计算理论
Theory of Computation

https://courses.zju.edu.cn/course/join/8P7F9K7S0KL

计算理论(周四第7… 内部

课程引论
Course Introduction

杨莹春

yyc@zju.edu.cn

该群属于“浙江大学”内部群，仅组织内部成员可以加入，如果组织外部人员收到此分享，需要先申请加入该组织。

浙江大学曹光彪西楼-201

2021年9月16日 **钉群 34658556**

课程用书
Course Books

- **教材：** **Elements of The Theory of Computation(Second Edition), Harry R. Lewis, Christos H. Papadimtriou, Prentice Hall, 清华大学出版社（影印版)**

- **参考书：**

  1.自动机理论、语言和计算导论（第**2**版）**, John E. Hopcroft, Rajeev Motwani, Jeffrey D. Ullman,** 清华大学出版社（影印版）**.**

  2. Introduction to the Theory of Computation(计算理论导引**) , Michael Sipser,** 机械工业出版社**.**

课程内容
Course Contents

- **Nature：**

  1.What is an algorithm?

  2.What can and what cannot be computed?

  3.When should an algorithm be considered feasible?

- **Purpose:**

  Introduction to the above fundamental ideas, models and results that permeate computer science, the *basic paradigms* of our fields, which are powerful and beautiful, excellent examples of mathematical modeling that is elegant, productive, and of lasting value. …… it is hard to understand computer science without first being exposed to them.

  This book is very much about algorithms and their formal foundations.

  ________ P3 of the course book

课程特点
Course Features

- **Languages:**

It probably comes as no surprise that these ideas and models are *mathematical* in nature.

It is generally discrete, the emphasis is on finite sets and sequences.

It is based on very few and elementary concepts, and draws its power and depth from the careful, patient, extensive, layer-by-layer manipulation of these concept ---just like the computer.

__________P1-2 of the course book

课程愿景
Course Views

- ## Hopes:

Our fervent hope is that the book will contribute to the intellectual development of the next generation of computer scientists by introducing them at an early stage of their education to crisp and methodical thinking about computational problems.

________ Preface to the first edition

Computation is essential, powerful, beautiful, challenging, ever-expanding --- and so is its theory. This book only tells the beginning of an exciting story. It is a modest introduction to a few basic and carefully selected topics from the treasure chest of the theory of computation. We hope that it will motivate its readers to seek out more; the references at the end of each chapter point to good places to start.

________ P4 of the course book

推荐阅读

1. 论文集： **Alan Turing: His Work and Impact, 2013**（**PDF**从群里下载）

ON COMPUTABLE NUMBERS, WITH AN APPLICATION TO THE ENTSCHEIDUNGSPROBLEM

By A. M. TURING

[Received 28 May, 1936.—Read 12 November, 1936.]

Review
Reviewed Work(s): On Computable Numbers, with an Application to the Entscheidungsproblem by A. M. Turing
Review by: Alonzo Church
Source: *The Journal of Symbolic Logic*, Vol. 2, No. 1 (Mar., 1937), pp. 42-43
Published by: Association for Symbolic Logic
Stable URL: https://www.jstor.org/stable/2268810
Accessed: 28-03-2019 02:29 UTC

On Computable Numbers, with an Application to the Entscheidungsproblem

(Proc. Lond. Math. Soc., ser. 2 vol. **42** (1936–37), pp. 230–265)

– A Correction

(ibid. vol. **43** (1937), pp. 544–546)

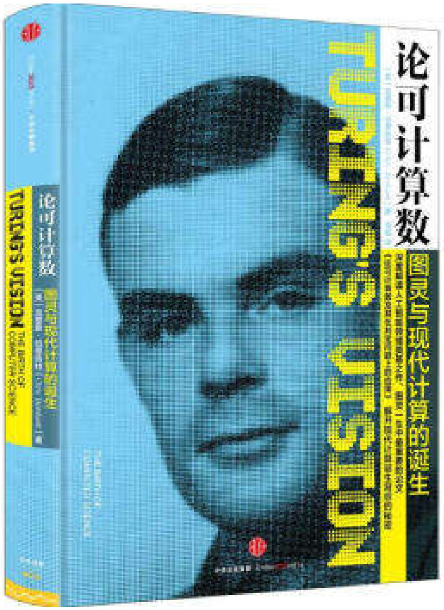
Christos Papadimitriou on —

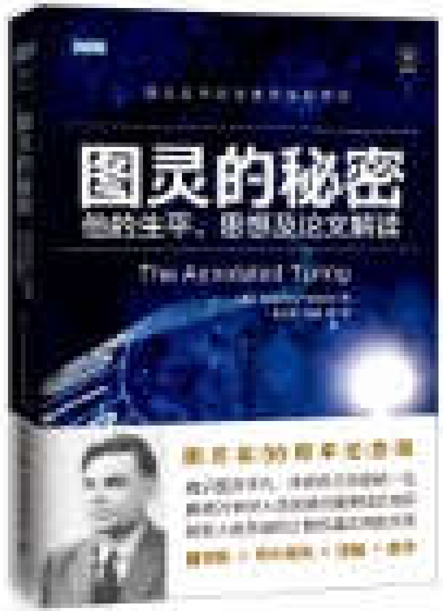
ALAN AND I

推荐阅读

1. 论可计算数：图灵与现代计算的诞生
[美] 克里斯·伯恩哈特（**Chris Bernhardt**） 著，雪曼 译，中信出版社，**2016.9.**

2. 图灵的秘密：他的生平、思想及论文解读
(美)Charles Petzold 著，杨卫东 朱皓 等 译，人民邮电出版社，**2012.11**





课程简介

- 考核

  作业　　10分

  测验　　20分

  笔试　　70分

- 答疑

  课堂、钉钉

内容安排

– 3 classes（9/16,9/23,9/30）
  Sets, Relations and Language（CH1）
– 3 classes(10/7,10/14,10/21)
  Regular Language and Finite Automata (CH2)
– 3 classes(10/28,11/4, 11/11)
  Context-free Languages (CH3)
– 3 classes (11/18,11/25,12/2)
  Turing machine (CH4)
– 2 classes(12/9,12/16)
  Undecidability（CH5）
– 2 classes(12/23,12/ 30)
  Review

Exam:2022/1

Keywords I

•Ch1．Sets, Relations and Language
Sets and Relations,
Finite and infinite sets,
Three fundamental proof techniques,
Alphabets and languages

Keywords II

- Ch2. Regular Language and Finite Automata

Finite automata and Nondeterministic finite automata,
Equivalence of finite automata and regular expressions,
Languages that are and are not regular

Keywords III

- Ch3. Context-free Languages

Context-free grammars, Pushdown automata, Equivalence of Pushdown automata and context-free grammars, Languages that are and are not context-free

Keywords IV

- Ch4.　Turing machine

The definition of Turing machines,
Computing with Turing machines,
Primitive recursive functions and Recursive functions

Keywords V

- Ch5. Undecidability

Church-Turing Thesis,
Universal Turing machines,
The halting problems

内容安排

- 3 classes（9/16,9/23,9/30 ）
  Sets, Relations and Language（CH1）
- 3 classes(10/7,10/14,10/21)
  Regular Language and Finite Automata (CH2)
- 3 classes(10/28,11/4, 11/11)
  Context-free Languages (CH3)
- 3 classes (11/18,11/25,12/2)
  Turing machine (CH4)
- 2 classes(12/9,12/16)
  Undecidability（CH5）
- 2 classes(12/23,12/ 30)
  Review

Exam:2021/1

计算理论

第1章 集合、关系、语言

Ch1．Sets, Relations and Language

Keywords I

- Ch1．Sets, Relations and Language

Sets and Relations,

Finite and infinite sets,

Three fundamental proof techniques,

Alphabets and languages

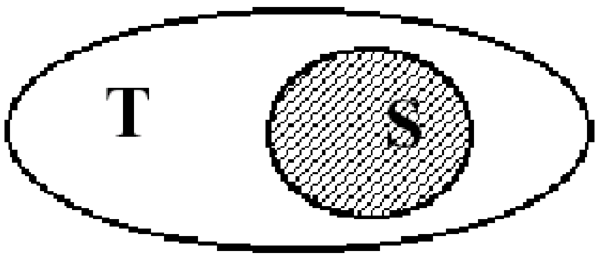
1.1 Sets

□ Set

– an unordered collection of elements

empty set $\emptyset$

□ Subsets and proper subsets



– Subset notation: $\subseteq$

$$S \subseteq T \Leftrightarrow (\forall x \in S \Rightarrow x \in T)$$

– Proper Subset: $\subset$

– Two sets are **equal** iff they contain the same elements

$$S = T \Leftrightarrow (S \subseteq T) \wedge (T \subseteq S)$$

Set Operations and Its Identities

– Union, Intersection, Difference, Symmetric difference, complement

– Commutative Law, Associative Law, Distributive law, Absorption, DeMorgan's Law, Idempotent law

Power Set

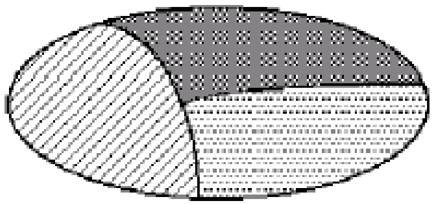
– 2^S = set of all subset of S

$$2^S = \{T \mid T \subseteq S\}$$

□ Partition

A **partition** of a nonempty set A is a subset Π of 2^A such that

1) $\emptyset \notin \Pi$;
2) $\forall S, T \in \Pi$, and $S \neq T, S \cap T = \emptyset$
3) $\bigcup \Pi = A$.



1.2 Relations and Functions

□ Ordered Pair and Binary Relation

– Ordered Pair: (a, b)

$$(a, b) = (c, d) \Leftrightarrow (a = c) \wedge (b = d)$$

– Cartesian Product: $A \times B$

$$A \times B = \{(a, b) | a \in A \wedge b \in B\}$$

– Binary Relation on A and B:

$$R \subseteq A \times B$$

□ Ordered Tuple and n-ary Relation (Omitted)

□ Operations of Relations

– Inverse

$$R \subseteq A \times B \Rightarrow R^{-1} \subseteq B \times A$$

– Composition

Let g be a function from the set A to the set B and let f be a function from the set B to the set C. The *composition* of the functions f and g, denoted for all $a \in A$ by $f \circ g$, is defined by

$$(f \circ g)(a) = f(g(a)).$$

When ideas fail, words come in very handy.
- Goethe (1749-1832)

□ Function

Definition: A **function** f: $A \rightarrow B$ must satisfy:

- $f \subseteq A \times B$
- $\forall a \in A, \exists$ exactly one $b \in B$ with $(a, b) \in f$

Note:

We write $(a, b) \in f$ as $f(a) = b$.

Domain, range

one-to-one function:

$\forall a, b \in A \wedge a \neq b \Rightarrow f(a) \neq f(b)$

onto function:

$\forall b \in B \ \exists a \in A$ such that $f(a) = b$

bijection function:

(one-to-one correspondence)

one-to-one + onto

1.3 Special Types of Binary Relations

□ Representation of Relations

– Directed graph: node, edge, path
– Matrix: Adjacency matrix

□ Properties of Relations $(R \subseteq A \times A)$

– reflexive: $\forall a \in A \Rightarrow (a,a) \in R$
– symmetric: $(a,b) \in R \wedge a \neq b \Rightarrow (b,a) \in R$
antisymmetric: $(a,b) \in R \Rightarrow (b,a) \notin R$
– transitive: $(a,b) \in R, (b,c) \in R \Rightarrow (a,c) \in R$

□ Equivalence Relation

– reflexive, symmetric, transitive

– equivalence classes

$$[a] = \{b|(a,b) \in R\}$$

Theorem Let R be an equivalence relation on a nonempty set A. Then the equivalence classes of R constitute a partition of A.

□ Partial Order

– reflexive, antisymmetric, transitive

– total order

– minimal element and maximal element

"Sometimes when reading Goethe, I have the paralyzing suspicion that he is trying to be funny.
- Guy Davenport

1.4 Finite and Infinite sets

□ **Equinumerous**

– Sets A and B equinumerous $\Leftrightarrow \exists$ bijection $f : A \to B$

– Cardinality and generalized Cardinality

– Finite and Infinite Sets

□ **Countable and Uncountable Infinite**

– A set is said to be **countably infinite** $\Leftrightarrow$ it is equinumerous with $\mathbb{N}$.

– S is an **uncountable** set $\Leftrightarrow |S| > |\mathbb{N}|$.

– The union of a countably infinite collection of countably infinite sets is countably infinite.

Example: Show that $\mathbb{N} \times \mathbb{N}$ is countably infinite.

Theorem: $|\mathbb{R}| > |\mathbb{N}|$.

Proof: See next section by diagonalization.

Question: Is $|\mathbb{R}| > |(0, 1)|$?

A set is said to be **countably infinite** if it is equinumerous with $\mathbf{N}$, and **countable** if it is finite or countably infinite. A set that is not countable is **uncountable**. To show that a set A is countably infinite we must exhibit a bijection f between A and $\mathbf{N}$; equivalently, we need only suggest a way in which A can be enumerated as

$$A = \{a_0, a_1, a_2, \ldots\},$$

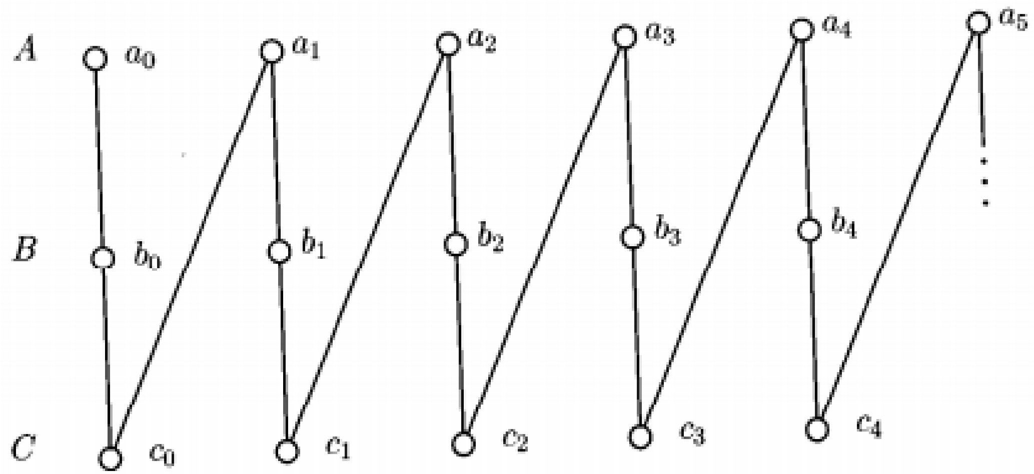
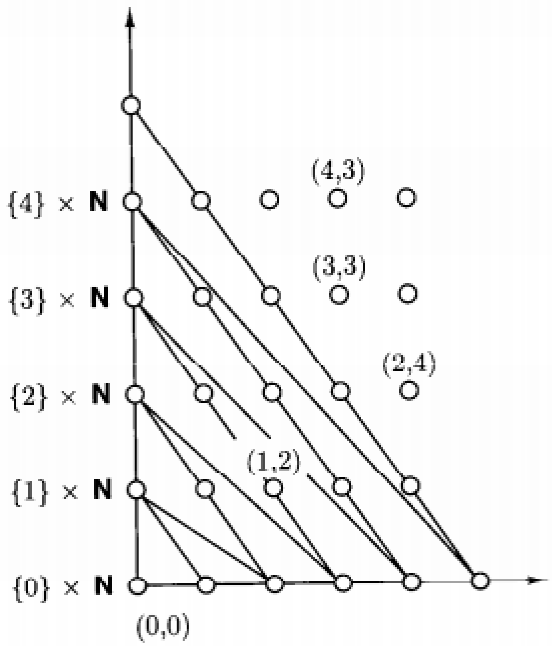


Figure 1-7

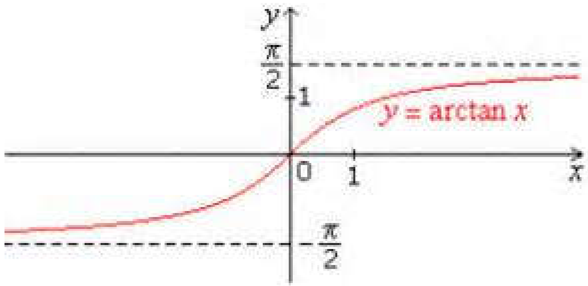
The same idea can be used to show that the union of a countably infinite collection of countably infinite sets is countably infinite. For example, let us show that $\mathbf{N} \times \mathbf{N}$ is countably infinite; note that $\mathbf{N} \times \mathbf{N}$ is the union of $\{0\} \times \mathbf{N}$, $\{1\} \times \mathbf{N}$, $\{2\} \times \mathbf{N}$, and so on, that is, the union of a countably infinite collection of countably infinite sets. Dovetailing must here be more subtle than in the



Another way of viewing this use of dovetailing is to observe that the pair (i, j) is visited mth, where $m = \frac{1}{2}[(i + j)^2 + 3i + j]$; that is to say, the function $f(i, j) = \frac{1}{2}[(i + j)^2 + 3i + j]$ is a *bijection* from $\mathbf{N} \times \mathbf{N}$ to $\mathbf{N}$ (see Problem 1.4.4).

Question: Is $|\mathbb{R}| > |(0,1)|$?

$$\boxed{f(x) = \frac{1}{\pi} arctan(x) + \frac{1}{2}}$$



If there is a one-to-one function from A to B, the cardinality of A is less than or the same as the cardinality of B and we write $|A| \leq |B|$. Moreover, when $|A| \leq |B|$ and A and B have different cardinality, we say that the cardinality of A is less than the cardinality of B and we write $|A| < |B|$.

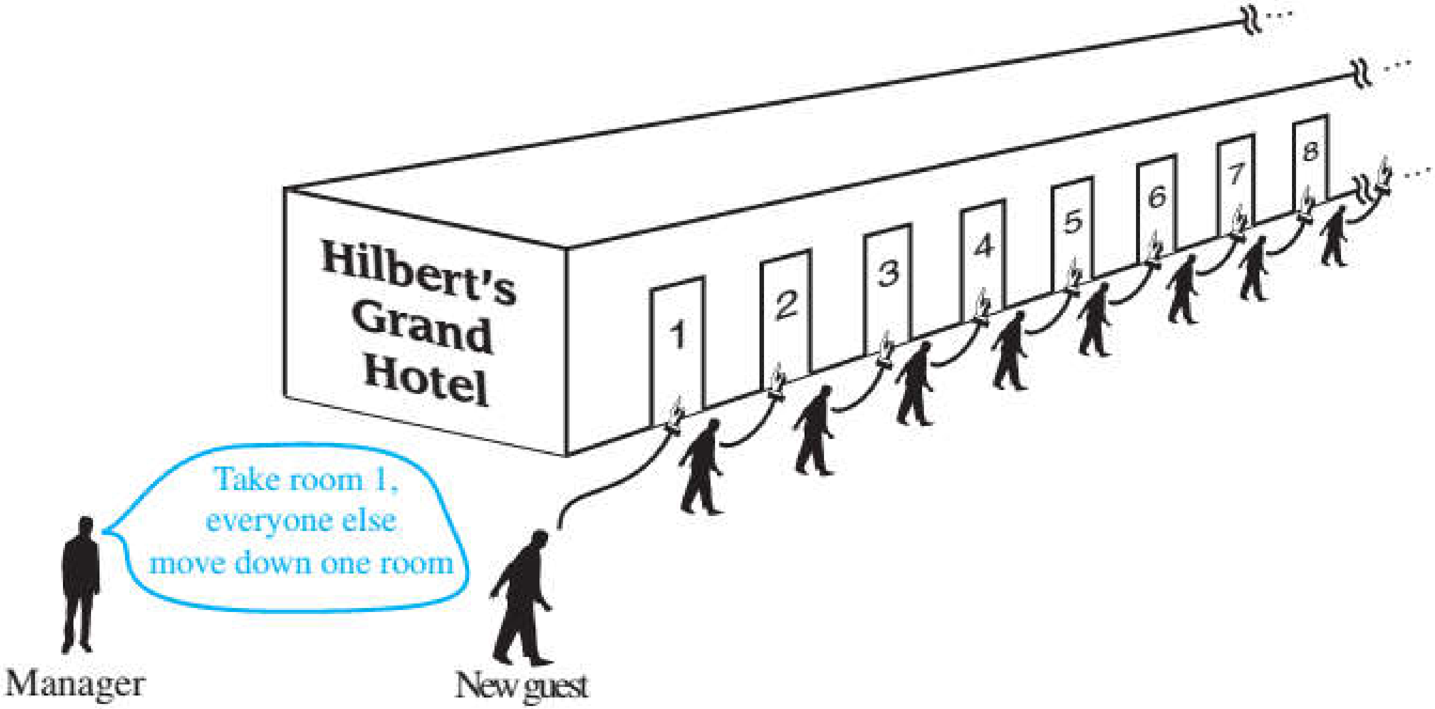


FIGURE 2 **A New Guest Arrives at Hilbert's Grand Hotel.**

Continuum Hypothesis

$|\mathbb{N}| = \aleph_0 \qquad |\mathbb{R}| = \aleph_1$

$\aleph_0 < \aleph_1$

$\exists ?\ w$ such that $\aleph_0 < w < \aleph_1$

Independent of the axioms ! [Cohen, 1966]

For an informal account on infinities, see e.g.: Rucker, *Infinity and the Mind* , Harvester Press, 1982.

THE CONTINUUM HYPOTHESIS We conclude this section with a brief discussion of a famous open question about cardinality. It can be shown that the power set of $\mathbf{Z}^+$ and the set of real numbers $\mathbf{R}$ have the same cardinality (see Exercise 38). In other words, we know that $|\mathcal{P}(\mathbf{Z}^+)| = |\mathbf{R}| = \mathfrak{c}$, where $\mathfrak{c}$ denotes the cardinality of the set of real numbers.

An important theorem of Cantor (Exercise 40) states that the cardinality of a set is always less than the cardinality of its power set. Hence, $|\mathbf{Z}^+| < |\mathcal{P}(\mathbf{Z}^+)|$. We can rewrite this as $\aleph_0 < 2^{\aleph_0}$, using the notation $2^{|S|}$ to denote the cardinality of the power set of the set S. Also, note that the relationship $|\mathcal{P}(\mathbf{Z}^+)| = |\mathbf{R}|$ can be expressed as $2^{\aleph_0} = \mathfrak{c}$.

This leads us to the famous **continuum hypothesis**, which asserts that there is no cardinal number X between $\aleph_0$ and $\mathfrak{c}$. In other words, the continuum hypothesis states that there is no set A such that $\aleph_0$, the cardinality of the set of positive integers, is less than $|A|$ and $|A|$ is less than $\mathfrak{c}$, the cardinality of the set of real numbers. It can be shown that the smallest infinite cardinal numbers form an infinite sequence $\aleph_0 < \aleph_1 < \aleph_2 < \cdots$. If we assume that the continuum hypothesis is true, it would follow that $\mathfrak{c} = \aleph_1$, so that $2^{\aleph_0} = \aleph_1$.

1.5 Three Fundamental Proof Techniques

□ The Principle of Mathematical Induction

Let A be a set of natural numbers such that

1) $0 \in A$, and

2) for each natural number n, if $\{0, 1, 2, \cdots, n\} \in A$, then $n + 1 \in A$.

□ The Pigeonhole Principle

If A and B are finite sets and $|A| > |B|$, then there is no one-to-one function from A to B.

Proof. *Basis step.*

$|B| = 0$ ($B = \emptyset$) $\Rightarrow$ no function from A to B

$\Rightarrow$ no one-to-one function.

Induction Hypothesis. Suppose $f : A \to B$, $|A| > |B|$, and $|B| \leq n$, where $n \geq 0 \Rightarrow f$ is not one-to-one.

Induction step. Suppose $f : A \to B$, and $|A| > |B| = n + 1$.

Choose some $a \in A$.

Case 1: If $\exists a' \in A$, such that $f(a) = f(a')$.

$\Rightarrow f$ is not one-to-one.

Case 2: a is the only element mapped by f to $f(a)$

Consider then the sets $A-\{a\}$ to $B-\{f(a)\}$, and the function $g : A - \{a\} \to B - \{f(a)\}$ that agree with f on all elements of $A - \{a\}$.

By induction hypothesis

$$|B - \{f(a)\}| = n$$

$$|A - \{a\}| = |A| - 1 > |B| - 1 = |B - \{f(a)\}|$$

$\Rightarrow \exists a, b \in A-\{a\}$, $a \neq b$, such that $g(a) = g(b)(f(a) = f(b))$.

$\Rightarrow$ f is not one-to-one

□ The Diagonalization Principle

Let R be a binary relation on a set A, and let D, the diagonal set for R, be $\{a : a \in A \wedge (a, a) \notin R\}$. For each $a \in A$, let

$$R_a = \{b : b \in A \wedge (a, b) \in R\}.$$

Then D is distinct from each R_a.

Example: Let us consider the relation $R = \{(a,b),(a,d),(b,b),(b,c),(c,c),(d,b),(d,c),(d,e),(d,f),(e,e),(e,f),(f,a),(f,c),(f,d),(f,e)\}$.

Notice that $R_a = \{b,d\}$, $R_b = \{b,c\}$,
$R_c = \{c\}$, $R_d = \{b,c,e,f\}$,
$R_e = \{e,f\}$, $R_f = \{a,c,d,e\}$.

The corresponds to the diagonal set is

$$D = \{a,d,f\}.$$

Example 1.5.3: Let us consider the relation $R = \{(a,b),(a,d),(b,b),(b,c),(c,c),(d,b),(d,c),(d,e),(d,f),(e,e),(e,f),(f,a),(f,c),(f,d),(f,e)\}$; notice that $R_a = \{b,d\}$, $R_b = \{b,c\}$, $R_c = \{c\}$, $R_d = \{b,c,e,f\}$, $R_e = \{e,f\}$ and $R_f = \{a,c,d,e\}$. All in all, R may be pictured like this:

	a	b	c	d	e	f
a		×		×		
b		×	×			
c			×			
d		×	×		×	×
e					×	×
f	×		×	×	×	

The sequence of boxes along the diagonal is

	×	×		×	

Its complement is

×			×		×

An Uncountable Set

Cantor diagonalization argument

Show that the set of real numbers is an uncountable set.

Solution: To show that the set of real numbers is uncountable, we suppose that the set of real numbers is countable and arrive at a contradiction. Then, the subset of all real numbers that fall between 0 and 1 would also be countable (because any subset of a countable set is also countable; see Exercise 16). Under this assumption, the real numbers between 0 and 1 can be listed in some order, say, $r_1, r_2, r_3, \ldots$. Let the decimal representation of these real numbers be

$$
\begin{aligned}
r_1 &= 0.d_{11}d_{12}d_{13}d_{14} \ldots \\
r_2 &= 0.d_{21}d_{22}d_{23}d_{24} \ldots \\
r_3 &= 0.d_{31}d_{32}d_{33}d_{34} \ldots \\
r_4 &= 0.d_{41}d_{42}d_{43}d_{44} \ldots \\
&\vdots
\end{aligned}
$$

where $d_{ij} \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. (For example, if $r_1 = 0.23794102\ldots$, we have $d_{11} = 2$, $d_{12} = 3$, $d_{13} = 7$, and so on.) Then, form a new real number with decimal expansion

$r = 0.d_1d_2d_3d_4\ldots$, where the decimal digits are determined by the following rule:

$$d_i = \begin{cases} 4 \text{ if } d_{ii} \neq 4 \\ 5 \text{ if } d_{ii} = 4. \end{cases}$$

(As an example, suppose that $r_1 = 0.23794102\ldots$, $r_2 = 0.44590138\ldots$, $r_3 = 0.09118764\ldots$, $r_4 = 0.80553900\ldots$, and so on. Then we have $r = 0.d_1d_2d_3d_4\ldots = 0.4544\ldots$, where $d_1 = 4$ because $d_{11} \neq 4$, $d_2 = 5$ because $d_{22} = 4$, $d_3 = 4$ because $d_{33} \neq 4$, $d_4 = 4$ because $d_{44} \neq 4$, and so on.)

Every real number has a unique decimal expansion (when the possibility that the expansion has a tail end that consists entirely of the digit 9 is excluded). Therefore, the real number r is not equal to any of $r_1, r_2, \ldots$ because the decimal expansion of r differs from the decimal expansion of r_i in the ith place to the right of the decimal point, for each i.

Because there is a real number r between 0 and 1 that is not in the list, the assumption that all the real numbers between 0 and 1 could be listed must be false. Therefore, all the real numbers between 0 and 1 cannot be listed, so the set of real numbers between 0 and 1 is uncountable. Any set with an uncountable subset is uncountable (see Exercise 15). Hence, the set of real numbers is uncountable. ◀

Theorem: (G. Cantor1845-1918) The set $2^{\mathbb{N}}$ is uncountable.

Proof: Suppose that $2^{\mathbb{N}}$ is countably infinite.

We assume that there is a way of enumerating all members of $2^{\mathbb{N}}$ as

$$2^{\mathbb{N}} = \{R_0, R_1, R_2, \cdots\}$$

Consider the relation

$$R = \{(i, j) : j \in R_i\}.$$

R_i — in the statement of the diagonalization principle.

Now consider the set $D = \{n \in \mathbb{N} : n \notin R_n\}$

—$D \subseteq \mathbb{N}$, should be appear somewhere in the enumeration $\{R_0, R_1, R_2, \cdots\}$

— D can not be the $R_i (i = 0, 1, 2, \cdots)$

Suppose that $D = R_k$ for some $k \geq 0$.

$k \in R_k?$

- $k \in R_k$: Since $D = \{n \in \mathbb{N} : n \notin R_n\} \Rightarrow k \notin D$; but $D = R_k$.
- $k \notin R_k \Rightarrow k \in D$. But $D = R_k$.

Contradiction!

Theorem: $(0, 1)$ is uncountable.

1.6 Closures

□ The Transitive Closure

the "smallest" relation that includes R and is transitive (usually called R^+)

e.g. If R is Parent-of, then the transitive closure of R is Ancestor-Of

More formally:
R^+ is a relation such that
* $R \subseteq R^+$
* R^+ is transitive
* $\forall R', R \subseteq R'$ and R' is transitive, $\Rightarrow R^+ \subseteq R'$

□ Closures of Relations

Given any binary relation R, one can form closures with respect to any combinations of the properties:

- reflexive
- symmetric
- transitive

e.g. Symmetric, transitive closure of "Parent-Of" is:

$$\cdots ? \cdots$$

Note:

Reflexive, transitive closure of R is usually denoted R^*.

1.7 Alphabet and Language

□ **Alphabet:** finite set of symbols

e.g. $\sum_1 = \{0, 1\}$, $\sum_2 = \{a, b, \cdots, x, y, z\}$

– String : finite symbol sequence

– Length: # of symbols

– Empty string : e

□ **Operations of Strings:**

- **Concatenation:** $x \circ y$ or xy

  Substring, suffix, prefix

Example: $\forall w, we = ew = w$

- **String exponentiation**

$w^0 = e$, the empty string
$w^{i+1} = w^i \circ w$, for each $i \geq 0$

—*definition by induction*

- **Reversal**

If w is a string of length 0, then $w^R = w = e$.
IF w is a string of length $n + 1 > 0$, then $w = ua$ for some $a \in \sum$, and $w^R = au^R$.

- **Language:** set of strings
  - $\sum$–alphabet, $\sum^*$–the set of all strings ($e \in \sum^*$)
  - Language $L \subseteq \sum^*$
  - $\emptyset$, $\sum$ and $\sum^*$ are languages.
  - Finite Language: by listing all the strings

    Infinite Language: specify by the following scheme

$$L = \{w \in \sum^* : w \text{ has property } P\}$$

Example: $L = \{ab, aabb, aaabbb, \cdots\} = \{a^n b^n \mid n \geq 1\}$

Theorem: If $\sum$ is a finite alphabet, then $\sum^*$ is countably infinite set.

Proof: Construct a bijection $f : \mathbb{N} \to \sum^*$.

Fix some ordering of the alphabet, say $\sum = \{a_1, a_2, \cdots, a_n\}$.

The member of $\sum^*$ can be enumerated in the following way:

1) For each $k \geq 0$, all string of length k are enumerated before all strings of length $k+1$.
2) The n^k strings of length exactly k are enumerated lexicographically.

If Σ is a finite alphabet, then Σ^* is certainly infinite; but is it a countably infinite set? It is not hard to see that this is indeed the case. To construct a bijection $f : \mathbf{N} \mapsto \Sigma^*$, first fix some ordering of the alphabet, say $\Sigma = \{a_1, \ldots, a_n\}$, where $a_1, \ldots, a_n$ are distinct. The members of Σ^* can then be enumerated in the following way.

(1) For each $k \geq 0$, all strings of length k are enumerated before all strings of length $k+1$.

(2) The n^k strings of length exactly k are enumerated **lexicographically**, that is, $a_{i_1} \ldots a_{i_k}$ precedes $a_{j_1} \ldots a_{j_k}$, provided that, for some m, $0 \leq m \leq k-1$, $i_\ell = j_\ell$ for $\ell = 1, \ldots, m$, and $i_{m+1} < j_{m+1}$.

For example, if $\Sigma = \{0, 1\}$, the order would be as follows:

$$e, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, \ldots$$

□ Operations of Languages:

- **Union, Intersection, Difference, Complement**

$(\overline{A} = \Sigma^* - A)$

$$\boxed{\begin{array}{l} L^0 = \{e\} \\ L^{i+1} = LL^i, \text{ for each } i \geq 0 \end{array}}$$

- **Concatenation:**

$$L_1 L_2 = \{w_1 w_2 \mid w_1 \in L_1 \wedge w_2 \in L_2\}$$

Example:

$L_1 = \{w \in \{0,1\}^* : w \text{ has an even number of 0's}\}$

$L_2 = \{w \in \{0,1\}^* : w \text{ starts with a 0, the rest symbol are 1's}\}$

$$L_1 L_2 = \{w \in \{0,1\}^* : w \text{ has an odd number of 0's}\}.$$

- **Kleene Star**

$$L^* = \{w \in \Sigma^* : w = w_1 \cdots w_k, k \geq 0, w_1, \cdots, w_k \in L\}$$

$$= L^0 \cup L^1 \cup L^2 \cup \cdots$$

$$L^+ = L^1 \cup L^2 \cup L^3 \cup \cdots$$

Example: $L = \{w \in \{0,1\}^* : w$ has an unequal number of 0's and 1's$\}$. Then $L^* = \{0,1\}^*$.

Hint: $L_1 \subseteq L_2 \Rightarrow L_1^* \subseteq L_2^*$ $\qquad$ $\{0,1\} \subseteq L$

As a final example, let us show that if L is the language $\{w \in \{0,1\}^* :$ w has an unequal number of 0's and 1's$\}$, then $L^* = \{0,1\}^*$. To see this, first note that for any languages L_1 and L_2, if $L_1 \subseteq L_2$, then $L_1^* \subseteq L_2^*$ as is evident from the definition of Kleene star. Second, $\{0,1\} \subseteq L$, since each of 0 and 1, regarded as a string, has an unequal number of 0's and 1's. Hence $\{0,1\}^* \subseteq L^*$; but $L^* \subseteq \{0,1\}^*$ by definition, so $L^* = \{0,1\}^*$.

Remark:

1) The use of $\sum^*$ to denote the set of all strings over $\sum$ is consistent with the notation for the Kleene star of $\sum$.

2) $\emptyset^* = \{e\}$

3) $L^+ = LL^*$

4) For any language L, $(L^*)^* = L^*; L\emptyset = \emptyset L = \emptyset$

We write L^+ for the language LL^*. Equivalently, L^+ is the language

$$\{w \in \Sigma^* : w = w_1 \circ w_2 \circ \cdots \circ w_k \text{ for some } k \geq 1 \text{ and some } w_1, \ldots, w_k \in L\}.$$

Notice that L^+ can be considered as the *closure* of L under the function of concatenation. That is, L^+ is the smallest language that includes L and all strings that are concatenations of strings in L.

1.8 Finite Representations of Languages

□ Finite Representations:

– must be a string

– different languages to have different representations

Representations (Σ^*, countable)	Languages (2^{Σ^*}, uncountable)

$\Rightarrow \exists$ undescribable languages!

Question: Prove that 2^{Σ^*} is uncountable.

□ Regular Expressions

Example Let $L = \{w \in \{0,1\}^* : w$ has two or three occurrences of 1, the first and second of which are not consecutive$\}$.

- The language can be described using only singleton sets and the symbols $\cup$, $\circ$, and $*$ as

$$\{0\}^* \circ \{1\} \circ \{0\}^* \circ \{0\} \circ \{1\} \circ \{0\}^*((\{1\} \circ \{0\}^*) \cup \emptyset^*)$$

- The language can be written simply as

$$0^*10^*010^*(10^* \cup \emptyset^*).$$

Definition: The regular expressions are all strings over the alphabet $\sum \cup \{(,),\cup,\star\}$ that can be obtained as follows.

1) Θ and $\{x\}(\forall x \in \sum)$ is a regular expression.

2) If α and β are regular expressions, then so are $(\alpha\beta)$, $(\alpha \cup \beta)$, $\alpha^\star$.

3) Nothing is regular expression unless it follows from 1) through 2).

Example: a^*b^* $\quad a^* \cup b^*$ $\quad a(a^* \cup b^*)$

$(a^* \cup b^*)a(a^* \cup b^*)$ $\quad aaaaa^*$

□ Regular expressions & languages.

The function $\mathscr{L}$ is defined as follows.

1) $\mathscr{L}(\Theta) = \emptyset$, and $\mathscr{L}(a) = \{a\}$ for each $a \in \sum$.
2) If α and β are regular expressions, then

$$\mathscr{L}(\alpha\beta) = \mathscr{L}(\alpha)\mathscr{L}(\beta)$$
$$\mathscr{L}(\alpha \cup \beta) = \mathscr{L}(\alpha) \cup \mathscr{L}(\beta)$$
$$\mathscr{L}(\alpha^{\star}) = \mathscr{L}(\alpha)^{*}$$

Example: What is $\mathscr{L}(((a\cup b)^{\star}a))$? ?

$$\begin{aligned}
\mathscr{L}(((a\cup b)^{\star}a)) &= \mathscr{L}((a\cup b)^{\star})\mathscr{L}(a)\\
&= \mathscr{L}((a\cup b)^{\star})\{a\}\\
&= \mathscr{L}((a\cup b))^{*}\{a\}\\
&= (\mathscr{L}(a)\cup\mathscr{L}(b))^{*}\{a\}\\
&= (\{a,b\})^{*}\{a\}\\
&= \{w\in\{a,b\}^{*}: w \text{ ends with an } a\}
\end{aligned}$$

Example: What language is represented by $(c^{\star}(a\cup(bc^{\star}))^{\star})$?

$L=\{w\in\{a,b,c\}^{*}:$ not have the substring $ac\}$.

□ Regular Expression Identities

- $SR \neq RS$
- $S \cup R = R \cup S$
- $R(ST) = (RS)T$
- $R(S \cup T) = RS \cup RT$, $(R \cup S)T = RT \cup ST$
- $\emptyset^* = \{e\}$
- $(R^*)^* = R^*$
- $(R^*S^*)^* = (R \cup S)^*$
- $(\{e\} \cup R)^* = R^*$

Remark:

1) Every language that can be represented by a regular expression can be represented by infinitely many of them.

2) The class of regular languages over an alphabet $\sum$ is defined to consist of all languages L such that $L = L(a)$ for some regular expression a over $\sum$. i.e. the class of regular languages over an alphabet $\sum$ is precisely the closure of the set of languages

$$\{\{\sigma\} : \sigma \in \sum\} \cup \{\emptyset\}$$

3) The regular expressions are an inadequate specification method in general.
For example, $\{0^n1^n : n \geq 0\}$ cannot be described by regular expressions.

4) Two important and useful means of representing languages:
– language recognition device

to answer questions of the form "Is string w a member of L?
– language generators

language recognition device.

$$L = \{w \in \{0,1\}^* : w \text{ does not have 111 as a substring}\}.$$

by reading strings, a symbol at a time, from left to right, might operate like this:

Keep a count, which starts at zero and is set back to zero every time a 0 is encountered in the input; add one every time a 1 is encountered in the input; stop with a No answer if the count ever reaches three, and stop with a Yes answer if the whole string is read without the count reaching three.

language generators

An alternative and somewhat orthogonal method for specifying a language is to describe how a generic specimen in the language is produced. For example, a regular expression such as $(e \cup b \cup bb)(a \cup ab \cup abb)^*$ may be viewed as a way of *generating* members of a language:

To produce a member of L, first write down either nothing, or b, or bb; then write down a or ab, or abb, and do this any number of times, including zero; all and only members of L can be produced in this way.

Homework 1:	
P46	
	1.7.4 (c)(d)
	1.7.6
P51	
	1.8.3(c)
	1.8.5