Lecture 5. The Incompressibility Method

- A key problem in computer science: analyze the average case performance of a program.
- Using the Incompressibility Method:
  - Give the program a random input (with high Kolmogorov complexity)
  - Analyze the program with respect to this single and fixed input. This is usually easier using the fact this input is incompressible.
  - The running time for this single input is the average case running time of all inputs!

Solutions to open questions

- We will tell the histories of some open questions and the ideas of how they were solved by the incompressibility method.
- We will not be able to give detailed proofs to these problems … but hopefully by telling you the ideas, you will be convinced enough and able to reconstruct the details on your own.
- Through these stories, we hope to convince you of the power of Kolmogorov complexity and hope you will extend it. As Fortnow puts it: May your proofs be short and elegant.

A selected list of results proved by the incompressibility method

- \( \Omega(n^2) \) for simulating 2 tapes by 1 (20 years)
- \( k \) heads > \( k-1 \) heads for PDAs (15 years)
- \( k \) one-ways heads can’t do string matching (13 yrs)
- 2 heads are better than 2 tapes (10 years)
- Average case analysis for heapsort (30 years)
- \( k \) tapes are better than \( k-1 \) tapes. (20 years)
- Many theorems in combinatorics, formal language/automata, parallel computing, VLSI
- Simplify old proofs (Hastad Lemma).
- Shellsort average case lower bound (40 years)

Boolean matrix rank (J. Seiferas and Y. Yesha)

- Consider matrix over GF(2): 0,1 elements, with usually Boolean x,+ operations. Such properties are often needed for example in proving tradeoff optimal bound \( TS=\Omega(n^3) \) for multiplying 2 matrices.

  Theorem. For each \( n \), there is an \( nxn \) matrix over GF(2) s.t. every submatrix of \( s \) rows and \( n-r \) columns has at least rank \( s/2 \), \( 2\log n < r, s < n/4 \).

  Proof. Take \( \mid x \mid=n^2 \). \( C(x)\geq n^2 \). Form an \( nxn \) matrix with \( x \), one bit per entry. For any submatrix \( R \) with \( s \) rows and \( n-r \) columns, if \( R \) does not have rank \( s/2 \), then \( s/2+1 \) rows can be linearly described by other rows. Then you can compress the original matrix, hence \( x \).

  QED
Sorting

- Given n elements (in an array). Sort them into ascending order.
- This is the most studied fundamental problem in computer science.

- Heapsort: Open for 40 years: Which is better in average case: Williams or Floyd?

- Shellsort (1959): P passes. In each pass, move the elements “in some stepwise fashion” (Bubblesort)
  - Open for over 40 years: a nontrivial general average case complexity lower bound of Shellsort?

Heapsort

- Immediately it was improved by RW Floyd.
- Worst case $O(n \log n)$.
- Open for 40 years: Which is better in average case: Williams or Floyd?

Heapsort average analysis (I. Munro)

- Average-case analysis of Heapsort.

- Heapsort: (1) Make Heap. $O(n)$ time.
  (2) Deletemin, restore heap, repeat.

- Williams
  \[ 2 \log n - 2d \]

- Floyd
  \[ \log n + d \]

Fix random heap $H$, $C(H) > n \log n$. Simulate Step (2). Each round, encode the red path in $\log n - d$ bits. The $n$ paths describe the heap! Hence, total $n$ paths, length $\geq n \log n$, $d$ must be a constant. Floyd takes $n \log n$ comparisons, and Williams takes $2n \log n$.

Shellsort Algorithm

- Using $p$ increments $h_1, \ldots, h_p$, with $h_p = 1$
- At $k$-th pass, the array is divided in $h_k$ separate sublists of length $n/h_k$ (taking every $h_k$-th element).
- Each sublist is sorted by insertion/bubble sort.

- Application: Sorting networks --- $n \log^2 n$ comparators.
Shellsort history

- Invented by D.L. Shell [1959], using \( p_k = \frac{n}{2^k} \) for step \( k \). It is a \( \Theta(n^2) \) time algorithm.
- Papernow & Stasevitch [1965]: \( O(n^{3/2}) \) time.
- Pratt [1972]: \( O(n \log^2 n) \) time.
- Incerpi-Sedgewick, Chazelle, Plaxton, Poonen, Suel (1980’s) – worst case, roughly, \( \Omega(n \log \log n / \log \log n)^2 \).
- Average case:
  - Knuth [1970’s]: \( \Theta(n^{5/3}) \) for \( p=2 \)
  - Yao [1980]: \( p=3 \)
  - Janson-Knuth [1997]: \( \Omega(n^{23/15}) \) for \( p=3 \).
  - Jiang-Li-Vitanyi [J.ACM, 2000]: \( \Omega(pn^{1+1/p}) \) for any \( p \).

Shellsort Average Case Lower bound

**Theorem.** \( p \)-pass Shellsort average case \( T(n) \geq pn^{1+1/p} \)

Proof. Fix a random permutation \( \Pi \) with Kolmogorov complexity \( n \log n \). I.e. \( C(\Pi) \geq n \log n \).

For pass \( i \), let \( m_{i,k} \) be the number of steps the \( k \)th element moves. Then \( T(n) = \Sigma_i m_{i,k} \).

From these \( m_{i,k} \)'s, one can reconstruct the input \( \Pi \), hence \( \Sigma \log m_{i,k} \geq C(\Pi) \geq n \log n \).

Maximizing the left, all \( m_{i,k} \) must be the same. Call it \( m \).

\( \Sigma \log m = pn \log m \geq \Sigma \log m_{i,k} \geq n \log n \Rightarrow m^p \geq n \).

So \( T(n) = pn^m > pn^{1+1/p} \).

**Corollary:**
- \( p=1 \): Bubblesort \( \Omega(n^2) \) average case lower bound.
- \( p=2 \): \( n^{1.5} \) lower bound. \( p=3 \), \( n^{2/3} \) lower bound.

The Incompressibility Method applied to Formal language theory

**Example:** Show \( L = \{0^k 1^k \mid k > 0 \} \) not regular. By contradiction, assume that DFA \( M \) accepts \( L \).

Choose \( k \) so that \( C(k) > 2|M| \). Simulate \( M \):

\[
\begin{array}{ccccccc}
0 & 0 & 0 & \ldots & 0 & 1 & 1 & 1 & \ldots & 1 \\
\end{array}
\]

\( \text{stop here} \)

\( M \)

\( Q \)

\( C(k) < |M| + q + O(1) < 2|M| \). Contradiction.

**Remark.** Generalize to iff condition: more powerful & easier to use than “pumping lemmas”.

A simple Turing machine lower bound

- Consider one-tape TM. Input tape is also work tape, allow read/write, two-way head.

**Theorem.** It takes \( \Omega(n^2) \) time for such TM \( M \) to accept \( L = \{ww \mid w \in \Sigma^* \} \).

Proof (W. Paul). Take \( w \) s.t. \( C(w) \geq |w| = n \). Consider \( M \)'s computation on input: \( 0^n w 0^n w \).

Consider the cross sequence below the middle 0.

- If it is \( \Omega(n) \), then the computation time is \( \Omega(n^2) \).
- If it is \( o(n) \), we can use this crossing sequence to find \( w \) by simulating on the “right side” of the crossing sequence by trying all the strings of length \( n \). Then \( C(w) = o(n) \), contradiction.

QED
More on formal language theory

Lemma (Li-Vitanyi) Let $L \subseteq V^*$, and $L_x = \{y: xy \in L\}$. Then $L$ is regular implies there is $c$ for all $x,y,n$, let $y$ be the $n$-th element in $L_x$, we have $C(y) \leq C(n)+c$.

Proof. Like example. QED.

Example 2. \{$1^p: p \text{ is prime}$\} is not regular.

Proof. Let $p_i, i=1,2 \ldots$, be the list of primes. Then $p_{k+1}$ is the first element in $L_{p_k}$, hence by Lemma, $C(p_{k+1}) \leq O(1)$. Impossible. QED

Characterizing regular sets

For any enumeration of $\Sigma^* = \{y_1, y_2, \ldots\}$, define characteristic sequence of $L_x = \{y_i: xy_i \in L\}$ by $X_i = 1$ iff $xy_i \in L$.

Theorem. $L$ is regular iff there is a $c$ for all $x, n$, $C(X_n|n) < c$.

1 tape vs 2 tape Turing machines

- Standard (on-line) TM Model:

  Input tape $\rightarrow$ one way

  Finite Control

  Work tape, two way

- Question since the 1960’s: Are two work tapes better than 1 work tape? How many works tapes are needed?

History

- 1969. Hartmanis & Stearns: 1 work tape TM can simulate $k>1$ tape TM in $O(n^2)$ time.
- 1963. Rabin: 2 work tapes are better than 1.
- 1966. Hennie-Stearns: 2 work tapes can simulate $k$ tapes in $O(n \log n)$ time.
- 1982. Paul: $\Omega(n(\log n)^{1/2})$ lower bound for 1 vs 2 work tapes.
- 1983. Duris-Galil: Improved to $\Omega(n \log n)$.
- 1985. Maass, Li, Vitanyi: $\Omega(n^2)$ tight bound, by incompressibility method, settling the 20 year effort.
Proof Sketch

Here is the language we have used to prove an Ω(n^{1.5}) lower bound:

\[ L = \{ x_1 \# x_2 \# \ldots \# x_k \# y_1 \# \ldots \# y_l \mid x_i = y_j \} \]

Choose random \( x, C(x) \geq |x| = n \), evenly break \( x \) into \( x_1 \ldots x_k \), \( k = \sqrt{n} \).

Then the two work tape machine can easily put \( x_i \) blocks on one tape and \( y_j \) blocks on the other. Then it accepts this language in linear time.

However, the one work tape machine has trouble where to put these blocks. Whichever way it does it, there bounds to be some \( x_i \) and \( y_j \) blocks that are far away, then our previous proof works. The proof needs to worry about not many blocks can be stored in a small region (they are non-compressible strings, hence intuitively we know they can't be). The nice thing about Kolmogorov complexity is that it can directly formulate your intuition into formal arguments.

To improve to \( \Omega(n^2) \) lower bound, we just need to make each \( x_i \) to be constant size. Then argue there are \( O(n) \) pairs of \( (x_i, y_j) \) need to be matched and they are \( O(n) \) away.

K-head PDA’s

Model: Normal finite or pushdown automaton with \( k \) one-way input heads. Thus \( k \)-FA or \( k \)-PDA.

These are natural extensions of our standard definition of FA and PDA.

Two conjectures:
- 1965, Rosenberg Conjecture: \((k+1)\)-FA > \( k \)-FA
- 1968, Harrison-Ibarra Conjecture: \((k+1)\)-PDA > \( k \)-PDA

A tale of twin conjectures

- 1965 Rosenberg actually claimed a proof for \((k+1)\)-FA > \( k \)-FA. But Floyd subsequently found error and the proof fail apart.
- 1971 (FOCS), Sudborough proved 3-FA > 2-FA.
- Ibarra-Kim: 3-FA > 2-FA
- 1976 (FOCS) Yao-Rivest: \((k+1)\)-FA > \( k \)-FA.
- 1973 Ibarra: both conjectures true for 2-way input. This is by diagonalization, does not work for 1-way machines.
- 1982, Miyano: If change pushdown store to counter, then Harrison-Ibarra conjecture is true.
- 1983, Miyano: If input is not bounded, then HI true.
- 1987 (FOCS), Chrobak-Li: Complete solution to Harrison-Ibarra conjecture, using incompressibility method. (The same argument also gives a cute simplification to Yao-Rivest proof.)

Proof Sketch

The language we have used is:

\[ L_b = \{ w_1 \# \ldots \# w_b \# w_{b+1} \# \ldots \# w_{b+k} \mid w_i \in \{0,1\}^* \} \]

Theorem. \( L_b \) can be accepted by a \( k \)-PDA iff \( b \leq k \) choose 2.

When \( b \leq k \) choose 2, then a \( k \)-FA can do it by pairing its \( k \) heads at right places at right time.

When \( b > k \) choose 2, then we can again choose random \( w \) and break it into \( w_b \) blocks. Then we say there must be a pair of \( (w_i, w_j) \) that are indirectly matched (via the pushdown store). But when storing into pushdown store, \( w_i \) is reversed, so it cannot be properly matched with its counter part \( w_j \). We will also need to argue information cannot be reversed, compressed etc. But these are all easy with Kolmogorov complexity.
String-matching by k-DFA

- String matching problem: \( L = \{ x \# y \mid x \text{ is a substring of } y \} \)
- This one of the most important problems in computer science (grep function for example)
- Hundreds of papers written.
- Many efficient algorithms – KMP, BM, KR. Main features of these algorithms:
  - Linear time
  - Constant space (not KMP, BM), i.e. multihead finite automaton. In fact, a two-way 6-head FA can do string matching in linear time (Galil-Seiferas, 1981, STOC)
  - No need to back up pointers in the text (e.g. KMP).
- Galil-Seiferas Conjecture: Can k-DFA for any k, do string matching?

History

- Li-Yesha: 2-DFA cannot.
- Gereb-Graus-Li: 3-DFA cannot
- Jiang-Li 1993 STOC: k-DFA cannot, for any k.

How we did it

- I will just tell you how we did it for 2-DFA.
- Remember the heads are one-way, and DFA does not remember much.
- We can play a game with the 2-DFA with input (of course with Kolmogorov random blocks):
  \[ xy \# y' x' \]
  such that \( x' \) can be \( x \) and \( y' \) can be \( y \), so if the 2-DFA decides to match \( x, x' \) directly, then it won’t be able to match \( y, y' \) directly (and vice versa), so then we simply make \( x' \) different from \( x \), but \( y' = y \). Then without the two heads simultaneously at \( y \) and \( y' \), we will argue, as before, that finite control cannot do it.

How far can we go?

- We have presented at least a dozen of problems that were solved by the incompressibility methods. There are many more … such problems (these include the important switching lemma in circuit complexity as well quantum complexity bounds).
- But can Kolmogorov complexity help to prove higher bounds? Or it is limited to linear, nlogn, \( n^2 \) bounds?
- Can we import some probabilistic method tools?
- If such a tool simply does not work for certain things, like \( \text{NP} \neq \text{P} \), can we be certain about it? (prove this?)