Theory of Computation Lecture Notes

Prof. Yuh-Dauh Lyuu

Dept. Computer Science & Information Engineering

and

Department of Finance

National Taiwan University

Class Information

- Papadimitriou. *Computational Complexity*. 2nd printing. Addison-Wesley. 1995.
  - We more or less follow the topics of the book.
  - More "advanced" materials may be added.
- You may want to review discrete mathematics.

Class Information (concluded)

- More information and lecture notes can be found at

  `www.csie.ntu.edu.tw/~lyuu/complexity.html`

  – Homeworks, exams, solutions and teaching assistants will be announced there.

- Please ask many questions in class.

  – The best way for me to remember you in a large class.[a]

---

[a] "[A] science concentrator [...] said that in his eighth semester of [Harvard] college, there was not a single science professor who could identify him by name." (*New York Times*, September 3, 2003.)

Grading

- Homeworks.
  - Do not copy others' homeworks.
  - Do not give your homeworks for others to copy.
- Two to three exams.
- You must show up for the exams in person.
- If you cannot make it to an exam, please email me or a TA beforehand (unless there is a legitimate reason).
- Missing the final exam will earn a "fail" grade.

Problems and Algorithms

I have never done anything "useful."
— Godfrey Harold Hardy (1877–1947),
A Mathematician's Apology (1940)

What This Course Is All About

Computation: What is computation?

Computability: What can be computed?

- There are *well-defined* problems that cannot be computed.
- In fact, "most" problems cannot be computed.

What This Course Is All About (concluded)

Complexity: What is a computable problem's inherent complexity?

- Some computable problems require at least exponential time and/or space.
  - They are said to be **intractable**.
- Some practical problems require superpolynomial resources unless certain conjectures are disproved.
- Resources besides time and space?
  - Circuit size, circuit layout area, program size, number of random bits, etc.

Tractability and Intractability

- Polynomial in terms of the input size n defines tractability.
  - n, $n \log n$, n^2, n^{90}.
  - Time, space, and circuit size.
- It results in a fruitful and practical theory of complexity.
- Few practical, tractable problems require a large degree.
- Exponential-time or superpolynomial-time algorithms are usually impractical.
  - $n^{\log n}$, $2^{\sqrt{n}}$,[a] 2^n, $n! \sim \sqrt{2\pi n}\,(n/e)^n$.

---

[a]Size of depth-3 circuits to compute the majority function (Wolfovitz (2006)).

Growth of Factorials

n	$n!$	n	$n!$
1	1	9	362,880
2	2	10	3,628,800
3	6	11	39,916,800
4	24	12	479,001,600
5	120	13	6,227,020,800
6	720	14	87,178,291,200
7	5040	15	1,307,674,368,000
8	40320	16	20,922,789,888,000

Growth of *E. Coli*[a]

- Under ideal conditions, *E. Coli* bacteria divide every 20 minutes.
- In two days, a single *E. Coli* bacterium would become 2^{144} bacteria.
- They would weigh 2,664 times the Earth!

[a]Nick Lane, *Power, Sex, Suicide: Mitochondria and the Meaning of Life* (2005).

Moore's Law to the Rescue?[a]

- Moore's law says the computing power doubles every 1.5 years.[b]
- So the computing power grows like
$$4^{y/3},$$
where y is the number of years from now.
- Assume Moore's law holds forever.
- Can you let the law take care of exponential complexity?
- Suppose a problem takes a^n seconds to solve now, where

[a] Contributed by Ms. Amy Liu (J94922016) on May 15, 2006. Thanks also to a lively discussion on September 14, 2010.

[b] Moore (1965).

n is the input length.

Moore's Law to the Rescue (concluded)?

- The same problem will take

$$\frac{a^n}{4^{y/3}}$$

seconds to solve y years from now.

- The hardware $3n \log_4 a$ years from now takes 1 second to solve it.

- The overall complexity is linear in n (years).

Turing Machines

Alan Turing (1912–1954)

What Is Computation?

- That can be coded in an **algorithm**.[a]
- An algorithm is a detailed step-by-step method for solving a problem.
  - The Euclidean algorithm for the greatest common divisor is an algorithm.
  - "Let s be the least upper bound of compact set A" is not an algorithm.
  - "Let s be a smallest element of a finite-sized array" can be solved by an algorithm.

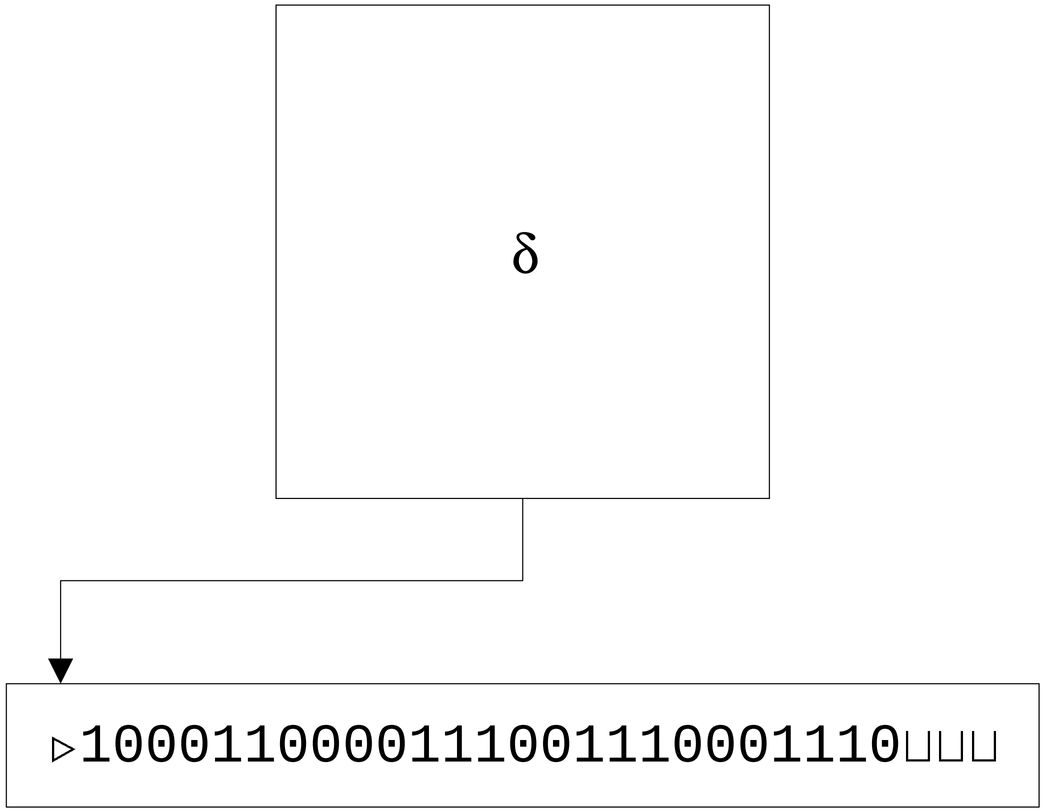
[a] Muhammad ibn Mūsā Al-Khwārizmī (780–850).

Turing Machines[a]

- A Turing machine (TM) is a quadruple $M = (K, \Sigma, \delta, s)$.
- K is a finite set of **states**.
- $s \in K$ is the **initial state**.
- Σ is a finite set of **symbols** (disjoint from K).
  - Σ includes $\sqcup$ (blank) and $\triangleright$ (first symbol).
- $\delta : K \times \Sigma \to (K \cup \{h, \text{"yes"}, \text{"no"}\}) \times \Sigma \times \{\leftarrow, \rightarrow, -\}$ is a **transition function**.
  - $\leftarrow$ (left), $\rightarrow$ (right), and $-$ (stay) signify cursor movements.

[a] Turing (1936).

A TM Schema



More about δ

- The program has the **halting state** (h), the **accepting state** ("yes"), and the **rejecting state** ("no").
- Given current state $q \in K$ and current symbol $\sigma \in \Sigma$,

$$\delta(q, \sigma) = (p, \rho, D).$$

  - It specifies:
    - The next state p;
    - The symbol ρ to be written over σ;
    - The direction D the cursor will move *afterwards*.
- We require $\delta(q, \rhd) = (p, \rhd, \rightarrow)$ so that the cursor never falls off the left end of the string.

More about δ (concluded)

- Think of the program as lines of codes:

$$\begin{aligned}\delta(q_1,\sigma_1) &= (p_1,\rho_1,D_1),\\ \delta(q_2,\sigma_2) &= (p_2,\rho_2,D_2),\\ &\vdots\\ \delta(q_n,\sigma_n) &= (p_n,\rho_n,D_n).\end{aligned}$$

- Given the state q and the symbol under the cursor σ, the machine finds the line that matches (q,σ).
- That line of code is then executed.

The Operations of TMs

- Initially the state is s.
- The string on the tape is initialized to a $\triangleright$, followed by a *finite-length* string $x \in (\Sigma - \{\sqcup\})^*$.
- x is the **input** of the TM.
  - The input must not contain $\sqcup$s (why?)!
- The cursor is pointing to the first symbol, always a $\triangleright$.
- The TM takes each step according to δ.
- The cursor may overwrite $\sqcup$ to make the string longer during the computation.

“Physical” Interpretations

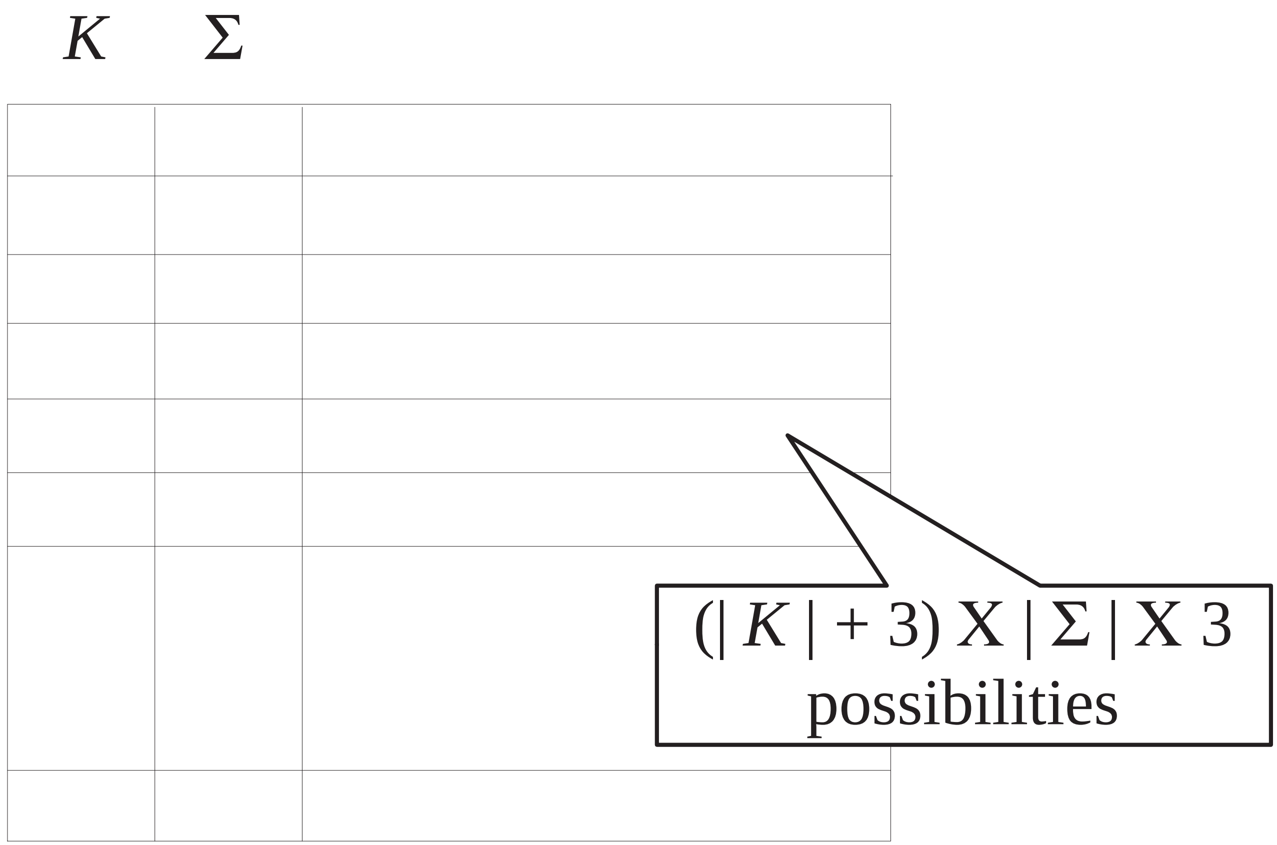
- The tape: computer memory and registers.
  - Except that the tape can be lengthened on demand.
- δ: program.
- K: instruction numbers.
- s: “`main()`” in C.
- Σ: **alphabet** much like the ASCII code.

Program Count

- A program has a *finite* size.
- Recall that
  $\delta : K \times \Sigma \rightarrow (K \cup \{h, \text{"yes"}, \text{"no"}\}) \times \Sigma \times \{\leftarrow, \rightarrow, -\}$.
- So $|K| \times |\Sigma|$ "lines" suffice to specify a program, one line per pair from $K \times \Sigma$ ($|\,x\,|$ denotes the length of x).
- Given K and Σ, there are

  $$((|K| + 3) \times |\Sigma| \times 3)^{|K| \times |\Sigma|}$$

  possible δ's (see next page).
  - This is a constant—albeit large.


K
Σ
$(|K| + 3) \times |\Sigma| \times 3$
possibilities

The Halting of a TM

- A TM M may **halt** in three cases.

  "yes": M **accepts** its input x, and $M(x) =$ "yes".

  "no": M **rejects** its input x, and $M(x) =$ "no".

  h**:** $M(x) = y$ means the string (tape) consists of a $\triangleright$, followed by a finite string y, whose last symbol is not $\sqcup$, followed by a string of $\sqcup$s.

  – y is the **output** of the computation.
  – y may be empty denoted by ϵ.

- If M never halts on x, then write $M(x) = \nearrow$.

Why TMs?

- Because of the simplicity of the TM, the model has the advantage when it comes to complexity issues.
- One can conceivably develop a complexity theory based on something similar to C++ or Java, say.
- But the added complexity does not yield additional fundamental insights.
- We will describe TMs in pseudocode.

Remarks

- A problem is computable if there is a TM that halts with the correct answer.
- If a TM (i.e., program) does not always halt, it does not solve a computable problem.[a]
- A computation model should be "physically" realizable.

[a]Contributed by Ms. Amy Liu (J94922016) on May 15, 2006. Control-C is not a legitimate way to halt a program.

Remarks (concluded)

- Although a TM requires a tape of infinite length, which is not realizable, it is not a major conceptual problem.[a]
  - Imagine you are living next to a paper mill, while carrying out the TM program using pencil and paper.
  - The mill will produce extra paper if needed.

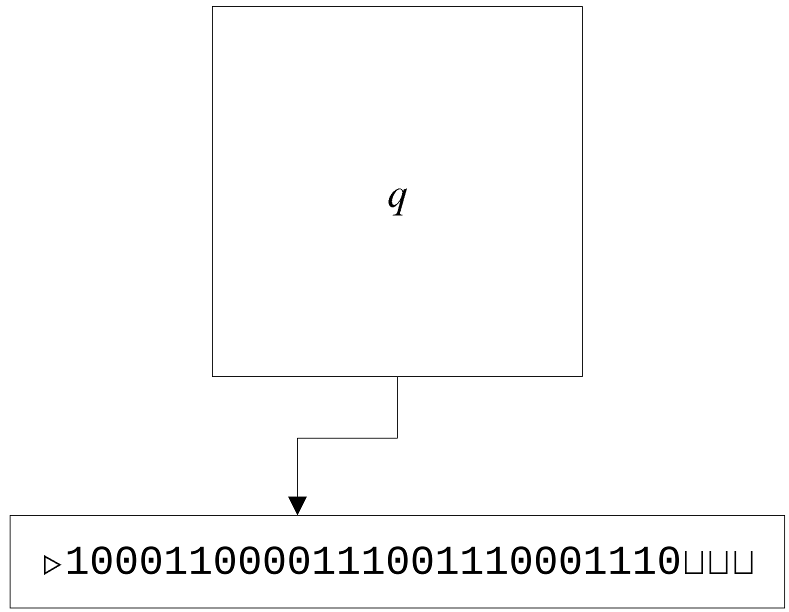
[a]Thanks to a lively discussion on September 20, 2006.

The Concept of Configuration

- A **configuration** is a complete description of the current state of the computation.
- The specification of a configuration is sufficient for the computation to continue as if it had not been stopped.
  - What does your PC save before it sleeps?
  - Enough for it to resume work later.
- Similar to the concept of state in Markov process.

Configurations (concluded)

- A configuration is a triple (q, w, u):
  - $q \in K$.
  - $w \in \Sigma^*$ is the string to the left of the cursor (inclusive).
  - $u \in \Sigma^*$ is the string to the right of the cursor.
- Note that (w, u) describes both the string and the cursor position.



- $w = \triangleright 1000110000$.
- $u = 111001110001110$.

Yielding

- Fix a TM M.
- Configuration (q, w, u) **yields** configuration (q', w', u') in one step,
$$(q, w, u) \xrightarrow{M} (q', w', u'),$$
if a step of M from configuration (q, w, u) results in configuration (q', w', u').
- $(q, w, u) \xrightarrow{M^k} (q', w', u')$: Configuration (q, w, u) yields configuration (q', w', u') in $k \in \mathbb{N}$ steps.
- $(q, w, u) \xrightarrow{M^*} (q', w', u')$: Configuration (q, w, u) yields configuration (q', w', u').

Example: How To Insert a Symbol

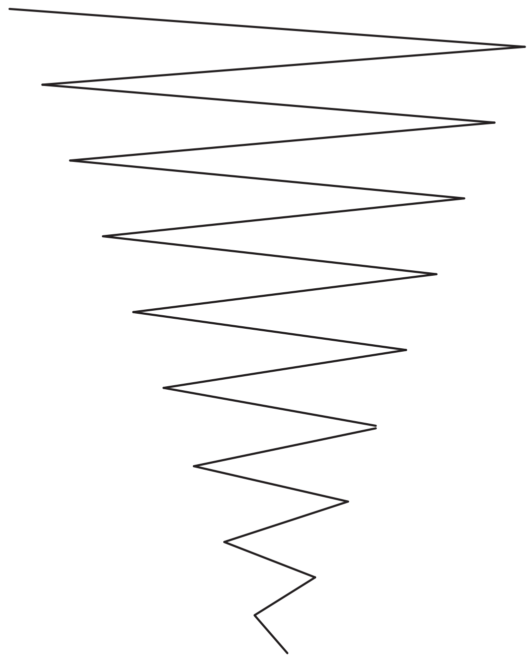
- We want to compute $f(x) = ax$.
  - The TM moves the last symbol of x to the right by one position.
  - It then moves the next to last symbol to the right, and so on.
  - The TM finally writes a in the first position.
- The total number of steps is $O(n)$, where n is the length of x.

Palindromes

- A string is a **palindrome** if it reads the same forwards and backwards (e.g., `001100`).
- A TM program can be written to recognize palindromes:
  - It matches the first character with the last character.
  - It matches the second character with the next to last character, etc. (see next page).
  - "yes" for palindromes and "no" for nonpalindromes.
- This program takes $O(n^2)$ steps.
- We cannot do better.[a]

[a] Hennie (1965).

$\boxed{100011000000100111}$



Comments on Lower-Bound Proofs

- They are usually difficult.
  - Worthy of a Ph.D. degree.
- An algorithm whose running time matches a lower bound means it is optimal.
  - The simple $O(n^2)$ algorithm for PALINDROME is optimal.
- This happens rarely and is model dependent.
  - Searching, sorting, PALINDROME, matrix-vector multiplication, etc.