Electronic Colloquium on Computational Complexity, Report No. 70 (2009)

Pseudorandomness for Width 2 Branching Programs

Andrej Bogdanov* Zeev Dvir† Elad Verbin‡ Amir Yehudayoff§

Abstract

Bogdanov and Viola (FOCS 2007) constructed a pseudorandom generator that fools degree k polynomials over $\mathbb{F}_2$ for an arbitrary constant k. We show that such generators can also be used to fool branching programs of width 2 and polynomial length that read k bits of inputs at a time. This model generalizes polynomials of degree k over $\mathbb{F}_2$ and includes some other interesting classes of functions, for instance k-DNF.

The construction of Bogdanov and Viola consists of summing k independent copies of a generator that ϵ-fools linear functions (an ϵ-biased set). Our second result investigates the limits of such constructions: We show that, in general, such a construction is not pseudorandom against bounded fan-in circuits of depth $O((\log(k \log 1/\epsilon))^2)$.

*andrejb@cse.cuhk.edu.hk. The Chinese University of Hong Kong. Work done while at Tsinghua University. Supported in part by the National Natural Science Foundation of China Grant 60553001, and the National Basic Research Program of China Grants 2007CB807900, 2007CB807901.

†zeev.dvir@weizmann.ac.il. Department of Computer Science, Weizmann institute of science, Rehovot, Israel. Research supported by Binational Science Foundation (BSF) grant, by Israel Science Foundation (ISF) grant and by Minerva Foundation grant.

‡eladv@tsinghua.edu.cn. ITCS, Tsinghua University, FIT Building 4-608-6. Supported in part by the National Natural Science Foundation of China Grant 60553001, and the National Basic Research Program of China Grants 2007CB807900, 2007CB807901.

§amir.yehudayoff@weizmann.ac.il. Department of Computer Science and Applied Mathematics, The Weizmann Institute of Science, Rehovot, 76100 Israel. Research supported by grants from the Binational Science Foundation (BSF), the Israel Science Foundation (ISF), the Minerva Foundation, and the Israel Ministry of Science (IMOS) - Eshkol Fellowship.

ISSN 1433-8092

1 Introduction

Bogdanov and Viola [BV07] suggested the following construction for a pseudorandom generator against degree k polynomials over a finite field:

$$G'(s_1, \ldots, s_k) = G(s_1) + \cdots + G(s_k), \tag{1}$$

where G is a pseudorandom generator against linear functions [NN90]. Building on their work and subsequent work by Lovett [Lov08], Viola [Vio08] recently proved that the generator is pseudorandom against low-degree polynomials: If G is ϵ-pseudorandom against linear functions, then G' is $O((\epsilon n^k)^{1/2^k})$-pseudorandom against degree k polynomials.

By Viola's analysis, this pseudorandom generator has seed length that is optimal up to constant factor, as long as the degree of the polynomial is constant. However, the seed length deteriorates exponentially with the degree, and the result becomes trivial when the degree exceeds $\log n$. It is natural to ask whether the Bogdanov-Viola construction is also pseudorandom for polynomials of, say, polylogarithmic degree. If the answer was yes, this would give a new pseudorandom generator against polynomial size constant depth circuits with modular gates, since such circuits can be approximated by polynomials of polylogarithmic degree in a very strong sense [Raz87, Smo87].

If this argument was correct, it would give an example where a pseudorandom generator that was designed for one class of functions (polynomials) would automatically yield a derandomization of a different class (small depth circuits). Conversely, if we believe that the Bogdanov-Viola generator is *not* pseudorandom against polynomials of polylogarithmic degree, we could try proving this by giving a constant depth circuit against which the generator is not pseudorandom.

More generally, we can ask: Given a probability distribution with some property (k-wise independence, small bias against linear functions, a convolution of several such distributions), is it pseudorandom against some class of functions (small circuits, space bounded computations)? This type of question has been considered before. For example, Linial and Nisan [LN90] showed that DNFs on n variables are fooled by $O(\sqrt{n} \log n)$-wise independent distributions and conjectured that the $\sqrt{n}$ factor can be removed. Recently, Bazzi [Baz07] made a breakthrough by proving the Linial-Nisan conjecture for depth 2 circuits.

We believe that this kind of study could reveal insights about the power and limitations of existing constructions of pseudorandom generators. In this paper, we begin such an investigation for the Bogdanov-Viola generator and prove two results, one positive and one negative:

- The Bogdanov-Viola generator can fool branching programs of width 2 and polynomial length that read k bits of inputs at a time (see below for precise definition of model). This model generalizes polynomials of degree k over $\mathbb{F}_2$ and includes some

other interesting classes of functions, for instance k-DNF.[1]

- The Bogdanov-Viola generator does not, in general, fool bounded fan-in circuits of depth $O((\log(k \log 1/\epsilon))^2)$. Here, ϵ is the bias of the generator G (see (1)).

1.1 Positive result

We say a distribution D on $\{0,1\}^n$ is *ϵ-pseudorandom* against a class C of functions from $\{0,1\}^n$ to $\{0,1\}$ if for every $f \in C$,

$$\left|\Pr_{x\sim D}[f(x)=1] - \Pr_{x\sim\{0,1\}^n}[f(x)=1]\right| \leq \epsilon$$

(where $x \sim \{0,1\}^n$ means that x is uniformly distributed in $\{0,1\}^n$). A function $G : \{0,1\}^m \to \{0,1\}^n$ is *an ϵ-pseudorandom generator* (PRG) against C if the distribution $G(s)$, $s \sim \{0,1\}^m$, is ϵ-pseudorandom against C. We call m the *seed length* of the generator.

Here we are interested in the class (k,t,n)-2BP of *width 2 branching programs* of length t that read k bits of input at a time and compute a function from $\{0,1\}^n$ to $\{0,1\}$. This device can be described by a layered directed acyclic graph, where there are t layers and each layer contains two nodes, which we label by 0 and 1. Each layer j is associated with an arbitrary k-bit substring $x|_j$ of the input x. Each node in layer j has 2^k outgoing edges labelled by possible values of the string $x|_j$. On input x, the computation starts in the first node in the first layer, then follows the edge labelled by $x|_1$ onto the second layer, and so on until a node in the last layer is reached. The identity of this last node is the outcome of the computation.

This type of branching program can represent a degree-k polynomial, a "space-bounded" computation with one bit of memory, as well as a k-DNF formula. We prove the following:

Theorem 1.1. *Let G be an ϵ-PRG against degree k polynomials in n variables over $\mathbb{F}_2$. Then G is an ϵ'-PRG against the class of functions computed by a (k,t,n)-2BP, with $\epsilon' = t \cdot \epsilon$.*

1.2 Negative result

Our second result is the following theorem which shows the limitations of the Bogdanov-Viola construction.

Theorem 1.2. *For every n, ϵ, and k there exists a distribution D such that D is ϵ-pseudorandom against linear functions over $\{0,1\}^n$, but the sum of k independent copies of D is not $1/3$-pseudorandom against bounded fan-in circuits (with and, or, and not gates) of depth $O((\log(k \log 1/\epsilon))^2)$.*

[1] In the special case of k-DNF, Trevisan [Tre04] has a better result: He shows that for every constant k there is an $\epsilon = 1/\mathrm{poly}(n)$ such that ϵ-biased generators against linear functions fool k-DNF over n variables.

It is known [MRRW77, FT05] that the seed length of an ϵ-biased generator against linear functions must be at least $\Omega(\log n + \log(1/\epsilon))$. Therefore, if we want the generator to be efficient, we are restricted to using $\epsilon = 1/\mathrm{poly}(n)$. For this setting of parameters, Theorem 1.2 tells us that the Bogdanov-Viola generator does not fool bounded fan-in circuits of depth $O((\log(k \log n))^2)$.

By Barrington's theorem [Bar89], any circuit of depth d can be simulated by a branching program of width 5 and size 4^d, so D^k is not pseudorandom against width 5, size $2^{O((\log(k \log 1/\epsilon))^2)}$ branching programs. So the Bogdanov-Viola generator fools branching programs of width 2, but not of width 5. We do not know what happens for width 3 or width 4.[2]

1.3 Proof overview for Theorem 1.1

It has been known for some time that *read-once* width 2 branching programs that read one bit at a time can be fooled by linear generators.[3] One way to argue this is to think of the computation of the branching program B as a boolean function over $\mathbb{F}_2^n$ and show inductively over the layers of B that the sum of the absolute values of the Fourier coefficients of B is bounded from above by t. It is easy to see that linear generators of bias ϵ are ϵL-pseudorandom against any boolean function whose sum of absolute values of Fourier coefficients is at most L, and the correctness follows from there.

For branching programs that read more than one bit at a time this argument cannot work, as there exist width 2 branching programs that read 2 bits at a time and that are not fooled by some small bias linear generator. One such branching program computes the inner product function

$$IP(x_1, \ldots, x_n) = x_1 x_2 + \cdots + x_{n-1} x_n \pmod 2 \quad (n \text{ even}).$$

Nevertheless, we argue along the same lines. Instead of using the Fourier transform of the branching program, we resort to "higher-order" representations of functions using low-degree polynomials. We show that every branching program B with length t and width 2 that reads k bits at a time admits a "representation of length t" in terms of degree k polynomials. By "representation of length t" we mean that B can be written as a sum *over the reals* of the form

$$(-1)^{B(x)} = \sum\nolimits_{p:\mathbb{F}_2^n \to \mathbb{F}_2} \alpha_p \cdot (-1)^{p(x)}$$

where p ranges over all degree k polynomials over $\mathbb{F}_2$, and α_p are real coefficients such that $\sum_p |\alpha_p| \leq t$. Unlike the Fourier transform, for degree 2 and larger this representation is not

[2] A correlation bound of Viola and Wigderson [VW07] shows that, in general, a distribution that is pseudorandom against degree d polynomials need not be pseudorandom against the "parity modulo 3" function, which is computable by a width 3 branching program. However, their counterexample is not a convolution of independent distributions with small bias against linear functions; i.e., it does not have the form of the Bogdanov-Viola generator.

[3] We are not aware of a published proof but have heard the result credited to Saks and Zuckerman.

unique. Once this representation is obtained, we argue that a pseudorandom generator for degree k polynomials is also pseudorandom for B by linearity of expectation.

While our proof is not technically difficult we find the application of "higher-order" Fourier type analysis conceptually interesting and potentially relevant for other computer science applications.

2 Fooling width 2 branching programs

Recall that we use (k, t, n)-2BP to denote width 2 branching programs of length t that read k bits of input at a time and compute a function from $\{0,1\}^n$ to $\{0,1\}$. For a function $f : \{0,1\}^n \to \{0,1\}$, we denote $\hat{f} = (-1)^f$, a map from $\{0,1\}^n$ to $\{1,-1\}$. Define $\deg(f)$ to be the degree of f when viewed as a multilinear polynomial in $\mathbb{F}_2[x_1, \ldots, x_n]$.

2.1 Width 2 Branching Programs as Sum of Polynomials

The following theorem is the basis for the proof of Theorem 1.1. It shows that width 2 branching programs have a "short representation by polynomials of small degree".

Theorem 2.1. *Let $f : \{0,1\}^n \to \{0,1\}$ be computed by a (k, t, n)-2BP. Then there exist $\alpha_1, \ldots, \alpha_s \in \mathbb{R}$ and $g_1, \ldots, g_s : \{0,1\}^n \to \{0,1\}$ such that*

1. *$\hat{f}(x) = \sum_{i=1}^{s} \alpha_i \cdot \hat{g}_i(x)$ for all $x \in \{0,1\}^n$ (where the sum is over the reals).*
2. *For all $i \in [s]$, $\deg(g_i) \leq k$.*
3. *$\sum_{i=1}^{s} |\alpha_i| \leq t$.*

We defer the proof of Theorem 2.1 to Section 2.2 and proceed by showing how it implies our main result.

Proof of Theorem 1.1

Let $G : \{0,1\}^m \to \{0,1\}^n$ be an ϵ-pseudorandom generator against degree k polynomials in n variables over $\mathbb{F}_2$. Let $f : \{0,1\}^n \to \{0,1\}$ be computed by a (k, t, n)-2BP. By Theorem 2.1, there exist $\alpha_1, \ldots, \alpha_s \in \mathbb{R}$ and $g_1, \ldots, g_s : \{0,1\}^n \to \{0,1\}$ such that

1. $\hat{f}(x) = \sum_{i=1}^{s} \alpha_i \cdot \hat{g}_i(x)$ for all $x \in \{0,1\}^n$.
2. For all $i \in [s]$, $\deg(g_i) \leq k$.

3. $\sum_{i=1}^{s} |\alpha_i| \leq t$.

For the rest of the proof $x \sim \{0,1\}^n$ and $s \sim \{0,1\}^m$ denote two independent random variables. First, note that for every function $f : \{0,1\}^n \to \{0,1\}$,

$$2 \cdot \left| \Pr[f(G(s)) = 1] - \Pr[f(x) = 1] \right| = \left| \mathbf{E}[\hat{f}(G(s)) - \hat{f}(x)] \right|.$$

Thus, using the properties above, and using the linearity of expectation,

$$\begin{aligned} 2 \cdot \left| \Pr[f(G(s)) = 1] - \Pr[f(x) = 1] \right| &= \left| \mathbf{E}\left[\hat{f}(G(s)) - \hat{f}(x)\right] \right| \\ &= \left| \sum_{i=1}^{s} \alpha_i \cdot \mathbf{E}\left[\hat{g}_i(G(s)) - \hat{g}_i(x)\right] \right| \\ &\leq \sum_{i=1}^{s} |\alpha_i| \cdot \left| \mathbf{E}\left[\hat{g}_i(G(s)) - \hat{g}_i(x)\right] \right| \\ &= \sum_{i=1}^{s} |\alpha_i| \cdot 2 \cdot \left| \Pr[g_i(G(s)) = 1] - \Pr[g_i(x) = 1] \right| \\ &\leq 2 \cdot t \cdot \epsilon, \end{aligned}$$

where the last inequality holds since G is an ϵ-pesudorandom generator against degree k polynomials. The theorem now follows. □

2.2 Proof of Theorem 2.1

Let f be a boolean function computed by a branching program B of width 2 and length t that reads k bits of input at a time. We will prove the theorem by induction on t.

Induction base: For the case $t = 1$, the theorem holds since $f(x)$ is a boolean function in k variables and so $\deg(f) \leq k$.

Induction step: Assume that the theorem holds for every function computed by a $(k, t-1, n)$-2BP. By the definition of such branching programs, there exists $P : \{0,1\}^{k+1} \to \{0,1\}$ such that

$$f(x) = P\left(f_{t-1}(x)\,,\, x|_{t-1}\right),$$

where f_{t-1} is the function computed at the $(t-1)$'th layer of B, and $x|_{t-1}$ is the k-bit substring of the input x associated with the $(t-1)$'th layer.

Let p_0 and p_1 be two maps from $\{0,1\}^k$ to $\{0,1\}$ defined as

$$p_0(y) = P(0,y) \quad \text{and} \quad p_1(y) = P(1,y).$$

Note that since both of p_0 and p_1 depend on at most k variables, then $\deg(p_0) \leq k$ and $\deg(p_1) \leq k$. In addition, for every $z \in \{0,1\}$ and $y \in \{0,1\}^k$,

$$\hat{P}(z,y) = \frac{1}{2}(\hat{p_0}(y) - \hat{p_1}(y)) \cdot (-1)^z + \frac{1}{2}(\hat{p_0}(y) + \hat{p_1}(y)).$$

We can now use the induction hypothesis. By the choice of P, for every $x \in \{0,1\}^n$,

$$\hat{f}(x) = \hat{P}(f_{t-1}(x), x|_{t-1}) = \frac{1}{2}(\hat{p_0}(x|_{t-1}) - \hat{p_1}(x|_{t-1})) \cdot \hat{f_{t-1}}(x) + \frac{1}{2}(\hat{p_0}(x|_{t-1}) + \hat{p_1}(x|_{t-1})).$$

By induction, there exist $\alpha_1, \ldots, \alpha_s \in \mathbb{R}$ and $g_1, \ldots, g_s : \{0,1\}^n \to \{0,1\}$ such that

1. $\hat{f_{t-1}}(x) = \sum_{i=1}^{s} \alpha_i \cdot \hat{g}_i(x)$ for all $x \in \{0,1\}^n$.
2. For all $i \in [s]$, $\deg(g_i) \leq k$.
3. $\sum_{i=1}^{s} |\alpha_i| \leq t-1$.

Thus, for all $x \in \{0,1\}^n$

$$\hat{f}(x) \;=\; \frac{1}{2}(\hat{p_0}(x|_{t-1}) - \hat{p_1}(x|_{t-1})) \cdot \sum_{i=1}^{s} \alpha_i \cdot \hat{g}_i(x) + \frac{1}{2}(\hat{p_0}(x|_{t-1}) + \hat{p_1}(x|_{t-1})).$$

We complete the proof by renaming the polynomials and the coefficients in the above sum. For $j = 1, \ldots, s$, set

$$\beta_j = \frac{\alpha_j}{2} \quad \text{and} \quad h_j(x) = p_0(x|_{t-1}) \oplus g_j(x)$$

and for $j = s+1, \ldots, 2s$, set

$$\beta_j = -\frac{\alpha_{j-s}}{2} \quad \text{and} \quad h_j(x) = p_1(x|_{t-1}) \oplus g_j(x)$$

(where $\oplus$ denotes summation in $\mathbb{F}_2$). Set $\beta_{2s+1} = \beta_{2s+2} = 1/2$, set $h_{2s+1}(x) = p_0(x|_{t-1})$, and set $h_{2s+2}(x) = p_1(x|_{t-1})$. Finally, set $s' = 2s+2$. Thus,

$$\hat{f}(x) = \sum_{j=1}^{s'} \beta_j \cdot \hat{h}_j(x)$$

for all $x \in \{0,1\}^n$. In addition, every h_j is of degree at most k (since addition in $\mathbb{F}_2$ does not increase the degree), and

$$\sum_{j=1}^{s'} |\beta_j| \leq 1 + 2 \cdot \sum_{i=1}^{s} \frac{|\alpha_i|}{2} \leq 1 + (t-1) = t.$$

□

3 Limitations of the Bogdanov-Viola construction

In this section we show that a sum of several copies of pseudorandom generators for linear functions fails to fool small depth bounded fan-in circuits (Theorem 1.2).

3.1 Proof of Theorem 1.2

Set $m = k\log(1/\epsilon) + 1$ and partition the input $x \in \mathbb{F}_2^n$ into n/m consecutive blocks $x|_1, \ldots, x|_{n/m} \in \mathbb{F}_2^m$. Consider the following distribution D.

1. Choose a random linear subspace S of $\mathbb{F}_2^m$ of dimension $(m-1)/k$.
2. For $1 \le i \le n$, choose each block $x|_i$ independently and uniformly from S.

To prove Theorem 1.2, we show the following two claims.

Claim 3.1. *The distribution D is ϵ-pseudorandom against linear functions.*

Claim 3.2. *The sum D^k of k independent samples from D is not $1/3$-pseudorandom against bounded fanin circuits of depth $O((\log m)^2)$.*

The theorem follows from these two claims.

Proof of Claim 3.1. Let $a(x) = \langle a, x\rangle$ be an arbitrary nonzero linear function over $\mathbb{F}_2^n$. We split a as a sum of linear functions a_i over the blocks of x as

$$a(x) = \sum\nolimits_{i=1}^{n/m} a_i(x|_i).$$

Without loss of generality, let's assume a_1 is nonzero. Conditioned on the choice of S, the values of the functions $a_i(x|_i)$ are independent:

$$\mathbf{E}_{x\sim D}[(-1)^{a(x)}] = \mathbf{E}_S\Big[\prod\nolimits_{i=1}^{m/n} \mathbf{E}_{x|_i\sim S}[(-1)^{a_i(x|_i)}]\Big].$$

Now for any fixed choice of S, the value $\mathbf{E}_{x|_i\sim S}[(-1)^{a_i(x|_i)}]$ is one if $a_i \in S^\perp$ and zero otherwise. Here

$$S^\perp = \{y : \langle y, x\rangle = 0 \text{ for all } x \in S\}.$$

Therefore

$$|\mathbf{E}_{x\sim D}[(-1)^{a(x)}]| = \mathbf{Pr}[\text{for all } i,\ a_i \in S^\perp] \le \mathbf{Pr}[a_1 \in S^\perp] = 2^{-(m-1)/k} = \epsilon$$

and so $|\mathbf{E}_{x\sim D}[a(x)] - 1/2| \le \epsilon/2 < \epsilon$. □

Proof of Claim 3.2. Let $X_1, \ldots, X_k$ be independent samples from the distribution D and $X = X_1 + \cdots + X_k$. Let S_i denote the subspace of $\mathbb{F}_2^m$ associated to the sample X_i. Since each block of X_i belongs to the subspace S_i, each block of X will belong to the sum of subspaces $S = S_1 + \cdots + S_k$. The subspace S has dimension at most $m-1$.

This suggests the following test for X: Arrange the first $2m$ blocks of X as rows in an $m \times 2m$ matrix M and compute the rank of M over $\mathbb{F}_2$. (By our choice of parameters, $2m^2 \le n$ so

this is always possible.) If the matrix has full rank output one, otherwise output zero. If X is chosen from D^k, then all the rows of M are chosen from the same subspace of dimension $m-1$ so M will never have full rank. If X is chosen from the uniform distribution, then M is a random $m \times 2m$ matrix and, by a union bound, the probability it doesn't have full rank is at most $2^{-m} < 1/3$.

It remains to observe that the above test, which is essentially a rank computation, can be implemented by a circuit of depth $O((\log m)^2)$ via Cook's theorem [Coo85]. □

4 Acknowledgments

We thank Anup Rao for helpful conversations on this problem. This work was done while the authors took part in "China Theory Week" workshop at Tsinghua University. We would like to thank the organizers of the workshop and in particular Andy Yao for their hospitality.

References

[Bar89] D. A. Barrington. Bounded-width polynomial-size branching programs recognize exactly those languages in NC^1. *J. Comput. Syst. Sci.*, 38(1):150–164, 1989.

[Baz07] L. Bazzi. Polylogarithmic independence can fool dnf formulas. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 63–73, 2007.

[BV07] A. Bogdanov and E. Viola. Pseudorandom bits for polynomials. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 41–51, 2007.

[Coo85] S. A. Cook. A taxonomy of problems with fast parallel algorithms. *Inf. Control*, 64(1-3):2–22, 1985.

[FT05] J. Friedman and J.-P. Tillich. Generalized Alon–Boppana theorems and error-correcting codes. *SIAM J. Discret. Math.*, 19(3):700–718, 2005.

[LN90] N. Linial and N. Nisan. Approximate inclusion-exclusion. *Combinatorica*, 10(4):349–365, 1990.

[Lov08] S. Lovett. Unconditional pseudorandom generators for low degree polynomials. *Proceedings of the IEEE Symposium on Foundations of Computer Science*, 2008.

[MRRW77] R. J. McEliece, E. R. Rodemich, H. Rumsey, and L. R. Welch. New upper bound on the rate of a code via the Delsarte-MacWilliams inequalities. *IEEE Trans. Inform. Theory*, 23(2):157–166, 1977.

[NN90] J. Naor and M. Naor. Small-bias probability spaces: Efficient constructions and applications. In *Proceedings of the Twenty Second Annual ACM Symposium on Theory of Computing*, pages 213–223, 1990.

[Raz87] A. Razborov. Lower bounds for the size of circuits of bounded depth with basis AND, XOR. Notes, 1987.

[Smo87] R. Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, pages 77–82, 1987.

[Tre04] L. Trevisan. A note on approximate counting for k-dnf. In *Proceedings of the 8th International Workshop on Randomization and Computation (RANDOM)*, pages 417–426, 2004.

[Vio08] E. Viola. The sum of d small-bias generators fools polynomials of degree d. In *Proceedings of the 23rd Annual IEEE Conference on Computational Complexity*, pages 124–127, 2008.

[VW07] E. Viola and A. Wigderson. Norms, XOR lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. In *IEEE Conference on Computational Complexity*, 2007.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de