Chapter 4 : Sizes of sets

In this chapter, we deal with the concept of how large a set is. The biggest implication of this study in Computer Science is the fact that computers cannot solve all problems. This is one the most fundamental realizations underlying the theory of computation.

4.1 The notion of size

The *size* or *cardinality* of a set is the number of elements in it. The size of a set A is denoted by $|A|$. The notation $|A| < \infty$ implies that A is a finite set. A finite set with n elements can be listed as $\{a_1, a_2, \ldots, a_n\}$, where a_i is the i-th element of A for $i = 1, 2, \ldots, n$.

The simplest example of an infinite set is the set $\mathbb{N} = \{1, 2, 3, \ldots\}$ of natural numbers. Consider the set $\mathbb{N}_n = \{1, 2, \ldots, n\}$. For every $n \in \mathbb{N}$ the set $\mathbb{N}_n$ is finite. However, their union $\mathbb{N} = \bigcup_{n\in\mathbb{N}} \mathbb{N}_n$ is not a finite set. Nonetheless, we can count the elements of $\mathbb{N}$ as $1, 2, 3, \ldots, n, \ldots$. This counting never stops, but every $n \in \mathbb{N}$, however large, is eventually covered in the counting process.

Two sets are called *equinumerous*, if they have the same size. For finite sets, this concept is easy to visualize. We run into trouble when we work with infinite sets. Clearly, no infinite set can be equinumerous with a finite set. What is more important is that two infinite sets need not be equinumerous. We will soon see that the set $\mathbb{Z}$ of all integers (positive, negative and zero) and the set $\mathbb{Q}$ of rational numbers are equinumerous with $\mathbb{N}$. That is surprising, since $\mathbb{N}$ is a strict subset of $\mathbb{Z}$ and can be easily visualized to be embedded inside $\mathbb{Q}$ (identify the natural number n with the rational number $n/1$). Since $\mathbb{Z}$ is equinumerous with $\mathbb{N}$, we can also count or enumerate the elements of $\mathbb{Z}$, so that for every $n \in \mathbb{N}$ we can identify an integer a_n as the n-th integer. For example, we may order the elements of $\mathbb{Z}$ as $0, 1, -1, 2, -2, 3, -3, \ldots, n, -n, \ldots$. Thus 0 is the *first* integer in the counting process, 1 is the *second* integer, -1 the *third* integer, 2 the *fourth* integer, and so on. It is clear that every $m \in \mathbb{Z}$, positive, negative, or zero, is eventually covered in the counting process.

The set $\mathbb{R}$ of all real numbers is not equinumerous with $\mathbb{N}$. In fact, $\mathbb{R}$ contains more elements than $\mathbb{N}$ (or $\mathbb{Z}$ or $\mathbb{Q}$). This implies that if we start counting real numbers as above and allow the counting process to continue *ad inifinitum*, we cannot exhaust the list of all real numbers. In other words, whatever way we count real numbers, we are sure to miss out some real number(s) in the counting process.

It is now time to make the concept of size mathematically concrete. We use the theory of functions to that effect. Since we will be dealing mostly with infinite sets, it is imperative that the reader is already quite comfortable with the concept of injective, surjective and bijective functions among infinite sets.

4.2 Comparing the sizes of two sets

Let A, B be two sets. We say that

$$|A| \leqslant |B|$$

if there exists an injective map $f : A \to B$. This notion is intuitively clear, since for every element $a \in A$ we can associate an element $b = f(a)$ of B in such a fashion that two different elements of A are not associated with the same element of B. The map f essentially produces an embedding of A in B. So B cannot be of size smaller than the size of A.

4.1 Example (1) Let $A \subseteq B$. The canonical inclusion map $\iota : A \to B$ taking $a \mapsto a$ is an injection, and so $|A| \leqslant |B|$. For example, $\mathbb{N} \subseteq \mathbb{Z}$ and so $|\mathbb{N}| \leqslant |\mathbb{Z}|$. Also let $\mathbb{Z}_{\text{odd}}$ (resp. $\mathbb{Z}_{\text{even}}$) denote the set of all odd

(resp. even) integers. We have $|\mathbb{Z}_{\text{odd}}| \leqslant |\mathbb{Z}|$ and $|\mathbb{Z}_{\text{even}}| \leqslant |\mathbb{Z}|$. Similarly, $|\mathbb{N}_{\text{odd}}| \leqslant |\mathbb{N}|$ and $|\mathbb{N}_{\text{even}}| \leqslant |\mathbb{N}|$ for natural numbers.

(2) The canonical inclusion map $\mathbb{Z} \to \mathbb{Q}$ that takes $a \mapsto a/1$ is an injection, and so $|\mathbb{Z}| \leqslant |\mathbb{Q}|$. Also $\mathbb{Z}$ is canonically embedded in $\mathbb{R}$ and so $|\mathbb{Z}| \leqslant |\mathbb{R}|$. Likewise, $|\mathbb{Q}| \leqslant |\mathbb{R}|$.

(3) What is initially confusing about infinite sets is that even when $A \subseteq B$, there may exist an injective map $B \to A$ implying that $|B| \leqslant |A|$. As an example, consider the injection $f : \mathbb{Z} \to \mathbb{N}$ defined as $f(0) = 1, f(1) = 2, f(-1) = 3, f(2) = 4, f(-2) = 5, \ldots, f(n) = 2n, f(-n) = 2n + 1, \ldots$. This implies $|\mathbb{Z}| \leqslant |\mathbb{N}|$.

Two sets A, B are called *equinumerous*, denoted $|A| = |B|$, if

$$|A| \leqslant |B| \text{ and } |B| \leqslant |A|,$$

or equivalently if there exist an injective map $f : A \to B$ and an injective map $g : B \to A$.

For example, we have proved that $|\mathbb{N}| \leqslant |\mathbb{Z}|$ and $|\mathbb{Z}| \leqslant |\mathbb{N}|$. It follows that:

$$|\mathbb{Z}| = |\mathbb{N}|\,,$$

i.e., $\mathbb{Z}$ is of the same size as $\mathbb{N}$. The inclusion function $\iota : \mathbb{N} \to \mathbb{Z}$ is injective but not surjective, whereas the function of Part (3) of Example 4.1 is a bijection. We may indeed propose an injective but non-surjective function $g : \mathbb{Z} \to \mathbb{N}$ as $g(0) = 1, g(1) = 2, g(-1) = 4, g(2) = 5, g(-2) = 7, \ldots, g(n) = 3n - 1, g(-n) = 3n + 1, \ldots$, the image of which does not include the positive multiples of 3.

It is natural to expect two sets A, B to be equinumerous if there exists a bijection $h : A \to B$ between them. Clearly, the existence of such a function implies $|A| \leqslant |B|$ (h is injective) and $|B| \leqslant |A|$ (h^{-1} is injective). The converse of this is not immediately clear, i.e., the existence of injective functions $f : A \to B$ and $g : B \to A$ does not immediately imply the existence of a bijection $h : A \to B$. This is, however, true, as is proved now.

4.2 Theorem [*Cantor-Schröder-Bernstein theorem*] Two sets A, B are equinumerous if and only if there exists a bijection $h : A \to B$ between them.

Proof [If] Obvious.

[Only if] Let $f : A \to B$ and $g : B \to A$ be injections. We need to construct a bijection $h : A \to B$. To that effect we construct a subset S of A and define h as

$$h(x) = \begin{cases} f(x) & \text{if } x \in S, \\ g^{-1}(x) & \text{if } x \notin S. \end{cases}$$

The construction of the subset S is quite tricky. If g is surjective, g^{-1} is already a bijection $A \to B$, and we take $S = \emptyset$. So we assume that g is not surjective.

The function g, surjective or not, yields a bijection $g^{-1} : g(B) \to B$. The elements of $A \setminus g(B)$ lie outside the domain of g^{-1}. So h needs to use f in order to map elements of $A \setminus g(B)$ and we start with $S_0 = A \setminus g(B)$.

Now suppose that there is an element $x \in S_0$ for which $f(x) = g^{-1}(y)$ for some element $y \in g(B)$ (i.e., $y \notin S_0$). We have already defined $h(x) = f(x)$ since $x \in S_0$. Since h is going to be injective, we cannot define $h(y) = g^{-1}(y)$, i.e., we must define $h(y) = f(y)$. Notice that $y = g(g^{-1}(y)) = g(f(x)) = (g \circ f)(x)$ with $x \in S_0$. It follows that h needs to use f for mapping elements of $S_1 = (g \circ f)(S_0)$.

Now take $x \in S_1$. There exists $y \in A$, $y \notin S_0 \cup S_1$, such that $f(x) = g^{-1}(y)$. Since we have already taken $h(x) = f(x)$, we must define $h(y) = f(y)$, i.e., we must use f in order to define h on elements of $S_2 = (g \circ f)(S_1)$.

Proceeding in this way we define S_k inductively as:

$$\begin{aligned} S_0 &= A \setminus g(B) = (g \circ f)^0(A \setminus g(B)), \\ S_k &= (g \circ f)(S_{k-1}) = (g \circ f)^k(A \setminus g(B)) \text{ for } k \geqslant 1. \end{aligned}$$

Finally, we take:

$$S = \bigcup_{k \geqslant 0} S_k = \bigcup_{k \geqslant 0} (g \circ f)^k(A \setminus g(B)).$$

I now formally establish that this construction works, i.e., h defined as above with respect to this S is indeed a bijection. First notice that $S \supseteq S_0 = A \setminus g(B)$, so that $A \setminus S \subseteq g(B)$, that is, $h(x)$ is defined for every $x \notin S$. It is also defined for every $x \in S$, i.e., h is well-defined.

Claim: h is injective.

Suppose that $h(x) = h(y)$ for some $x, y \in A$. If both $x, y \in S$, then $x = y$ by the injectivity of f. If both $x, y \in A \setminus S$, then the injectivity of g^{-1} implies $x = y$. So assume that one of x, y is in S, the other in $A \setminus S$. Suppose $x \in S$ and $y \notin S$. By definition of S we have $x \in S_k$ for some $k \geqslant 0$. Also since $y \notin S$, we have $y \in g(B)$. Therefore, $y = g(g^{-1}(y)) = g(h(y)) = g(h(x)) = g(f(x)) = (g \circ f)(x)$, i.e., $y \in S_{k+1}$, i.e., $y \in S$, a contradiction. So it is not possible to have $x \in S$ and $y \notin S$.

Claim: h is surjective.

Take an arbitrary $b \in B$. We need to produce an $a \in A$ for which $h(a) = b$. Consider $x = g(b) \in A$. If $x \notin S$, then $h(x) = g^{-1}(x) = g^{-1}(g(b)) = b$, i.e., we take $a = x$. So suppose that $x \in S$. By construction of S, we have $x \in S_k$ for some $k \geqslant 0$, i.e., $x = (g \circ f)^k(y)$ for some $y \in S_0$. Since $x = g(b) \in g(B)$, we cannot have $k = 0$, i.e., $k \geqslant 1$. Take $a = (g \circ f)^{k-1}(y) \in S_{k-1} \subseteq S$. But then $h(a) = f(a) = g^{-1}(g(f(a)) = g^{-1}((g \circ f)((g \circ f)^{k-1}(y))) = g^{-1}((g \circ f)^k(y)) = g^{-1}(x) = b$. ◀

4.3 Example As an illustration of the Cantor-Schröder-Bernstein theorem, take $A = \mathbb{N}$ and $B = \mathbb{N}_{\text{even}}$ (the set of even natural numbers). Also take $f : A \to B$ as $f(a) = 4a$ and $g : B \to A$ as $g(b) = b$. Both f and g are injective, but neither of them is bijective. We have $g(B) = \{2, 4, 6, 8, 10, \ldots, 2n, \ldots\}$, and so

$$\begin{aligned} S_0 &= \{1, 3, 5, 7, \ldots, 2n-1, \ldots\} = \{2n-1 \mid n \in \mathbb{N}\}, \\ S_1 &= \{2^2(2n-1) \mid n \in \mathbb{N}\}, \\ S_2 &= \{2^4(2n-1) \mid n \in \mathbb{N}\}, \\ &\ldots \\ S_k &= \{2^{2k}(2n-1) \mid n \in \mathbb{N}\}. \end{aligned}$$

Therefore, S comprises positive integers in which the multiplicities of 2 are even. The values of $h(a)$ are listed below for some small values of a.

a	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	$\cdots$
$h(a)$	4	2	12	16	20	6	28	8	36	10	44	48	52	14	60	64	68	18	$\cdots$

This bijection is different from the standard bijection $\mathbb{N} \to \mathbb{N}_{\text{even}}$ that maps $n \mapsto 2n$.

4.3 Countable sets

Not all infinite sets have the same size. Every infinite set does not have bijective correspondence with the set $\mathbb{N}$ of natural numbers. Indeed $\mathbb{N}$ (or $\mathbb{Z}$) happens to be the smallest infinite set, i.e., any infinite set must be at least as big as $\mathbb{N}$ is. For a proof, take any infinite set A and define a map $f : \mathbb{N} \to A$ as follows. Pick an element a_1 from A and set $f(1) = a_1$. Then pick another element a_2 from A and set $f(2) = a_2$. Assume that for some $n \in \mathbb{N}$ pairwise distinct elements $a_1, a_2, \ldots, a_n$ have been chosen from A. Since A is infinite, there remains an element of A not chosen so far. Pick any such element a_{n+1} from A and set $f(n+1) = a_{n+1}$. By induction, we then have a well-defined injective function $f : \mathbb{N} \to A$, and consequently $|\mathbb{N}| \leqslant |A|$.

It turns out that for some infinite sets A the function f constructed as above cannot be surjective irrespective of how we choose the elements $a_1, a_2, \ldots, a_n, \ldots$. This happens because A has a size strictly bigger than that of $\mathbb{N}$, and so there cannot exist any surjective function from $\mathbb{N}$ onto A.

Elements of a finite set A can be counted, i.e., we can build an injective function $f : \mathbb{N}_n \to A$ as above, where $n = |A|$. This process stops after all of the n elements are picked from A. So finite sets are called *countable*.

An infinite set A for which a bijection $f : \mathbb{N} \to A$ can be established is also called *countable*. Since f is bijective (in particular surjective), every element $a \in A$ is the image of some $n \in \mathbb{N}$ under f. This means that a has been picked during the n-th step of the counting process. The counting process here is infinite, but the guarantee that every element of A is considered in finite time prompts us to treat A as countable.

In order to prove an infinite countable set A to be so, one may produce a bijection $f : \mathbb{N} \to A$. In other words, one may supply a way of numbering elements of A so that every element of A is covered in the process and no element is counted more than once in the process. Since A is infinite, we anyway have $|\mathbb{N}| \leqslant |A|$. It then suffices to show that $|A| \leqslant |\mathbb{N}|$ (See the Cantor-Schröder-Bernstein theorem). That is, it suffices to produce an injective map $A \to \mathbb{N}$. To sum up, an infinite set A is countable if and only if there exists an injective map $f : A \to \mathbb{N}$. In this assertion, $\mathbb{N}$ can be replaced by any set that is already known to be countable.

I will now furnish some countability proofs.

4.4 Proposition Any subset of a countable set is again countable.

Proof Let A be a countable set and $B \subseteq A$. If B is finite, it is countable. Otherwise, consider the inclusion map $B \to A$, $b \mapsto b$, which is injective. ◀

This result implies that the sets $\mathbb{Z}_{\text{odd}}$, $\mathbb{Z}_{\text{even}}$, $\mathbb{N}_{\text{odd}}$ and $\mathbb{N}_{\text{even}}$ are countable.

4.5 Proposition The union of two countable sets is again countable. More generally, the union of any finite number of countable sets is again countable.

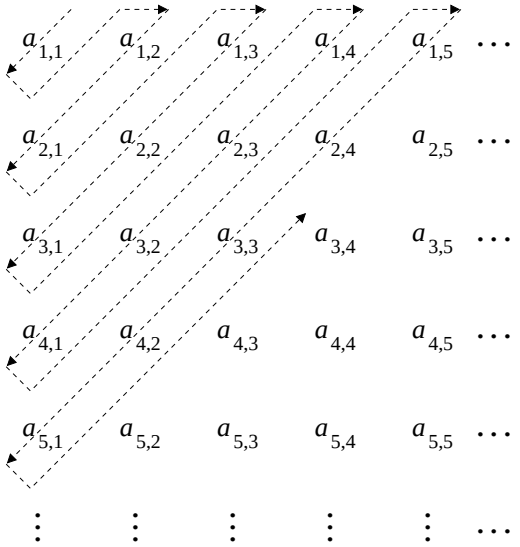
Proof Let A, B be two countable sets. If one or both of the sets is/are finite, then $A \cup B$ is evidently countable. So assume that both A and B are infinite. Let $a_1, a_2, \ldots, a_n, \ldots$ and $b_1, b_2, \ldots, b_n, \ldots$ be exhaustive listings of the elements of A and B respectively. I need to produce an exhaustive listing of the elements of $A \cup B$. If we first list the elements of A, followed by the elements of B, we encounter a trouble. Here the listing of the elements of A does not terminate after finitely many steps, and so the elements of B do not receive a chance of getting listed. A proper listing of the elements of $A \cup B$ can be $a_1, b_1, a_2, b_2, a_3, b_3, \ldots, a_n, b_n, \ldots$. There is a small catch here: the sets A and B need not be disjoint. In case of a repetition, the second occurrence of an element is skipped from the list.

For proving the generalized assertion, let k countable sets $A_1, \ldots, A_k$ be provided. We need to show that $A = \bigcup_{i=1}^{k} A_i$ is countable. We proceed by induction on k. The result trivially holds for $k = 0, 1$. So

suppose that the union B of $k-1$ sets $A_1, \ldots, A_{k-1}$ is countable. But then $A = B \cup A_k$ is the union of two countable sets and is countable too. ◄

4.6 Proposition The union of a countable number of countable sets is again countable.

Proof Let A_n, $n \in \mathbb{N}$, be a family of countable sets. We plan to show that $A = \bigcup_{n\in\mathbb{N}} A_n$ is countable. Once again I need to supply a listing of the elements of A in which every element appears after finitely many steps. Let $a_{i,j}$ be the j-th element in a given listing of A_i (each A_i is countable). The following figure depicts a way of combining the lists.



We first list $a_{1,1}$. We then list $a_{1,2}, a_{2,1}$, then $a_{1,3}, a_{2,2}, a_{3,1}$, and so on. More concretely, we list $a_{i,j}$ in the increasing order of $i + j$. For a fixed value of $i + j$, we list the elements $a_{i,j}$ in the increasing order of i. Of course, we exclude all repetitions in the list, i.e., if some element $a_{i,j}$ has already been encountered as $a_{i',j'}$, then we do not insert $a_{i,j}$ again in the list. Also notice that some of the sets A_n may be finite and so have only finite listings. In that case the elements $a_{n,i}$ are not defined after all elements of A_n are exhausted. During the above diagonal-wise listing, we skip all empty locations where no defined elements reside. ◄

4.7 Proposition The Cartesian product $A \times B$ of two countable sets A, B is countable. More generally, the Cartesian product of a finite number of countable sets is again countable.

Proof For each $a \in A$ the set

$$B_a = \{(a, b) \mid b \in B\}$$

is in bijective correspondence with B and so is countable. Therefore, $A \times B = \bigcup_{a\in A} B_a$ is the union of countably many countable sets and is countable.

The general statement can be proved easily by induction on the number of sets in the given collection. ◄

But that is all. The Cartesian product of countably many countable sets is, in general, not countable.

4.8 Proposition The set $\mathbb{Q}$ of rational numbers is countable.

Proof Every element of $\mathbb{Q}$ has a normalized representation of the form a/b with $a \in \mathbb{Z}$, $b \in \mathbb{N}$, and $\gcd(a, b) = 1$. Thus each rational number is identified by a pair of integers, i.e., we can view $\mathbb{Q}$ as a subset of $\mathbb{Z} \times \mathbb{N}$. The result then follows from Propositions 4.4 and 4.7. ◄

Maths-savvy students may note that *the* countable infinity is denoted by the symbol $\aleph_0$ pronounced "aleph-not". We have essentially proved the following assertions about $\aleph_0$.

$$\begin{aligned} \aleph_0 + \aleph_0 &= \aleph_0. \\ k\aleph_0 &= \aleph_0 \text{ for every } k \in \mathbb{N}. \\ \aleph_0 \times \aleph_0 &= \aleph_0. \\ \aleph_0^k &= \aleph_0 \text{ for every } k \in \mathbb{N}. \end{aligned}$$

4.4 Proving uncountability using diagonalization

A set A is called *uncountable* if it is not countable, i.e., if A cannot have any bijection with $\mathbb{N}$. Proving that no bijection can exist between A and $\mathbb{N}$ is not as easy task. A technique called *diagonalization* often helps us here. We start with the assumption that A does possess a bijective correspondence with $\mathbb{N}$. Then using diagonalization we arrive at a contradiction implying that our assumption about the countability of A is false. I now present some proofs based on diagonalization.

4.9 Proposition The set $\mathbb{R}$ of real numbers is uncountable.

Proof I will prove that the interval

$$[0, 1) = \{x \in \mathbb{R} \mid 0 \leqslant x < 1\}$$

is uncountable. If $\mathbb{R}$ were countable, the interval $[0, 1)$ would be countable too (Proposition 4.4). So the following proof suffices.

Notice first that every real number has a decimal expansion. For real numbers in the interval $[0, 1)$, the integer part is 0 and it suffices to concentrate only on the expansion following the decimal point. We consider infinite decimal expansions only. For example, $1/3 = 0.3333333333\ldots$, $\pi - 3 = 0.1415926535\ldots$, etc. If the decimal expansion of some real number is terminating, we append an infinite number of 0's in the expansion. For example, $3/8 = 0.3750000000\ldots$. Some real numbers do have two infinite expansions, like $3/8 = 0.3750000000\ldots = 0.3749999999\ldots$. In such a case we pick one of these two expansions arbitrarily.

Assume that $[0, 1)$ is countable. Then there exists a bijection $f : \mathbb{N} \to [0, 1)$. We write down the decimal expansion of each $f(n)$ (resolving ambiguities arbitrarily, if necessary) as follows. Here each $a_{n,i}$ is a decimal digit (an integer between 0 and 9).

$$\begin{array}{rl} f(1) = & 0\,.\,\underline{a_{1,1}}\; a_{1,2}\; a_{1,3}\; a_{1,4}\; a_{1,5} \ldots a_{1,n} \ldots \\ f(2) = & 0\,.\,a_{2,1}\; \underline{a_{2,2}}\; a_{2,3}\; a_{2,4}\; a_{2,5} \ldots a_{2,n} \ldots \\ f(3) = & 0\,.\,a_{3,1}\; a_{3,2}\; \underline{a_{3,3}}\; a_{3,4}\; a_{3,5} \ldots a_{3,n} \ldots \\ f(4) = & 0\,.\,a_{4,1}\; a_{4,2}\; a_{4,3}\; \underline{a_{4,4}}\; a_{4,5} \ldots a_{4,n} \ldots \\ f(5) = & 0\,.\,a_{5,1}\; a_{5,2}\; a_{5,3}\; a_{5,4}\; \underline{a_{5,5}} \ldots a_{5,n} \ldots \\ \ldots & \\ f(n) = & 0\,.\,a_{n,1}\; a_{n,2}\; a_{n,3}\; a_{n,4}\; a_{n,5} \ldots \underline{a_{n,n}} \ldots \\ \ldots & \\ \hline b = & 0\,.\; b_1 \quad b_2 \quad b_3 \quad b_4 \quad b_5 \quad \ldots \quad b_n \quad \ldots \end{array}$$

We now look at the diagonal digits $a_{1,1}, a_{2,2}, \ldots, a_{n,n}, \ldots$ in order to construct the decimal expansion $0.b_1b_2 \ldots b_n \ldots$ of a real number b. We take the n-th digit b_n of b as:

$$b_n = \begin{cases} 2 & \text{if } a_{n,n} = 1, \\ 1 & \text{otherwise.} \end{cases}$$

Clearly, $b \in [0, 1)$. Since f is surjective, we must have $b = f(n)$ for some $n \in \mathbb{N}$. But by the construction of b, the n-th digits of b and $f(n)$ are different. Moreover, the decimal expansion of b consists only of 1's and 2's, and so b has a unique expansion. Therefore, we cannot have $b = f(n)$, a contradiction. ◀

The above example illustrates the construction of an element b of an uncountable set A, which is forced to differ from each $f(n)$ at at least one point. So we cannot have $b = f(n)$ for any $n \in \mathbb{N}$, i.e., no map $f : \mathbb{N} \to A$ can be surjective. Usually, the point at which b differs from $f(n)$ lies on the diagonal in a pictorial representation. This is why the name diagonalization is used.

A slight modification of the above proof implies the following result. Here Σ is any finite set of size at least 2. For example, Σ may be the binary alphabet $\{0, 1\}$, or the decimal alphabet $\{0, 1, 2, \ldots, 9\}$ or the Roman alphabet $\{a, b, c, \ldots, z\}$.

4.10 Proposition The set of infinite sequences over Σ is uncountable. ◀

Thus the Cartesian product of countably many finite sets need not be countable. In fact,

$$k^{\aleph_0} > \aleph_0 \text{ for any integer } k \geqslant 2\,.$$

Now I will prove another important result using diagonalization.

4.11 Proposition For any set A, there cannot exist a bijection between A and its power set $\mathcal{P}(A)$. In particular, the number of subsets of any countably infinite set is uncountable.

Proof Take any set A, and assume that $f : A \to \mathcal{P}(A)$ is a bijection. Construct a subset B of A as follows:

$$B = \{a \in A \mid a \notin f(a)\}.$$

Since f is surjective, there exists $a \in A$ such that $B = f(a)$. If $a \in f(a)$, then $a \notin B$ (This is how B is constructed). But $B = f(a)$, and so $a \notin f(a)$. On the other hand, if $a \notin f(a)$, then $a \in B$ (by the construction of B), i.e., $a \in f(a)$ (since $B = f(a)$). Thus we have proved that $a \in f(a) \iff a \notin f(a)$. This is absurd. ◀

In the above proof, the set B is forced to differ from each $f(a)$ with respect to the inclusion of a. If a belongs to $f(a)$, we do not include a in B. On the other hand, if a does not belong to $f(a)$, we include a in B. In this way B is forced to lie outside the range of f. If you still wonder what is diagonal in the above argument, here is a visualization for a special case. Take $A = \mathbb{N}$. For $B \subseteq A$, define the characteristic function $C_B : A \to \{0, 1\}$ as:

$$C_B(a) = \begin{cases} 1 & \text{if } a \in B, \\ 0 & \text{if } a \notin B. \end{cases}$$

Now I make a two-dimensional listing of $C_{f(n)}(i)$, where n runs over the rows, and i over the columns.

		$C_{f(n)}(i)$					
n	$f(n)$	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$\cdots$
1	$\emptyset$	$\underline{0}$	0	0	0	0	$\cdots$
2	$\{2, 4, 6, 8\}$	0	$\underline{1}$	0	1	0	$\cdots$
3	$\{2, 3, 5, 7, 11, \ldots\}$	0	1	$\underline{1}$	0	1	$\cdots$
4	$\{1, 3, 5, 7, 9, \ldots\}$	1	0	1	$\underline{0}$	1	$\cdots$
5	$\{1, 2, 3, 5, 8, 13, \ldots\}$	1	1	1	0	$\underline{1}$	$\cdots$
$\cdots$							$\cdots$
B	$\{1, 4, \ldots\}$	1	0	0	1	0	$\cdots$

Here B is constructed to force that $C_B(n)$ is different from the diagonal value $C_{f(n)}(n)$.

The set $\mathbb{R}$ of real numbers is uncountable. Its size is denoted by c, as $\mathbb{R}$ is often referred to as the *continuum*. We have proved that

$$c > \aleph_0 .$$

There exist infinite sets whose sizes are bigger than c. For example, the power set of $\mathbb{R}$ is of size strictly bigger than c.

Does there exist a set of size strictly between $\aleph_0$ and c? The answer is not known. Georg Cantor conjectured that no such set exists. This famous unproven conjecture is referred to as the *continuum hypothesis*.

4.5 Application: Computers cannot solve all problems

Computer theoreticians have devised a way to characterize computational problems mathematically. This characterization requires the introduction of a series of definitions.

An *alphabet* is a finite set. Members of an alphabet are called *symbols*. We use upper-case Greek letters like Σ, Γ for naming alphabets. An alphabet is typically used to represent languages. For example, the Roman alphabet ('a' through 'z', space, and punctuation marks) is used to express the language of English in written form. Similarly, the alphabet consisting of digits (0 through 9), the decimal point, and the signs ($+$ and $-$) can be used to represent the language of real numbers. If we want to write complex numbers, we use an additional symbol i (some people use j or ι instead) in the representation alphabet. In all these examples, we describe infinite sets using finitely many symbols.

A *string* over an alphabet Σ is a finite sequence of symbols from Σ. For example, 'abracadabra', 'zyzzyva', 'Madam, I'm Adam!' are English strings, $+0.1123581321$, -435 are numeric strings. A string is finite in length by definition, and is different from a set in that the order of the symbols in the sequence is important, and repetitions of symbols are allowed in the sequence. For example, the numeric strings 1231, $+1231$, 01231 are distinct from one another (although they refer to the same value). Moreover, 1231 is different from 1123 and also from 123.

The set of all strings over an alphabet Σ is denoted by Σ^*. An important fact about Σ^* is the following:

4.12 Proposition Σ^* is countably infinite.

Proof The *length* of a string is the number of symbols in the string. For example, the length of 1231 is 4, and the length of $+01231$ is 6. We can write Σ^* as the union of Σ^l for $l = 0, 1, 2, \ldots$, where Σ^l is the set of all strings over Σ having length l. Each Σ^l has finite size (namely $|\Sigma|^l$), and so is countable. Thus Σ^* is the union of countably many countable sets. ◄

The above proposition remains valid even if we allow Σ to be countably infinite.

A *language* over Σ is a subset of Σ^*. For example, the language of English consists precisely of those strings (over the Roman alphabet) that has an interpretation in English. Thus, 'Madam, I'm Adam!' is in the English language, whereas 'I Adam, Madam am!' is not in the English language. Similarly, $+0.12.345$ is not a numeric string (i.e., not in the language of real numbers).

The set of all languages over Σ is precisely the power set $\mathcal{P}(\Sigma^*)$ of Σ^*.

4.13 Proposition $\mathcal{P}(\Sigma^*)$ is uncountable.

Proof No set can have a bijective correspondence with its power set. ◄

Now suppose that we have an alphabet Σ and we want to represent a language over Σ. The representation is done by a *finite* description called a *grammar* for the language. For example, the English language is described by the English grammar. The language of real numbers is described by the rule: "optionally a sign, followed by zero or more (but finitely many) digits, followed optionally by the decimal point and another finite sequence of zero or more decimal digits". Of course, you may disallow exceptions like '$.$', '$+.$', and '$-.$'.

Examples of languages with better computational flavor include the language of primes (those decimal strings representing positive primes), the language of valid C programs (the C language), and so on. The language of primes have a mathematical description (something like $(a \mid p) \Rightarrow (a = 1) \vee (a = p)$). The C language is specified by a grammar that compilers should follow while compiling a program.

In all these examples, a language is specified by a grammar which itself is a string over some alphabet Γ. Notice that the representation alphabet Γ may be different from the alphabet Σ of the language being described. That is not a big issue as long as both Σ and Γ remain finite (or countable). What is more important here is that any grammar has to be finite. For example, your C compiler cannot follow an infinite grammar, for if so, some programs would require infinite time during compilation.

A (computational) *problem* is defined as follows: Given the finite representation of a language L over Σ and a string α over Σ, determine whether $\alpha \in L$. Each language is, therefore, a problem, and conversely!

If L itself is finite, we can list the strings in L one by one. The list fits in a finite amount of space, and is a grammar for L. One can exhaustively search for α in the list. On the other hand, if L is infinite, an exhaustive listing of the elements of L is not finite, and a grammar for L has to resort to other means. But then some procedure must also be specified in order to identify whether α can be generated using the rules in this grammar. This procedure is precisely what we mean by *computation*.

Suppose that we are interested in languages over the alphabet Σ. Suppose also that we use the representation alphabet Γ for writing grammars. (Since symbols in any alphabet can be encoded in binary, you can take $\Sigma = \Gamma = \{0, 1\}$.) The set of languages over Σ is $\mathcal{P}(\Sigma^*)$ which is proved earlier to be uncountable. On the other hand, Γ^* is countable, so we can write grammars for only countably many languages. Ability to express a language by a grammar does not immediately lead to an algorithm to solve the corresponding problem. That is, the set of problems that have algorithms (that is, that can be solved by computers) is a (potentially strict) subset of the set of languages that can be represented by grammars, and so is countable too. It turns out that we can solve only countably many problems using computers, whereas the number of problems is uncountable. It then follows that computers cannot solve uncountably many problems (in fact, more problems than they can solve).

This dark reality follows from simple counting (well, countability) arguments. Locating problems that cannot be solved by computers (and proving them to be unsolvable) is, however, not an easy task. We often use diagonalization proofs to this effect. All these are topics to be discussed in a course on formal languages and automata theory (or in an advanced sequel to that course).

Exercises

4.1 Prove that the sets S_k, $k \geqslant 0$, used in the proof of the Cantor-Schröder-Bernstein theorem are pairwise disjoint.

4.2 Determine the set S and the corresponding bijection for the maps $f : \mathbb{N} \to \mathbb{Z}$ and $g : \mathbb{Z} \to \mathbb{N}$, where f is the inclusion map, and g is defined as $g(0) = 1, g(1) = 2, g(-1) = 4, g(2) = 5, g(-2) = 7, \ldots, g(n) = 3n - 1, g(-n) = 3n + 1, \ldots$.

4.3 Determine the set S and the corresponding bijection for the maps $f : \mathbb{N}_{\text{even}} \to \mathbb{N}$ and $g : \mathbb{N} \to \mathbb{N}_{\text{even}}$ defined by $f(a) = a$ and $g(b) = 4b$. Argue that this bijection is the inverse of the map $\mathbb{N} \to \mathbb{N}_{\text{even}}$ constructed in Example 4.3.

4.4 Prove that a set is countable if and only if it has bijective correspondence with a subset of $\mathbb{N}$.

4.5 Let A be a countable set. Prove that:

(a) The set of all finite subsets of A is countable.

(b) The set of all infinite subsets of A is uncountable.

4.6 **(a)** Let $\mathbb{Z}[X]$ denote the set of polynomials in one indeterminate X and with integer coefficients. Prove that $\mathbb{Z}[X]$ is countable.

(b) Let k be a fixed positive integer. Prove that the set $\mathbb{Z}[X_1, X_2, \ldots, X_k]$ of multivariate polynomials with integer coefficients is countable.

(c) Prove that the set $\mathbb{Z}[X_1, X_2, \ldots, X_k, \ldots]$ of polynomials with countably infinite indeterminates and with integer coefficients is countable.

4.7 Prove that the set $\mathbb{Z}[[X]]$ of power series with integer coefficients is uncountable.

4.8 A real or complex number α is called *algebraic* if $f(\alpha) = 0$ for some non-zero polynomial $f(X)$ with integer coefficients. Let $\mathbb{A}$ denote the set of all algebraic numbers. (We have $\mathbb{A} \subseteq \mathbb{C}$.)

(a) Prove that $\mathbb{A}$ is countable.

(b) Conclude that there are uncountably many transcendental numbers.

4.9 Let A be a countably infinite set and B a finite set. Prove that:

(a) The set of all functions $A \to B$ is uncountable.

(b) The set of all functions $B \to A$ is countable.

4.10 **(a)** Let a, b be real numbers with $a < b$. Supply an explicit bijection between the intervals $[0, 1)$ and $[a, b)$.

(b) Suggest an explicit bijection between the interval $[0, 1)$ and the entire real line $\mathbb{R}$.

4.11 Let A, B be sets, where A is equinumerous with $\mathbb{R}$ and B is equinumerous with $\mathbb{N}$. Prove that $A \cup B$ is equinumerous with $\mathbb{R}$. (This means $c + \aleph_0 = c$.)

4.12 Let A be a countable set. Prove that the set of all functions $A \to \{0, 1\}$ is equinumerous with $\mathbb{R}$ (i.e., $2^{\aleph_0} = c$). Conclude that the power set $\mathcal{P}(A)$ is equinumerous with $\mathbb{R}$.

4.13 Prove that the set of all permutations of a countable set is not countable. (One can show $\aleph_0! = c$.)

4.14 Prove that the union of two sets each equinumerous with $\mathbb{R}$ is again equinumerous with $\mathbb{R}$ (i.e., $c + c = c$).

4.15 Prove that the union of countably many sets each equinumerous with $\mathbb{R}$ is again equinumerous with $\mathbb{R}$ (i.e., $\aleph_0 \times c = c$).

4.16 Prove that the real interval $[0, 1)$ is equinumerous with the two-dimensional square $[0, 1) \times [0, 1)$ (i.e., $c \times c = c$).

4.17 **(a)** Prove that we can represent every integer in finite space using only finitely many symbols (digits and signs).

(b) Prove that we can represent every rational number in finite space using only finitely many symbols.

(c) Prove that the set of real numbers that have finite decimal expansions is countable. (This set is a subset of $\mathbb{Q}$.)

(d) Prove that every rational number has terminating or repeating decimal expansion. Enclose the repeating part in a decimal expansion by a pair of curly braces. For example, $1/3 = 0.\{3\}$, $1.2\{142857\} = 1 + (2/10) + (1/70) = 85/70 = 17/14$. Conclude that the inclusion of the extra symbols { and } lets us represent each rational number in finite space only.

(e) Prove that the set of finite arithmetic expressions involving rational numbers represented as in the previous part, arithmetic operators ($+$, $-$, $\times$, and $/$), and parentheses is countable.

(f) Argue that allowing countably many symbols representing square, cube, ... roots in arithmetic expressions leaves the set of finitely representable numbers countable.

(g) Now allow well-known transcendental numbers like π, e in arithmetic expressions. Prove that the set of finitely representable numbers still remains countable.