T065001: Introduction to Formal Languages

Lecture 1: Introduction

Chapter 0 in Sipser's textbook

2025-04-14

(Lecture slides by Yih-Kuen Tsay)

What It Is

- The central question:

  What are the fundamental capabilities and limitations of computers?

- Three main areas:
  - *Automata Theory*
  - *Computability Theory*
  - *Complexity Theory*

Automata Theory

- The theories of computability and complexity require a precise, formal definition of a *computer*.
- *Automata theory* deals with the definitions and properties of mathematical models of computation.
- Two basic and practically useful models:
  - *Finite-state*, or simply *finite*, *automaton*
  - *Context-free grammar* (pushdown automaton)

Computability Theory

- Alan Turing, among other mathematicians, discovered in the 1930s that certain basic problems cannot be solved by computers.
- One example is the problem of determining whether a mathematical statement is true or false.
- Theoretical models of computers developed at that time eventually lead to the construction of actual computers.
- The theories of computability and complexity are closely related.
- *Complexity theory* seeks to classify problems as easy ones and hard ones, while in *computability theory* the classification is by whether the problem is solvable or not.

Complexity Theory

- Some problems are easy and some hard.
  For example, sorting is easy and scheduling is hard.
- The central question of complexity theory:
  What makes some problems computationally hard and others easy?
- We don't have the answer to it.
- However, researchers have found a scheme for classifying problems according to their computational difficulty.
- One practical application: cryptography/security.

Sets

- A *set* is a group of objects represented as a unit.
  The individual objects in a set are called its *elements* or *members*.
  Example: Let $S = \{7, 21, 57\}$ be a set. Then $7 \in S$ and $8 \notin S$.

Sets

- A *set* is a group of objects represented as a unit.
  The individual objects in a set are called its *elements* or *members*.

  Example: Let $S = \{7, 21, 57\}$ be a set. Then $7 \in S$ and $8 \notin S$.

- For two sets A and B, A is a *subset* of B ($A \subseteq B$) if every element in A is also an element in B. A is a *proper subset* of B ($A \subsetneq B$) if A is a subset of B and A is not equal to B.

  Example: Let $S = \{7, 21, 57\}$. Then $\{7, 21, 57\} \subseteq S$, $\{7, 57\} \subseteq S$, $\{7, 57\} \subsetneq S$, and $\{8, 57\} \not\subseteq S$.

Sets

- A *set* is a group of objects represented as a unit.
  The individual objects in a set are called its *elements* or *members*.
  Example: Let $S = \{7, 21, 57\}$ be a set. Then $7 \in S$ and $8 \notin S$.
- For two sets A and B, A is a *subset* of B ($A \subseteq B$) if every element in A is also an element in B. A is a *proper subset* of B ($A \subsetneq B$) if A is a subset of B and A is not equal to B.
  Example: Let $S = \{7, 21, 57\}$. Then $\{7, 21, 57\} \subseteq S$, $\{7, 57\} \subseteq S$, $\{7, 57\} \subsetneq S$, and $\{8, 57\} \not\subseteq S$.
- Special case: *The empty set* $\emptyset$ has zero elements. Written as $\emptyset = \{\ \}$.

Sets

- A *set* is a group of objects represented as a unit.
  The individual objects in a set are called its *elements* or *members*.

  Example: Let $S = \{7, 21, 57\}$ be a set. Then $7 \in S$ and $8 \notin S$.

- For two sets A and B, A is a *subset* of B ($A \subseteq B$) if every element in A is also an element in B. A is a *proper subset* of B ($A \subsetneq B$) if A is a subset of B and A is not equal to B.

  Example: Let $S = \{7, 21, 57\}$. Then $\{7, 21, 57\} \subseteq S$, $\{7, 57\} \subseteq S$, $\{7, 57\} \subsetneq S$, and $\{8, 57\} \not\subseteq S$.

- Special case: *The empty set* $\emptyset$ has zero elements. Written as $\emptyset = \{\,\}$.

- Examples of *infinite sets*:
  $\mathbb{N} = \{1, 2, 3, \dots\}$ ("the set of natural numbers")
  $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ ("the set of integers")
  $\{n \mid n = m^2 \text{ for some } m \in \mathbb{Z}\}$ ("the set of perfect squares")

Sets (cont.)

- For any two sets X and Y, the *union* of X and Y is the set of elements that belong to at least one of X and Y,

Sets (cont.)

- For any two sets X and Y, the *union* of X and Y is the set of elements that belong to at least one of X and Y, the *intersection* of X and Y is the set of elements that belong to both X and Y,

Sets (cont.)

- For any two sets X and Y, the *union* of X and Y is the set of elements that belong to at least one of X and Y, the *intersection* of X and Y is the set of elements that belong to both X and Y, and the *set difference* of X and Y is the set of elements that belong to X but not to Y.

Sets (cont.)

- For any two sets X and Y, the *union* of X and Y is the set of elements that belong to at least one of X and Y, the *intersection* of X and Y is the set of elements that belong to both X and Y, and the *set difference* of X and Y is the set of elements that belong to X but not to Y.

  Example: Let $X = \{a, b, c\}$ and $Y = \{c, d\}$. Then $X \cup Y = \{a, b, c, d\}$, $X \cap Y = \{c\}$, $X \setminus Y = \{a, b\}$, and $Y \setminus X = \{d\}$.

Sets (cont.)

- For any two sets X and Y, the *union* of X and Y is the set of elements that belong to at least one of X and Y, the *intersection* of X and Y is the set of elements that belong to both X and Y, and the *set difference* of X and Y is the set of elements that belong to X but not to Y.

  Example: Let $X = \{a, b, c\}$ and $Y = \{c, d\}$. Then $X \cup Y = \{a, b, c, d\}$, $X \cap Y = \{c\}$, $X \setminus Y = \{a, b\}$, and $Y \setminus X = \{d\}$.

- The *complement* of a set X, written in this textbook as $\overline{X}$, is the set of all elements under consideration that are not in X.

  Example: Suppose the underlying set is $U = \{a, b, \ldots, z\}$ and let $Y = \{c, d\}$. Then $\overline{Y} = U \setminus Y = \{a, b, e, f, \ldots, z\}$.

Sets (cont.)

- For any two sets X and Y, the *union* of X and Y is the set of elements that belong to at least one of X and Y, the *intersection* of X and Y is the set of elements that belong to both X and Y, and the *set difference* of X and Y is the set of elements that belong to X but not to Y.

  Example: Let $X = \{a, b, c\}$ and $Y = \{c, d\}$. Then $X \cup Y = \{a, b, c, d\}$, $X \cap Y = \{c\}$, $X \setminus Y = \{a, b\}$, and $Y \setminus X = \{d\}$.

- The *complement* of a set X, written in this textbook as $\overline{X}$, is the set of all elements under consideration that are not in X.

  Example: Suppose the underlying set is $U = \{a, b, \ldots, z\}$ and let $Y = \{c, d\}$. Then $\overline{Y} = U \setminus Y = \{a, b, e, f, \ldots, z\}$.

- The *power set* $\mathcal{P}(S)$ of a set S is the set of all subsets of S.

  Example: Let $Q = \{q_0, q_1, q_2\}$. Then $\mathcal{P}(Q) = \big\{\emptyset, \{q_0\}, \{q_1\}, \{q_2\}, \{q_0, q_1\}, \{q_0, q_2\}, \{q_1, q_2\}, \{q_0, q_1, q_2\}\big\}$.

Sets (cont.)

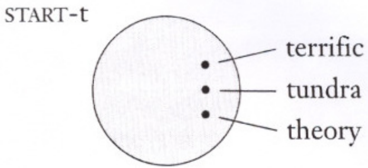


FIGURE 0.1
Venn diagram for the set of English words starting with "t"

Sets (cont.)

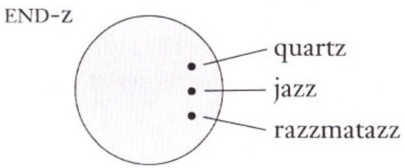


FIGURE 0.2
Venn diagram for the set of English words ending with "z"

Sets (cont.)

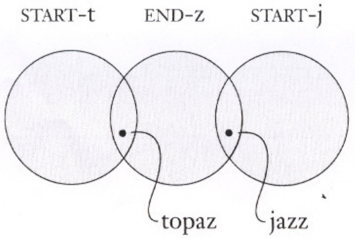


FIGURE 0.3
Overlapping circles indicate common elements

Sets (cont.)

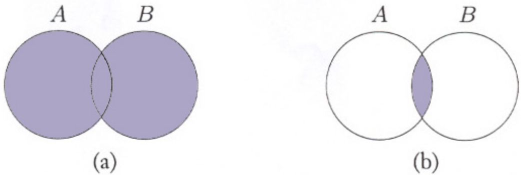


FIGURE 0.4
Diagrams for (a) $A \cup B$ and (b) $A \cap B$

Sequences and Tuples

- A *sequence* of objects is a list of these objects in some order. Order is essential and repetition is also allowed.
- Finite sequences are often called *tuples*. A sequence with k elements is a *k-tuple*; a 2-tuple is also called a *pair*.
- The *Cartesian product*, or cross product, of A and B, written as $A \times B$, is the set of all pairs (x, y) such that $x \in A$ and $y \in B$.

EXAMPLE 0.5

If $A = \{1, 2\}$ and $B = \{x, y, z\}$,

$$A \times B = \{\,(1,x),\ (1,y),\ (1,z),\ (2,x),\ (2,y),\ (2,z)\,\}.$$

- Cartesian products generalize to k sets, A_1, A_2, ..., A_k, written as $A_1 \times A_2 \times \ldots \times A_k$. A^k is a shorthand for $A \times A \times \ldots \times A$ (k times).

Strings and Languages

- An *alphabet* is any finite set of *symbols*.
- A *string* over an alphabet is a finite sequence of symbols from that alphabet.
- The *length* of a string w, written as $|w|$, is the number of symbols that w contains.
- The string of length 0 is called the *empty string*, written as ε.
- The *concatenation* of x and y, written as xy, is the string obtained from appending y to the end of x.

Strings and Languages

- An *alphabet* is any finite set of *symbols*.
- A *string* over an alphabet is a finite sequence of symbols from that alphabet.
- The *length* of a string w, written as $|w|$, is the number of symbols that w contains.
- The string of length 0 is called the *empty string*, written as ε.
- The *concatenation* of x and y, written as xy, is the string obtained from appending y to the end of x.
- A *formal language* is a set of strings. (Referred to as a *language* from now on.) A language can be finite or infinite.

Strings and Languages

- An *alphabet* is any finite set of *symbols*.
- A *string* over an alphabet is a finite sequence of symbols from that alphabet.
- The *length* of a string w, written as $|w|$, is the number of symbols that w contains.
- The string of length 0 is called the *empty string*, written as ε.
- The *concatenation* of x and y, written as xy, is the string obtained from appending y to the end of x.
- A *formal language* is a set of strings. (Referred to as a *language* from now on.) A language can be finite or infinite.

Example: $\{\mathtt{a}^n\,\mathtt{b}^n \mid n \geq 0\}$ is a language over the alphabet $\{\mathtt{a}, \mathtt{b}\}$. It consists of all strings of the form $\mathtt{aa}\ldots\mathtt{abb}\ldots\mathtt{b}$ with an equal number of as and bs. Note that ε also belongs to this language.

Functions

- A *function* sets up an *input-output* relationship, where the same input always produces the same output.
- If f is a function whose output is b when the input is a, we write $f(a) = b$.
- A function is also called a *mapping*; if $f(a) = b$, we say that f maps a to b.

Functions

- A *function* sets up an *input-output* relationship, where the same input always produces the same output.
- If f is a function whose output is b when the input is a, we write $f(a) = b$.
- A function is also called a *mapping*; if $f(a) = b$, we say that f maps a to b.
- The set of possible inputs to a function is called its *domain*; the outputs come from a set called its *range*.
- The notation $f : D \longrightarrow R$ says that f is a function with domain D and range R.

Functions (cont.)

EXAMPLE 0.8

Consider the function $f\colon \{0,1,2,3,4\} \longrightarrow \{0,1,2,3,4\}$.

n	$f(n)$
0	1
1	2
2	3
3	4
4	0

Functions (cont.)

EXAMPLE 0.8

Consider the function $f\colon \{0,1,2,3,4\} \longrightarrow \{0,1,2,3,4\}$.

n	$f(n)$
0	1
1	2
2	3
3	4
4	0

This function adds 1 to its input and then outputs the result modulo 5. A number modulo m is the remainder after division by m. For example, the minute hand on a clock face counts modulo 60. When we do modular arithmetic, we define $\mathcal{Z}_m = \{0,1,2,\ldots,m-1\}$. With this notation, the aforementioned function f has the form $f\colon \mathcal{Z}_5 \longrightarrow \mathcal{Z}_5$.

Functions (cont.)

EXAMPLE 0.9

Sometimes a two-dimensional table is used if the domain of the function is the Cartesian product of two sets. Here is another function, $g\colon \mathcal{Z}_4 \times \mathcal{Z}_4 \longrightarrow \mathcal{Z}_4$. The entry at the row labeled i and the column labeled j in the table is the value of $g(i, j)$.

g	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

Functions (cont.)

EXAMPLE 0.9

Sometimes a two-dimensional table is used if the domain of the function is the Cartesian product of two sets. Here is another function, $g\colon \mathcal{Z}_4 \times \mathcal{Z}_4 \longrightarrow \mathcal{Z}_4$. The entry at the row labeled i and the column labeled j in the table is the value of $g(i, j)$.

g	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

The function g is the addition function modulo 4. ∎

Relations

- A *predicate*, or property, is a function whose range is $\{\text{TRUE,FALSE}\}$.
- A predicate whose domain is a set of k-tuples $A \times \ldots \times A$ is called a (k-ary) *relation* on A.
- A 2-ary relation is also called a *binary relation*.

Relations

EXAMPLE 0.10

In a children's game called Scissors–Paper–Stone, the two players simultaneously select a member of the set {SCISSORS, PAPER, STONE} and indicate their selections with hand signals. If the two selections are the same, the game starts over. If the selections differ, one player wins, according to the relation *beats*.

beats	SCISSORS	PAPER	STONE
SCISSORS	FALSE	TRUE	FALSE
PAPER	FALSE	FALSE	TRUE
STONE	TRUE	FALSE	FALSE

From this table we determine that SCISSORS *beats* PAPER is TRUE and that PAPER *beats* SCISSORS is FALSE.

Graphs

- Undirected graph, node (vertex), edge (link), degree
- Description of a graph: $G = (V, E)$
- Labeled graph
- Subgraph, induced subgraph
- Path, simple path, cycle, simple cycle
- Connected graph
- Tree, root, leaf
- Directed graph, outdegree, indegree
- Strongly connected graph

Graphs (cont.)

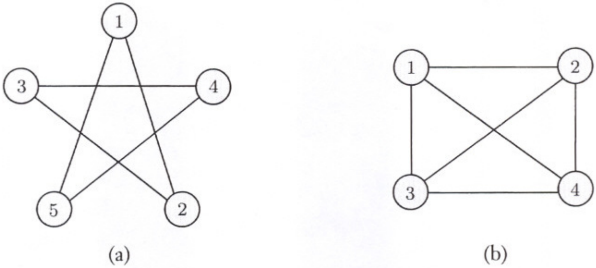


FIGURE **0.12**
Examples of graphs

Graphs (cont.)

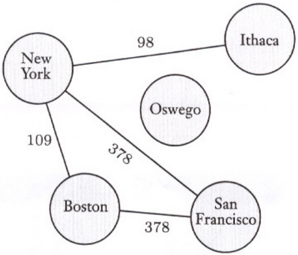


FIGURE 0.13
Cheapest nonstop air fares between various cities

Graphs (cont.)

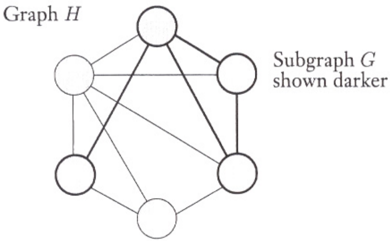
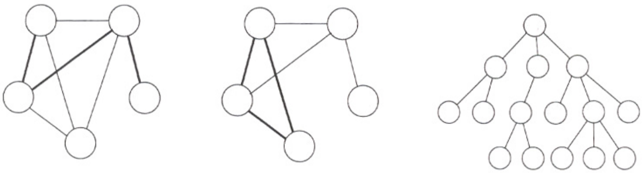


FIGURE 0.14
Graph G (shown darker) is a subgraph of H

Graphs (cont.)



(a) (b) (c)

FIGURE **0.15**
(a) A path in a graph, (b) a cycle in a graph, and (c) a tree

Graphs (cont.)

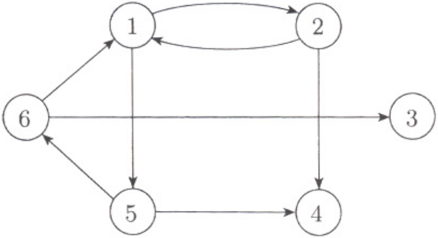


FIGURE **0.16**
A directed graph

Graphs (cont.)

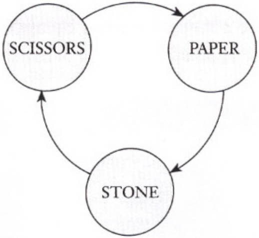


FIGURE 0.18
The graph of the relation *beats*

Definitions, Theorems, and Proofs

- *Definitions* describe the objects and notions that we use. Precision is essential to any definition.
- After we have defined various objects and notions, we usually make *mathematical statements* about them. Again, the statements must be precise.
- A *proof* is a convincing logical argument that a statement is true. The only way to determine the truth or falsity of a mathematical statement is with a mathematical proof.
- A *theorem* is a mathematical statement proven true. *Lemmas* are proven statements for assisting the proof of another more significant statement.
- *Corollaries* are statements seen to follow easily from other proven ones.

An Example Proof

Theorem

For any two sets A and B, $\overline{A \cup B} = \overline{A} \cap \overline{B}$.

An Example Proof

Theorem

For any two sets A and B, $\overline{A \cup B} = \overline{A} \cap \overline{B}$.

Proof. We show that every element of $\overline{A \cup B}$ is also an element of $\overline{A} \cap \overline{B}$ and vice versa.

Forward ($x \in \overline{A \cup B} \rightarrow x \in \overline{A} \cap \overline{B}$):

$$
\begin{array}{lll}
 & x \in \overline{A \cup B} & \\
\rightarrow & x \notin A \cup B & \text{, def. of complement} \\
\rightarrow & x \notin A \text{ and } x \notin B & \text{, def. of union} \\
\rightarrow & x \in \overline{A} \text{ and } x \in \overline{B} & \text{, def. of complement} \\
\rightarrow & x \in \overline{A} \cap \overline{B} & \text{, def. of intersection}
\end{array}
$$

Reverse ($x \in \overline{A} \cap \overline{B} \rightarrow x \in \overline{A \cup B}$): ...

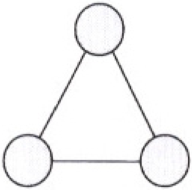
Another Example Proof

Definition: The number of edges at a node u is called the *degree* of u.

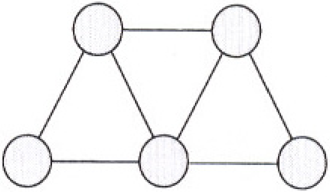
Another Example Proof

Definition: The number of edges at a node u is called the *degree* of u.

Now take a look at the following two graphs:



$$\text{sum} = 2+2+2 = 6$$

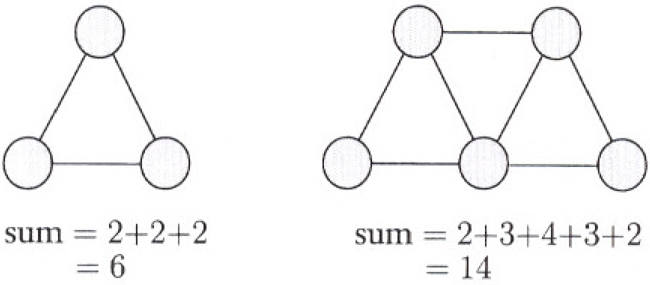


$$\text{sum} = 2+3+4+3+2 = 14$$

Another Example Proof

Definition: The number of edges at a node u is called the *degree* of u.

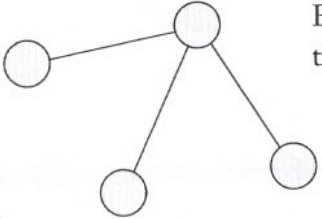
Now take a look at the following two graphs:



Question: Is the sum of the degrees of the nodes in a graph always an even number?

Answer: Yes! Let's try to prove it.

Another Example Proof (cont.)



Every time an edge is added, the sum increases by 2.

Another Example Proof (cont.)

Theorem

In any graph G, the sum of the degrees of the nodes of G is an even number.

Proof.

- Every edge in G connects two nodes, contributing 1 to the degree of each.
- Therefore, each edge contributes 2 to the sum of the degrees of all the nodes.
- If G has e edges, then the sum of the degrees of the nodes of G is $2e$, which is even.

Types of Proof

Many types of proofs exist. Some of the most common ones are:

- *Proof by construction*:
  prove that a particular type of object exists, by showing how to construct the object.
- *Proof by contradiction*:
  prove a statement by first assuming that the statement is false and then showing that the assumption leads to an obviously false consequence, called a contradiction.
- *Proof by induction*:
  prove that all elements of an infinite set have a specified property, by exploiting the inductive structure of the set.

Proof by Construction

Definition: A graph G is *k-regular* if every vertex in G has degree k.

Theorem

For each even number n greater than 2, there exists a 3-regular graph with n nodes.

Proof by Construction

Definition: A graph G is *k-regular* if every vertex in G has degree k.

Theorem

For each even number n greater than 2, there exists a 3-regular graph with n nodes.

Proof. Construct a graph $G = (V, E)$ with n $(= 2k \geq 2)$ nodes as follows.

Let V be $\{0, 1, \ldots, n-1\}$ and E be defined as

$$\begin{aligned} E \;=\; & \{\{i, i+1\} \mid \text{for } 0 \leq i \leq n-2\} \cup \\ & \{\{n-1, 0\}\} \cup \\ & \{\{i, i+n/2\} \mid \text{for } 0 \leq i \leq n/2-1\}. \end{aligned}$$

Proof by Contradiction

Definition: A real number is *irrational* if it cannot be written as $\frac{m}{n}$, where m and n are integers.

Theorem

$\sqrt{2}$ *is irrational.*

Proof by Contradiction

Definition: A real number is *irrational* if it cannot be written as $\frac{m}{n}$, where m and n are integers.

Theorem

$\sqrt{2}$ is irrational.

Proof. Assume toward a contradiction that $\sqrt{2}$ is rational, i.e., $\sqrt{2} = \frac{m}{n}$ for some integers m and n, which *cannot both be even*.

$\sqrt{2} = \frac{m}{n}$	, from the assumption
$n\sqrt{2} = m$	, multipl. both sides by n
$2n^2 = m^2$	, square both sides
m is even	, m^2 is even
$2n^2 = (2k)^2 = 4k^2$	, from the above two
$n^2 = 2k^2$	, divide both sides by 2
n is even	, n^2 is even

Now both m and n are even, a contradiction.

Example: Home Mortgages

P: the *principle* (amount of the original loan).
I: the yearly *interest rate*.
Y: the monthly payment.
M: the *monthly* multiplier $= 1 + I/12$.
P_t: the amount of loan outstanding after the t-th month; $P_0 = P$ and $P_{k+1} = P_k M - Y$.

Theorem

For each $t \geq 0$,

$$P_t = PM^t - Y\left(\frac{M^t - 1}{M - 1}\right).$$

Proof by Induction

Theorem

For each $t \geq 0$,

$$P_t = PM^t - Y(\frac{M^t - 1}{M - 1}).$$

Proof. The proof is by induction on t.

- *Basis*: When $t = 0$, $PM^0 - Y(\frac{M^0-1}{M-1}) = P = P_0$.

Proof by Induction (cont.)

- *Induction step*: When $t = k + 1$ ($k \geq 0$),

$$
\begin{array}{cl}
 & P_{k+1} \\
= & \quad \{\text{definition of } P_t\} \\
 & P_k M - Y \\
= & \quad \{\text{the induction hypothesis}\} \\
 & (PM^k - Y(\frac{M^k - 1}{M-1}))M - Y \\
= & \quad \{\text{distribute } M \text{ and rewrite } Y\} \\
 & PM^{k+1} - Y(\frac{M^{k+1} - M}{M-1}) - Y(\frac{M-1}{M-1}) \\
= & \quad \{\text{combine the last two terms}\} \\
 & PM^{k+1} - Y(\frac{M^{k+1} - 1}{M-1})
\end{array}
$$