled numbers are fractions which are not in lowest terms. We will snake through this array along the path indicated by arrows, assigning values of f as we go. Start with $f(1) = 1/1$, define $f(2) = 1/2$, then follow the arrows to define $f(3) = 2/1$, $f(4) = 3/1$. When we get to $2/2$ in the array, we will skip it, since $2/2 = 1/1$ and we have already assigned $f(1) = 1/1$. Therefore we set $f(5) = 1/3$. Continue in this way to define $f(n)$ for all $n \in \mathbf{N}$, remembering to skip over all fractions in the array which are not in lowest terms. You should convince yourself that f is one-to-one and onto the positive elements of $\mathbf{Q}$. It would be difficult to write down a single formula for this function, but nonetheless it represents a one-to-one correspondence. This shows that the set of positive rationals is countable. To finish the proof we use the same argument to count all the negative rationals (throwing in 0 to get them all), thereby showing that the remaining rationals are also countable. The theorem then follows because the union of two countable sets is also countable. (see Exercise 3). $\square$

Surprisingly, we have shown that both $\mathbf{Z}$ and $\mathbf{Q}$ are countable sets, having the same size as $\mathbf{N}$, even though they contain many more elements than $\mathbf{N}$.

Proposition 1.6 *The set of real algebraic numbers is countable.*

Proof: By definition, a real number is **algebraic** if it is a solution of an equation of the form $P(x) = 0$, where

$$P(x) = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n \tag{1}$$

where $a_0, a_1, ..., a_n$ are integers, $a_n \neq 0$, and n is a natural number. The *index* of a nonconstant polynomial P of the form (1) is defined to be

$$|a_0| + |a_1| + ... + |a_n| + n$$

For example, the index of $P(x) = x^3 - 3x + 7$ is $7 + 3 + 1 + 3 = 14$. We can imagine arranging all possible polynomials by their index. There are no polynomials of index less than 2, and there is only polynomial of index 2, namely $P(x) = x$. There are five polynomials of index 3: $P(x) = \pm 2x, P(x) = x \pm 1$, and $P(x) = x^2$, and so on. For a given index, we will collect all the roots of all possible polynomials of that index. For example, starting with index two, and setting $P(x) = 0$, we have a single root, namely $x = 0$. For index three we have the three roots $x = 0, -1, +1$. Using this heirarchical ordering of roots, we can then describe a function f from $\mathbf{N}$ to the algebraic numbers which is both one-to-one and onto, as follows: define $f(1) = 0$, since that is our first root from index 2 polynomials. Then go to the roots for index 3 polynomials. Ignore 0, since we have already encountered it as a root. Set $f(2) = -1$ and $f(3) = 1$. This finishes off the roots of index 3 polynomials, so we go to the index four polynomials, (problem 3 asks you to carry this out explicitly) and so forth, remembering to discard roots which we have already seen. For each index let's agree to list the new roots in increasing order of magnitude when we assign the values of f - this is the rule followed in choosing the root -1 before the root +1. The function f is clearly onto, since if x is an algebraic number, it is the root of some polynomial P, and we will find x when we list all roots of polynomials with the same index as P. It is one-to-one, since we agree to disregard roots which have already appeared on our list. □

Now we are almost ready to make the jump to a comparison of all the real numbers with $\mathbf{N}$. We will use the fact that real numbers can be specified by **decimal expansions** of the form

$$x = b_0.b_1b_2b_3...$$

where $b_0 \in \mathbf{Z}$ and $b_1, b_2, ...$ are in the set $\{0, 1, ..9\}$. The fact that such expansions exist and have meaning is actually closely connected with the Least Upper Bound Axiom (Completeness Axiom); we will discuss this issue in more detail when we discuss sequences and series. For now, let's assume that every real number has a decimal expansion and that every decimal expansion represents a real number. The next result literally shook the foundations of mathematics. It is one of the "classic" proofs in Mathematics, ranking in importance and elegance with the proof that the $\sqrt{2}$ is irrational and the proof that the number of primes is infinite (due to Euclid originally).

Theorem 1.7 *(Cantor) The set of all real numbers is uncountable.*

Proof: This is a proof by contradiction: we assume that the real numbers are countable, and derive a contradiction. The famous technique is often referred to as Cantor's diagonalization proof. Before we start, you might be interested to know why people thought the real numbers were actually countable, and even got into heated arguments with Cantor over his proof. The naysayers simply said, "Put all the real numbers in a hat. Reach in and pull out a number, this is $f(1)$. Reach in and pull out another number, this is $f(2)$. Continue the process ad infinitum. In this way we can 'count' the real numbers." Cantor was quick to point out that you could never guarantee that you would count all the real numbers, but it took some time for his proof to be accepted into the main stream of mathematics.

We will actually prove that the interval $(0, 1)$ is not countable. Since $(0, 1)$ is a subset of $\mathbf{R}$, it follows easily that $\mathbf{R}$ is not countable. (why?) If the interval $(0, 1)$ is countable, there exists a function $f : \mathbf{N} \to (0, 1)$ which is both one-to-one and onto. This means we can list all the real numbers in $(0, 1)$ as $\{x_1, x_2, x_3, x_4, \ldots\}$. Put each number in its decimal expansion and denote the jth decimal place of the element x_i by x_{ij}. So our list might look something like this:

$$
\begin{aligned}
x_1 &= \quad 0.3418\ ... \\
x_2 &= \quad 0.4106\ ... \\
x_3 &= \quad 0.1119\ ... \\
x_4 &= \quad 0.3244\ ... \\
x_5 &= \quad \ldots
\end{aligned}
$$

Every real number in (0,1) is supposed to appear somewhere on the right-hand side of this list, since f is onto. However, and this was Cantor's wonderful insight, we can produce a number $y \in (0,1)$ which **cannot** be on the list. We will construct the decimal places of y using the following formula. For notation, let b_k be the kth decimal place of y, so $y = 0.b_1b_2b_3b_4\ldots$. Now define

$$
b_k = \begin{cases} 4 & \text{if } x_{kk} \neq 4 \\ 5 & \text{if } x_{kk} = 4 \end{cases}
$$

So in the example above, y will begin as $y = 0.4445\ldots$. No matter what follows in the list of x's, the number y will never appear on the right-hand side of the list, since the kth decimal place of y is never equal to the kth decimal place of x_k. However, y is certainly a real number between 0 and 1. Thus the function f is not onto which is a contradiction. □

The fact that the set of real numbers is not countable has far-reaching consequences, some of which we will see later. An immediate consequence is that **transcendental** numbers exist. A transcendental number is a real number which is not algebraic. The proof is left to the reader. This proof made some mathematicians very unhappy, since it gives no clue as to how you might construct a single transcendental number! Such "pure existence" proofs gradually became more accepted in mathematics, although there are still those who reject them as having little meaning.

Definition 1.8 *The cardinality of* **R** *is denoted by c, standing for continuum. Since* **R** *is uncountable, its size is bigger than* **N**. *In other words,*

$$
\aleph_0 < c
$$

Question: Is there an infinite set with cardinality between $\aleph_0$ and c?

Cantor conjectured that no such set exists. This conjecture became known as **Cantor's continuum hypothesis**. Hilbert listed this in 1900 as his first problem to be solved for the 20th century. (Remember Hilbert and his infinite hotel?) It turns out that this is problem is harder than it seems. In 1938, Gödel proved that assuming Cantor's continuum hypothesis does not contradict any of the axioms of set theory. This sounded promising until in 1963, Cohen proved that denial of Cantor's continuum hypothesis does not contradict any axioms either. The problem is somewhat unsatisfactorily called **formerly undecidable** — neither provable nor disprovable under the current axioms of set theory. It will rest with future generations of mathematicians (perhaps you) to decide the issue by coming up with a new set of axioms.

Exercises:

1. Explain in your own words Cantor's Diagonalization argument proving that the set of real numbers is uncountable. Fill in the detail about using $(0,1)$ instead of **R**. Give a detailed example different than the one in the text. Is there anything special about the numbers 4 and 5?

2. Prove that the set of even natural numbers is countable.

3. Show that the union of two countable sets is countable. Note: One WRONG proof is to say, "Since both sets are countable, list the elements of one set and after you are finished, list the elements of the other." *Hint:* How did we show that the integers were countable?

4. Show that the countable union of countable sets is countable. In other words, suppose that you have a collection of sets A_n, with $n \in \mathbf{N}$, such that A_n is a countable set for each n. Show that the infinite (but countable) union of all the sets A_n is countable. *Hint:* Try to find an explicit method of counting this set which insures that every element will be counted.

5. Using proof by contradiction, show that the set of irrational numbers, Q^{C}, is **uncountable**. Conclude that the set of irrational numbers is much, much "bigger" than the set of rationals.

6. Show that the open intervals $(0, 1)$ and (a, b) have the same cardinality by finding a bijection between them. You may assume that $a < b$. Your bijection will be a linear function.

7. Show that the open interval $(0, 1)$ has the same cardinality as $\mathbf{R}$ by finding a bijection between them. Based on the previous problem, conclude that the entire real number line and any open interval, no matter how small, have the same size! *Hint:* Use a function from Trigonometry.

8. What is wrong with the following "proof" that the interval $(0, 1)$ is countable? Define a one-to-one correspondence $f : \mathbf{N} \to \mathbf{R}$ as follows: let k be the number of digits in n and set $f(n) = n/10^k$. For example, $f(342) = \frac{342}{10^3} = .342$, $f(51698) = \frac{51698}{10^5} = .51698$, etc. Every real number in $(0, 1)$ has a pre-image in $\mathbf{N}$ by simply taking the decimal expansion of the real number and removing the decimal place. QED.