# Average-case complexity of randomized computations with bounded error

Dmitry Itsykson

Steklov Institute of Mathematics at St. Petersburg

EWSCS
March 2, 2009

# Outline

1. Worst-case complexity
   - $\mathbf{P}, \mathbf{NP}, \mathbf{BPP}$
   - Structural properties: time hierarchy and complete problems
2. Average-case complexity
   - Distributional problems
   - Average-case tractability
   - Class **AvgBPP** and cryptography
3. Results: structural properties of **AvgBPP**.

# P, NP, BPP

| Class | Problem | Turing machine | Time | Error |
|---|---|---|---|---|
| **P** | language $L$ | deterministic $M$ | poly | no error $\forall x\ M(x) = L(x)$ |
| **NP** | language $L$ | nondeter-ministic $M$ | poly | no error $\forall x\ M(x) = L(x)$ |
| **BPP** | language $L$ | randomized $M$ | poly | bounded error $\forall x\, \Pr[M(x) = L(x)] \geq \frac{3}{4}$ |

# Time hierarchy

- A time hierarchy theorem states that a given computational model can decide more languages if it is allowed to use more time.
- (Hartmanis and Stearns, 1965) $\mathbf{DTime}[n^a] \subsetneq \mathbf{DTime}[n^{a+\epsilon}]$.
- (Cook, 1972) $\mathbf{NTime}[n^a] \subsetneq \mathbf{NTime}[n^{a+\epsilon}]$.
- (Karpinski and Verbeek, 1987) $\mathbf{BPTime}[n^{\log n}] \subsetneq \mathbf{BPTime}[2^{n^\epsilon}]$
- Main technique: diagonalization.
    - $M$ is $n^a$-time machine, $\exists x_M$ such that $M(x_M) \neq L(x_M)$.
    - To solve $L$ on $x_M$ in $n^{a+\epsilon}$-time: simulate $M$ and negate answer.

# Complete problems

$\mathbf{P} \overset{?}{\underset{?}{=}} \mathbf{NP}$
conjectured NO

$\mathbf{P} \overset{?}{\underset{?}{=}} \mathbf{BPP}$
conjectured YES

Turing reduction from $A$ to $B$: a polynomial-time algorithm $\mathcal{R}^B$ that solves $A$ with oracle access to $B$.
$B$ is a complete problem in the class $\mathbf{C}$ if $B \in \mathbf{C}$ and $\forall A \in \mathbf{C}$, $A$ reduces to $B$.

- (Cook, Levin, 1971) **NP**-complete problems: Bounded Halting, Tiling, SAT, TSP,...
- Bounded Halting
  - $BH = \{(M, x, 1^t) | \text{NTM } M \text{ accepts } x \text{ for } \le t \text{ steps}\}$
  - $L \in \mathbf{NP}$ is solved by NTM $M$ in p(n)-time;
  - Reduction: oracle request $(M, x, 1^{p(n)})$.
- Complete problem for **BPP** is not known.

# Structure in **BPP**

- Time hierarchies and complete problems usually require enumeration of (correct) machines in the respective computational model.
- How to enumerate machines that have bounded error?

Known facts:

1. (folklore) **BPP**-complete language $\implies$ time hierarchy for **BPP**;
2. (Hartmanis and Hemachandra, 1986) $\exists$ oracle $A$, such that $\mathbf{BPP}^A$ doesn't have complete languages.
3. (Barak, Fortnow, Santhanam, Trevisan, van Melkebeek, Pervyshev) Time hierarchy for **BPP** with one bit of nonuniform advice
4. (Fortnow, Santhanam, 2004 , Pervyshev 2007) Time hierarchy for heuristic **BPP**.

# Distributional problems

- Distribution $D = \{D_n\}_{n=1}^{\infty}$ where $D_n : \{0,1\}^n \to \mathbb{R}_+$ such that $\sum_{a \in \{0,1\}^n} D_n(a) = 1$.
- Distributional problem $(L, D)$, where $L$ is a language, $D$ is a distribution.
- Polynomial-time samplable distribution $\exists$ polynomial time algorithm (sampler) $S$ such that $S(1^n)$ is distributed according $D_n$.

# Average-case tractability

Levin (1986):

$T(x)$ is working time on input $x$;

$T(x)$ is polynomial on the average if

$\exists \epsilon > 0 : \mathrm{E}_{x \leftarrow D_n} T^{\epsilon}(x) = O(n)$

Typical situation:

- $\frac{1}{\mathrm{exp}}$: exponential time
- $1 - \frac{1}{\mathrm{exp}}$: polynomial time

# AvgP, AvgBPP

| Class | Problem | Turing machine | Time | Error |
|---|---|---|---|---|
| **P** | language $L$ | deterministic $M$ | poly | no error $\forall x\ M(x) = L(x)$ |
| **BPP** | language $L$ | randomized $M$ | poly | bounded error $\forall x \Pr[M(x) = L(x)] \geq \frac{3}{4}$ |
| **AvgP** | distr. problem $(L, D)$ | deterministic $M$ | avg. poly | no error $\forall x\ M(x) = L(x)$ |
| **Avg-BPP** | distr. problem $(L, D)$ | randomized $M$ | avg. poly | bounded error $\forall x \Pr[M(x) = L(x)] \geq \frac{3}{4}$ |

## **AvgBPP** and cryptography

- If $(\mathbf{NP}, U) \in \mathbf{AvgBPP}$, then there are no one-way functions.
- (Hirsch, Itsykson, 2007) If there exists $f$, such that problem $(f^{-1}, f(U)) \notin \mathbf{FAvgBPP}$ then there exists i.o. one-way function.
- Informally **AvgBPP** is the class of problems solved by succsessful cryptographical adversary.

# Results

1. Construction of distributional problem $(C, R)$ that is complete in $(\mathbf{AvgBPP}, \mathbf{PSamp})$ under deterministic Turing reduction.
   - If $(C, R) \in \mathbf{AvgP}$, then $(\mathbf{AvgP}, \mathbf{PSamp}) = (\mathbf{AvgBPP}, \mathbf{PSamp})$
   - $R$ is enough complicated samplable distribution.
   - Existence of complete problem with uniform (or uniform-like) distribution implies some derandomization.
2. Time hierarchy theorem for $(\mathbf{AvgBPP}, \mathbf{PSamp})$.
3. Proper inclusions:
   - $\mathbf{P} \subsetneq \mathbf{AvgP} \subsetneq \mathbf{EXP}$;
   - $\mathbf{BPP} \subsetneq \mathbf{AvgBPP} \subsetneq \mathbf{BPEXP}$.

# Intuition: complete problem $(C, R)$

Sampler $\mathcal{R}(1^n)$:

1. $|(M, y, r, S, l)| = n$ is generated at random;
2. $x \leftarrow S^{\leq |l|}(1^{|y|})$;
3. $n^2$ times execute $M^{\leq |r|}(x)$: $p$ answers 1, $q$ answers 0;
4. If $\frac{\max\{p,q\}}{p+q} \geq 0.9$ return, $(M, x, 1^{|r|}, S, 1^{|l|})$;
5. Else return $0^n$.

Algorithm $\mathcal{C}(M, x, 1^t, S, 1^s)$:

1. $n^2$ times execute $M^{\leq t}(x)$: $p$ answers 1, $q$ answers 0;
2. If $\frac{\max\{p,q\}}{p+q} \geq 0.85$, return the most frequent answer;
3. Else execute $M(x)$ with all random sequences and return the most frequent answer.

## Reduction

$(L, D) \in \mathbf{AvgBPP}$, $M$ solves $(L, D)$ in average time $p(n)$. $D$ is generated by sampler $S$ in time $s(n)$.

Oracle request: $(M, x, 1^{p(n)}, S, 1^{s(n)})$.