0368-4159: First Course in Derandomization 24/12/18 – Lecture 11

The Forbes-Kelley PRG for unordered BP

Lecturer: Amnon Ta-Shma

Scribe: ??

In this lecture we present a recent PRG against small space probabilistic machines (and small width branching programs).

1 Branching programs

1.1 A non-uniform model of small space probabilistic machines

Definition 1. *(*$\mathsf{BPSPACE}(s(n)), \mathsf{BPTimeSPACE}(t(n), s(n))$*)*

Definition 2. *(PRG against* $\mathsf{BPTimeSPACE}(t(n), s(n))$*)*
We say $G : \{0,1\}^{\ell(n)} \to \{0,1\}^{t(n)}$ *is an* $\varepsilon(n)$ *pseudorandom generator (PRG) against* $\mathsf{BPTimeSPACE}(t(n), s(n))$ *if for every machine* $M \in \mathsf{BPTimeSPACE}(t(n), s(n))$ *and every input* $x \in \{0,1\}^n$*:*

$$|\Pr_{y \in \{0,1\}^{t(n)}} M(x,y) = 1 - \Pr_{z \in \{0,1\}^{\ell(n)}} M(x, G(z)) = 1| \leq \varepsilon(n).$$

We want to replace a random string from $\{0,1\}^{t(n)}$ with a pseudo-random string. Thus, we may view x as hardcoded into the machine, giving the function $M(x, \cdot)$. Thus, x may be seen as a non-uniform advice to the machine acting on the input y. Thus, on input $t(n)$ the advice length is n. To simplify things, we will construct the stronger object that fools machine of small space no matter how long the advice is. For that we define branching programs which are the non-uniform analogue of small space uniform machines.

Definition 3. *(Branching programs)*

Branching programs which are the non-uniform analogue of small space uniform machines, the same way circuits are the non-uniform analogue of time-bounded machines.

1.2 A matrix representation of a BP

Let B be a (w, n) BP. We can represent the action of B by the transition matrix, mapping state a at the initial layer to vertex b at the final layer. This gives a $w \times w$ stochastic matrix. We can similarly represent the transition matrix between any two layers $i < j$ by a $w \times w$ stochastic matrix. Furthermore, we can also define the transition matrix for a given input $y = (y_1, \ldots, y_{t(n)})$. We define:

Definition 4. *Let* $0 \leq i \leq t(n)$*,* $b \in \{0,1\}$*.* $M_{i,b}$ *is the* $w \times w$ *transition matrix from layer* i *to layer* $i+1$ *when the input bit is* b*.*

Notice that $M_{i,b}$ is a deterministic transition matrix. Also,

$$M_{[i,j]}(b_i, \dots, b_j) = M_{j,b_j} \cdot \dots \cdot M_{i,b_i}$$

is the deterministic transition matrix from level i to j given the input $b_i, \dots, b_j$. The transition matrix under the uniform distribution from the first layer to the last is $E_{b \in \{0,1\}^t} M_{[0,t]}(b)$, and the transition matrix under the PRG from the first layer to the last is $E_{z \in \{0,1\}^\ell} M_{[0,t]}(G(z))$. In particular,

$$\begin{aligned} |\Pr_{y \in \{0,1\}^t} B(y) = 1 - \Pr_{z \in \{0,1\}^\ell} M(x, G(y)) = 1| &= E_{b \in \{0,1\}^t} M_{[0,t]}(b)[1,1] - E_{z \in \{0,1\}^\ell} M_{[0,t]}(G(z))[1,1] \\ &\le \left\| E_{b \in \{0,1\}^t} M_{[0,t]}(b) - E_{z \in \{0,1\}^\ell} M_{[0,t]}(G(z)) \right\|_F . \end{aligned}$$

A paragraph about Frobenius norm, $\|A\|_F = \sqrt{\operatorname{Tr}(A^\dagger A)}$.

1.3 The Fourier representation of a BP

Given a function $F : \{0,1\}^t \to M_{w \times w}\mathbb{C}$. We may view it as w^2 boolean functions, and apply Fourier transform on each. For the entry (i,j),

$$F_{i,j} : \{0,1\}^t \to \{0,1\}$$

and

$$F_{i,j} = \sum_{S \subseteq [t]} \widehat{F_{i,j}}_S \chi_S$$

where

$$\widehat{F_{i,j}}_S = \langle \chi_S, F_{i,j} \rangle = \mathop{\mathbb{E}}_{b \in \{0,1\}^t} F_{i,j}(b) \chi_S(b).$$

Let us define $\widehat{F}_S$ the $w \times w$ matrix where

$$\widehat{F}_S[i,j] = \widehat{F_{i,j}}_S.$$

Then, this gives us:

$$\begin{aligned} \widehat{F}_S &= \mathop{\mathbb{E}}_{b \in \{0,1\}^t} F(b) \chi_S(b), \text{ and,} \\ F &= \sum_{S \subseteq [t]} \widehat{F}_S \cdot \chi_S. \end{aligned}$$

Claim 5. $\mathbb{E}_{b \in \{0,1\}^t} \|F(b)\|_F^2 = \sum_{S \subseteq [t]} \left\| \widehat{F}_S \right\|_F^2$.

Proof.

$$\begin{aligned}
\mathop{\mathbb{E}}_{b\in\{0,1\}^t} \|F(b)\|_F^2 &= \mathop{\mathbb{E}}_{b\in\{0,1\}^t} \mathrm{Tr}(F(b)F(b)^\dagger) = \mathop{\mathbb{E}}_{b\in\{0,1\}^t} \mathrm{Tr}(\sum_{S_1,S_2\subseteq[t]} \overline{\chi_{S_1}(b)}\chi_{S_2(b)}\widehat{F}_{S_1}^\dagger \widehat{F}_{S_2}) \\
&= \sum_{S_1,S_2\subseteq[t]} \mathop{\mathbb{E}}_{b\in\{0,1\}^t} \overline{\chi_{S_1}(b)}\chi_{S_2(b)}\mathrm{Tr}(\widehat{F}_{S_1}^\dagger \widehat{F_{S_2}}) \\
&= \sum_{S\subseteq[t]} \mathrm{Tr}(\widehat{F}_S^\dagger \widehat{F_S}) = \sum_{S\subseteq[t]} \left\|\widehat{F}_S\right\|_F^2 .
\end{aligned}$$

□

Lemma 6. *If $M : \{0,1\}^t \to M_{w\times w}[0,1]$ is computed by a branching program, then:*

- $\sum_{S\subseteq[t]} \left\|\widehat{F}_S\right\|_F^2 = w.$
- $\sum_{S\subseteq[t]:|S|=k} \left\|\widehat{F}_S\right\|_F \le \sqrt{\binom{n}{k}w}.$

Proof. For the first item, we know $\sum_{S\subseteq[t]} \left\|\widehat{F}_S\right\|_F^2 = \mathbb{E}_b \|F(b)\|_F^2$. However, for every b, $F(b)$ is a $w\times w$ deterministic stochastic matrix, hence $\|F(b)\|_F^2 = \sum_{i,j} F(b)[i,j]^2 = w$.

For the second item, by Cauchy-Schwarz, $\sum_{S\subseteq[t]:|S|=k} \left\|\widehat{F}_S\right\|_F \le \sqrt{\binom{n}{k}}\sqrt{\sum_{S\subseteq[t],|S|=k} \left\|\widehat{F}_S\right\|_F^2}$. □

2 The generator

We define distributions over $\{0,1\}^t$. First we sample independently:

- $2k$-wise independent distributions $D_0, D_1, \ldots, D_r$ over $\{0,1\}^t$,
- k-wise independent distributions $T_1, \ldots, T_r$ over $\{0,1\}^t$.

We define PRG $G_0, \ldots, G_r$ with increasing accuracy (and growing seed length). We let $G_0 = D_0$. We let

$$G_{i+1} = T_{i+1} \wedge D_{i+1} + \overline{T_{i+1}} \wedge G_i,$$

where $\wedge$ is coordinate-wise and, $+$ is coordinate-wise XOR, and $\overline{S} = [t]\setminus S$.

Opening the recursion we get, e.g.,

$$\begin{aligned}
G_2 &= T_2 \wedge D_2 + \overline{T_2} \wedge (T_1 \wedge D_1 + \overline{T_1} \wedge G_0) = T_2 \wedge D_2 + \overline{T_2} \wedge T_1 \wedge D_1 + \overline{T_2} \wedge \overline{T_1} \wedge D_0 \\
G_3 &= T_3 \wedge D_3 + \overline{T_3} \wedge T_2 \wedge D_2 + \overline{T_3} \wedge \overline{T_2} \wedge T_1 \wedge D_1 + \overline{T_3} \wedge \overline{T_2} \wedge \overline{T_1} \wedge D_0
\end{aligned}$$

The main lemma analyses one step bin the recursion:

Lemma 7. *Suppose $F : \{0,1\}^t \to M_{w\times w}[0,1]$ encodes a branching program, D is $2k$-wise independent over $\{0,1\}^t$, and T is k-wise independent over $\{0,1\}^t$, then*

$$\left\| \mathop{\mathbb{E}}_{b\in\{0,1\}^t} F(b) - \mathop{\mathbb{E}}_{d\in D, t\in T, b\in\{0,1\}^t} F(t\wedge d + \bar{t}\wedge b) \right\|_F \le tw2^{-k/2}.$$

We will prove it in the next section. With that,

Theorem 8. *Suppose $r \ge 2\log t$. If $F : \{0,1\}^t \to M_{w\times w}[0,1]$ encodes a branching program, then*

$$\left\|\mathop{\mathbb{E}}_{b\in\{0,1\}^t} F(b) - \mathop{\mathbb{E}}_{z\in\{0,1\}^\ell} F(G_r(z))\right\|_F \le rtw2^{-k/2} + 2^{-\Omega(r)}.$$

Proof.

$$\begin{aligned} & \|\mathbb{E}\, F(U) - \mathbb{E}\, F(G_{r+1}(U))\|_F \\ = \quad & \left\|\mathbb{E}\, F(U) - \mathop{\mathbb{E}}_{D_r,T_r,U} F(T_r \wedge D_r + \overline{T_r} \wedge G_r(U))\right\|_F \\ \le \quad & \left\|\mathbb{E}\, F(U) - \mathbb{E}\, F(T_r \wedge D_r + \overline{T_r} \wedge U)\right\|_F + \left\|\mathbb{E}\, F(T_r \wedge D_r + \overline{T_r} \wedge G_r(U)) - \mathbb{E}\, F(T_r \wedge D_r + \overline{T_r} \wedge U\right\|_F \end{aligned}$$

The first term is bounded by Lemma 7. For the second term:

$$\begin{aligned} & \left\|\mathop{\mathbb{E}}_{D_r,T_r,U} F(T_r \wedge D_r + \overline{T_r} \wedge G_r(U)) - \mathop{\mathbb{E}}_{D_r,T_r,U} F(T_r \wedge D_r + \overline{T_r} \wedge U\right\|_F \\ \le \quad & \mathop{\mathbb{E}}_{T_r,D_r} \left\|\mathop{\mathbb{E}}_{U} F(T_r \wedge D_r + \overline{T_r} \wedge G_r(U)) - \mathbb{E}\, F(T_r \wedge D_r + \overline{T_r} \wedge U\right\|_F \end{aligned}$$

Continuing like this we are left with the term

$$\mathop{\mathbb{E}}_{T_r,D_r,\ldots,T_1,D_1} \left\|\mathop{\mathbb{E}}_{U} F_*(\overline{T_r} \wedge \ldots \overline{T_1} \wedge U) - F_*(\overline{T_r} \wedge \ldots \overline{T_1} \wedge D_0)\right\|_F ,$$

where F_* is the restricted function. However, except for probability $2^{-\Omega(r)}$, the number of bits kept alive in F_* is at most k, in which case the difference contributes zero. □

In particular for error ε it is enough to take $r = O(\log \frac{t}{\varepsilon})$ and $k = O(\log \frac{tw}{\varepsilon})$ and get a PRG of length $O(rk\log t) = O(\log^3 \frac{tw}{\varepsilon})$. For $w = poly(t)$ and $\varepsilon = t^{-\Theta(1)}$ we get seed length $O(\log^3 t)$.

We remark that the generator is symmetric (i.e., for any permutation $\pi \in S_t$ we get the same distribution, hence the PRG works even if the BP chooses the order of out bits fed to it.

3 Analysing the recursion

Lemma 9. *Suppose $F : \{0,1\}^t \to M_{w\times}\mathbb{C}$ is a product function $F(b) = F_t(b_t) \cdot \ldots F_1(b_1)$. For $i \in [n]$ let*

$$\begin{aligned} F^{\le i}(b_1,\ldots,b_i) &= F_i(b_i) \cdot \ldots F_1(b_1), \\ F^{>i}(b_{i+1},\ldots,b_t) &= F_t(b_t) \cdot \ldots F_{i+1}(b_{i+1}). \end{aligned}$$

Let k be an integer. Then

$$F = \widehat{F}_\emptyset + \sum_{S\subseteq[t]:|S|<k} \widehat{F}_S \chi_S + \sum_{i=1}^{t} \sum_{\substack{A\subseteq[i]:\\ |A|=k, i\in A}} \widehat{F^{\le i}}_A \cdot \chi_A \cdot F^{>i}.$$

Proof. Before we start notice that because F is product, i.e., $F(b) = F_t(b_t) \cdot \ldots F_1(b_1)$, we have

$$\begin{aligned}\widehat{F}_S &= \mathop{\mathbb{E}}_{b} F(b)\chi_S(b) = \mathop{\mathbb{E}}_{b_1,\ldots,b_t} F_t(b_t) \cdot \ldots F_1(b_1)\chi_{S_1}(b_1)\ldots\chi_{S_t}(b_t) \\ &= \mathop{\mathbb{E}}_{b_t} \chi_{S_t}(b_t)F_t(b_t) \cdot \ldots \cdot \mathop{\mathbb{E}}_{b_1} \chi_{S_1}(b_1)F_t(b_1) = \widehat{F}_{ts_t} \cdot \ldots \widehat{F_1}_{s_1}.\end{aligned}$$

$F = \sum_S \widehat{F}_S \chi_S$. $\widehat{F}_\emptyset + \sum_{S\subseteq[t]:|S|<k}$ take care of the low cardinality set terms and we are only left with $\widehat{F}_S\chi_S$ where $|S| \geq k$. For each such set there is a first $i \in [t]$ where $|S \cap [i]| = k$ and we associate S to this i. We get:

$$\begin{aligned}F &= \widehat{F}_\emptyset + \sum_{S\subseteq[t]:1\leq|S|<k} \widehat{F}_S\chi_S + \sum_{\substack{S\subseteq[t]:\\|S|\geq k}} \widehat{F}_S \cdot \chi_S \\ &= \widehat{F}_\emptyset + \sum_{S\subseteq[t]:1\leq|S|<k} \widehat{F}_S\chi_S + \sum_{\substack{S\subseteq[t]:\\|S|\geq k}} \widehat{F}_{ts_t} \cdot \ldots \widehat{F_1}_{s_1} \cdot \chi_S \\ &= \widehat{F}_\emptyset + \sum_{S\subseteq[t]:1\leq|S|<k} \widehat{F}_S\chi_S + \sum_{i=1}^{t} \sum_{B\subseteq[i+1,t]} \widehat{F}_{tb_t} \cdot \ldots \widehat{F_{i+1}}_{b_{i+1}}\chi_B \sum_{\substack{A\subseteq[i]:\\|A|=k,i\in A}} \widehat{F}_{ia_i} \cdot \ldots \widehat{F_1}_{a_1} \cdot \chi_A \cdot \\ &= \widehat{F}_\emptyset + \sum_{S\subseteq[t]:1\leq|S|<k} \widehat{F}_S\chi_S + \sum_{i=1}^{t} \sum_{\substack{A\subseteq[i]:\\|A|=k,i\in A}} \widehat{F^{\leq i}}_A \cdot \chi_A \cdot \sum_{B\subseteq[i+1,t]} \widehat{F^{>i}}_B\chi_B \\ &= \widehat{F}_\emptyset + \sum_{S\subseteq[t]:1\leq|S|<k} \widehat{F}_S\chi_S + \sum_{i=1}^{t} \sum_{\substack{A\subseteq[i]:\\|A|=k,i\in A}} \widehat{F^{\leq i}}_A \cdot \chi_A \cdot F^{>i}.\end{aligned}$$

□

The idea now is the following. $\widehat{F}_\emptyset$ is the correct term. The low order terms are fooled by the k-wise independent distribution. We are left with the high order terms. For each $i \in [n]$ we have a term that looks like the high order terms of $F^{\leq i}$ product $F^{>i}$. This expression is hit hard by the uniform part (because the first element in the product has only high order terms) Formally,

Lemma 10. *Let* $H : \{0,1\}^t \to M_{w\times w}\mathbb{C}$ *be supported on cardinality* k *sets, i.e.,* $H = \sum_{A\subseteq[t]:|A|=k} \widehat{H}_A \cdot \chi_A$. *Let* D *be* $2k$*-wise independent distribution over* $\{0,1\}^t$, T *be* k*-wise independent distribution over* $\{0,1\}^t$ *and* U *the uniform distribution over* $\{0,1\}^t$. *Then,*

$$\begin{aligned}E_{D,T}\left\|\mathop{\mathbb{E}}_{U} H(T\wedge D + \overline{T}\wedge U)\right\|_F &\leq (E_{D,T}\left\|\mathop{\mathbb{E}}_{U} H(T\wedge D + \overline{T}\wedge U)\right\|_F^2)^{1/2} \\ E_{D,T}\left\|\mathop{\mathbb{E}}_{U} H(T\wedge D + \overline{T}\wedge U)\right\|_F^2 &\leq 2^{-k} \sum_{A\subseteq[t]:|A|=k} \left\|\widehat{H}_A\right\|_F^2\end{aligned}$$

Proof. The first inequality is because $\sum p_i|c_i| \leq \sqrt{\sum p_i \sum p_i |c_i|^2}$.

The second part is where we use the k-wise independence. We have:

$$\begin{aligned}
\mathop{\mathbb{E}}_{U} H(T \wedge D + \overline{T} \wedge U) &= \mathop{\mathbb{E}}_{U} \sum_{A \subseteq [t] : |A| = k} \widehat{H}_A \cdot \chi_A(T \wedge D + \overline{T} \wedge U) \\
&= \sum_{A \subseteq [t] : |A| = k} \widehat{H}_A \cdot \mathop{\mathbb{E}}_{U} \chi_A(T \wedge D + \overline{T} \wedge U) \\
&= \sum_{A \subseteq [t] : |A| = k} \widehat{H}_A \cdot \chi_{A \cap T}(D) \cdot \mathop{\mathbb{E}}_{U} \chi_{A \cap \overline{T}}(U) \\
&= \sum_{A \subseteq T : |A| = k} \widehat{H}_A \cdot \chi_{A \cap T}(D).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left\| \mathop{\mathbb{E}}_{U} H(T \wedge D + \overline{T} \wedge U) \right\|_F^2 &= \mathrm{Tr}[(\mathop{\mathbb{E}}_{U} H(T \wedge D + \overline{T} \wedge U)) \cdot (\mathop{\mathbb{E}}_{U} H(T \wedge D + \overline{T} \wedge U))^\dagger] \\
&= \sum_{A, B \subseteq T, |A| = |B| = k} \chi_{A \oplus B}(D) \cdot \mathrm{Tr}(\widehat{H}_A \widehat{H}_B^\dagger)
\end{aligned}$$

Now $|A \oplus B| \leq 2k$ and D is $2k$-wise independent. Hence all terms vanish except for $A \oplus B = \emptyset$, i.e., $A = B$. Hence,

$$\left\| \mathop{\mathbb{E}}_{U} H(T \wedge D + \overline{T} \wedge U) \right\|_F^2 = \sum_{A \subseteq T, |A| = k} \left\| \widehat{H}_A \right\|_F^2 .$$

Taking expectations over T we get:

$$\begin{aligned}
\mathop{\mathbb{E}}_{T} \left\| \mathop{\mathbb{E}}_{U} H(T \wedge D + \overline{T} \wedge U) \right\|_F^2 &= \mathop{\mathbb{E}}_{T} \sum_{A \subseteq T, |A| = k} \left\| \widehat{H}_A \right\|_F^2 \\
&= \sum_{A : |A| = k} \left\| \widehat{H}_A \right\|_F^2 \Pr_{T}[A \subseteq T] \\
&= 2^{-k} \sum_{A : |A| = k} \left\| \widehat{H}_A \right\|_F^2 .
\end{aligned}$$

□

We now complete the proof:

Proof. (Of Lemma 7)

$$\begin{aligned}
&\left\| \mathop{\mathbb{E}}_{D,T,U} F(T \wedge D + \overline{T} \wedge U) - \mathbb{E}\, F(U) \right\|_F \\
\leq\ & \left\| \mathop{\mathbb{E}}_{D,T,U} \sum_{\substack{S \subseteq [t]: \\ 1 \leq |S| < k}} \widehat{F}_S \chi_S (T \wedge D + \overline{T} \wedge U) \right\|_F \\
+\ & \sum_{i=1}^{t} \left\| \mathop{\mathbb{E}}_{D,T,U} \sum_{\substack{A \subseteq [i]: \\ |A| = k, i \in A}} \widehat{F^{\leq i}}_A \cdot \chi_A (T \wedge D + \overline{T} \wedge U) \right\|_F \cdot \left\| \mathop{\mathbb{E}}_{D,T,U} F^{>i} (T \wedge D + \overline{T} \wedge U) \right\|_F \\
\leq\ & \sqrt{w} \sum_{i=1}^{t} 2^{-k/2} \sqrt{ \sum_{\substack{A \subseteq [i]: \\ |A| = k, i \in A}} \left\| \widehat{F^{\leq i}}_A \right\|_F^2 } \\
\leq\ & \sqrt{w} \sum_{i=1}^{t} 2^{-k/2} \sqrt{ \sum_{A \subseteq [i]} \left\| \widehat{F^{\leq i}}_A \right\|_F^2 } \leq wt2^{-k/2}.
\end{aligned}$$

where in the first inequality we have used $\mathbb{E}\, F(U) = \widehat{F}_\emptyset$ and $\mathbb{E}_{D,U}\, \chi_S(T \wedge D + \overline{T} \wedge U) = 0$. In the second inequality we have used $\left\| \mathbb{E}_{D,T,U}\, F^{>i}(T \wedge D + \overline{T} \wedge U) \right\|_F \leq \mathbb{E}_{D,T} \left\| F^{>i}(T \wedge D + \overline{T} \wedge U) \right\|_F \leq \sqrt{w}$ and Lemma 10. Finally we use Lemma 6. □