Derandomization (III)

Yijia Chen

Shanghai Jiao Tong University

November 26, 2012

The Construction of Nisan and Wigderson

For a string $z \in \{0,1\}^{\ell}$ and a subset $I \subseteq [\ell]$, we define z_I to be $|I|$-length string that is the projection of z to the coordinates in I.

Definition

Let $\mathcal{I} = \{I_1, \ldots, I_m\}$ be a family of subsets of $[\ell]$ with $|I_j| = n$ for each $j \in [m]$, and let $f : \{0,1\}^n \to \{0,1\}$. Then the $(\mathcal{I}, f)$-*NW generator* is the function $\mathrm{NW}_{\mathcal{I}}^{f} : \{0,1\}^{\ell} \to \{0,1\}^m$ with

$$\mathrm{NW}_{\mathcal{I}}^{f}(z) = f(z_{I_1}) \circ f(z_{I_2}) \circ \cdots \circ f(z_{I_m}).$$

Combinatorial design

Definition

Let $\ell > n > d$. A family $I = \{I_1, \ldots, I_m\}$ of subsets of $[\ell]$ is an (ℓ, n, d)-*design* if $|I_j| = n$ for every $j \in [m]$ and $|I_j \cap I_k| \le d$ for every $1 \le j < k \le m$.

Lemma

There is an algorithm $\mathbb{A}$ such that on input $\ell, d, n \in \mathbb{N}$ with $n > d$ and $\ell > 10 \cdot n^2/d$, runs for $2^{O(\ell)}$ steps and outputs an (ℓ, n, d)-design I containing $\lceil 2^{d/10} \rceil$ subsets of $[\ell]$.

We will choose ℓ, n, d, m in such a way that

$$\ell, d = \Theta(n) \quad \text{and} \quad m = 2^{\Theta(n)}.$$

Proof (1)

We consider the following greedy algorithm $\mathbb{A}$:

`Start with` $I = \emptyset$`, and after constructing` $I = \{I_1, \ldots, I_m\}$ `for` $m < 2^{d/10}$`, search all subsets of` $[\ell]$ `and add to` I `the first` n`-sized set` J `satisfying` $|J \cap I_j| \leq d$ `for every` $j \in [m]$`.`

Clearly, the running time of $\mathbb{A}$ is bounded by $2^{O(\ell)}$.

Proof (2)

Claim: $\mathbb{A}$ will not stop until $m \geq 2^{d/10}$, i.e., when $m < 2^{d/10}$, there exists a J such that $|J \cap I_j| \leq d$ for every $j \in [m]$.

We choose J at random by choosing independently every element $x \in [\ell]$ to be in J with probability $2 \cdot n/\ell$.

Then

$$E[|J|] = 2 \cdot n \quad \text{and} \quad E\left[|J \cap I_j|\right] = 2 \cdot n^2/\ell < d/5.$$

which implies,

$$\Pr\left[|J| \geq n\right] \geq 0.9 \quad \text{and} \quad \Pr\left[|J \cap I_j| \geq d\right] \leq 0.5 \cdot 2^{-d/10} \text{ for every } j \in [m]$$

by the Chernoff bound.

Therefore,

$$\Pr\left[|J| \geq n \text{ and } |J \cap I_j| \leq d \text{ for every } j \in [m]\right] \geq 0.4.$$

□

Pseudorandomness using the NW generator

Lemma

If I is an (ℓ, n, d)-design with $|I| = 2^{d/10}$ and $f : \{0,1\}^n \to \{0,1\}$ a function satisfying $H_{\mathrm{avg}}(f)(n) \geq 2^{2 \cdot d}$, then the distribution $NW_I^f(U_\ell)$ is $\big(H_{\mathrm{avg}}(f)(n)/4, 1/10\big)$-pseudorandom.

Proof (1)

Let $S := \mathsf{H}_{\mathrm{avg}}(f)(n)$. By Yao's Theorem, we need to prove that for every $i \in \left[2^{d/10}\right]$ there does not exist an $S/2$-sized circuit C such that

$$\Pr_{\substack{Z \sim U_\ell, \\ R = \mathsf{NW}^f_I(Z)}} \left[C(R_1, \ldots, R_{i-1}) = R_i\right] \geq \frac{1}{2} + \frac{1}{10 \cdot 2^{d/10}}.$$

Assume that for some C and i,

$$\Pr_{Z \sim U_\ell} \left[C(f(z_{I_1}) \circ \cdots \circ f(z_{I_{i-1}})) = f(z_{I_i})\right] \geq \frac{1}{2} + \frac{1}{10 \cdot 2^{d/10}}.$$

Proof (2)

Let Z_1 and Z_2 be two independent variables corresponding to the coordinates of Z in I_i and $[\ell] \setminus I_i$, respectively.

Then

$$\Pr_{\substack{Z_1 \sim U_n \\ Z_2 \sim U_{\ell-n}}} \left[C(f_1(Z_1, Z_2) \circ \cdots \circ f_{i-1}(Z_1, Z_2)) = f(Z_1) \right] \geq \frac{1}{2} + \frac{1}{10 \cdot 2^{d/10}}.$$

where for every $j \in \left[2^{d/10}\right]$, the function f_j applies f to the coordinates of Z_1 corresponding to $I_j \cap I_i$ and the coordinates of Z_2 corresponding to $I_j \setminus I_i$.

By the averaging principle, then there exists a string $z_2 \in \{0,1\}^{\ell-n}$ such that

$$\Pr_{Z_1 \sim U_n} \left[C(f_1(Z_1, z_2) \circ \cdots \circ f_{i-1}(Z_1, z_2)) = f(Z_1) \right] \geq \frac{1}{2} + \frac{1}{10 \cdot 2^{d/10}}.$$

Proof (3)

Since $|I_j \cap I_i| \leq d$ for $j \neq i$, the function $Z_1 \mapsto f_j(Z_1, z_2)$ depends at most d coordinates of Z_1 and hence can be computed by a $d \cdot 2^d$-sized circuit.

Thus for a circuit B of size $2^{d/10} \cdot d \cdot 2^d + S/2 \leq S$ we have

$$\Pr_{Z_1 \sim U_n}\left[B(Z_1) = f(Z_1)\right] \geq \frac{1}{2} + \frac{1}{10 \cdot 2^{d/10}} \geq \frac{1}{2} + \frac{1}{S}.$$

It contradicts $\mathrm{H}_{\mathrm{avg}}(f)(n) = S$. □

A slightly weaker version of NW's theorem

Theorem

For every time-constructible and nondecreasing function $S : \mathbb{N} \to \mathbb{N}$, if there exists a function $f \in \mathbf{DTIME}(2^{O(n)})$ such that $H_{\text{avg}}(f) \geq S$, then we can construct an $S'(\ell)$-pseudorandom generator, where $S'(\ell) = S(n)^{\varepsilon}$ for some $\varepsilon > 0$ and $n \in \mathbb{N}$ satisfying $n \geq \varepsilon \cdot \sqrt{\ell \cdot \log S(n)}$.

Recall we need a $2^{\lfloor \ell/a \rfloor}$-pseudorandom generator to show $\mathbf{BPP} = \mathbf{P}$. In order to achieve that, we choose $2^{\lfloor \ell/a \rfloor} = S'(\ell) = S(n)^{\varepsilon}$ with $n \geq \varepsilon \cdot \sqrt{\ell \cdot \log S(n)}$. Therefore,

$$(n/\varepsilon)^2/\ell \geq \ell/(\varepsilon \cdot a) \qquad \text{i.e., } \ell \leq n \cdot \sqrt{a/\varepsilon}.$$

So we can take

$$S(n) = 2^{\lceil n/\sqrt{\varepsilon \cdot a} \rceil/\varepsilon}.$$

Proof (1)

On input $z \in \{0,1\}^\ell$ the generator operates as follows:

1. Set n to be the largest number such that $\ell > 100 \cdot n^2 / \log S(n)$.
2. Set $d = \log S(n)/2$.
3. Compute an (ℓ, n, d)-design $I = \left\{ I_1, \ldots, I_{2^{d/10}} \right\}$.
4. Output the first $S(n)^{1/40}$ bits of $\mathrm{NW}_I^f(z)$.

By Line 1, we conclude

$$\ell \leq \frac{100 \cdot (n+1)^2}{\log S(n+1)} \leq \frac{200 \cdot n^2}{\log S(n)},$$

and hence $n \geq \sqrt{\ell \cdot \log S(n)/200}$.

Proof (2)

The generator makes $2^{d/10}$ invocations of f, taking a total of $2^{O(n+d)}$ steps. By possibly reducing n by a constant factor, we can ensure the running time is bounded by 2^{ℓ}.

Moreover, since $d = \log S(n)/2$, we conclude

$$2^{2 \cdot d} \geq S(n).$$

So by Nisan and Wigderson's Lemma, the distribution $\mathrm{NW}^f(U_\ell)$ is $\big(S(n)/4.1/10\big)$-pseudorandom. □

Thank you