Size of Sets and Definable Languages

Size of a Set

- We start our discussion about the size of a set. It has some purpose in connection to description of languages.
- A set is finite, if it has finite number of elements and its size is the number of its elements
- The size of $A = \{a, b, c, d\}$ is 4, the size of the set of all prime numbers between 1 to 100 is 25 etc.

Size of a Set

- The size of a finite set A is larger than the size of a finite set B, if A has more elements than B.
- But how do we compare two infinite sets e.g. the set of natural numbers, $\mathbb{N} = \{0, 1, 2, \cdots\}$ and the set of integers, $\mathbb{Z} = \{\cdots, -1, 0, 1, \cdots\}$?

Size of a Set

- In an obvious sense the set of integers is larger than the set of natural numbers, as $\mathbb{N} \subset \mathbb{Z}$.
- But in some other sense we can establish a one-to-one correspondence between the elements of these two sets.

Size of a Set

$$\begin{array}{cccccccc} \cdots & 6 & 4 & 2 & 1 & 3 & 5 & \cdots \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ \cdots & -3 & -2 & -1 & 0 & 1 & 2 & \cdots \end{array}$$

We have a bijection $f : \mathbb{N} \rightarrow \mathbb{Z}$,

$$f(n) = \begin{cases} (n-1)/2 & \text{if } n \text{ is odd}, \\ -n/2 & \text{if } n \text{ is even}. \end{cases}$$

The set of integers is equinumerous to the set of natural numbers ($\mathbb{N} \simeq \mathbb{Z}$).

Size of a Set

- Similarly we can establish a bijection between the set of integers ($\mathbb{Z}$) and the set of rational numbers ($\mathbb{Q}$).
- We have a bijection $f_1 : \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N}$, $f_1(n) = (a+1, b+1)$, where $n = 2^a(2b+1)$ and a, b are non-negative integers.

Size of a Set

$$
\begin{array}{cccc}
(1,1)1 & (1,2)3 & (1,3)5 & \cdots \\
(2,1)2 & (2,2)6 & (2,3)10 & \cdots \\
(3,1)4 & (3,2)12 & (3,3)20 & \cdots \\
(4,1)8 & (4,2)24 & (4,3)40 & \cdots \\
\vdots & \vdots & \vdots & \vdots
\end{array}
$$

Size of a Set

The functions f and f_1 can be used to construct a bijection from $\mathbb{N} \to \mathbb{Z} \times \mathbb{N}$.

$$(f \times id_{\mathbb{N}}) \circ f_1 : \mathbb{N} \to \mathbb{N} \times \mathbb{N} \to \mathbb{Z} \times \mathbb{N}.$$

As an example

$$(f \times id_{\mathbb{N}})((f_1(40)) = (f \times id_{\mathbb{N}})(4, 3) = (-2, 3),$$

as $40 = 2^3(2 \times 2 + 1)$, so
$f_1(40) = (3 + 1, 2 + 1) = (4, 3)$, and $f(4) = -1$.

Size of a Set

- As $(f \times id_{\mathbb{N}}) \circ f_1$ is a bijection between $\mathbb{N}$ and $\mathbb{Z} \times \mathbb{N}$, its inverse is also a bijection and therefore is an injective function from $\mathbb{Z} \times \mathbb{N} \rightarrow \mathbb{N}$.
- According to the SchröderBernstein theorem, if there are injective functions $f : A \rightarrow B$ and $g : B \rightarrow A$, there is a bijection between A and B i.e. $A \simeq B$.

Size of a Set

- The set of rational numbers $\mathbb{Q} \subseteq \mathbb{Z} \times \mathbb{N}$. An injective function from $\mathbb{Z} \times \mathbb{N} \rightarrow \mathbb{N}$ gives an injective function from $\mathbb{Q} \rightarrow \mathbb{N}$.
- In the other direction, an injective function from $\mathbb{N} \rightarrow \mathbb{Q}$ is $n \mapsto (n, 1)$ i.e. $n/1$.
- So by SchröderBernstein theorem there is a bijection from $\mathbb{N} \rightarrow \mathbb{Q}$ i.e. they are equinumerous ($\mathbb{N} \simeq \mathbb{Q}$).

Size of a Set

- The obvious question is - are all infinite sets equinumerous?
- The famous theorem of Cantor[a] gives the negative answer.
- No set is equinumerous to its power set i.e. there cannot be a bijection from a set $A \to \mathcal{P}A$.

[a] Georg Ferdinand Ludwig Philipp Cantor, German mathematician, invented set theory, 1845-1918

Size of a Set

- The statement is trivially true if $A = \emptyset$ as its power set $\{\emptyset\}$ has one element.
- If the A is non-empty, there is an obvious injective function from $A \rightarrow \mathcal{P}A : a \mapsto \{a\}$.
- So the actual claim of Cantor's theorem is that there cannot be any surjective function from $A \rightarrow \mathcal{P}A$.

Size of a Set

- The proof of the theorem is by contradiction - we assume that there is a surjective function $g : A \rightarrow \mathcal{P}A$ and show that it leads to a contradiction.
- Note that for each $a \in A$, the image $g(a)$ is a subset of A, an element of the power set of A.

Size of a Set

- Consider the subset $B = \{a \in A : \ a \notin g(a)\}$ of A. The subset B is the collection of all elements of A such that they are not element of their images under g.
- As $B \in \mathcal{P}A$ and g is a surjective map, there is an element $a_0 \in A$ so that $g(a_0) = B$.
- The question is whether a_0 is an element of B.

Size of a Set

- If we assume $a_0 \in B = g(a_0)$, we have to conclude that $a_0 \notin B$, by definition of B.
- But if we assume $a_0 \notin B = g(a_0)$, we have to conclude that $a_0 \in B$, by definition of B.
- So it is a contradiction - $a_0 \in B$ if and only if $a_0 \notin B$.
- Hence the assumption that $g : A \to \mathcal{P}A$ is a surjective map is false.

Size of a Set

There are infinite sets that are not equinumerous.

$$A \not\approx \mathcal{P}A \not\approx \mathcal{P}\mathcal{P}A \not\approx \mathcal{P}\mathcal{P}\mathcal{P}A \cdots$$

There is a hierarchy of infinite sets.

Size of a Set

In a more concrete terms we shall demonstrate that

- the set of natural numbers $\mathbb{N}$ is not equinumerous to the collection of all functions from $\mathbb{N}$ to itself, and
- the set of natural numbers $\mathbb{N}$ is not equinumerous to the collection of the set of real numbers $\mathbb{R}$.

Size of a Set

- Given a subset B of natural numbers, we consider a function $\mu_B : \mathbb{N} \rightarrow \{1, 2\}$ defined as

$$\mu_B(n) = \begin{cases} 1 & \text{if } n \in B \\ 2 & \text{if } n \notin B \end{cases}$$

- This shows that $\mathcal{P}A$ is equinumerous to all functions from $\mathbb{N}$ to $\{1, 2\}$ ($\{1, 2\}^{\mathbb{N}}$).

Size of a Set

- So $\mathbb{N}$ is not equinumerous $\{1,2\}^{\mathbb{N}}$.
- Again all functions from $\mathbb{N}$ to $\{1,2\}$ ($\{1,2\}^{\mathbb{N}}$) is a subset of all functions from $\mathbb{N}$ to $\mathbb{N}$ ($\mathbb{N}^{\mathbb{N}}$).
- So $\mathbb{N}$ cannot be equinumerous to $\mathbb{N}^{\mathbb{N}}$.
- In fact it is not difficult to show that $\mathbb{N}^k$ is also not equinumerous to $\mathbb{N}^{\mathbb{N}}$, where $k \in \mathbb{N}$.

Size of a Set

- This result is very important in connection to effectively definable or computable functions from $\mathbb{N}^k \rightarrow \mathbb{N}$, where k is a positive integer.
- There are functions that cannot be effectively defined.

Size of a Set

- To prove that $\mathbb{N} \not\simeq \mathbb{R}$ we first show that the interval $(0,1) \subseteq \mathbb{R}$ is equinumerous to $\mathbb{R}$.
- There is a bijection $\tan : (-\pi/2, \pi/2) \rightarrow (-\infty, \infty)$.
- Also there is a bijection $f_2 : (0,1) \rightarrow (-\pi/2, \pi/2) : x \mapsto \frac{\pi(2x-1)}{2}$.
- So $\tan \circ f_2$ is a bijection from $(0,1) \rightarrow \mathbb{R}$, i.e. $(0,1) \simeq \mathbb{R}$.

Size of a Set

- Finally we demonstrate that $\mathbb{N} \not\simeq (0, 1)$. The proof is again by contradiction, well known as diagonalization.
- Suppose there is a bijective map $h : \mathbb{N} \to (0, 1)$. So for every natural number i there is a non-zero proper fraction $h(i) \in (0, 1)$ and that exhausts all such fractions.

Size of a Set

- Each $h(i)$ can be written as an infinite decimal fraction, $0.h_{i1}h_{i2}\cdots h_{ij}\cdots$, where h_{ij} is a decimal digit.
- We construct a fraction $d = 0.d_1d_2\cdots d_i\cdots$ as follows.

$$d_i = \begin{cases} 4 & \text{if } h_{ii} = 5 \\ 5 & \text{otherwise.} \end{cases}$$

Size of a Set

- The fraction d is not equal to any $h(i)$ (by construction) and this contradicts our assumption that h is a bijection.
- So $\mathbb{N} \not\simeq \mathbb{R}$. It is possible to show that $\mathcal{P}\mathbb{N} \simeq \mathbb{R}$.

Size of a Set

- A set A is called finite if there is $n \in \mathbb{N}$ such that $A \simeq \{1, \cdots, n\}$.
- A set A is called countably infinite if $A \simeq \mathbb{N}$.
- A set A is called countable if it is either finite or countably infinite.
- A set A is called uncountable if it is not countable.

Alphabet, Words and Languages

- Any language has a finite set of primitive symbols known as the alphabet of the language.
- The alphabet of decimal number system is $\{0, 1, \cdots, 9, +, -, \cdot\}$. The English language alphabet has more symbols including a, $\cdots$, z, A, $\cdots$, Z, punctuation marks etc.

Alphabet, Words and Languages

- For our discussion we shall often take small size of alphabet e.g. $\{0, 1\}, \{a, b, c\}$. Symbols like Σ, Γ are used to denote an alphabet.
- A finite sequence (possibly empty) of the elements (called letters) of the alphabet Σ is called a finite word.
- Similarly an infinite sequence of letters is called an infinite word.

Alphabet, Words and Languages

- Let our alphabet be $\Sigma = \{a, b\}$. Finite words over Σ are $\varepsilon(\text{empty word}), a, b, aa, aba, \cdots$.
- A finite word $x = \sigma_1 \cdots \sigma_n$, where $\sigma_i \in \Sigma$ is of length n and we write $|x| = n$.
- The word x may be viewed as an element of Σ^n or a map $x : \{1, \cdots, n\} \to \Sigma$, where $x(i) = \sigma_i$.

Alphabet, Words and Languages

- Similarly infinite words over the alphabet $\Sigma = \{a, b\}$ is an infinite sequence of a's and b's e.g. $u = aabaabaabaab\cdots$.
- It may be viewed as a map $u : \mathbb{N} \rightarrow \Sigma$, where $u(i) = \sigma_i \in \Sigma$.

Alphabet, Words and Languages

- All finite words of length n are elements of Σ^n. So the collection of all finite words over Σ is $\bigcup_{n\geq 0} \Sigma^n$.
- This set is called Σ^*. If $\Sigma = \{a, b\}$, then first few elements Σ^* are $\{\varepsilon, a, b, aa, ab, ba, bb, aaa, \cdots\}$.

Alphabet, Words and Languages

- The collection of infinite words or ω-words over an alphabet Σ is denoted by Σ^ω.
- If there is only one element in the alphabet, then there is only one element in Σ^ω!

Alphabet, Words and Languages

- We claim that Σ^* is countably infinite.
- If $a \in \Sigma$, then the map $f_3 : \mathbb{N} \to \Sigma^* : n \mapsto a^n = \overbrace{a \cdots a}^{n}$ is injective.

Alphabet, Words and Languages

- If there are k symbols in Σ, then words of Σ^* may be viewed as numerals of base-(k+1) number system.
- As an example, let $\Sigma = \{a, b\}$. We may view $a \mapsto 1$ and $b \mapsto 2$, and $\Sigma^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, \cdots\}$ as $\{0, 1, 2, 11, 12, 21, 22, 111, \cdots\}$, numerals of base-3 number system.

Alphabet, Words and Languages

- Viewing Σ^* as numerals gives us a natural injective map from $\Sigma^* \to \mathbb{N}$. And by Schröder-Bernstine theorem there is a bijection between Σ^* and $\mathbb{N}$.
- On the other hand, if there are at least two elements in Σ, the number of elements of Σ^ω is uncountable.

Alphabet, Words and Languages

Given an alphabet Σ, the size of

- the collection of all finite words, Σ^* is countably infinite, but
- the size of the collection of all ω-words, Σ^ω is uncountably infinite when $|\Sigma| > 1$.
- A language of finite words (ω-words) over Σ is a subset (non-empty subset) of Σ^* (Σ^ω).

Alphabet, Words and Languages

We consider languages of finite words over Σ i.e. the subsets of Σ^*.

- As the size of Σ^* is countably infinite, the size of its power set i.e. the size of the collection of all possible languages of finite words over Σ is uncountably infinite.

Alphabet, Words and Languages

- The question is how to give an effective description of a language.
- If the language is a finite collection of finite words, we can prepare a list of its words. This is possible in principle.
- But if the language is an infinite collection, or any collection of ω-words, we need some formalism to describe them.

Alphabet, Words and Languages

- One simple but important need is to test membership of a word in a language, $x \stackrel{?}{\in} L$, where $x \in \Sigma^*$ and $L \subseteq \Sigma^*$.
- We may have to test whether a positive integer is a prime, or a undirected graph is connected, or a text is a valid C program.

Alphabet, Words and Languages

- In fact any 'precisely defined' decision problem can be formulated as a membership problem of some language. So it is important to describe a language.
- But any description will use some alphabet Γ[a].
- A description of a language $L \subseteq \Sigma^*$ is a sequence or a finite word over Γ.

[a]The alphabet of the language being described (object language) is Σ.

Alphabet, Words and Languages

- As size of Γ^* is countably infinite, the number of valid descriptions are countable.
- But the size of the collection of languages over Σ, $\mathcal{P}\Sigma^*$, is uncountable.
- So there is no surjective map from $\Gamma^* \rightarrow \mathcal{P}\Sigma^*$.

Alphabet, Words and Languages

- All languages over Σ cannot be described.
- In fact only countably many can be described.