Pseudo-Random Pseudo-Distributions (for Read-Once Branching Programs)

Gil Cohen
joint work with **Mark Braverman** and **Sumegha Garg**

December 16, 2019

Outline

1. The **BPL** vs. **L** Problem
2. The $\mathcal{B}$-Polynomial
3. Our Contribution
4. Nisan's Construction
5. Some Ideas From Our Work

The **BPL** vs. **L** Problem

The Problem

Derandomize with minimal overhead in space.

The **BPL** vs. **L** Problem

The Problem

Derandomize with minimal overhead in space.

Given a randomized algorithm with space complexity S, devise a deterministic algorithm with a comparable space complexity S'.

The **BPL** vs. **L** Problem

The Problem

Derandomize with minimal overhead in space.

Given a randomized algorithm with space complexity S, devise a deterministic algorithm with a comparable space complexity S'.

- Given how?

The **BPL** vs. **L** Problem

The Problem

Derandomize with minimal overhead in space.

Given a randomized algorithm with space complexity S, devise a deterministic algorithm with a comparable space complexity S'.

- Given how? Black-box access.

The **BPL** vs. **L** Problem

The Problem

Derandomize with minimal overhead in space.

Given a randomized algorithm with space complexity S, devise a deterministic algorithm with a comparable space complexity S'.

- Given how? Black-box access.
- In the regime $S(n) = \Omega(\log n)$, whether or not a derandomization with constant overhead in space $S' = O(S)$ is possible is the **BPL** = **L** question.

What is known?

- Savitch's Theorem (1970) implies $\mathbf{RL} \subseteq \mathbf{L}^2$.
- Borodin-Cook-Pippenger (1983) established $\mathbf{BPL} \subseteq \mathbf{L}^2$.
- Nisan (1992, 94) proved that $\mathbf{BPL} \subseteq \mathbf{SC}$.
- The state-of-the-art result concerning space only is due to Saks and Zhou (1999) who proved $\mathbf{BPL} \subseteq \mathbf{L}^{3/2}$.
- The $\mathbf{BPL}$ vs. $\mathbf{L}$ problem has been studied extensively with a fantastic array of results (e.g., Reingold's $\mathbf{SL} = \mathbf{L}$ (2005)). An extensive pseudorandom machinery was developed motivated by this problem (e.g., Impagliazzo-Nisan-Wigderson (1994), Nisan-Zuckerman (1996), and Raz-Reingold (1999)).

Outline

The $\mathcal{B}$-Polynomial

Derandomization is typically executed using pseudorandom generators. In our setting, the PRG are constructed for read-once branching programs. We will take a somewhat different perspective on this.

The $\mathcal{B}$-Polynomial

Derandomization is typically executed using pseudorandom generators. In our setting, the PRG are constructed for read-once branching programs. We will take a somewhat different perspective on this.

Definition (The $\mathcal{B}$-polynomial)

$$\mathcal{B}(\bar{x}, \bar{y}) = 2^{-n} \prod_{i=1}^{n} (x_i + y_i) \in \mathbb{R}[\bar{x}, \bar{y}].$$

$\mathcal{B}$ has sparsity 2^n with respect to the "natural" basis $\mathcal{M} = \{x_1 \cdots x_n,\ x_1 \cdots x_{n-1} y_n, \ldots,\ y_1 \cdots y_n\}$.

We can think of $\mathcal{B}$ as encoding the uniform distribution over n-bit strings by identifying the elements of $\mathcal{M}$ with n-bit strings and the coefficients with the respective probabilities.

Approximating the $\mathcal{B}$-polynomial

Definition

A polynomial $P(\bar{x}, \bar{y})$ is said to (w, ε)-approximate $\mathcal{B}$ if for every sequence of zero-one $w \times w$ stochastic matrices $X_1, \ldots, X_n, Y_1, \ldots, Y_n$, it holds that

$$\|\mathcal{B}(\bar{X}, \bar{Y}) - P(\bar{X}, \bar{Y})\| \leq \varepsilon.$$

Approximating the $\mathcal{B}$-polynomial

Definition

A polynomial $P(\bar{x}, \bar{y})$ is said to (w, ε)-approximate $\mathcal{B}$ if for every sequence of zero-one $w \times w$ stochastic matrices $X_1, \ldots, X_n, Y_1, \ldots, Y_n$, it holds that

$$\|\mathcal{B}(\bar{X}, \bar{Y}) - P(\bar{X}, \bar{Y})\| \leq \varepsilon.$$

Lemma (The Naïve Derandomization Lemma)

Let P be a polynomial that $(w = n, \varepsilon = 1/3)$-approximate $\mathcal{B}$. Assume P has the following properties:

Sparsity. *P has sparsity s with respect to $\mathcal{M}$; and*

Explicitness. *Every coefficient of P is computable in space $O(\log s)$.*

Then, $\mathbf{BPL} \subseteq \mathbf{DSPACE}(\log s)$.

Approximating the **BPL** polynomial

Theorem (easy)

For every n, w, ε $\exists P \in \mathbb{R}[\bar{x}, \bar{y}]$ *with sparsity* $s = (nw/\varepsilon)^{O(1)}$ *that* (w, ε)*-approximate* $\mathcal{B}$.

Approximating the BPL polynomial

Theorem (easy)

For every n, w, ε $\exists P \in \mathbb{R}[\bar{x}, \bar{y}]$ with sparsity $s = (nw/\varepsilon)^{O(1)}$ that (w, ε)-approximate $\mathcal{B}$.

Had the explicitness condition been met, $\mathbf{BPL} = \mathbf{L}$.

Approximating the **BPL** polynomial

Theorem (easy)

For every n, w, ε $\exists P \in \mathbb{R}[\bar{x}, \bar{y}]$ with sparsity $s = (nw/\varepsilon)^{O(1)}$ that (w, ε)-approximate $\mathcal{B}$.

Had the explicitness condition been met, $\mathbf{BPL} = \mathbf{L}$.

Theorem (Nisan'92)

For every n, w, ε there exists an explicit polynomial $P \in \mathbb{R}[\bar{x}, \bar{y}]$ with sparsity $s = (nw/\varepsilon)^{O(\log n)}$ that (w, ε)-approximate $\mathcal{B}$. As a corollary, $\mathbf{BPL} \subseteq \mathbf{L}^2$.

Nisan's polynomial has all coefficients equal (to $1/s$). This corresponds to a pseudorandom distribution.

Approximating the **BPL** polynomial

Theorem (easy)

For every n, w, ε $\exists P \in \mathbb{R}[\bar{x}, \bar{y}]$ with sparsity $s = (nw/\varepsilon)^{O(1)}$ that (w, ε)-approximate $\mathcal{B}$.

Had the explicitness condition been met, $\mathbf{BPL} = \mathbf{L}$.

Theorem (Nisan'92)

For every n, w, ε there exists an explicit polynomial $P \in \mathbb{R}[\bar{x}, \bar{y}]$ with sparsity $s = (nw/\varepsilon)^{O(\log n)}$ that (w, ε)-approximate $\mathcal{B}$. As a corollary, $\mathbf{BPL} \subseteq \mathbf{L}^2$.

Nisan's polynomial has all coefficients equal (to $1/s$). This corresponds to a pseudorandom distribution.

Despite much success studying restricted settings, there has been no progress on improving Nisan's PRG.

Outline

1. The **BPL** vs. **L** Problem
2. The $\mathcal{B}$-Polynomial
3. Our Contribution
4. Nisan's Construction
5. Some Ideas From Our Work

Our contribution

Theorem (Main result)

For every n, w, ε there exists an explicit polynomial $P \in \mathbb{R}[\bar{x}, \bar{y}]$ with sparsity

$$s = (nw)^{\tilde{O}(\log n)} \cdot (1/\varepsilon)^{O(\log \log(1/\varepsilon))}$$

that (w, ε)-approximate $\mathcal{B}$.

Our contribution

Theorem (Main result)

For every n, w, ε there exists an explicit polynomial $P \in \mathbb{R}[\bar{x}, \bar{y}]$ with sparsity

$$s = (nw)^{\tilde{O}(\log n)} \cdot (1/\varepsilon)^{O(\log\log(1/\varepsilon))}$$

that (w, ε)-approximate $\mathcal{B}$.

The polynomial we construct has positive as well as negative coefficients, and these can be large in absolute value. Hence, the polynomial does not correspond to a distribution but rather to what we call a pseudo-distribution. This is perfectly fine for the purpose of derandomization and we view this as a feature.

Why care about ε?

At the end of the day, when applying the Naïve Derandomization Lemma, we set $\varepsilon = 1/3$, so...

Why care about ε?

At the end of the day, when applying the Naïve Derandomization Lemma, we set $\varepsilon = 1/3$, so...

- As we will see, the $n^{\log n}$ factor is due to the way the error is aggregated, and so a better understanding of the error is crucial.

Why care about ε?

At the end of the day, when applying the Naïve Derandomization Lemma, we set $\varepsilon = 1/3$, so...

- As we will see, the $n^{\log n}$ factor is due to the way the error is aggregated, and so a better understanding of the error is crucial.
- We observe that sparsity $s = n^{\log n}(w/\varepsilon)^{O(1)}$ would yield $\mathbf{BPL} \subseteq \mathbf{L}^{4/3}$ via the Saks-Zhou framework. A conditional result of Raz-Reingold gives $s = (n/\varepsilon)^{\log n} w^{O(1)}$ (in the white-box model).

Why care about ε?

At the end of the day, when applying the Naïve Derandomization Lemma, we set $\varepsilon = 1/3$, so...

- As we will see, the $n^{\log n}$ factor is due to the way the error is aggregated, and so a better understanding of the error is crucial.
- We observe that sparsity $s = n^{\log n}(w/\varepsilon)^{O(1)}$ would yield $\mathbf{BPL} \subseteq \mathbf{L}^{4/3}$ via the Saks-Zhou framework. A conditional result of Raz-Reingold gives $s = (n/\varepsilon)^{\log n} w^{O(1)}$ (in the white-box model).
- Pseudo-random pseudo-distributions readily yield hitting sets (suitable for derandomizing **RL**). Thus, our work gave the first improved hitting set over Nisan's in the general setting. A substantially simpler construction was obtained afterwards by Hoza and Zuckerman (2018).

Outline

Nisan's construction

Nisan's construction is recursive. Recall that

$$\mathcal{B}(\bar{x}, \bar{y}) = 2^{-n} \prod_{i=1}^{n} (x_i + y_i),$$

Factor $\mathcal{B} = \mathcal{B}_L \mathcal{B}_R$, where

$$\mathcal{B}_L(\bar{x}, \bar{y}) = 2^{-n/2} \prod_{i=1}^{n/2} (x_i + y_i),$$
$$\mathcal{B}_R(\bar{x}, \bar{y}) = 2^{-n/2} \prod_{i=n/2+1}^{n} (x_i + y_i).$$

Say we recursively obtained P_L, P_R that $\varepsilon(n/2)$-approximate $\mathcal{B}_L$ and $\mathcal{B}_R$, respectively, each having sparsity $s(n/2)$.

Nisan's construction

Nisan's construction is recursive. Recall that

$$\mathcal{B}(\bar{x}, \bar{y}) = 2^{-n} \prod_{i=1}^{n} (x_i + y_i),$$

Factor $\mathcal{B} = \mathcal{B}_L \mathcal{B}_R$, where

$$\mathcal{B}_L(\bar{x}, \bar{y}) = 2^{-n/2} \prod_{i=1}^{n/2} (x_i + y_i),$$

$$\mathcal{B}_R(\bar{x}, \bar{y}) = 2^{-n/2} \prod_{i=n/2+1}^{n} (x_i + y_i).$$

Say we recursively obtained P_L, P_R that $\varepsilon(n/2)$-approximate $\mathcal{B}_L$ and $\mathcal{B}_R$, respectively, each having sparsity $s(n/2)$. How can we approximate the product $P_L P_R$ of approximations?

Nisan's construction

Taking the naïve product $P_L P_R$ will result in sparisty $s(n/2)^2$ which will get us nowhere.

Nisan's construction

Taking the naïve product $P_L P_R$ will result in sparisty $s(n/2)^2$ which will get us nowhere.

Definition (Samplers (Bellare-Rompel 1994))

A bipartite graph $G = (L, R, E)$ is an (ε, δ)-sampler if $\forall f : R \to [0,1]$ there is a set $B \subseteq L$ of size at most $|B| \le \delta |L|$ such that $\forall v \in L \setminus B$,

$$|\mathbb{E}[f(\Gamma(v))] - \mathbb{E}[f(R)]| \le \varepsilon.$$

Taking the naïve product $P_L P_R$ will result in sparisty $s(n/2)^2$ which will get us nowhere.

Definition (Samplers (Bellare-Rompel 1994))

A bipartite graph $G = (L, R, E)$ is an (ε, δ)-sampler if $\forall f : R \to [0, 1]$ there is a set $B \subseteq L$ of size at most $|B| \le \delta |L|$ such that $\forall v \in L \setminus B$,

$$|\mathbb{E}[f(\Gamma(v))] - \mathbb{E}[f(R)]| \le \varepsilon.$$

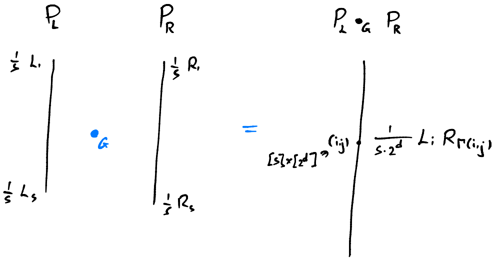
Theorem (Goldreich-Wigderson 1997)

For every integer n and $\varepsilon, \delta > 0$ there exists an explicit (ε, δ)-sampler with $|L| = |R| = n$ and left degree $2^d = O(\varepsilon^{-2}\delta^{-1})$.

Nisan's construction

Write $P_L = \mathbb{E}[L_i]$ and $P_R = \mathbb{E}[R_i]$. Take an $(\varepsilon_S, \delta_S)$-sampler G with $s(n/2)$ vertices on each side, and define

$$P_L \bullet_G P_R = \mathbb{E}_i \left[L_i \, \mathbb{E}_{j \sim \Gamma(i)} R_j \right].$$



Nisan's construction

Lemma (The Derandomized Product Lemma)

For all zero-one $w \times w$ stochastic matrices $X_1, \ldots, X_n, Y_1, \ldots, Y_n$,

$$\|(P_L \bullet_G P_R)(\bar{X}, \bar{Y}) - P_L(\bar{X})P_R(\bar{Y})\| = O((\varepsilon_S + \delta_S)w).$$

Nisan's construction

Lemma (The Derandomized Product Lemma)

For all zero-one $w \times w$ stochastic matrices $X_1, \ldots, X_n, Y_1, \ldots, Y_n$,

$$\|(P_L \bullet_G P_R)(\bar{X}, \bar{Y}) - P_L(\bar{X})P_R(\bar{Y})\| = O((\varepsilon_S + \delta_S)w).$$

Taking $\varepsilon_S = \delta_S \sim 2^{-d}$ and opening the recursion,

$$s(n) = s(n/2)2^d = \cdots = 2^{d \log n},$$
$$\varepsilon(n) \leq 2\varepsilon(n/2) + 2^{-d}w = \cdots = 2^{-d}nw = \varepsilon,$$

and so $s(n) = (nw/\varepsilon)^{O(\log n)}$.

Proof of the Derandomized Product Lemma

We will prove that $\forall a_1, \ldots, a_s, b_1, \ldots, b_s \in [0,1]$ with $\mathbb{E}_i[a_i] = \alpha$, $\mathbb{E}_i[b_i] = \beta$, it holds that

$$\left| \mathbb{E}_i \left[a_i \, \mathbb{E}_{j \sim \Gamma(i)} b_j \right] - \alpha\beta \right| = O(\varepsilon_S + \delta_S).$$

Proof of the Derandomized Product Lemma

We will prove that $\forall a_1, \ldots, a_s, b_1, \ldots, b_s \in [0,1]$ with $\mathbb{E}_i[a_i] = \alpha$, $\mathbb{E}_i[b_i] = \beta$, it holds that

$$\left| \mathbb{E}_i \left[a_i \, \mathbb{E}_{j \sim \Gamma(i)} b_j \right] - \alpha\beta \right| = O(\varepsilon_S + \delta_S).$$

If i is "good" then $b_{\Gamma(i)} = \mathbb{E}_{j \sim \Gamma(i)} b_j \in [\beta - \varepsilon_S, \beta + \varepsilon_S]$. Thus,

$$\begin{aligned}
\mathbb{E}_i \left[a_i b_{\Gamma(i)} \right] &\leq \mathbb{E}_i \left[a_i b_{\Gamma(i)} \mid i \text{ good} \right] + \Pr[i \text{ not good}] \\
&\leq (\beta + \varepsilon_S) \mathbb{E}_i \left[a_i \mid i \text{ good} \right] + \delta_S \\
&\leq (\beta + \varepsilon_S) \frac{\alpha}{1 - \delta_S} + \delta_S \\
&= \alpha\beta + O(\alpha\varepsilon_S + \delta_S).
\end{aligned}$$

Outline

An Observation

The error term we got is $O(\alpha\varepsilon_S + \delta_S)$. Can we exploit the α factor?

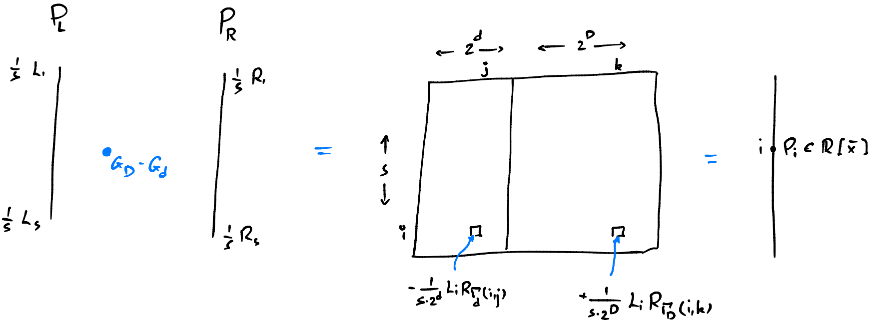
An Observation

The error term we got is $O(\alpha\varepsilon_S + \delta_S)$. Can we exploit the α factor? Well...

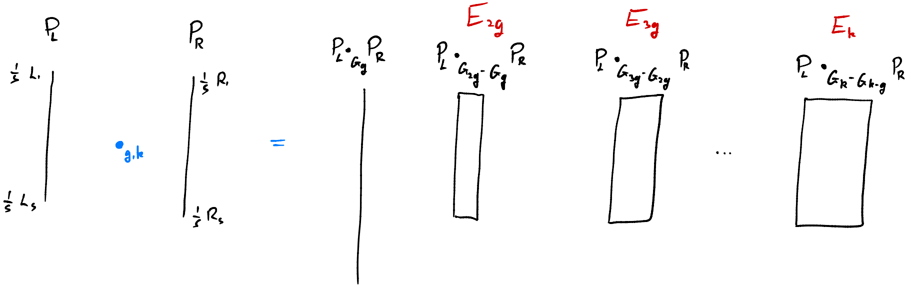
- α is not small (either 1 or increasing with w, depending on the choice of norm); Furthermore,
- δ_S is not multiplied by α and has the same effect on the degree d.

Delta of samplers

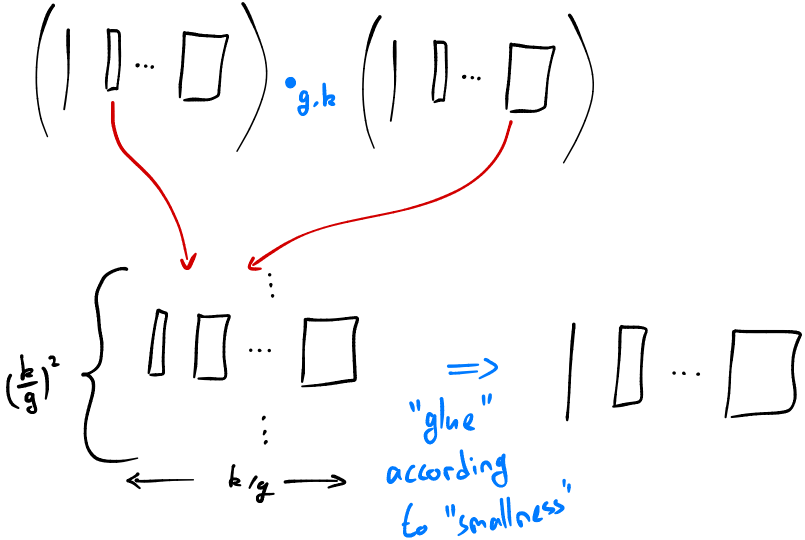
Take two samplers G_D and G_d with $D \gg d$.



Graded representations



Product of graded representations



Outline

1. The **BPL** vs. **L** Problem
2. The $\mathcal{B}$-Polynomial
3. Our Contribution
4. Nisan's Construction
5. Some Ideas From Our Work