Streaming Algorithms

Difference v. DFA

- → Can output >1 bits
- → Memory can increase as longer strings are read
  - → can recognize non-regular languages
- → Sometimes can allow making multiple passes over data
- → Can be randomized

Space complexity

Theorem: Suppose a language L can be recognized by a DFA M with $\leq 2^p$ states.
Then L is computable by a streaming algorithm A using $\leq p$ bits.

The DE (distinct elements) problem

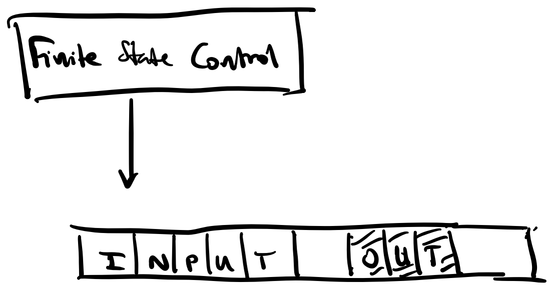
Input: $x \in \{0, 1, \dots, 2^k\}^n$, $2^k > |x|^2$

Output: # distinct elements in x

Theorem: Every algorithm for DE requires $\Omega(kn)$ space

Theorem: There is a randomized streaming algorithm that can approximate DE to $\sim 0.1\%$ error, using $O(k + \log n)$ space

Turing Machines

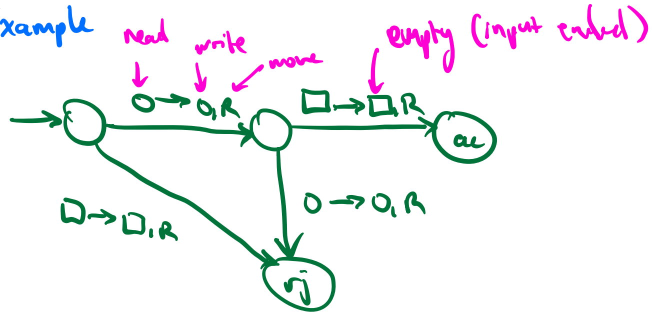


Infinite rewritable tape

TM vs DFA

→ TM can both read/write tape
→ TM head can move left/right
→ Input doesn't need to be read entirely
Computation can run (even forever) after all input has been read
→ Accept, Reject take immediate effect

Example



Decides vs Recognizes

A TM <u>decides</u> a language L if it explicitly accepts all strings in L and rejects all strings not in L

A TM <u>recognizes</u> a language L if it accepts all strings in L, and rejects or infinitely loops for all strings not in L

A language L is <u>recognizable</u> (recursively enumerable, <u>RE</u>) if there exists a TM that recognizes it

A language L is decidable (recursive, R) if there exists a TM that decides it

Formalization

A Turing Machine is a 7-tuple

$$T = (Q, \Sigma, \Gamma, \delta, q_0, q_{acc}, q_{rej})$$

where

Q: finite set of states

Σ: Input alphabet, where $\square \notin \Sigma$

Γ: Tape alphabet, where $\square \in \Gamma$, $\Sigma \subseteq \Gamma$

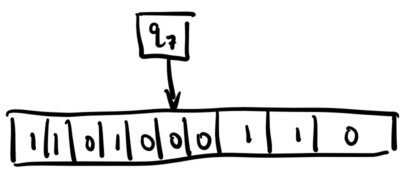
δ: $Q \times \Gamma \to Q \times \Gamma \times \{L, R\}$

$q_0 \in Q$: start state

$q_{accept} \in Q$: accept state

$q_{reject} \in Q$: reject state, $q_{accept} \neq q_{reject}$

TM configuration



corresponds to the configuration

$11010q_7 00110 \in \{Q \cup \Gamma\}^*$

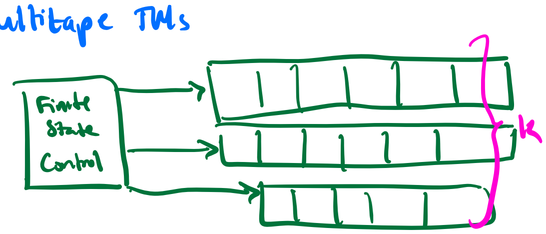
TM acceptance and rejection

Let C_1, C_2 be configs of a TM M

Definition: C_1 <u>yields</u> C_2 if M is in C_2 after running M in C_1 for one step

TM variants

Multitape TMs



Note: every multitape TM can be converted into a single-tape TM

The Church-Turing thesis

"Everyone's intuitive notion of algorithms = Turing machines"

(note: hypothesis, not theorem)