Computational Complexity

Lecture 12: Average-case complexity and Impagliazzo’s Five Worlds

Ronald de Haan
me@ronalddehaan.eu

University of Amsterdam

May 21, 2024
Recap

- Subexponential-time complexity
- Exponential-Time Hypothesis (ETH)
What will we do today?

- Average-case complexity
- One-way functions
- Impagliazzo’s Five Worlds
A problem $L \subseteq \{0, 1\}^*$ can be solved in \textit{worst-case running time} $T(n)$ if there exists an algorithm $A$ that solves $L$ and that halts within time $T(|x|)$ for each $x \in \{0, 1\}^*$.

In other words, the worst-case running time $T(n)$ is the maximum of the running times for all inputs of size $n$. 
A *distributional problem* $\langle L, D \rangle$ consists of a language $L \subseteq \{0, 1\}^*$ and a sequence $D = \{D_n\}_{n \in \mathbb{N}}$ of probability distributions, where each $D_n$ is a probability distribution over $\{0, 1\}^n$. 
The class distP / avgP

Definition (distP)

\(\langle L, \mathcal{D}\rangle\) is in the class distP (also called: avgP) if there exists a deterministic TM \(M\) that decides \(L\) and a constant \(\epsilon > 0\) such that for all \(n \in \mathbb{N}\):

\[
\mathbb{E}_{x \in \mathcal{D}_n} \left[ \text{time}_M(x)^\epsilon \right] \text{ is } O(n).
\]

- The \(\epsilon\) is there for technical reasons—to invert a polynomial to \(O(n)\).
Definition (P-computable distributions)

A sequence \( \mathcal{D} = \{ \mathcal{D}_n \}_{n \in \mathbb{N}} \) of distributions is P-computable if there exists a polynomial-time TM that, given \( x \in \{0, 1\}^n \), computes:

\[
\mu_{\mathcal{D}_n}(x) = \sum_{y \in \{0, 1\}^n \atop y \leq x} \mathbb{P}[y],
\]

where \( y \leq x \) if the number represented by the binary string \( y \) is at most the number represented by the binary string \( x \).
Definition (P-samplable distributions)

A sequence $\mathcal{D} = \{\mathcal{D}_n\}_{n \in \mathbb{N}}$ of distributions is P-samplable if there exists a polynomial-time probabilistic TM $\mathbb{M}$ such that for each $n \in \mathbb{N}$, the random variables $\mathbb{M}(1^n)$ and $\mathcal{D}_n$ are equally distributed.
The class distNP and sampNP

**Definition (distNP)**

A problem $\langle L, D \rangle$ is in distNP if $L \in \text{NP}$ and $D$ is P-computable.

**Definition (sampNP)**

A problem $\langle L, D \rangle$ is in sampNP if $L \in \text{NP}$ and $D$ is P-samplable.

- The questions “distNP $\not=} \text{distP}$” and “sampNP $\not=} \text{distP}$” are average-case analogues of the question “NP $\not=} \text{P}$”
**Definition (one-way functions)**

A polynomial-time computable function $f : \{0, 1\}^* \rightarrow \{0, 1\}^*$ is a *one-way function* if for every polynomial-time probabilistic TM $M$ there is a negligible function $\epsilon : \mathbb{N} \rightarrow [0, 1]$ such that for every $n \in \mathbb{N}$:

$$\Pr_{x \in \mathbb{R} \{0, 1\}^n, y = f(x)} \left[ M(y) = x' \text{ such that } f(x') = y \right] \leq \epsilon(n)$$

where a function $\epsilon : \mathbb{N} \rightarrow [0, 1]$ is *negligible* if $\epsilon(n) = \frac{1}{n^{\omega(1)}}$, that is, for every $c$ and sufficiently large $n$, $\epsilon(n) < \frac{1}{n^c}$.

- Conjecture: there exist one-way functions (implying $P \neq NP$)
- OWFs can be used to create private-key cryptography
Definition

An *encryption scheme* is a pair \((E, D)\) of algorithms, each taking a key \(k\) and a message \(x\), such that \(D_k(E_k(x)) = x\).

The scheme is *perfectly secret*, for messages of length \(m\) and keys of length \(n\), if for every pair \(x, x' \in \{0, 1\}^m\) of messages, the distributions \(E_U^n(x)\) and \(E_U^n(x')\) are identical.

The scheme is *computationally secure* if for every probabilistic polynomial-time algorithm \(A\), there is a negligible function \(\epsilon : \mathbb{N} \rightarrow [0, 1]\) such that

\[
\mathbb{P}_{k \in_R \{0, 1\}^n, x \in_R \{0, 1\}^m} \left[ A(E_k(x)) = (i, b) \text{ s.t. } x_i = b \right] < \frac{1}{2} + \epsilon(n).
\]

Suppose that OWFs exist. Then for every \(c \in \mathbb{N}\) there exists a computationally secure encryption scheme \((E, D)\) using \(n\)-length keys for \(n^c\)-length messages.
Impagliazzo’s Five Worlds (1995)

Five possible situations regarding the status of various complexity-theoretic assumptions:

- Algorithmica
- Heuristica
- Pessiland
- Minicrypt
- Cryptomania

P = NP (or NP ⊆ BPP)

- Say, SAT is linear-time solvable
- This is a computational utopia
- There exist efficient algorithms for creative tasks, e.g., writing proofs
- Essentially no cryptography possible (private-key nor public-key)
- $P \neq NP$, but $\text{distNP, sampNP} \subseteq \text{distP}$

- Breakthroughs of $P = NP$ work almost all the time

- So cryptography breaks too
distNP, sampNP $\not\subseteq$ distP (so $P \neq NP$)

- one-way functions do not exist

- No computational breakthroughs, and most cryptography schemes do not work
One-way functions exist (so $P \neq NP$ and $\text{distNP} \not\subseteq \text{distP}$)

- No “$P = NP$”-type breakthroughs

- Private-key cryptography works

- All “highly structured” problems in NP, such as integer factoring, are solvable in polynomial-time

- Public-key cryptography might not work
Factoring large integers takes exponential time on average (or a corresponding result for a similar problem)

- No general-purpose efficient algorithms ($P \neq NP$)
- Private-key and public-key cryptography works
Impagliazzo’s Five Worlds (1995)

- Five worlds:
  - Algorithmica – efficient general-purpose algorithms
  - Heuristica
  - Pessiland – worst of all worlds
  - Minicrypt
  - Cryptomania – all kinds of cryptography possible

- (Technically, these cases are not exhaustive—there are some “weirdland” scenarios, e.g., the case where SAT \( \in \mathsf{P} \), but the fastest algorithm takes time \( \Theta(n^{100}) \).)
Recap

- Average-case complexity
- One-way functions
- Impagliazzo’s Five Worlds
Recent research directions: Meta Complexity