CS 830: Randomness and Complexity 2/24/05

Lecture 12: Deterministic Amplification

instructor: Jin-Yi Cai Scribe: Rakesh Kumar, Yunpeng Li

Last time we introduce the randomized algorithm for the st-connectivity problem by [AKL$^+$79]. It is known that the directed version of the problem, STCONN, is NL-complete.

We will focus on the undirected version, USTCONN, i.e., given a undirected graph $G = (V, E)$ and two distinct nodes $s, t \in V$, whether s is connected to t in G. As defined in [LP82], this problem is SL-complete. Recently, Reingold [Rei04] solves the long standing open problem $SL = ?L$ by gives a deterministic logarithmic space algorithm for USTCONN.

Before Reingold's algorithm, the research on USTCONN focus on the derandomization of the random walk algorithm by [AKL$^+$79]. Building on a sequence of work [Nis92] [Nis94] [NSW92] [SZ99] [CCvM04], the best known derandomization result is $SL \subseteq L^{4/3}$ by [ATSWZ00]. We will explore the derandomization method in this course. We begin this from deterministic amplification.

1 Preliminaries

We consider the randomized computation. The model we use is the randomized Turing machine M, which is defined as an ordinary Turing machine together with a special tape called random tape. During the computation, machine can access the cells of the tape one by one and it is restricted that each cell can be only read once. We can define the complexity class, RP, by

Definition 1.1 (RP). *If $L \in RP$, there exist a polynomial time bounded randomized Turing machine M such that*

$$x \in L \quad \Leftrightarrow \quad \Pr_r[M(x,r) accepts] \geq \frac{1}{2}$$
$$x \notin L \quad \Leftrightarrow \quad \Pr_r[M(x,r) accepts] = 0$$

where r is the bits read from the random tape of the machine called random input.

Let r be the n-bits string. We define the witness set W_x as a subset of $\{0,1\}^n$ such that

$$X_x = \{r \mid M(x,r) accepts\}$$

Let $L \in RP$, $\mu(W_x)$, the measure of W_x, is greater than 1/2 if $x \in L$, $\mu(W_x) = 0$ otherwise.

1.1 Naive Amplification

In the definition of RP, the error will happen only when $x \in L$ and the error probability is bounded from above by 1/2. We would like diminishing the error probability deterministically and we call this process *deterministic amplification*.

The naive way to do the amplification if simply repeat the computation of M using independent random inputs r's. Let k be the number of trials, the error probability is reduced to

$$\Pr_{r_1,\cdots,r_k \in \{0,1\}^n}[r_1, \cdots, r_k \in \overline{W_x}] \leq \frac{1}{2^k}$$

Note that we achieve exponentially small error probability using $k \cdot n$ number of random bits. Can we do this more efficiently, i.e., using less random bits and reach the exponentially small error probability?

2 Chor-Goldreich Generator

Chor and Goldreich [CG85] use the property of pairwise independence give the the following easy efficient amplification. Let $\mathbf{H}$ be a family of universal hash function mapping from $\{0,1\}^n$ to $\{0,1\}^n$ with index set S. Note that universal hash function can be explicitly constructed as

$$h_{a,b}(x) = ax + b \qquad \text{for } x, a, b \in GF[2^n]$$

i.e., each hash function can be specified by $2n$ bits. We define $r_i = h_{a,b}(i)$ for $1 \leq i \leq k < 2^n$. And run the randomized machine using the r_i's. If any of the k trials accept,otherwise we reject. For our RP machine, if $x \notin L$, $M(x, r_i$ will always reject for all i's. Suppose $x \in L$, we estimate the error probability.

Theorem 2.1. *If $x \in L$, then* $\Pr[$*C-G Algorithm reject*$] \leq 1/k$

Proof. Let Z_i be the indicator random variable variable defined as

$$Z_i = \begin{cases} 0, & \text{if } M(x, r_i) \text{ accepts;} \\ 1, & \text{if } M(x, r_i) \text{ rejects.} \end{cases}$$

And let $Z = \sum_{i=1}^{k} Z_i$. Note that the expectation $E[Z_i] = \mu W_x$ and variance $\sigma^2(Z_i) \leq 1/4$. By the linearity of expectation,

$$E[Z] = E[\sum_{i=1}^{k} Z_i] = \sum_{i=1}^{k} E[Z_i] = k\mu(W_x)$$

And since those r_i's are pairwise independent,

$$\sigma^2(Z) = \sigma^2(\sum_{i=1}^{k} Z_i] = \sum_{i=1}^{k} \sigma^2(Z_i) \leq k/4$$

Therefore

$$\begin{aligned} & \Pr[\text{C-G Algorithm reject}] \\ = \;& \Pr[r_1, \cdots, r_k \in \overline{W_x}] \\ = \;& \Pr[Z = 0] \\ \leq \;& \Pr[|Z - E[Z]| \geq k/2] \leq \frac{4\sigma^2(Z)}{k^2} \leq 1/k \end{aligned}$$

where the last inequality is the Chebyshev inequality.

□

Note that this amplification does not give us exponentially small error probability, but we do reduce the random bits greatly.

3 Hash Mixing Lemma and Nisan's Generator

The following generator is originally used for space bounded computation by Nisan [Nis92]. It turns out it is also useful for deterministic amplification.

Let $\mathbf{H}$ be a family of universal hash functions mapping from $\{0,1\}^n$ to $\{0,1\}^n$. We consider the following recursive construction

$$\begin{aligned}
G_0(y) &= y \\
G_1(y, s_1) &= y \cdot h_{s_1}(y) \\
G_2(y, s_1, s_2) &= y \cdot h_{s_1}(y) \cdot h_{s_2}(y) \cdot h_{s_2}(h_{s_1}(y)) \\
&\vdots \\
G_{i+1}(y) &= G_i(y, s_1, \cdots, s_i) \cdot G_i(h_{s_{i+1}}(y), s_1, \cdots, s_i)
\end{aligned}$$

This generator exponentially expands its input. For G_k, it stretches $(2k+1)n$ to $2^k n$ bit. And hash mixing lemma tells us that those bits are fairly good for amplification.

Lemma 3.1 (Hash Mixing Lemma). *Let $\epsilon = 2^{-n/3}$ and for all subset $E \subseteq (\{0,1\}^n)^2$. for all but a fraction of $\epsilon/4$ many s*

$$|\Pr_y[y \cdot h_s(y) \in E] - \mu(E)| < \epsilon$$

where $\mu(E) = \Pr_{y,z}[(y, z) \in E]$.

Proof. We want to estimate the probability over s such that

$$|\Pr_y[y \cdot h_s(y) \in E] - \mu(E)| \geq \epsilon \tag{1}$$

The key idea of this proof is to express $|\Pr_y[y \cdot h_s(y) \in E]$ as a sum of pairwise independent random variables with expectation $\mu(E)$, then apply the Chebyshev inequality. For any fixed y, we define the indicator variable $Z_y^{h_s}$ as

$$Z_y^{h_s} = \begin{cases} 1, & \text{if } y \cdot h_s(y) \in E; \\ 0, & \text{otherwise.} \end{cases}$$

with $E[Z_y^{h_2}] = \mu(E)$ and $sigma^2(Z_y^{h_2}) \leq 1/4$. Hence we can write

$$Z = \Pr_y[y \cdot h_s(y) \in E] = \frac{1}{2^n} \sum_y Z_y^{h_2}$$

by the linearity of expectation, we have $E[Z] = \mu(E)$. Therefore the probability of Equation (1) happens is bounded from above by

$$\frac{\sigma^2(R)}{\epsilon^2} = \frac{1}{4 \cdot 2^{2n}} 2^n = \frac{1}{\epsilon^2 2^{n+2}} < \epsilon/4$$

□

Using the hashing mixing lemma we will prove nexttime that

Theorem 3.2. *If $x \in L$, then*

$$G_k(y, s_1, \cdot, s_k) \subseteq \overline{W_x}] \leq (\mu(\overline{W_x}))^{2^k} + (k/4 + 2)\epsilon$$

References

[AKL+79] R. Aleliunas, R. Karp, R. Lipton, L. Lovász, and C. Rackoff. Random walks, uniersal traversal sequencec and the complexity of maze problems. In *Proceddings of the 20th FOCS*, pages 218–223, 1979.

[ATSWZ00] R. Armoni, A. Ta-Shma, A. Wigderson, and S. Zhou. An $o(\log^{4.3} n)$ space algorithm for (s,t)-connectivity in undirected graphs. *Journal of the ACM*, 47(2):294–311, 2000.

[CCvM04] J. Cai, V. Chakaravarthy, and D. van Melkebeek. Time-space tradeoff in derandomizing probabilistic logspace. *21st Annual Symposium on Theoretical Aspects of Computer Science*, pages 571–583, 2004.

[CG85] B. Chor and O. Goeldreich. On the power of two-points based sampling. *Technical Report, MIT*, 1985.

[LP82] H. Lewis and C. Papadimitriou. Symmetric space-bounded computation. *Theoretical Computer Science*, 19:161–187, 1982.

[Nis92] N. Nisan. Pseudorandom generators for space-bounded computation. *Combinatorica*, 12(4):449–461, 1992.

[Nis94] N. Nisan. $\mathbf{RL} \subseteq \mathbf{SC}$. *Computational Complexity*, 4:1–11, 1994.

[NSW92] N. Nisan, E. Szemerédi, and A. Wigderson. Undirected connectivity in $\mathbf{O}(\log^{1.5} n)$ space. *In proceedings of the 33rd IEEE Symposium on Foundations of Computer Science*, pages 24–29, 1992.

[Rei04] Omer Reingold. Undirected st-connectivity in log-space. *ECCC*, 11 2004.

[SZ99] M. Saks and S. Zhou. $\mathbf{BP_H SPACE(S)} \subseteq \mathbf{DSPACE(S^{3/2})}$. *Journal of Computer and System Sciences*, 58:376–403, 1999.