Turing Machines

* Finite State Machines
  Regular Languages
* Pushdown Automata
  Context-Free Languages
* TM: Turing Machines

A new model of computation.
Not much more elaborate.
A "model" for all computers.
"Decidable" Languages
"Turing Recognizable" Languages
Languages that are
"Not Turing Recognizable"
Classes of Languages (The "Language Onion")

All Languages (Some are not Turing Recognizable)

Turing Recognizable

Decidable

Context-Free

Regular

This is a Venn Diagram. The subset relationships are all "proper subset"
Turing Machine Definition

Note: There are variations in the exact definition from textbook to textbook. All variations are equivalent!

We'll discuss this later.

Data Structure

FSM
- The input string

PDA
- The input string
- A stack

TM
- A "tape"

Symbols from an alphabet \( \Sigma \)
A special blank symbol \( \_ \)
Infinite in one direction — but filled with blanks.
Current position
Tape Alphabet

Typical: \[ \Sigma = \{0, 1\} \]

But also common:
\[ \Sigma = \{0, 1, a, b, x, \#, \$\} \]

The "blank" symbol is special
\[ \_ \notin \Sigma \]

Initial configuration:

```
011011011_...
```

The "input" string

Blanks out to infinity.

The current position ("the tape head")

Initially at the leftmost cell.
Can move left or right.
Can read ("scan") the current symbol
Can write the current symbol
Finite State Machine

The control portion. Similar to a FSM or PDA. The "PROGRAM." Deterministic.

The READ/WRITE Tape "HEAD"

101101010...

Portion of the "TAPE" that has been used so far

Unused portion of the tape. Infinite in length. Filled with a special "BLANK" symbol——
RULES OF OPERATION

At each step of the computation:

• Read the current symbol
• Update (i.e., write) the same cell.
• Move exactly one cell either LEFT or RIGHT.

(If we are at the left end of the tape and trying to move left, then do not move; stay at left end.)

Don't want to update the cell? Just write the same symbol.

States

Symbol to read
Symbol to write
Direction to move: "L" or "R"
Rules of Operation - 2

- Control is with a sort of finite state machine.
- Initial state
- Final states
  - The "Accept" state
  - The "Reject" state

- Computation can...
  - Halt and "Accept"
    (Whenever the machine enters the Accept state, computation immediately halts.
  - Halt and "Reject"
    (Whenever the machine enters the Reject state, computation immediately halts.
  - "Loop"
    (The machine fails to halt.

- The TM is deterministic.
An Example

$L = 01^*0$

Is it deterministic?

Henceforth:
If an edge is missing... Assume it leads to **reject**.
EXAMPLE
\[ L = \{0^n 1^n \} \]

INPUT
ALPHABET:
\[ \Sigma = \{0, 1\} \]

ALGORITHM

CHANGE "0" TO "x"
MOVE RIGHT TO FIRST "1"
IF NONE: REJECT.
CHANGE "1" INTO "y"
MOVE LEFT TO LEFTMOST "0"
REPEAT UNTIL NO MORE "0"s
MAKE SURE NO MORE "1"s REMAIN

A COMPUTATION HISTORY

\[
\begin{array}{cccccccc}
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
x & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
x & 0 & 0 & 0 & x & y & 1 & 1 \\
x & x & x & 0 & 0 & y & y & y \\
\vdots \\
x & x & x & x & x & y & y & y \\
\end{array}
\]

TAPE
ALPHABET:
\[ \Gamma = \{0, 1, x, y, \_\} \]
QUESTION:

Is this machine correct?
Does it work?
Does it contain bugs?

TM's model computers.
In this way they are similar!
\[(Q, \Sigma, \Gamma, \delta, q_0, q_{\text{accept}}, q_{\text{reject}})\]

- \(Q = \text{Set of States.}\)
- \(\Sigma = \text{Input Alphabet.}\)
- \(\Gamma = \text{Tape Alphabet} \quad \Sigma \subseteq \Gamma\)
  
  \[\text{THE INPUT CANNOT CONTAIN A BLANK.} \quad \_ \notin \Sigma \text{ AND } \_ \in \Gamma\]
  
  \(q_0 = \text{Initial State} \quad q_0 \in Q\)
  
  \(q_{\text{accept}} \in Q \quad \text{WE ONLY NEED ONE}\)
  
  \(q_{\text{reject}} \in Q \quad \text{ACCEPT AND ONE}\)
  
  \(\text{REJECT STATE.}\)

\[\delta : Q \times \Gamma \rightarrow Q \times \Gamma \times \{L, R\}\]

\(\text{Transition function}\)
"Configuration" gives the entire state of the machine. A snapshot of execution at some step.

Need:
- Contents of the tape.
- Location of the "tape head".
- Current state.

A configuration is a string like this:

0 1 1 1 0 1 1

A sequence of configurations, starting with the "start configuration," and ending with an [accepting] configuration, and containing only legal transitions provides a "computation history."
DECIDABLE LANGUAGES

When given a string as input, the TM will always halt.
The TM will ACCEPT if it is in L.
The TM will REJECT if it is not in L.

TURING RECOGNIZABLE LANGUAGES

When given a string that is in the language, the TM will always HALT and ACCEPT.
When given a string that is not in the language, the TM will either REJECT or LOOP.

NOT TURING RECOGNIZABLE LANGUAGES

CAN'T EVEN RECOGNIZE MEMBERS RELIABLY!

Also:
"RECURSIVE"
"COMPUTABLE"
"SOLVABLE"

Also:
"RECURSIVELY ENUMERABLE," RE
"PARTIALLY DECIDABLE" "SEMI-DECIDABLE"

Also:
NOT RECURSIVELY ENUMERABLE
NOT R.E.
NOT PARTIALLY DECIDABLE
Turing machine uses
* To "decide" a language
* To "recognize" a language
* To compute a function

"Computable" = Decidable
"Totally computable" defined on all inputs.

Partially computable functions
Undefined on some inputs
"Semi-decidable functions"
THE CHURCH-TURING THESIS

1930's: What does "computable" mean?
Alonzo Church: LAMBDA CALCULUS
Alan Turing: TMs.

Several variations on Turing Machines.
- One tape or many?
- Infinite on both ends?
- Tiny alphabet 0, 1 or not?
- Can the head also stay in the same place?
- Allow nondeterminism.

All variations are equivalent in computing capability!

TM's and Lambda Calculus are also equivalent in power.

Conclusion: (Or Definition?)
"Algorithmically computable" equals "Computable by a TM."
A Venn Diagram

ALL PROBLEMS
(ALL LANGUAGES)

NOT DECIDABLE

TURING
RECOGNIZABLE
(RECURSIVELY
ENUMERABLE)

DECIDABLE
(RECURSIVE)

C.F.L.
(REG)

NOT TURING
RECOGNIZABLE
(Not R.E.)
How can we recognize the left end of the tape?

**GOAL:** Want to put a special symbol $\#$ on the left end and shift the input over 1 cell to the right.

Assume $\Sigma = \{a, b, \#$}
Q: How much TM programming do you need to do?
A: Just enough to get the idea and to convince yourself that all programs/algorithms can be implemented on a TM.

Machine Code
0110,1100

Assembly Code
ADD R1,R2,R3

C code
i = (2+k)*n;

Algorithms
If SNT={}...
No implementation details.

TURING MACHINES
STATES,
TRANSITION FUNCTION
(complete TM specification)

Outline of Algorithm
Still talking about Tape head movement,
Data representation

High-level specification of algorithm
No TM-specific details
If SNT={}...
EXAMPLE

Build a TM to recognize the language $0^n1^n0^n$.

"Build a TM to decide the language."

This language is not context-free. So this will prove

$\text{CONTEXT-FREE LANGUAGES} \subset \text{DECIDABLE LANGUAGES}$

proper subset
We already have a TM to turn $0^n1^n$ into $x^n y^n$ and to recognize decide that language.

**IDEA**

Use that TM as a **SUBROUTINE**!

**STEP 1:**

```
000000 111111 000000
```

```
\downarrow
```

```
xxxxxx yyyyyy 000000
```

Reject if problems.

**STEP 2:**

Build a similar TM to recognize $y^n0^n$

**STEP 3:**

Build the final TM by "gluing" these smaller TM's together into one larger TM.
**Problem:**
Compare two strings.
A TM to decide $\{w \neq \#w \mid w \in \{a, b, c\}^*\}$

**Solution:**
Use a new symbol, such as "x".
Turn each symbol into an x after it has been examined.

```
a b a c \# a a b a c
\downarrow
x a b a c \# x a b a c
\downarrow
x x b a c \# x x b a c
\downarrow*
xx xx xx xx xx xx xx xx xx xx
```
**Problem:**
Do it nondestructively, without losing the original strings. (Perhaps this task is part of a larger task.)

**Solution:**
"MARK" each symbol to keep track of what we've already done. Add some new symbols to help.

\[
\begin{align*}
a & \rightarrow x \\
b & \rightarrow y \\
c & \rightarrow z \\
\end{align*}
\]

\[
\begin{align*}
aabbcc & \rightarrow #aaabca \\
\downarrow & \\
xxyz & \rightarrow #xxyyza \\
\end{align*}
\]

Later, restore the strings, if we need to:

\[
\begin{align*}
x & \rightarrow a, R \\
yz & \rightarrow b, R \\
z & \rightarrow c, R \\
\end{align*}
\]
PROGRAMMING TECHNIQUE: 

**MARKING SYMBOLS**

"Mark each symbol with a dot."
\[ \Gamma = \{ a, b, c, \cdot, \cdot, \cdot, \# \} \]

"Put a dot under this symbol."

\[
\begin{array}{c}
{\mathbf{a}}^3 \\
\Rightarrow \\
{\mathbf{a}}^4
\end{array}
\]

"Remember this location."

\[ \Rightarrow \text{Mark that symbol with a dot.} \]

We could have several different types of marks, such as \[ \uparrow \uparrow \uparrow \]

```
0110100101110001
```

"Let \( P \) point to the beginning of the second string and let \( q \) point to..."
**Theorem**

Every multitape Turing machine has an equivalent single-tape Turing machine.

"Equivalent" means it decides/recognizes the same languages. It's not about speed, efficiency or ease of programming.

**Proof**

Given a multitape TM, show how to build a single-tape TM.

* Need to store all tapes on a single tape.
  => Show data representation.
* Each tape has a tape "head".
  => Show how to store that info.
* Need to transform moves in the multitape TM into one or more moves in the single-tape TM.
MULTITAPE TM:

F.S.M.

\[ \text{SINGLE-TAPE TM:} \]

\[ \text{A TM with one tape to simulate it.} \]

\[ \text{AN EXAMPLE MACHINE WITH K=3 TAPES.} \]

\[ \text{ADD "DOTS" TO SHOW WHERE HEAD "K" IS.} \]

\[ \text{To simulate a transition from state Q, we must scan our tape to see which symbols are "UNDER" the K tape heads.} \]

\[ \text{Once we determine this and are ready to "MAKE" the transition, we must scan across the tape again to update the cells and move the dots.} \]

\[ \text{Whenever one head moves off the right end, we must shift our tape so we can insert a } \_ \text{.} \]
Non-deterministic Turing Machines

Transition Function

\[ \delta : Q \times \Gamma \rightarrow P(Q \times \Gamma \times \{L, R\}) \]

Deterministic

Non-deterministic
A "CONFIGURATION" is...

- A way to represent the entire state of a TM at one moment during a computation.
- A string which captures
  
  THE CURRENT STATE
  
  THE CURRENT POSITION OF HEAD
  
  THE ENTIRE TAPE CONTENTS.

\[
\begin{array}{c}
  \text{ABBACAA}_- \\
  \text{ABBB}_37\text{ACA}_a \\
\end{array}
\]

WITH NONDETERMINISM:

At each moment in the computation there can be MORE THAN ONE successor configuration.
A tree shows the computation of a non-deterministic TM.

Configurations of the computation

Initial configuration

Choices – More than one next configuration

Some branches of computation may reach ACCEPT and Halt.

Some branches may die out (i.e., REJECT and HALT)

Some branches may never halt.
Outcomes of a Nondeterministic Computation:

**Accept**

If any branch of the computation accepts, then the nondeterministic TM will accept.

**Reject**

If all branches of the computation halt and reject (i.e., no branches accept, but all computation halts), then the nondeterministic TM rejects.

**Loop**

Computation continues, but "accept" is never encountered.

Some branches in the computation history are infinite.
EVERY NONDETERMINISTIC TM HAS AN EQUIVALENT DETERMINISTIC TM.

Proof

- Given a non-deterministic TM, \(N\), show how to construct an equivalent deterministic TM \(D\).
- If \(N\) accepts (on any branch), then \(D\) will accept.
- If \(N\) halts on every branch without any "accepts", then \(D\) will halt and reject.

Approach: Simulate \(N\); simulate all branches of computation; search for any way \(N\) can accept.
The "Computation History" is a tree showing all possible branches/choices in a nondeterministic computation:

- A path to any node is given by a number.
- Search the tree, looking for "Accept."
- Search order?
  - Depth-first? < No!
  - Breadth-first! < Yes!
- To examine a node:
  - Perform the entire computation from scratch.
  - The path numbers tell which of the many nondeterministic choices to make.
How many choices at each step in the computation?

Examine the nondeterministic machine; there will be some maximum.

**Example:**

```
    9
   /\
  1-2
 /   \
3    3
```

Breadth-first search order:

```
  1 2 3
 /   \
1 1 1
```

In any particular computation there will be fewer than 3 choices at most of the computation steps.

Some points will have zero choices ⇒ these branches of the nondeterminism halt and reject.
TM
Finite State Control

INPUT TAPE
... Initial input; never modified.

SIMULATION TAPE
Used like the tape of a deterministic TM to perform the simulation.

ADDRESS TAPE
Used to control the breadth-first search. Tells which choices to make during a simulation.
Algorithm

Initially: Tape 1 contains the input.
Tapes 2 and 3 are empty.

1. Copy Tape 1 to Tape 2.
2. Run the simulation.
   Use Tape 2 as "The Tape."
3. When choices occur (i.e., when non-deterministic branch points are encountered) consult Tape 3.
   Tape 3 contains a "Path." Each number tells which choice to make.
4. Run the simulation all the way down the branch, as far as the address/path goes.
   (Or the computation "dies out.")
5. Try the next branch increment the address on Tape 3.
6. Repeat

If accept is ever encountered,
   halt and "Accept."
If all branches reject or die out,
   then halt and "Reject."
A LANGUAGE is "TURING RECOGNIZABLE" IFF SOME NON-DETERMINISTIC TURING MACHINE RECOGNIZES IT.

HALTING?
If a nondeterministic TM halts on ALL branches without ACCEPTING, then it REJECTS.

A LANGUAGE is "DECIDABLE" IFF SOME NON-DETERMINISTIC TURING MACHINE DECIDES IT.

DECIDES?
- Will always halt!
- Will always ACCEPT or REJECT!
- Will never LOOP!
Any arbitrary problem can be expressed as a language.

Any instance of the "problem" is encoded into a string.

The string is in the language $\equiv$ The answer is "yes"

The string is not in the language $\equiv$ The answer is "no"
**Example**

**Undirected Graphs**

**Is this graph connected?**

We must encode the problem into a language.

\[ A = \{ \langle G \rangle \mid G \text{ is a connected graph} \} \]

We would like to find a TM to decide this language.

**Accept** = "YES," this is a connected graph.

**Reject** = "NO," this is not a connected graph [or this is not a valid representation of a graph].

**Loop** = ... This problem is DECIDABLE. Our TM will always halt.
ONE REASONABLE REPRESENTATION:

\[ \langle G \rangle = \]

\[ (1, 2, 3, 4) \left\{ (1, 2), (2, 3), (1, 3), (1, 4) \right\} \]

- List of node "names"
- List of edges
- An edge from 1 to 3

\[ \Sigma = \{ (\_, \_), 1, 2, 3, 4, \ldots, 9 \} \]

\[ (1, 2, 3, 3, \ldots) \]
REPRESENTING NUMBERS ON A TAPE

DECIMAL

$$\Sigma = \{0, 1, 2, \ldots, 9, \#, \$, \ldots, \}$$

$$\cdots, 1439, 1\ldots$$

BINARY

$$\Sigma = \{0, 1\}$$

$$\cdots, 1101101\ldots$$

= 22

UNARY

$$\Sigma = \{1\}$$

$$\cdots, 1111111\ldots$$

= 8

(This is no more than a programming detail.)
Algorithm = Turing Machine

- High-level specification
  Pseudo-code
  Expressed in programming language

- Implementation-level
  Contents of the tape.
  Data representation.
  Motion of the tape head.
  More detail, but still abstract

- TM specification
  States
  Alphabets
  Transition function
  Fully detailed (and incomprehensible?)

Once we are comfortable with TM specification details, we'll start to give more abstract algorithms, with the understanding that we could build the exact TM whenever necessary.
**High-Level Algorithm**

Select a node and mark it.

Repeat

For each node N ...

If N is unmarked and there is an edge from N to an already marked node then mark node N.

End

Until no more nodes can be marked

For each node N ...

If N is unmarked then "Reject"

End

"Accept"
IMPLEMENTATION - LEVEL ALGORITHM

- Check that input describes a valid graph
  - Check node list
    - Scan "C", followed by digit,...
  - Check that all nodes are different, i.e., no repeats.
  - Check edge list...
    etc.

- Mark first node.
  - Place a dot under the first node in node list.

- Scan the node list to find a node that is not marked...
  etc., etc.
ENUMERATORS

LIKE A TURING MACHINE:
  INFINITETIME
  FINITE STATE CONTROL

Plus...
  A PRINTER

OPERATION

• THE TAPE IS INITIALLY EMPTY
  (i.e., NO INPUT).

• THE PRINTER PRODUCES A SERIES
  OF STRINGS.

• THE MACHINE ENUMERATES
  (i.e., IT "LISTS OUT"/"PRINTS")
  THE STRINGS IN A LANGUAGE.

• IT MAY HALT OR LOOP.

• THE LANGUAGE MAY BE INFINITE.

• IT MAY PRINT OUT Duplicates
  (JUST IGNORE DUPLICATE STRINGS).

• IT MAY PRINT IN ANY ORDER.
THEOREM

A LANGUAGE IS TURING-RECOGNIZABLE IFF SOME ENUMERATOR ENUMERATES IT.

PROOF

GIVEN "E" CONSTRUCT A TM "M".
ON INPUT w...
RUN "E"
COMPARE EACH OUTPUT STRING TO w.
IF WE FIND A MATCH, ACCEPT.

GIVEN A TM "M", CONSTRUCT AN ENUMERATOR "E".

CONSTRUCT E, USING M AS A "SUBROUTINE".

RUN M ON ALL POSSIBLE STRINGS. A ^
IF M EVER ACCEPTS A STRING, THEN PRINT IT OUT.

PROBLEM?
M MIGHT LOOP ON SOME PARTICULAR STRING.
WE MUST RUN ALL THESE SIMULATIONS IN PARALLEL!
GOAL
Run a TM on all strings in $\Sigma^*$ simultaneously (i.e., "in parallel")

APPROACH
We can list out all strings in
$\Sigma^* = \exists s_1, s_2, s_3, \ldots$ 3

EXAMPLE: $\epsilon, 0, 1, 00, 01, 10, 11, 000, \ldots$

The computation on any string $s_i$ may be infinite!
We must not get stuck on some string.

EXAMPLE: $s_4$ might infinite loop,
but $s_7$ might ACCEPT

:: INTERLEAVE THE COMPUTATIONS.
  Work on $s_1$ a little bit.
  Work on $s_2$ a little bit.
  
  Eventually, we must go back to $s_1$ and do a little more work.
Algorithm

FOR i = 1, 2, 3, ... (infinite loop)
  FOR j = 1 TO i
    Simulate M, the Turing Machine.
    Use sj as input.
    Run simulation for i steps.
    If M accepts sj (within i steps)
    Then PRINT sj
  END
END

Note: Every (i, j) pair will eventually be encountered.
Example

Assume

$S_1$ is ACCEPTED after 3 steps.
$S_3$ is ACCEPTED after 4 steps.

Output:

$S_i$ $S_i$ $S_3$ $S_i$ $S_3$ ...