Small Random Sets for Affine Spaces and Better Explicit Lower Bounds for Branching Programs

Alexander E. Andreev
University of Moscow
andreev@mntn.msk.su

Juri L. Baskakov
University of Moscow
baskakov@mntn.msk.su

Andrea E. F. Clementi
University of Rome *La Sapienza*
clementi@dsi.uniroma1.it

José D. P. Rolim
University of Geneva
rolim@cui.unige.ch

Abstract

We show the following *Reduction Lemma*: any ϵ-biased sample space with respect to (Boolean) linear tests is also 2ϵ-biased with respect to any *system* of independent linear tests. Combining this result with the previous constructions of ϵ-biased sample space with respect to linear tests, we obtain the first efficient construction of *discrepancy sets* (i.e. two-sided pseudo-random sets) for *Boolean affine spaces*. We also give a different version of the *powering construction* of ϵ-biased sample spaces given by Alon *et al* in order to obtain k-wise ϵ-biased sample spaces with respect to any affine spaces.

The second main contribution of this paper is a new direct connection between the efficient construction of pseudo-random (both two-sided and one-sided) sets for Boolean affine spaces and the explicit construction of Boolean functions having hard *branching program* complexity.

In the case of 1-read branching programs (1-*Br.Pr.*), this connection relies on a different interpretation of Simon and Szegedy's Theorem in terms of Boolean linear systems. Our constructions of non trivial (i.e. of cardinality $2^{o(n)}$) *discrepancy sets* for Boolean affine spaces of dimension greater than $n/2$ yield a set of Boolean functions in PH having very hard 1-*Br.Pr.* size. In particular, we obtain a Boolean function in $\mathsf{NP}^{\mathsf{NP}} \cap \mathsf{P/poly}$ having 1-*Br.Pr.* size not smaller than $2^{n-4\log n}$. This bound is tight and improves over the best previously known bound which was 2^{n-s} where $s = O(n^{2/3}\log^{1/3} n)$ [22].

For the more general case of non deterministic, syntactic k-read branching programs (k-*Br.Pr.*), we introduce a new method to derive explicit, exponential lower bounds that envolves the efficient construction of *hitting sets* (i.e. one-sided pseudo-random sets) for affine spaces of dimension $o(n/2)$. Using an "orthogonal" representation of small Boolean affine spaces and again the Reduction Lemma, we give the required construction of hitting sets thus obtaining an explicit Boolean function that belongs to PH and has k-*Br.Pr.* size not smaller than $2^{n^{1-o(1)}}$ for any $k = o\left(\frac{\log n}{\log\log n}\right)$. This asymptotically improves the exponents of the previous known lower bounds given in [8, 12, 20] for some range of k.

1 Introduction

This paper presents two sets of new results in complexity theory. The first set concerns the efficient construction of small spaces that looks random to a certain class of Boolean functions. We give a natural generalization of Naor and Naor's definition of ϵ-biased sample space for linear tests [17] to the case of *system* of independent linear tests and then we present some efficient constructions of small sets that satisfy this stronger property. The second set of results instead concerns the explicit construction of Boolean functions having hard branching programs complexity. We indeed use the first set of results to derive new explicit lower bounds for this computational model.

Small Random Sets for Affine Spaces

The efficient construction of small sample spaces that approximate uniform probability distributions turns out to be an important method for the aim of de-randomization. A nice construction of "good" random sample spaces is that introduced by Naor and Naor [17]. We quote this result in terms of *discrepancy sets.* A Boolean function $f : \{0,1\}^n \rightarrow \{0,1\}$ is said *linear* if it can be written as a linear combination of the input variables with coefficients in $\{0,1\}$.

Definition 1.1 *A subset $S \subseteq \{0,1\}^n$ is* ϵ-discrepant *for* linear functions *if, for any Boolean linear function $f : \{0,1\}^n \rightarrow \{0,1\}$, $|\mathbf{Pr}_S(f(\vec{x}) = 0) - \mathbf{Pr}_S(f(\vec{x}) = 1)| \leq \epsilon$. Further, S is said to be k-wise ϵ-discrepant if the "test" linear functions in the above definition can have at most k non-zero coefficients.*

(Note that, according to the definition in [17], a subset is an ϵ-*biased* sample space w.r.t linear tests iff it is ϵ-discrepant for linear functions).

The main motivation in constructing ϵ-discrepancy sets for linear functions relies on the fact that several probabilistic algorithms (see [17]) work efficiently when the sample space has just the ϵ-discrepancy property for small values of ϵ. Furthemore, the efficient construction of ϵ-discrepancy sets for linear functions also yields small spaces which are k-wise δ-*dependent* [7, 24], a weaker notion of k-wise independent sample spaces that still works for certain classes of probabilistic algorithms.

The main result of [17] is the efficient construction of an ϵ-discrepancy set for linear functions of size $O((n/\epsilon)^{O(1)})$, and the consequent k-wise ϵ-discrepancy set for linear functions of size $O((k \log n)/\epsilon)^{O(1)})$. Such constructions are then exploited, in the same paper, to obtain a small sample space yielding k-wise δ-dependence. Three simpler and better (in some parameter ranges) constructions of ϵ-discrepancy sets for linear functions have been introduced by Alon *et al* in [1]. All of them yield sample spaces of size $O((n/\epsilon)^2)$. Their third method is called the *powering construction* and will be strongly used in our paper (a formal description is given in Section 2.1).

Still motivated by the general goal of de-randomization, much recent research [4, 5, 10, 15] has been devoted to the efficient construction of *hitting sets*, i.e. the *one-sided* version of discrepancy sets. In the general definition, given a class $\mathcal{F}$ of Boolean functions of n inputs, a subset $H \subseteq \{0,1\}^n$ is a *hitting set* for $\mathcal{F}$ if, for any non-zero function $f \in \mathcal{F}$, H contains at least one inputs on which f outputs 1.

Our Results. A general question often investigated in complexity theory is whether the complexity of a certain problem significantly increases when more mutually independent instances have to be solved "in parallel" (this kind of problems is known in literature as the *direct-sum* problem [9, 16]). A natural way to introduce the *direct-sum* problem in the context of the discrepancy sets for linear functions is clearly that of considering the case of *system of linear functions.* We can thus address the problem of constructing small sample spaces which are ϵ-discrepant w.r.t. such systems. More formally, let $\mathcal{AFF}(n,k,s)$ be the set of all n variables linear systems $A\vec{x}$ ($A \in \{0,1\}^{s \times n}$ and $\vec{x} \in \{0,1\}^n$) of at most s linear functions in which at most k variables appear with a non-zero coefficient (such variables are said *essential*). Note that the restriction on the number of essential variables is the generalization of the definition of k-wise ϵ-discrepancy set for linear tests given by Naor and Naor. Indeed for $s = 1$, we get exactly the latter concept.

Definition 1.2 *Let $\epsilon > 0$. A (multi)set $S \subseteq \{0,1\}^n$ is said to be ϵ-discrepant for $\mathcal{AFF}(n,k,s)$ if for any $A\vec{x} \in \mathcal{AFF}(n,k,s)$ with $\mathrm{rank}(A) = s$, and for any $\vec{b} \in \{0,1\}^s$ such that the linear system $A\vec{x} = \vec{b}$ is feasible, it holds $|\mathbf{Pr}_{\vec{x} \in S}(A\vec{x} = \vec{b}) - 2^{-s}| \leq \epsilon$.*

Note that 2^{-s} equals the probability that $\vec{x}$ is a solution of $A\vec{x} = \vec{b}$ when x is chosen uniformly at random from $\{0,1\}^n$. Our main result can be stated in the following way

Theorem 1.1 *Let $n > 0$. For any $k, s \leq n$ and for any $\epsilon > 0$ it is possible to efficiently construct an ϵ-discrepancy set $D(n,k,s)$ for $\mathcal{AFF}(n,k,s)$ of size $|D(n,k,s)| \leq (2k(\log n + 1))/\epsilon)^2$.*

The above theorem implies that the size of the sample space does not surprisingly depend on the number s of independent linear functions that the sample space must satisfy. We thus have an interesting solution of this particular "instance" of the direct-sum problem. This fact relies on a *Reduction Lemma* which is one of the main technical contributions of this paper.

Lemma 1.1 (Reduction Lemma.) *Let $\epsilon > 0$. If $S \subseteq \{0,1\}^n$ is ϵ-discrepant for $\mathcal{AFF}(n,n,1)$ then it is 2ϵ-discrepant for $\mathcal{AFF}(n,n,n)$.*

Thanks to this result, any previous construction of ϵ-discrepancy sets for linear functions turns out to be an ϵ-discrepancy set also for $\mathcal{AFF}(n,n,n)$. Further, the result in Theorem 1.1 concerning $\mathcal{AFF}(n,k,s)$ (i.e. when a non trivial bound on the number of essential variable is imposed) is achieved by constructing an ϵ-discrepancy set for $\mathcal{AFF}(n,k,s)$ which is based on the *powering* construction given by Alon *et al* [1].

As already observed, another important problem in the theory of de-randomization is the efficient construction of the *one-sided* version of discrepancy sets, i.e., hitting sets. It is easy to verify that Theorem 1.1 yields a hitting set for $\mathcal{AFF}(n,k,s)$ if and only if $\epsilon < 1/2^s$. In fact, in this case we have $|D(n,k,s)| \leq (2k(\log n + 1))/\epsilon)^2 = 2k(\log n + 1)2^{2s}$ that gives a hitting set of non trivial size (i.e. of size smaller than 2^n) only when $s < n/2$, i.e., when the affine space is large. However, when the affine space is small there is a more efficient way to represent it by using the corresponding *orthogonal* space (which is large). Combining this idea with the Reduction Lemma, we derive an efficient construction of non trivial hitting sets in the case of small affine spaces.

Theorem 1.2 *For any $n > 0$ and $m \leq n$ it is possible to efficiently construct a hitting set $\mathcal{H}(n,m)$ for $\mathcal{AFF}(n,n,n-m)$ such that $|\mathcal{H}(n,m)| \leq 2^{n-m+m(O(1)\log n)/\sqrt{m}}$.*

The above construction thus gives a non trivial hitting set when $m \geq \frac{1}{o(1)} \log^2 n$.

The construction of small pseudo-random sets finds its natural application in decreasing the use of randomness in some classes of probabilistc algorithms. In this paper, we will show a different application of our previous results: as we will describe in the next section, both discrepancy and hitting sets for affine spaces can be used to de-randomize the construction of some Boolean functions having exponential branching programs size.

Branching Programs

In several previous works [2, 3, 6, 11, 18, 19] the explicit constructon of a computationally-hard function with respect to a certain class $\mathcal{C}$ of algorithms has been used to derive small sample spaces able to derandomize a somewhat probabilistic version of the algorithms in $\mathcal{C}$. Here, we revert this connection and we use our sample spaces to define a set of explicit functions having hard complexity with respect to *read-1-time-branching programs* and *non deterministic syntactic read-k-times branching programs.*

Branching programs (*Br.Pr.*'s) are a general model to compute Boolean functions. In what follows we adopt notations and terminology of [22]. A *Br.Pr.* is a directed acyclic graph where one of the nodes, called *source*, has fan-in 0 and some other nodes, called *terminals*, have fan-out 0. Each non terminal nodes is labelled by the index of an input Boolean variable and has fan-out 2. The two arcs leaving a non terminal node are respectively labelled 0 and 1. A *Br.Pr.* computes a Boolean function $f : \{0,1\}^n \to \{0,1\}$ on input $(x_1, \ldots, x_n)$ as follows. Starting the computation from the source node, if a generic node is reached, the corresponding input variable is tested and the computation chooses the arc corresponding to the actual value of this input variable. The process terminates when a sink node is reached and its label represents the output of $f(x_1, \ldots, x_n)$. The *size* of a *Br.Pr.* is the number of its nodes. In order to investigate the computing power of branching programs, several restrictions have been considered, in particular, that of assuming a fixed bound on the maximum number of reading the input variables. A *read-k-times Br.Pr.* is allowed to read each variable at most k times along any valid computation and, in a *syntactic read-k-times Br.Pr.*, this reading restriction holds to any path from the source to any sink. Notice that while a *read-1-time Br.Pr.* is always a syntactic *read-1-time Br.Pr.*, for $k \geq 2$ this does not hold. Clearly, the branching program model can also be made *nondeterministic* by allowing the existence of more arcs labeled with the same value that leave a same node.

The research in branching programs is extremly active from both theoretical and practical point of view [22, 13, 21, 25]. In complexity theory, among other problems, that of finding lower bounds for the size of *read-k-times Br.Pr.*'s for explicit Boolean functions has been deeply studied (see [8, 12, 21, 22]). One of the "challenging" goals in this field is that of obtaining a superpolynomial lower bound for a Boolean function computable in polynomial time since it would immediatly imply $\mathsf{P} \neq \mathsf{LOG}$. To our knowledge, the best known explicit lower bound for *read-1-time*-branching programs (1-*br.pr.*'s) is due to Savicky and Zak [22] that derive a Boolean function in P that, for any sufficiently large n, has 1-*br.pr.* size not smaller than 2^{n-s}, where $s = O(n^{2/3} \ln^{1/3} n)$. We emphasize that, for this model, there are no known harder explicit functions even in classes larger than P, while there is a non constructive proof that some functions require 1-*Br.Pr.*'s of size $O(2^{n-\log n})$. On the other hand, the general upper bound on the size of 1-*Br.Pr.* is $O(2^{n-\log n})$.

For non-deterministic syntactic read-k-times *Br.Pr.*'s (k-*Br.Pr.*) when $k \geq 2$, there are explicit lower bounds of exponential size when $k = O(\log n)$ [8, 12, 20]. Borodin *et al* [8] showed an explicit family of Boolean functions such that any k-*Br.Pr.* computing it must have a number of labeled edges not smaller than $\mathsf{exp}(\Omega(n)/(4^k * k^3)))$ when $k \leq c \log n$ for a fixed constant $0 < c < 1$. A lower bound of size $\mathsf{exp}(\Omega(\sqrt{n}/k^{2k}))$ have been independently proved by Okolnishnikova [20]. More recently, Jukna studied the Okolnishnikova's function and obtained an exponential gap between the size of *Br.Pr.* required for such a function and that required for its complement. For more general cases, i.e. either when k growes faster than any logarithmic function or when the branching programs are non syntactic, finding exponential lower bound is still an open question.

Our Results. In the case of 1-read branching programs, Simon and Szegedy [23] provided a nice combinatorial theorem that, given any Boolean function $f : \{0,1\}^n \to \{0,1\}$ and any fixed input dimension

$r \leq n$, establishes a relationship between the maximum number of pairwise different subfunctions of f having input dimension r and the 1-*br.pr.* size required for f. This theorem gives a very useful method to obtain explicit lower bounds for this computational model [22]. We first give a new interpretation of Simon and Szegedy's Theorem in terms of Boolean linear systems. Then, by applying this interpretation to some appropriate versions of the generator functions of our *discrepancy sets* for Boolean affine spaces we achieve the following lower bounds that have asymptotically larger exponents than that obtained by Savicky and Zak [22].

Theorem 1.3 *It is possible to construct:*
a) a Boolean function in $\mathsf{NP}^{\mathsf{NP}} \cap \mathsf{P/poly}$ *having* 1-*Br.Pr. size not smaller than* $2^n/4n$ *(observe that this bound is optimal);*
b) a Boolean function in $\mathsf{DTIME}(2^{O(\log^2 n)}) \cap \mathsf{P/poly}$ *having* 1-*Br.Pr. size not smaller than* $2^n/n^{O(1)}$*;*
c) a Boolean function in $\mathsf{DTIME}(2^{O(\log^2 n)}) \cap \mathsf{NP}$ *having* 1-*Br.Pr. size not smaller than* $2^n/n^{O(\log n)}$*.*

For the more general case of *k-Br.Pr.*'s, we introduce a new method to derive explicit, exponential lower bounds that uses concepts introduced by Borodin *et al.* Our method relies on the following result. Given any Boolean function $f : \{0,1\}^n \to \{0,1\}$, let $N_f^1 = \{\vec{x} \in \{0,1\}^n \ : f(\vec{x}) = 1\}$ and $L_{k-br}(f)$ be the size of smallest *k-Br.Pr.* that computes f.

Theorem 1.4 *Let* $t \geq \log^2 n$, $k = o(\log n / \log t)$ *and let* $f(x_1, \ldots, x_n)$ *be a boolean function such that* $|N_f^1| \geq 2^{n-1}$ *and* $L_{k-br}(f) \leq 2^{n^{1-\epsilon}}$ *for some constant* $0 < \epsilon < 1$*. Then, for sufficiently large* n*, the set* N_f^1 *contains an affine space of dimension not smaller than* $t/4$*.*

We then consider the complement function $F_n^4(x_1 \ldots x_n)$ of the characteristic function of the hitting set $\mathcal{H}(n,m)$ in Theorem 1.2 (for an appropriate choice of $\log n \leq m < n/2$). By the definition of hitting sets, it should be clear that F_n^4 cannot contain any affine subspace of dimension not smaller than m. We will thus show the following lower bound that improves over those in [8] and [12] for some range of k.

Theorem 1.5 *Let* $k = o(\frac{\log n}{\log \log n})$*. There exists a Boolean function* $F^4 = \{F_n^4 : \{0,1\}^n \to \{0,1\}, n > 0\}$ *that belongs to* PH *and, for sufficiently large* n*,* $L_{k-br}(F_n^4) \geq 2^{n^{1-o(1)}}$*.*

1.1 Preliminaries

Given a subset $W \subseteq \{0,1\}^n$, its size is denoted by $|W|$, and its probability with respect to the uniform distribution in $\{0,1\}^n$ is denoted by $\mathbf{Pr}(W)$; the notation $\mathbf{Pr}_S(W)$ instead refers to case in which the uniform distribution is defined on the sample space S.
A Boolean function $f : \{0,1\}^N \to \{0,1\}$ is said *linear* if a vector $\vec{a} = (a_1, \ldots, a_N) \in \{0,1\}^N$ exists such that f can be written as $f(x_1, \ldots, x_N) = a_1 x_1 \oplus \ldots \oplus a_N x_N$. Given two binary vectors $\vec{a} = (a_1, \ldots, a_N)$ and $\vec{b} = (b_1, \ldots, b_N)$ from $\{0,1\}^N$, we define $\vec{c} = \vec{a} \oplus \vec{b}$ as the vector whose the i-th component is given by the xor of the i-th components of $\vec{a}$ and $\vec{b}$. The *inner* product is defined as $< \vec{a}, \vec{b} > = a_1 b_1 \oplus \ldots \oplus a_N b_N$.
The finite field $GF(2^N)$ will be represented by the standard one-to-one mapping bin : $GF(2^N) \to \{0,1\}^N$ such that $\text{bin}(a+b) = \text{bin}(a) \oplus \text{bin}(b)$ and $\text{bin}(0) = (0, \ldots, 0)$.
Furthermore, given any element $a \in GF(2^N)$ and an integer $k > 0$, we will make use of the concatenation of powers $U(a,k) \in \{0,1\}^{kN}$ where $U(a,k) = \text{bin}(a^0)\text{bin}(a^1)\ldots\text{bin}(\vec{a}^{k-1})$. For the sake of brevity, the improper notation $\vec{a}^j$ will replace the term $\text{bin}(\vec{a}^j)$.

2 The Reduction Lemma

For any subset $S \subseteq \{0,1\}^N$ and for any $\vec{\alpha} \in \{0,1\}^N$ we define

$$\varepsilon(S,\vec{\alpha}) = |\ 1/2\ -\ (\ \sum_{\vec{x}\in S} < \vec{\alpha}, \vec{x} > /|S|\ )\ |$$

and the "discrepancy" degree of S as $\varepsilon(S) = \max\{\varepsilon(S,\vec{\alpha})\ :\ \vec{\alpha} \in \{0,1\}^N,\ \vec{\alpha} \neq \vec{0}\}$. Note that $S \subseteq \{0,1\}^N$ is ϵ-discrepant for $\mathcal{AFF}(N,N,1)$ iff $\varepsilon(S) \leq \epsilon$. We now introduce the discrepancy degree w.r.t. systems of linear functions. For any choice of $\vec{\sigma} \in \{0,1\}^s$ and $\vec{\alpha}_i \in \{0,1\}^N$ for $i = 1, \ldots s$, we define

$$\pi(S,\vec{\alpha}_1,\ldots,\vec{\alpha}_s,\sigma_1,\ldots,\sigma_s) = \frac{|\{\vec{x} \in S\ :\ < \vec{\alpha}_i, \vec{x} >= \sigma_i; i = 1,\ldots,s\}|}{|S|}$$

and we let $\varepsilon_s(S)$ be the maximum of the difference $|2^{-s} - \pi(S,\vec{\alpha}_1,\ldots,\vec{\alpha}_s,\sigma_1,\ldots,\sigma_s)|$, over all possible linearly independent s-ple of vectors $\vec{\alpha}_i$'s and vectors $\vec{\sigma} \in \{0,1\}^s$.
Our interest in the above definition relies on the fact that a set S is ϵ-discrepant for $\mathcal{AFF}(N,N,s)$ iff $\varepsilon_s(S) \leq \epsilon$. We can now present the Reduction Lemma according to the above definitions (for the proof see Section A.1).

Lemma 2.1 (The Reduction Lemma) *Let $S \subseteq \{0,1\}^N$. If $\vec{\alpha}_1,\ldots,\vec{\alpha}_s$ are linearly independent and $\vec{\sigma} \in \{0,1\}^s$, then*

$$|2^{-s} - \pi(S,\vec{\alpha}_1,\ldots,\vec{\alpha}_s,\sigma_1,\ldots,\sigma_s)| \leq \left(2 - \frac{1}{2^{s-1}}\right)\varepsilon(S).$$

Hence if S is ϵ-discrepant w.r.t. $\mathcal{AFF}(N,N,1)$ then it is 2ϵ-discrepant w.r.t. $\mathcal{AFF}(N,N,s)$, for any $s \leq N$.

2.1 Applying the Reduction Lemma to Alon *et al*'s Powering Construction

In this section we first describe the powering construction of the ϵ-discrepancy set $\mathcal{L}^*_{N,z} \subseteq \{0,1\}^N$ given by Alon *et al* in [1], and we then combine it with our Reduction Lemma in order to get a good discrepancy set for the general class $\mathcal{AFF}(N,N,N)$.

Definition 2.1 (The Powering Sample Space [1]) *The generic vector $\vec{l}$ in the sample space $\mathcal{L}^*_{N,z}$ is specified by two vectors $\vec{x},\vec{y} \in \{0,1\}^z$. The i-th bit of $\vec{l}$ is the inner product of the i-th power of $\vec{x}$ and y. Clearly, we have that $|\mathcal{L}^*_{N,z}|\ =\ 2^{2z}$.*

Theorem 2.1 ([1]) *For any $N > 0$ and $z \leq N$, it holds $\varepsilon(\mathcal{L}^*_{N,z}) \leq \frac{N}{2^{z+1}}$. Hence $\mathcal{L}^*_{N,z}$ is $\frac{N}{2^{z+1}}$-discrepant for $\mathcal{AFF}(N,N,1)$.*

Notice also that Alon *et al*'s ϵ-discrepancy set for linear functions is the set $\mathcal{L}^*_{N,z}$ for $z = \log(N/\epsilon)$.
By applying the Reduction Lemma to Theorem 2.1 we can easily prove the following

Corollary 2.1 *If $\vec{\alpha}_1,\ldots,\vec{\alpha}_s \in \{0,1\}^N$ are linearly independent then, for any $\vec{\sigma} \in \{0,1\}^s$,*

$$\left|2^{-s} - \pi(\mathcal{L}^*_{N,z},\vec{\alpha}_1,\ldots,\vec{\alpha}_s,\sigma_1,\ldots,\sigma_s)\right| \leq N \cdot 2^{-z}\ .$$

Hence $\mathcal{L}^*_{N,z}$ is $(N2^{-z})$-discrepant for $\mathcal{AFF}(N,N,N)$.

The above combination of the Alon *et al*'s powering construction and our Reduction Lemma can be slightly modified in order to obtain a discrepancy set for $\mathcal{AFF}(N,k,s)$ that takes in consideration the number k of essential variables (this construction is described in Section A.2).

Theorem 2.2 (Discrepancy set for AFF(N,k,N)) *For any* $N > 0$*, and* $0 \leq k, z \leq N$*, the set* $\mathcal{D}_2(N,k,z)$ *is* $(k\lceil \log N \rceil 2^{-z})$*-discrepant for* $\mathcal{AFF}(N,k,N)$*.*

3 Hitting Sets for Small Affine Spaces

In this section we give an efficient construction of non trivial hitting sets for small affine spaces, i.e., for large systems of linear equations. The key idea of the construction is the fact that any affine subspace $\mathcal{S}$ can be represented by using its corresponding orthogonal subspace $\mathcal{S}^{\perp}$ in $\{0,1\}^n$, and if $\mathcal{S}$ has small dimension then $\mathcal{S}^{\perp}$ has a large dimension. The orthogonal subspace is thus described by a small linear system and so we can apply the powering construction of the previous section[1]. Our first main task will thus be the construction of a different representation of small affine spaces using their respective orthogonal subspaces.

If $A \subseteq \{0,1\}^n$ then $\text{rank}(A)$ denotes the maximal number of linear independent vectors (A will be also considered as a Boolean matrix). Given $\vec{a} \in \{0,1\}^n$, the term $[\vec{a}]^j$ denotes the the first j bits of $\vec{a}$. Furthermore, we will use the notation $[A]^j = \{[\vec{a}]^j \ : \ \vec{a} \in A\}$. and define $\text{rank}_j(A) = \text{rank}([A]^j)$.
The small affine space that we want to hit is described by the following linear system

$$\{< \vec{a}_i, \vec{x} > = \ b_i \ , \qquad i = 1,2,\ldots,n-m \ , \tag{1}$$

where $\vec{a}_i \in \{0,1\}^n$, $i = 1,2,...,n-m$ and $b_i \in \{0,1\}$. We let $A = \{\vec{a}_1,\ldots,\vec{a}_{n-m}\}$ and assume that $\text{rank}(A) = n-m$. As stated before we are interested in the case in which m is small, i.e., $0 \leq m < n/2$. Our first step is to partition the matrix A into $s \geq 16$ smaller submatrices having almost equal rank. Let $\mathcal{N}(n,s)$ be the set of vectors $\vec{N} = (n_1,\ldots,n_s)$ with positive integer components such that $n_1 + \ldots + n_s = n$. Let $k = m/s$ and *wlog* assume that $k \geq 2$. Then, for sufficiently large n, given any $A = \{\vec{a}_1,\ldots,\vec{a}_{n-m}\}$, for any fixed $s \geq 16$, we can always choose a vector $\vec{N} = (n_1,n_2,\ldots,n_s) \in \mathcal{N}(n,s)$ such that

$$\text{rank}_{n_1+\ldots+n_i}(A) - \text{rank}_{n_1+\ldots+n_{i-1}}(A) = n_i - m_i \text{ and } \ k \leq m_i \leq k+1 \ , \ \text{ for any } i = 1,\ldots,s. \tag{2}$$

Note that $m_1 + m_2 + \ldots + m_s = m$. Let $\vec{x} = (\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_s)$, where the length of $\vec{x}_i$ is n_i. Then, by using a standart diagonalization method for linear systems, we can represent the solution space of the system in Eq.(1) with a set of smaller linear systems:

$$\Big\{< \vec{b}_1, \vec{x}_1 > \ = \ u_1^{\vec{b}_1}, \ \forall \vec{b}_1 \in B_1 \ \ldots \ \Big\{< \vec{b}_s, \vec{x}_s > \ = u_s^{\vec{b}_s} \ \forall \ \vec{b}_s \in B_s$$

for some $B_i \subseteq \{0,1\}^{n_i}$ such that $|B_i| = n_i - m_i$ and $\text{rank}(B_i) = n_i - m_i$, (i=1,...,s) and for some $u_i^{\vec{b}_i} \in \{0,1\}$, $i = 1 \ldots s$ (clearly, both B_i's and $u_i^{\vec{b}_i}$'s depend on the initial system in Eq. (1).

Let $t = \lceil \log s \rceil + 1$ and consider any subset $D = \{\vec{d}_1, \ldots, \vec{d}_{k-t}\}$ of vectors from $\{0,1\}^n$. Then the orthogonal space (w.r.t. $[D]^j$) $\mathcal{S}(D,j)$ is the set of solutions of system $< [\vec{d}_i]^j, (x_1, x_2, \ldots, x_j) >= 0$ for $i = 1, \ldots, k-t$. For any $N \in \mathcal{N}(n,s)$, we then consider the sets of prefix combinations from $\mathcal{S}(D,n_j)$'s,

[1] This new representation of small affine spaces does not allow to preserve the discrepancy property of our sample space but only their hitting property.

i.e., $\mathcal{S}(D,\vec{N}) = \mathcal{S}(D,n_1) \times \ldots \mathcal{S}(D,n_s)$. The next result provides the orthogonal condition that a set of vectors D must satisfy in order to let $\mathcal{S}(D,\vec{N})$ contain at least one solution of the linear system in Eq. (1) (for a formal proof see Section A.3.1).

Lemma 3.1 *Let $D = \{\vec{d_1},\ldots,\vec{d_{k-t}}\} \subseteq \{0,1\}^n$ and $N \in \mathcal{N}(n,s)$. If for any $i = 1,2,\ldots,s$ $\mathrm{rank}(B_i \cup [D]^{n_i}) = n_i - m_i + k - t$, then $\mathcal{S}(D,\vec{N})$ contains at least one solution of the system in Eq. (1).*

Lemma 3.1 suggests us the way to construct a hitting set for the system in Eq. (1): it suffices to find a vector set D that satisfies the "orthogonal" condition in Lemma 3.1 for any possible choices of B_i's, $i = 1,\ldots,s$. It is not hard to see that a random subset D from $\{0,1\}^n$ of size $k-t$ satisfies, with high probability, the above condition. Our idea is to de-randomize the process by choosing our hitting set as the union of all subsets of size $k-t$ from the powering discrepancy set $\mathcal{L}^*_{n,(\lceil\log n\rceil+t)}$; given any $16 \le s \le n$, for any subset $\vec{N} \in \mathcal{N}(n,s)$ we let

$$\mathcal{H}(\vec{N}) = \bigcup_{D \subseteq \mathcal{L}^*_{n,(\lceil\log n\rceil+t)} \;:\; |D|=k-t} \mathcal{S}(D,\vec{N}) \ ,$$

and we define our hitting set $\mathcal{H}(n,s)$ as the union of $\mathcal{H}(\vec{N})$ over all possible choices of $\vec{N}$ from $\mathcal{N}(n,s)$.

The hitting property of $\mathcal{H}(n,s)$ is a consequence of the fact that, from Corollary 2.1, $\mathcal{L}^*_{n,(\lceil\log n\rceil+t)}$ contains at least one solution of any linear systems in $\mathcal{AFF}(n,n,o(n))$. This implies the following result (for a proof see Section A.4).

Lemma 3.2 *Let $t = \lceil\log s\rceil + 1$, $\vec{N} = (n_1,\ldots,n_s) \in \mathcal{N}(n,s)$. Let $B_1,\ldots,B_s$ be such that for any $i = 1,\ldots,s$ $B_i \subseteq \{0,1\}^{n_i}$ with $|B_i| = n_i - m_i$ where $k \le m_i \le k+1$, and $\mathrm{rank}(B_i) = n_i - m_i$. Then a vector subset $D \subseteq \mathcal{L}^*_{n,(\lceil\log n\rceil+t)}$ exists such that $|D| = k-t$ and $\mathrm{rank}(B_i \cup [D]^{n_i}) = n_i - m_i + k - t$, for any $i = 1,\ldots,s$.*

We have now at hand all the ingredients to derive a hitting set for small affine subspaces.

Theorem 3.1 *For any $1 \le m \le n$, for any $16 \le s \le n$, the set $\mathcal{H}(n,s) \subseteq \{0,1\}^n$ is a hitting set for $\mathcal{AFF}(n,n,n-m)$. Furthermore, If $s = \Theta(\sqrt{m})$ then*

$$|\mathcal{H}(n,s)| \le 2^{n-m+m\frac{O(1)\log n}{\sqrt{m}}} \ .$$

Proof. Consider the system in Eq. (1). As before mentioned, for any $s \ge 16$ (and in particular for $s = \lceil\sqrt{m}\rceil$), we can choose vector $\vec{N} \in \mathcal{N}(n,s)$, , that satisfies Condition (2). Lemma 3.2 then implies that a subset D from $\mathcal{L}^*_{n,(\lceil\log n\rceil+t)}$ exists that satisfies the orthogonal condition in Lemma 3.1. This lemma implies that $\mathcal{S}(D,\vec{N})$ contains at least one solutions of the system in Eq. (1). The first claim of the theorem thus follows from the fact that $\mathcal{S}(D,\vec{N}) \subseteq \mathcal{H}(n,s)$.

Concerning the size of the hitting set, assume that $s = \lceil\sqrt{m}\rceil$; then by definition of k and t we have (here we omit some standard calculations)

$$|\mathcal{H}(\vec{N})| \ \le \ \left(2^{2(\lceil\log n\rceil+t)}\right)^s \prod_{i=1}^{s} 2^{n_i-(k-t)} \ \le \ 2^{n-m+m\frac{20\log n}{\sqrt{m}}} \ .$$

It follows that

$$|\mathcal{H}(n,s)| \leq \left(\begin{array}{c} n \\ s-1 \end{array} \right) 2^{n-m+m\frac{20\log n}{\sqrt{m}}} \leq 2^{n-m+m\frac{20\log n}{\sqrt{m}+s\log n}} \leq 2^{n-m+m\frac{22\log n}{\sqrt{m}}} \ .$$

□

Finally, we remark that the set $\mathcal{H}(n,\lceil\sqrt(m)\rceil)$ turns out to be a non trivial hitting set (i.e. of size asymptotically smaller than 2^n) for $\mathcal{AFF}(n,n,n-m)$ for any $m \geq \frac{1}{o(1)}\log^2 n$.

4 New Explicit Lower Bounds for Branching Programs

4.1 Lower Bound for 1-Read Branching Programs

We adopt notations and terminolgy introduced in [22]. In particular, for a *partial* input we mean any element from $\{0,1,*\}$ where the positions containing 0 or 1 mean that the corresponding input bit is fixed, while the notation $*$ mean that the corresponding input bit is free. We say that a partial input v is defined on the set $I \subseteq \{0,1\}^n$ if $v_i \in \{0,1\}$ iff $i \in I$. Partial inputs define subfunctions of a given Boolean functions $f : \{0,1\}^n \to \{0,1\}$ in the following natural way. For any set $I \subseteq \{1,2,...,n\}$, let $\mathcal{B}(I)$ be the set of all partial inputs defined on I. Given any partial input $v \in \mathcal{B}(I)$, the subfunction $f|_v$ of $f(x_1,\ldots,x_n)$ is obtained by setting $x_i = v_i$ for any $i \in I$. The set of inputs on which $f|_v$ is defined consists of the Boolean rectangle $R(v)$ of dimension $n-|I|$, $R(v) = \{(a_1,\ldots,a_n) : \ a_i = v(i), \ i \in I\}$. Given any function $f : \{0,1\}^n \to \{0,1\}$ and any subset $I \subseteq \{1,2,...,n\}$, let

$$\nu(f,I) \ = \ \max_{v\in\mathcal{B}(I)} |\{u \in \mathcal{B}(I) \ : \ f|_u = f|_v\}| \ .$$

In the case of 1-*br.pr.*'s, Simon and Szegedy introduced a nice technique to derive explicit lower bounds that enjoys of the following theorem.

Theorem 4.1 ([23]) *Let f be a Boolean variables of n variables and let $r \leq n$. The size of any* 1-*Br.Pr. computing f is at least* $2^{n-r}/(\max\{\nu(f,I) : \ |I| = n-r\})$.

Our next goal is to give a suitable interpretation of this theorem for the family of n-inputs Boolean functions $\mathsf{powF} = \{f^{\vec{\alpha}}_{n,k} : \vec{\alpha} \in \{0,1\}^{kn}\}$ where $f^{\vec{\alpha}}_{n,k}(\vec{x})$ is the inner product between $\vec{\alpha}$ and the concatenation of the first k powers of $\vec{x}$, i.e. $f^{\vec{\alpha}}_{n,k}(\vec{x}) =< \vec{\alpha}, U(\vec{x},k) >$. Let $I \subseteq \{1,2,...,n\}$ such that $|I| = n-r$; for any $u,v \in \mathcal{B}(I)$ we consider the *xor* vector $\vec{df}(u,v) \in \{0,1\}^n$ where $\vec{df}(u,v)_i = 1$ iff $i \in I$ and $v_i \neq u_i$. Then the condition (implicitly required by Theorem 4.1)

$$f^{\vec{\alpha}}_{n,k}|_v \ = \ f^{\vec{\alpha}}_{n,k}|_u \ . \tag{3}$$

is equivalent to require that for all $\vec{x} \in R(v)$, $f^{\vec{\alpha}}_{n,k}(\vec{x}) \oplus f^{\vec{\alpha}}_{n,k}(\vec{x} \oplus \vec{df}(v,u)) = 0$. Therefore, by definition of $f^{\vec{\alpha}}_{n,k}$, it is equivalent to require that vector $\vec{\alpha} \in \{0,1\}^{kn}$ is a solution of the following Boolean system of $|R(v)| = 2^r$ linear equations

$$< \vec{\alpha}, (U(\vec{x},k) \oplus \{U(\vec{x} \oplus df(v,u),k)) >= 0 \ , \qquad \vec{x} \in R(v) \ . \tag{4}$$

So, in order to obtain a lower bound as large as possible, we need to minimize the number of α's that satisfy the above system. This is achieved by choosing r and k in order to make the equations linearly independent. If indeed we set $r = \lceil \log n \rceil + 1$, and $k = 2^{r+1}$ then, by definitions of $U(\vec{x}, k)$ and $\vec{df}(v, u)$, vectors $U(\vec{x}, k) \oplus U(\vec{x} \oplus df(v, u), k)$ are linearly independent for any $\vec{x} \in R(v)$ and for any different $v, u \in \mathcal{B}(I)$. This implies that for any fixed choice of u and v such that $u \neq v$, the number of $\vec{\alpha}$ satisfying the linear system is $2^{kn}2^{-h}$ where $h = 2^r$. By applying standard counting arguments, it would be possible to show that a vector $\vec{\alpha}$ selected uniformly at random from $\{0,1\}^{kn}$ will not satisfy the linear system in Eq. (4) for any I such that $|I| = n - r$ and for any different $u, v \in \mathcal{B}(I)$ with high probability. This however will not give an explicit and efficient construction of a Boolean function having asymptotically maximal 1-*Br.Pr.* complexity. The key idea is to use the sample space $\mathcal{L}^*_{nk,t}$ that can efficiently de-randomize the probabilistic construction described above. Indeed, let $t = 2^r$ and let $G(y_1, \ldots, y_{2t})$ be the Boolean generator of the set $\mathcal{L}^*_{nk,t}$ described in Section 2.1. By applying Corollary 2.1 with $N = kn$, $s = 2^r$, and $z = t$ we easily have that the system in Eq. (4) for $\vec{\alpha} = G(\vec{\beta})$ is satisfied for not more than $2^{2t}(2^{-2^r} + 2^{r+1}n2^{-t})$ different β's. It follows that the number of β's for which the system is satisfied for some (at least one) I, with $|I| = n - r$, and some (at least one pair) different $v, u \in \mathcal{B}(I)$ is at most

$$2^{2t}(2^{-2^r} + 2^{r+1}n2^{-t}) \begin{pmatrix} n \\ r \end{pmatrix} 2^{n-r} \;\leq\; 2^{2t}2^{\log n+3}n2^{-2n}n^{\log n+2}2^n \;\leq\; 2^{2t}o(1) \;.$$

Consequently, an element $\beta_0 \in \{0,1\}^{2t}$ exists such that $\vec{\alpha}_0 = G(\vec{\beta}_0)$ does not satisfy System (4) for any I with $|I| = n - r$, and for any $v \neq u$ from $\mathcal{B}(I)$. By applying the above determinististc construction and Theorem 4.1, we easily prove the following result.

Theorem 4.2 *There exists a function $F^1 = \{F^1_n : \{0,1\}^n \to \{0,1\}, n > 0\}$ that belongs to $\mathsf{NP}^{\mathsf{NP}} \cap \mathsf{P/poly}$ and such that, for almost every $n > 0$, $L_{1-br}(F^1_n) \geq \frac{2^n}{4n}$.*

Proof. For any $n > 0$, we choose with a non-determistic algorithm a vector $\vec{\beta} \in \{0,1\}^{2t}$ such that condition (3) does not hold for any I with $|I| = n - r$ and for any $v \neq u$ from $\mathcal{B}(I)$. Then we let $F^1_n = F^{G(\beta)}_{n,k}$; it is easy to verify that the function can be computed with circuits of polynomial size provided that the correct β is given and also that $F^1 \in \mathsf{NP}^{\mathsf{NP}}$. Finally, by applying Theorem 4.1, we have that $L_{1-br}(F^1_n) \geq 2^{n-r} \geq 2^n/4n$. □

Using the same de-randomization technique based on sampling from the discrepancy set $\mathcal{L}^*_{nk,t}$, it is possible to show the other explicit lower bounds stated in Theorem 1.3 (a complete proof will be given in the full version of this paper).

4.2 A Lower Bound for k-Branching Programs

In this section, we make use of definitions and results from [8].

Definition 4.1 ([8]) *A Boolean function $g(x_1, \ldots, x_n)$ is a (k,a)-rectangle if g can be represented in the form*

$$g \;=\; \bigwedge_{i=1}^{ka} g_i(X_i)$$

where g_i is a Boolean function depending only on variables from $X_i \subseteq \{x_1, \ldots, x_n\}$, $|X_i| \leq \lceil n/a \rceil$ and each variable belongs to at most k of the sets $\{X_1, \ldots, X_{ka}\}$.

Let $L_{k-br}(f)$ be the size of the smallest (nondeterministic syntactic) *k-Br.Pr.* that computes f.

Lemma 4.1 *[8] Let $f(x_1 \dots, x_n)$ be a Boolean function such that $L_{k-br}(f) \leq M$ and let a and k be positive integers. Then f is an OR of at most $(2M)^{2ka}$ (k,a)-rectangles.*

The next lemma provides a method to represent any (k,a)-rectangle (with any $k > 1$) as an OR of a certain number of $(\mathbf{1}, a(k))$-rectangles where $a(k)$ is exponential in k (the proof is given in Section B).

Lemma 4.2 *Let $g(x_1, \dots, x_n)$ be a (k,a)-rectangle and assume that for some $t \geq 2$ it holds*

$$(4t^{k+1})/n \;+\; (4t^{2k+1}k)/a \;<\; 1 \;.$$

Then g is an OR of at most $2^{n-n/(2t^{k-1})}$ $(1, 2 \cdot t^k)$-rectangles.

Our interest in the above lemma is given by the following connection between (1,T)-rectangles and Boolean affine subspaces (the proof is given in Section C). Let N_g^1 be the set of $\vec{x}$ such that $g(\vec{x}) = 1$.

Lemma 4.3 *If $g(x_1, \dots, x_n)$ is a $(1,T)$-rectangle then it contains an affine space of dimension at least $(|N_g^1|)/((n/T)+1)$.*

By combining the above three lemmas it is not hard to derive a new connetction between the *k-Br.Pr.* size of a given function $f : \{0,1\}^n \to \{0,1\}$ and its "behaviour" w.r.t. to the class of affine spaces in $\{0,1\}^n$.

Theorem 4.3 *Let $t \geq \log^2 n$ and $k = o\left(\frac{\log n}{\log t}\right)$. Let $f : \{0,1\}^n \to \{0,1\}$ be such that $|N_f^1| \geq 2^{n-1}$ and $L_{k-br}(f) = M \leq 2^{n^{1-\epsilon}}$ for some constant $0 < \epsilon < 1$. Then, for sufficiently large n, the subset N_f^1 contains a (at least one) Boolean affine subspace in $\{0,1\}^n$ of dimension at least $t/4$.*

Sketch of the proof. If we choose $a = 16kt^{2k+1}$ then, for sufficiently large n, we obtain

$$(4t^{k+1})/n \;+\; (4t^{2k+1}k)/a \;\leq\; o(n) + 1/4 < 1 \;.$$

By combining Lemma 4.1 and Lemma 4.2, we have that f is an OR of at most r $(1, 2 \cdot t^k)$-rectangles, where $r \leq (2M)^{2ka} 2^{n-n/(2t^{k-1})} \leq 2^{n-n/(4t^{k-1})}$. Since $|N_f^1| \geq 2^{n-1}$, the above inequality implies that N_f^1 contains at least one $(1, 2 \cdot t^k)$-rectangle of size not smaller than $(2^{n-1})/(2^{n-n/(4t^{k-1})}) = 2^{n/(4t^{k-1})-1}$. From Lemma 4.3 this rectangle contains an affine space of dimension at least

$$\frac{\log\left(2^{n/(4t^{k-1})-1}\right)}{(n/(2t^k))+1} \;=\; \frac{n/(4t^{k-1})-1}{(n/(2t^k))+1} \;\geq\; \frac{1}{2}\frac{n/(4t^{k-1})}{(n/(2t^k))} \;=\; \frac{t}{4} \;.$$

□

For any $n > 0$, we define the Boolean function $F_n^4(x_1, \dots, x_n)$ as the complement of the characteristic function of the hitting set $\mathcal{H}(n, \sqrt{m})$ given by Theorem 3.1 where we set $m = \lceil \log^4 n \rceil$. From the construction of $\mathcal{H}(n,s)$ shown in Section 3, it is not hard to show that the family $F^4 = \{F_n^4,\ n > 0\}$ belongs to PH. Furthermore, from Theorem 3.1 we have that $|\mathcal{H}(n, O(\log^2 n))| = o(2^n)$, so, by applying Theorem 4.3, we obtain the following lower bound.

Corollary 4.1 *If $k = o(\frac{\log n}{\log\log n})$ then, for sufficiently large n, $L_{k-br}(F_n^4) \geq 2^{n^{1-o(1)}}$.*

Acknowledgements. We would like to thank Oded Goldreich and Luca Trevisan for several comments on a preliminary version of this paper. We also thank Sasha Razborov and Martin Sauerhoff for helpful pointers.

References

[1] N. Alon, O. Goldreich, J. Hastad, and R. Peralta (1990), "Simple Constructions of Almost k-wise Independent Random Variables", *Proc. of IEEE-FOCS*, Vol. 2, pp. 544-553.

[2] Andreev A., Clementi A., and Rolim J. (1996), "A New General De-randozation Method", *J.ACM* to appear (available at the *J.ACM* Web site: `http://www.acm.org/jacm`).

[3] Andreev A., Clementi A., and Rolim J. (1997), "Worst-case Hardness Suffices for Derandomization: a New method for Hardness-Randomness Trade-Offs", in Proc. of *ICALP*, LNCS, 1256, pp. 177-187.

[4] A. Andreev, A. Clementi, J. Rolim, and L. Trevisan (1997), "Weak Random Sources, Hitting Sets, and BPP Simulations" Proc. of *IEEE FOCS*, pp. 264-273. (also in *ECCC*, TR97-011).

[5] R. Armoni, M. Saks, A. Wigderson, and S. Zhou (1996 "Discrepancy Sets and Pseudorandom Generators for Combinatorial Rectangles", Proc. of *IEEE-FOCS*, pp. 412-421.

[6] Blum M., and Micali S. (1984), "How to generate cryptographically strong sequences of pseudorandom bits", *SIAM J. of Computing*, 13(4), pp. 850-864.

[7] R. Ben-Natan (1990), "On independent random variables over small sample spaces", *M.Sc. Thesis*, Computer Science Dept., Hebrew University, Jerusalem, Israel, Feb.

[8] A.Borodin, A.Razborov and R.Smolensky (1993), *On lower bounds for read-k times branching programs*, *Computational Complexity*, 3, pp. 1-18.

[9] Bshouty N. H. (1989), "On the Extended Direct Sum Problem Conjecture", Proc. of *STOC*, pp. 177-185.

[10] G. Even, O. Goldreich, M. Luby, N. Nisan and B. Velickovic (1992), "Approximations of general independent distributions", Proc. of *ACM-STOC*, pp 10-16.

[11] R. Impagliazzo, and A. Wigderson (1997), "P= BPP if E requires exponential circuits: Derandomizing the XOR lemma" Proc. of *ACM STOC*, pp. 220-229.

[12] S. Jukna (1995), "A Note on Read-k-times Branching programs", *RAIRO Theoretical Informatics and Applications*, 29 (1), pp.75-83. (also in *ECCC TR94-027*).

[13] S. Jukna, A. Razborov, P. Savicky, I. Wegener (1997), "On P versus $NP \cap co-NP$ for Decision Trees and Read-Once Branching Programs", *ECCC*, TR97-023.

[14] J. Justesen (1972), "A Class of Constructive Asymptoticaly Good Algebraic Codes", *IEEE Transactions on Information Theory,* 18, pp. 652–656.

[15] D. Karger and D. Koller (1994), "(De)randomized construction of small sample spaces in NC", Proc. of *IEEE-FOCS*, pp. 252-263.

[16] Karchmer M., Raz R., and Wigderson A. (1991), "On Proving Super-Logarithmic Depth Lower Bounds via the Direct Sum in Communication Complexity", Proc. of *IEEE Structure in Complexity Theory*, pp. 299-304.

[17] J.Naor, and M.Naor (1990), "Small-bias probability spaces: efficient constructions and aplications", Proc. of *ACM STOC*, pp. 213-223.

[18] N.Nisan (1990), "Pseudo-random generators for Space-Bounded Computation", Proc. of *ACM STOC*, pp. 204-212.

[19] Nisan N., and Wigderson A. (1994), "Hardness vs Randomness", *J. Comput. System Sci.* 49, pp. 149-167.

[20] E.A. Okolnishnikova (1993), "On lower bounds for branching programs", *SiberianAdvances in Mathematics*, 3(1), pp. 152-166.

[21] M. Sauerhoff (1997), "A Lower Bound for Randomized Read-k-Times Branching Programs", *ECCC*, TR97-019.

[22] P. Savicky and S. Zak (1996), "A large lower bound for 1-branching programs", *ECCC*, TR96-036.

[23] J. Simon and M. Szegedy (1993), "A new lower bound theorem for read only once branching programs and its aplplications", *Advances in Computational Complexity Theory (J.Cai, editor)*, DIMACS Series, 13, AMS, pp.183-193.

[24] U. Vazirani (1986), "Randomness, Adversaries and Computation", *PhD Thesis*, EECS, UC Berkeley.

[25] I. Wegener (1988), "On the Complexity of Branching Programs and Decision Trees for Clique Functions, *J.ACM*, 35, pp. 461-477.

A Proofs

A.1 Proof of the Reduction Lemma

Fix $S \subseteq \{0,1\}^N$ and a set of s linearly independent vectors $\vec{\alpha}_1, \ldots, \vec{\alpha}_s$ from $\{0,1\}^N$. For any $\vec{\sigma} = (\sigma_1, \ldots, \sigma_s)$ we define the "discrepancy degree" of S as $q(\vec{\sigma}) = \pi(S, \vec{\alpha}_1, \ldots, \vec{\alpha}_s, \sigma_1, \ldots, \sigma_s)$; for any $\vec{b} = (b_1, \ldots, b_s) \in \{0,1\}^s$, we define the vector $\vec{\alpha}(\vec{b}) \;=\; b_1\vec{\alpha}_1 \oplus \ldots \oplus b_s\vec{\alpha}_s$. For the sum

$$Q(\vec{b}) \;=\; \sum_{\vec{l} \in S} < \vec{l}, \vec{\alpha}(\vec{b}) > \;,$$

we have that

$$Q(\vec{b}) \;=\; \sum_{\vec{l} \in S} < \vec{l}, b_1\vec{\alpha}_1 \oplus \ldots \oplus b_s\vec{\alpha}_s > \;=\; \sum_{\vec{l} \in S} (b_1 < \vec{l}, \vec{\alpha}_1 > \oplus \ldots \oplus b_s < \vec{l}, \vec{\alpha}_s >) \;=$$

$$= \sum_{\vec{\gamma} = (\gamma_1, \ldots, \gamma_s) \in \{0,1\}^s} \left(\sum_{\vec{l} \in S \;:\; <\vec{l}, \vec{\alpha}_1> = \gamma_1, \ldots, <\vec{l}, \vec{\alpha}_s> = \gamma_s} < \vec{b}, \vec{\gamma} > \right) ;$$

so,

$$Q(\vec{b}) = \sum_{\vec{\gamma} \in \{0,1\}^s} q(\vec{\gamma}) < \vec{b}, \vec{\gamma} > \;=\; \sum_{\vec{\gamma} \in \{0,1\}^s \setminus \{\vec{0}\}} q(\vec{\gamma}) < \vec{b}, \vec{\gamma} > \;. \tag{5}$$

Two cases may arise depending on whether or not $\vec{\sigma} = \vec{0}$.

1.) For $\vec{\sigma} \neq \vec{0}$, Eq. (5) implies that

$$\sum_{\vec{b} \;:\; <\vec{b}, \vec{\sigma}> = 1} Q(\vec{b}) \;=\; \sum_{\vec{b} \;:\; <\vec{b}, \vec{\sigma}> = 1} \; \sum_{\vec{\gamma} \in \{0,1\}^s \setminus \{\vec{0}\}} q(\vec{\gamma}) < \vec{b}, \vec{\gamma} > \;=$$

$$= \sum_{\vec{\gamma} \in \{0,1\}^s \setminus \{\vec{0}\}} q(\vec{\gamma}) \sum_{\vec{b} \;:\; <\vec{b}, \vec{\sigma}> = 1} < \vec{b}, \vec{\gamma} > \;=\; q(\vec{\sigma})2^{s-1} + \sum_{\vec{\gamma} \in \{0,1\}^s \setminus \{\vec{0}, \vec{\sigma}\}} q(\vec{\gamma})2^{s-2} \tag{6}$$

where the last equality is due to the fact the the number of $\vec{b} \neq \vec{0}$ such that $< \vec{b}, \vec{\sigma} >= 1$ and $< \vec{b}, \vec{\gamma} >= 1$ for non zero $\vec{\sigma} \neq \vec{\gamma}$ is equal to 2^{s-2}. Similarly we have

$$\sum_{\vec{b} \;:\; <\vec{b}, \vec{\sigma}> = 0 \;,\; \vec{b} \neq \vec{0}} Q(\vec{b}) = \sum_{\vec{\gamma} \in \{0,1\}^s \setminus \{\vec{0}\}} q(\vec{\gamma}) \sum_{\vec{b} \;:\; <\vec{b}, \vec{\sigma}> = 0 \;,\; \vec{b} \neq \vec{0}} < \vec{b}, \vec{\gamma} >= \sum_{\vec{\gamma} \in \{0,1\}^s \setminus \{\vec{0}, \vec{\sigma}\}} q(\vec{\gamma})2^{s-2}. \tag{7}$$

Combining Eq.s (6) and (7), we get

$$q(\vec{\sigma})2^{s-1} \;=\; \sum_{\vec{b} \;:\; <\vec{b}, \vec{\sigma}> = 1} Q(\vec{b}) \;-\; \sum_{\vec{b} \;:\; <\vec{b}, \vec{\sigma}> = 0 \;,\; \vec{b} \neq \vec{0}} Q(\vec{b}) \tag{8}$$

By definition of $\varepsilon(S)$, for any $\vec{b} \neq \vec{0}$, it is not hard to prove that

$$|S| \left(\frac{1}{2} - \varepsilon(S) \right) \leq Q(\vec{b}) \leq |S| \left(\frac{1}{2} + \varepsilon(S) \right) \;. \tag{9}$$

Then from Eq.s (8) and (9), we have that

$$2^{s-1}|S|\left(\frac{1}{2}-\varepsilon(S)\right)-(2^{s-1}-1)|S|\left(\frac{1}{2}+\varepsilon(S)\right)\leq q(\vec{\sigma})2^{s-1}\ \leq$$

$$\leq\ 2^{s-1}|S|\left(\frac{1}{2}+\varepsilon(S)\right)-(2^{s-1}-1)|S|\left(\frac{1}{2}-\varepsilon(S)\right)$$

It follows that

$$2^{-s}-(2-2^{1-s})\varepsilon(S)\leq\frac{q(\vec{\sigma})}{|S|}\leq 2^{-s}+(2-2^{1-s})\varepsilon(S)\ .$$

2.) For $\vec{\sigma}=\vec{0}$, we can obtain the same bounds in the following way.

$$\sum_{\vec{b}\ :\ \vec{b}\neq\vec{0}}Q(\vec{b})=\sum_{\vec{\gamma}\in\{0,1\}^s\setminus\{\vec{0}\}}q(\vec{\gamma})\sum_{\vec{b}}<\vec{b},\vec{\gamma}>=\sum_{\vec{\gamma}\in\{0,1\}^s\setminus\{\vec{0}\}}q(\vec{\gamma})2^{s-1}\ =2^{s-1}(|S|-q(\vec{0})).$$

This implies that

$$\frac{q(\vec{0})}{|S|}\ =\ 1-\frac{1}{2^{s-1}|S|}\sum_{\vec{b}\ :\ \vec{b}\neq\vec{0}}Q(\vec{b})\ . \tag{10}$$

By using the same argument yielding Eq. (9), we get

$$(2^s-1)|S|\left(\frac{1}{2}-\varepsilon(S)\right)\ \leq\ \sum_{\vec{b}\ :\ \vec{b}\neq\vec{0}}Q(\vec{b})\ \leq\ (2^s-1)|S|\left(\frac{1}{2}+\varepsilon(S)\right). \tag{11}$$

From Eq. 10 and 11, we obtain

$$\frac{q(\vec{0})}{|S|}\leq 1-\frac{2^s-1}{2^{s-1}}\left(\frac{1}{2}-\varepsilon(S)\right)=1-(2-2^{1-s})\left(\frac{1}{2}-\varepsilon(S)\right)=2^{-s}+(2-2^{1-s})\varepsilon(S)$$

In the same way, we can derive the lower bound for $\frac{q(\vec{0})}{|S|}$:

$$\frac{q(\vec{0})}{|S|}\ \geq\ 2^{-s}-(2-2^{1-s})\varepsilon(S).$$

A.2 A Discrepancy Set for Linear Systems with a Limited Number of Essential Variables

The aim of this section is to modify the powering construction given in the previous section in order to obtain a discrepancy set for $\mathcal{AFF}(N,k,s)$ that takes in consideration the number k of essential variables.

Let $n=\log N$ and $1\leq z\leq n$. Consider the N elements $\vec{a}_1,\ldots,\vec{a}_N$ of $GF(2^n)$. For any $\vec{a}\in GF(2^n)$, we define the concatenation of powers $U(\vec{a},k)=\vec{a}^0\vec{a}^1\ldots\vec{a}^{k-1}$ and, given a vector $\vec{l}\in\mathcal{L}^*_{kn,z}$, we construct the following vector of N bits

$$(<\vec{l},U(\vec{a}_1,k)>,<\vec{l},U(\vec{a}_2,k)>,\ldots,<\vec{l},U(\vec{a}_N,k)>)$$

The new powering discrepancy set is then defined as

$$\mathcal{D}(N,k,z) \;=\; \bigcup_{\vec{l}\in\mathcal{L}^*_{kn,z}} \{(<\vec{l},U(\vec{a}_1,k)>,<\vec{l},U(\vec{a}_2,k)>,\ldots,<\vec{l},U(\vec{a}_N,k)>)\} \;.$$

Clearly, we have $|\mathcal{D}(N,k,z)| = |\mathcal{L}^*_{kn,z}| = 2^{2z}$. Informally speaking, the discrepancy property of $D(N,k,z)$ is a consequence of the fact that given any k pairwise different $\vec{a}_i$, $i = 1,\ldots,k$ from $GF(2^n)$, the corresponding k vectors $U(\vec{a}_i,k)$ in $GF(2^{nk})$ are linearly independent

Theorem A.1 (Discrepancy set for AFF(N,k,N)) *For any $N > 0$, and $0 \le k,z \le N$, the set $\mathcal{D}_2(N,k,z)$ is $(k\lceil\log N\rceil 2^{-z})$-discrepant for $\mathcal{AFF}(N,k,N)$.*

Proof. Consider the set W of all solutions of a fixed linear system $<\vec{l_1},\vec{x}>= \beta_1,\ldots,<\vec{l_s},\vec{x}>= \beta_s$ where vectors $\vec{l}_1,\ldots,\vec{l}_s$ from $\{0,1\}^N$ are linearly independent. Assume also that the number of essential variables is at most k, i.e., the number of 1's in the *or* vector $\vec{l}_1 \vee \ldots \vee \vec{l}_s$ is at most k. Consider a fixed $\vec{l} \in \mathcal{L}^*_{kn,z}$ and its corresponding string in $\mathcal{D}(N,k,z)$:

$$\vec{v}_l \;=\; (<\vec{l},U(\vec{a}_1,k)>,<\vec{l}(U(\vec{a}_2,k)>,\ldots,<\vec{l},U(\vec{a}_N,k)>)$$

For $\vec{l_i} = (\lambda^i_1,\ldots,\lambda^i_N)$ $(i = 1,\ldots,s)$, we define vectors $\vec{R}_i \in GF(2^{kn})$ as

$$\vec{R}_i \;=\; \lambda^i_1 U(\vec{a}_1,k) \oplus \lambda^i_2 U(\vec{a}_2,k) \oplus \ldots \oplus \lambda^i_N U(\vec{a}_N,k).$$

It is possible to prove that $\vec{R}_1,\ldots,\vec{R}_s$ are linearly independent. This indeed follows from three facts: 1) any k vectors of type $U(\vec{a}_j,k)$ are linearly independent; 2) at most k of such vectors appear in each $\vec{R}_i$. 3) The linear combinations of them remain linearly independent.

Then, for any $i = 1,\ldots,s$, we have

$$<\vec{l}_i,\vec{v}_l> \;=\; <(\lambda^i_1,\ldots,\lambda^i_N),(<\vec{l},U(\vec{a}_1,k)>,\ldots,<\vec{l},U(\vec{a}_N,k)>)> =$$

$$= \lambda^i_1 <\vec{l},U(\vec{a}_1,k)> \oplus \ldots \oplus \lambda^i_N <\vec{l},U(\vec{a}_N,k)>)> =$$

$$= <\vec{l},\lambda^i_1 U(\vec{a}_1,k) \oplus \ldots \oplus \lambda^i_N U(\vec{a}_N,k)> \;=\; <\vec{l},\vec{R}_i> \;\;.$$

Our final step is to reduce to conditions required by Corollary 2.1. Indeed, the equation $<\vec{l}_i,\vec{v}_l>= \beta_i$ is equivalent to $l(\vec{R}_i) = \beta_i$. It follows that, for any $\vec{l}$ such that $\vec{v}_l \in W \cap \mathcal{D}(N,k,z)$, we have

$$<\vec{l},\vec{R}_1>= \beta_1 \;\ldots,<\vec{l},\vec{R}_s>= \beta_s \;.$$

where $\vec{R}_1,\ldots,\vec{R}_s$ are linearly independent. Corollary 2.1 thus implies the thesis. □

A.3 Proofs of Lemmas 3.1 and 3.2

A.3.1 Proof of Lemma 3.1

From the hypothesis on rank$(B_i \cup [D]^{n_i})$, if we assume that $u_i^{\vec{b}_i} = 0$ for any $\vec{b}_i \in [D]^{n_i}$, it easily follows that the system

$$\Big\{<\vec{b}_i,\vec{x}_i> \;=\; u_i(\vec{b}_i) \;\;,\forall\vec{b}_i \in B_i \cup [D]^{n_i}$$

admits a solution and $B_i \cap [D]^{n_i} = \emptyset$. This immediatly implies the lemma.

A.4 Proof of Lemma 3.2

We show an iterative method to construct the set D by adding one new element to it in order to satisfy the orthogonal property specified by the lemma. Assume we already have the set

$$D_j \subseteq \mathcal{L}^*_{n,(\lceil \log n \rceil + t)}$$

with $|D| = j$ $(0 \leq j < k - t)$ such that

$$\mathrm{rank}(B_i \cup [D_j]^{n_i}) = n_i - m_i + j \ , \qquad i = 1, \dots, s \ .$$

We then consider the homogeneous system

$$\Big\{ < \vec{b}_i, \vec{x}_i > \; = \; 0 \ , \quad \text{for any } \vec{b}_i \in B_i \cup [D_j]^{n_i} \ ,$$

whose solution space has dimension $m_i - j$. Let $\{\vec{v}^i_1, \dots, \vec{v}^i_{m_i - j}\}$ be a vector basis of this space, and define $\vec{V}^i_j \in \{0,1\}^n$ as the concatentation of $\vec{v}^i_j$ and dummy 0's. Observe that for any $\vec{a} \in \{0,1\}^n$, we have

$$< \vec{V}^i_j, \vec{a} > \; = \; < \vec{v}^i_j, [\vec{a}]^{n_i} > \ .$$

For any fixed $i \in \{1, \dots, s\}$, the vector $[\vec{a}]^{n_i}$ cannot be represented as a linear combination of vectors from $(B_i \cup [D_j]^{n_i})$ iff the following system is not satisfied

$$\Big\{ < \vec{V}^i_t, \vec{a} > = 0 \ \ , \ t = 1, \dots, m_i - j \ . \tag{12}$$

For any fixed i, Corollary 2.1 implies that the probability that a vector $\vec{a}$ from $\mathcal{L}^*_{n,(\lceil \log n \rceil + t)}$ satisfies the system in Eq. (12) is at most

$$2^{j - m_i} + 2^{-(\lceil \log n \rceil + t)} n \leq 2^{(k-t-1)-k} + 2^{-t} = \frac{3}{2} 2^{-t} = \frac{3}{2} 2^{-(\lceil \log s \rceil + 1)} \leq \frac{3}{4} \frac{1}{s} \ .$$

It follows that the probability that a vector $\vec{a}$ chosen uniformly at random from $\mathcal{L}^*_{n,(\lceil \log n \rceil + t)}$ satisfies Condition (12) is true for at least one i from $\{1, \dots, s\}$ is at most

$$s \frac{3}{4} \frac{1}{s} \; = \; \frac{3}{4} \ ,$$

So, there exists $\vec{a}$ from $\mathcal{L}^*_{n,(\lceil \log n \rceil + t)}$ such that for all $i \in \{1, \dots, s\}$ the system in Eq. (12) is not satisfied. We then define $D_{j+1} \; = \; D_j \cup \{\vec{a}\}$; the construction terminates when $j = k - t$ and we set $D = D_{k-t}$.

B Proof of Lemma 4.2

For any fixed $i \in \{1, \dots, ka\}$, consider the subset $X_i \subseteq \{x_1, \dots, x_n\}$ where $|X_i| \leq \lceil n/a \rceil$ and assume that each variable belongs to at most k subsets from $\{X_1, \dots, X_{ka}\}$. Let $q(i)$ be the number of sets X_j such that $x_i \in X_j$.

Let us fix an integer $t \geq 2$ and then select randomly a function $\phi : \{1, \dots, ka\} \to \{1, \dots, t\}$ with uniform distribution. Let us consider the Boolean function $\xi_{s,i}(\phi)$ defined over all possible choices of ϕ

such that $\xi_{s,i}(\phi) = 1$ if for all j for which $x_i \in X_j$ it holds $\phi(j) = s$, and $\xi_{s,i}(\phi) = 0$ otherwise. Then, for any fixed s, we let

$$\Xi_s(\phi) \; = \; \sum_{i=1}^{n} \xi_{s,i}(\phi) \; .$$

Our first goal is study the expected behaviour of the above random variable. In particular, if $\mathbf{E}\,(*)$ denotes the expected value of a given random variable, then using Chebichev's inequaltiy, it is possible to prove the following bound. For any fixed $s = \{1, \ldots, t\}$ and for any $0 \leq \epsilon \leq 1$

$$\mathbf{Pr}\,(|\Xi_s - \mathbf{E}\,(\Xi_s)| \geq \epsilon \mathbf{E}\,(\Xi_s)) \; \leq \; \frac{1}{\epsilon^2}\left(\frac{t^k}{n} + \frac{t^{2k}k}{a}\right) \; . \tag{13}$$

Now, let $g(x_1, \ldots, x_n)$ be a (k, a)-rectangle and assume that for some $t \geq 2$

$$\frac{4t^{k+1}}{n} + \frac{4t^{2k+1}k}{a} \; < \; 1 \; .$$

From Eq. (13), for any fixed s, we get

$$\mathbf{Pr}\left(\Xi_s \leq \frac{1}{2}\frac{n}{t^k}\right) \; \leq \; \mathbf{Pr}\left(\Xi_s \leq \frac{1}{2}\mathbf{E}\,(\Xi_s)\right) \; \leq \; \mathbf{Pr}\left(|\Xi_s - \mathbf{E}\,(\Xi_s)| \geq \frac{1}{2}\mathbf{E}\,(\Xi_s)\right) \; \leq \; 4\left(\frac{t^k}{n} + \frac{t^{2k}k}{a}\right) \; .$$

Consequently

$$\mathbf{Pr}\left(\exists s \in \{1, \ldots, t\} \; : \; \Xi_s \leq \frac{1}{2}\frac{n}{t^k}\right) \; \leq \; t \cdot 4\left(\frac{t^k}{n} + \frac{t^{2k}k}{a}\right) \; \leq \; \frac{4t^{k+1}}{n} + \frac{4t^{2k+1}k}{a} \; < \; 1 \; .$$

Therefore, a function ϕ exists such that

$$\forall s \in \{1, \ldots, t\} : \; \Xi_s(\phi) > \frac{1}{2}\frac{n}{t^k}$$

For any s, we can thus choose X_s^* such that

$$|X_s^*| = \lceil \frac{n}{2t^k} \rceil \; , \qquad \text{and } X_s^* \subseteq \bigcap_{j \; , \; \phi(j)=s} X_j \; .$$

We then define

$$X_0^* \; = \; \{x_1, \ldots, x_n\} \setminus \left(\bigcup_{s=0}^{t} X_s^*\right) \; .$$

Since g is a (k, a)-rectangle, we can write g as follows

$$g(x_1, \ldots, x_n) \; = \; \bigwedge_{s=1}^{t}\left(\bigwedge_{j \; : \; \phi(j)=s} g_j(X_j)\right) \; .$$

Let v be a partial input from $\mathcal{B}(X_0^*)$ then, for any $j = 1, \ldots, ka$, if $\phi(j) = s$ then $X_j \setminus X_0^* = X_s^*$. It follows that the function

$$\left(\bigwedge_{i \in X_0^*} (x_i \oplus v(x_i)) \right) \left(\bigwedge_{j \,:\, \phi(j)=s} g_j(X_j) \right)$$

depends only on the variables from X_s^*, so there exist functions $g_s^v(X_s^*)$, $s = 1, \ldots, t$, such that

$$\left(\bigwedge_{i \in X_0^*} (x_i \oplus v(x_i)) \right) g(x_1, \ldots, x_n) = \left(\bigwedge_{s=1}^{t} g_j^v(X_s^*) \right) ,$$

and g can be written in the following form

$$g(x_1, \ldots, x_n) = \bigvee_{v \in \mathcal{B}(X_0^*)} \left(\bigwedge_{i \in X_0^*} (x_i \oplus v(x_i)) \right) \left(\bigwedge_{s=1}^{t} g_j^v(X_s^*) \right) , \tag{14}$$

Observe now that if $s \neq v$ then $X_s^* \cap X_v^* = \emptyset$ and, moreover, for any $s = 1, \ldots, t$, $|X_s^*| \leq \lceil n/(2t^k) \rceil$. Consequently the representation in Eq. (14) of g is an OR of $2^{|X_0^*|}$ of $(1, 2t^k)$-rectangles. The Lemma is proved by observing that

$$|X_0^*| \leq n - t \left(\frac{n}{2t^k} \right) = n - \frac{n}{2t^{k-1}} .$$

C Proof of Lemma 4.3

Let

$$g = \bigwedge_{i=1}^{a} g_i(X_i) .$$

Since $k = 1$, the sets X_i's are pairwise disjoint. Further, we have that

$$|N_g^1| = \prod_{i=1}^{a} |N_{g_i}| .$$

Let r be the number of i's such that $|N_{g_i}^1| > 1$. Since

$$|N_g^1| \leq \left(2^{\lceil \frac{n}{T} \rceil} \right)^r ,$$

then

$$r \geq \frac{|N_g^1|}{(n/T) + 1} .$$

Let A_i be the affine space of maximum dimension d_i such that $A_i \subseteq N_{g_i}^1$. Observe that if $|N_{g_i}^1| > 1$ then $d_i \geq 1$ since any two different points yield an affine space of dimension 1. We thus consider the affine space $A = A_1 \times \ldots \times A_a$. The dimension of A is at least r.