Electronic Colloquium on Computational Complexity, Report No. 67 (2019)

Sign rank vs Discrepancy

Hamed Hatami *
McGill University
hatami@cs.mcgill.ca

Kaave Hosseini†
University of California, San Diego
skhossei@ucsd.edu

Shachar Lovett‡
University of California, San Diego
slovett@ucsd.edu

Abstract

Sign-rank and discrepancy are two central notions in communication complexity. The seminal work of Babai, Frankl, and Simon from 1986 initiated an active line of research that investigates the gap between these two notions. In this article, we establish the strongest possible separation by constructing a Boolean matrix whose sign-rank is only 3, and yet its discrepancy is $2^{-\Omega(n)}$. We note that every matrix of sign-rank 2 has discrepancy $n^{-O(1)}$.

Our result in particular implies that there are Boolean functions with $O(1)$ *unbounded error* randomized communication complexity while having $\Omega(n)$ *weakly unbounded error* randomized communication complexity.

1 Introduction

The purpose of this article is to prove a strong gap between the sign-rank and discrepancy of Boolean matrices, or equivalently, between unbounded error randomized communication complexity and weakly unbounded error randomized communication complexity.

Sign-rank and discrepancy are arguably the most important analytic notions in the area of communication complexity. Let A be a matrix with $\{-1, 1\}$ entries (we refer to these matrices as Boolean matrices in this paper). The *discrepancy* of A is the maximal correlation it has with a rectangle (for a formal definition see Section 2). The *sign-rank* of A is the minimal rank of a real matrix whose entries have the same sign pattern as A. This natural and fundamental notion was first introduced by Paturi and Simon [PS86] in the context of the unbounded error communication complexity. Since then, its applications have extended beyond communication complexity to areas such as circuit complexity [RS10,BT16], learning theory [LMSS07,LS09b,KS04], and even algebraic geometry [War68].

The extremely relaxed requirement in the definition of sign-rank makes it one of the most elusive analytic notions in complexity theory, and despite several decades of interest, many basic questions about this natural notion have remained unanswered. One of the earlier results regarding

*Supported by an NSERC grant.
†Supported by NSF grant CCF-1614023.
‡Supported by NSF grant CCF-1614023.

ISSN 1433-8092

sign-rank is due to Alon, Frankl, and Rödl [AFR85] who showed that every $N \times N$ matrix has sign-rank at most $\frac{N}{2} + o(N)$, and furthermore using a clever counting argument showed that there are matrices of sign-rank at least $\frac{N}{32}$. Almost two decades passed until finally Forster [For02] discovered a method for proving lower-bounds for *explicit* matrices, and applied it to show that the Hadamard matrix of dimension N has sign-rank at least $\sqrt{N}$. To this day, Forster's theorem and its various generalizations are the only known methods for proving lower-bounds on sign-rank. In this article we are interested in the other end of the spectrum, i.e. matrices of very small sign-rank. Since we are interested in applications to communication complexity, it is more natural to state our result for functions $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,1\}$, corresponding to $2^n \times 2^n$ Boolean matrices.

Theorem 1.1 (Main Theorem)**.** *There exists a function $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,1\}$ of sign-rank 3 and discrepancy $2^{-\Omega(n)}$.*

The sign-rank 3 in Theorem 1.1 is tight. As we show in Section 3, functions of sign-rank 1 or 2 are very simple combinatorially, and in particular have discrepancy $n^{-O(1)}$.

The function f in Theorem 1.1 is simple to define. Let $N \approx 2^{n/4}$. The input to Alice is $[x, z]$, where $x = (x_1, x_2) \in [N]^2$ and $z \in [-3N^2, 3N^2]$. The input to Bob is $y = (y_1, y_2) \in [N]^2$. We assume below that $\text{sign} : \mathbb{R} \to \{-1, 1\}$ maps positive inputs to 1 and zero or negative inputs to -1. Define

$$f([x, z], [y]) = \text{sign}(z - x_1 y_1 - x_2 y_2).$$

It is obvious from the definition that f has sign-rank 3, as f has the same sign pattern as the matrix $A_{[x,z],y} = z - \langle x, y\rangle$ which has rank 3. The main technical contribution of the paper is the following theorem.

Theorem 1.2. *Let f be as above. Then* $\text{Disc}(f) = O(n \cdot 2^{-n/8})$.

1.1 Connections to communication complexity

Theorem 1.1 is motivated by its applications in communication complexity. Consider a communication problem $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,1\}$ in Yao's two party model. Given an error parameter $\epsilon \in [0, 1/2]$, let $R_\epsilon(f)$ be the smallest communication cost of a *private-coin* randomized communication protocol that on *every* input produces the correct answer with probability at least $1 - \epsilon$. Here private-coin refers to the assumption that players each have their own unlimited *private* source of randomness. Three natural complexity measures arise from $R_\epsilon(f)$.

1. The quantity $R_{1/3}(f)$ is called the *bounded-error randomized communication complexity* of f. The particular choice of $1/3$ is not important as long as one is concerned with an error that is bounded away from both 0 and $1/2$ since in this case the error can be reduced by running the protocol constantly many times and outputting the majority answer.

2. The *weakly unbounded error randomized communication complexity* of f is defined as

$$\text{PP}(f) = \inf_{0 \le \epsilon \le 1/2} \left\{ R_\epsilon(f) + \log \frac{1}{1 - 2\epsilon} \right\},$$

   that includes an additional penalty term, which increases as ϵ approaches $\frac{1}{2}$. The purpose of this error term is to capture the range where ϵ is "moderately" bounded away from $\frac{1}{2}$.

3. Finally the *unbounded error communication complexity* of f is defined as the smallest communication cost of a private-coin randomized communication protocol that computes every entry of f with an error probability that is *strictly* smaller than $\frac{1}{2}$. In other words, the protocol only needs to outdo a random guess, which is always correct with probability $\frac{1}{2}$. We have

$$\mathrm{UPP}(f) = \lim_{\epsilon \nearrow \frac{1}{2}} R_\epsilon(f).$$

In their seminal paper, Babai, Frankl and Simon [BFS86] associated a complexity class to each measure of communication complexity. While in the theory of Turing machines, a complexity that is polynomial in the size of input bits is considered efficient, in the realm of communication complexity, poly-logarithmic complexity plays this role, and communication complexity classes are defined accordingly. Here, the communication complexity classes $\mathrm{BPP}^{\mathrm{cc}}$, $\mathrm{PP}^{\mathrm{cc}}$, and $\mathrm{UPP}^{\mathrm{cc}}$ correspond to the class of communication problems $\{f_n\}_{n=0}^{\infty}$ with polylogarithmic $R_{1/3}(f_n)$, $\mathrm{PP}(f_n)$, and $\mathrm{UPP}(f_n)$, respectively.

Note that while $\mathrm{BPP}^{\mathrm{cc}}$ requires a strong bounds on the error probability, and $\mathrm{UPP}^{\mathrm{cc}}$ only requires an error that beats the random guess, $\mathrm{PP}^{\mathrm{cc}}$ corresponds to the natural requirement that the protocol beats the $\frac{1}{2}$ bound by a margin that is quasi-polynomially large. That is, $\mathrm{PP}^{\mathrm{cc}}$ is the class of communication problems f_n that satisfy $R_{\frac{1}{2}-n^{-c}}(f_n) \leq \log^c(n)$ for some positive constant c. We have the containment $\mathrm{BPP}^{\mathrm{cc}} \subseteq \mathrm{PP}^{\mathrm{cc}} \subseteq \mathrm{UPP}^{\mathrm{cc}}$.

It turns out that both $\mathrm{UPP}(f)$ and $\mathrm{PP}(f)$ have elegant algebraic formulations. Paturi and Simon [PS86] proved that UPP essentially coincides with the sign-rank of f:

$$\log \mathbf{rk}_\pm(f) \leq \mathrm{UPP}(f) \leq \log \mathbf{rk}_\pm(f) + 2.$$

Similar to the way that sign-rank captures the complexity measure $\mathrm{UPP}(f)$, discrepancy captures $\mathrm{PP}(f)$. The classical result relating randomized communication complexity and discrepancy, due to Chor and Goldreich [CG88], is the inequality

$$R_\epsilon(f) \geq \log \frac{1-2\epsilon}{\mathrm{Disc}(f)}.$$

This in particular implies $\mathrm{PP}(f) \geq -\log \mathrm{Disc}(f)$. Klauck [Kla01] showed that the opposite is also true; more precisely, that

$$\mathrm{PP}(f) = O\left(-\log \mathrm{Disc}(f) + \log(n)\right).$$

Thus, a direct corollary of Theorem 1.1 is the following separation between unbounded error and weakly bounded error communication complexity.

Corollary 1.3. *There exists a function* $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,1\}$ *with* $\mathrm{UPP}(f) = O(1)$ *and* $\mathrm{PP}(f) = \Omega(n)$.

Another closely related notion to sign-rank is approximate rank. Given $\alpha > 1$, the α-approximate rank of a boolean matrix A is the minimal rank of a real matrix B, of the same dimensions as A, that satisfies $1 \leq A_{i,j}B_{i,j} \leq \alpha$ for all i, j. The α-approximate rank of a boolean function $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,1\}$ is the α-approximate rank of the associated $2^n \times 2^n$ boolean matrix. Observe that

$$\mathbf{rk}_\pm(f) = \lim_{\alpha \to \infty} \mathbf{rk}^\alpha(f).$$

Given this, a natural question is whether sign-rank can be separated from α-approximate rank. This is also a consequence of Theorem 1.1.

Corollary 1.4. *There exists a function* $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,1\}$ *with* $\mathbf{rk}_{\pm}(f) = 3$ *and* $\mathbf{rk}^{\alpha}(f) = \Omega(2^{n/4}/(\alpha n)^2)$ *for any* $\alpha > 1$.

Corollary 1.4 follows from Theorem 1.1 and the fact that

$$\mathbf{rk}^{\alpha}(f) \geq \Omega\left(\alpha^{-2}\mathrm{Disc}(f)^{-2}\right),$$

which is a combination of the results of Linial and Shraibman [LS09c, Theorem 18] and Lee and Shraibman [LS09a, Theorem 1].

1.2 Related works

The question of separating sign-rank from discrepancy (or equivalently, separating unbounded from weakly unbounded communication complexity) has been studied in a number of papers.

When Babai et al. [BFS86] introduced the complexity classes $\mathrm{BPP}^{\mathrm{cc}} \subseteq \mathrm{PP}^{\mathrm{cc}} \subseteq \mathrm{UPP}^{\mathrm{cc}}$, they noticed that the set-disjointness problem separates $\mathrm{BPP}^{\mathrm{cc}}$ from $\mathrm{PP}^{\mathrm{cc}}$, but they left open the question of separating $\mathrm{UPP}^{\mathrm{cc}}$ from $\mathrm{PP}^{\mathrm{cc}}$, or equivalently sign-rank from discrepancy. This question remained unanswered for more than two decades until finally Buhrman et al. [BVdW07] and independently Sherstov [She08] showed that there are n-bit Boolean function f such that $\mathrm{UPP}(f) = O(\log n)$ but $\mathrm{PP}(f) = \Omega(n^{1/3})$ and $\mathrm{PP}(f) = \Omega(\sqrt{n})$, respectively. The bounds on $\mathrm{PP}(f)$ were strengthened in subsequent works [She11, She13, Tha16, She19] with the final recent separation from [She19] giving a function f with $\mathrm{UPP}(f) = O(\log n)$ and maximal possible $\mathrm{PP}(f) = \Omega(n)$. Despite this line of work, no improvement was made on the $O(\log(n))$ bound on $\mathrm{UPP}(f)$. In fact, to the best of our knowledge, prior to this work, it was not even known whether there are functions with $\mathrm{UPP}(f) = O(1)$ and $R_{1/3}(f) = \omega(\log(n))$. To recall, Corollary 1.3 gives a function f with $\mathrm{UPP}(f) = O(1)$ and $\mathrm{PP}(f) = \Omega(n)$.

A different aspect is the study of sign-rank of matrices. Matrices of sign-rank 1 and 2 are simple combinatorially, while matrices with sign-rank 3 seem to be much more complex. First, it turns out that deciding whether a matrix has sign-rank 3 is NP-hard, a result that was shown by Basri et al. [BFG+09] and independently by Bhangale and Kopparty [BK15]. In fact, deciding if a matrix has sign-rank 3 is $\exists\mathbb{R}$-complete, where $\exists\mathbb{R}$ is the existential first-order theory of the reals, a complexity class lying between NP and PSPACE. This $\exists\mathbb{R}$-completeness result is implicit in both [BFG+09] and [BK15], as observed by [AMY16].

1.3 Proof overview

We give a proof overview of Theorem 1.2. It is useful to think about our discrepancy bound in the language of communication complexity. First, we define a hard distribution ν. Alice and Bob receive uniformly random integers $\mathbf{x}, \mathbf{y} \in [N]^2$ respectively where $N \approx 2^{n/4}$. The inner product $\langle \mathbf{x}, \mathbf{y} \rangle$ is a random variable over $[2N^2]$. Alice also receives another random variable $\mathbf{z}$ over $[-3N^2, 3N^2]$, whose distribution we will explain shortly. The players want to decide whether $\langle \mathbf{x}, \mathbf{y} \rangle \geq \mathbf{z}$. We define $\mathbf{z}$ in such a way that

- $\langle \mathbf{x}, \mathbf{y} \rangle - \mathbf{z} \in [-2N, 2N)$,
- $\langle \mathbf{x}, \mathbf{y} \rangle \geq \mathbf{z}$ happens with probability $\frac{1}{2}$,
- $\langle \mathbf{x}, \mathbf{y} \rangle - \mathbf{z}$ is extremely close in total variation distance to $\langle \mathbf{x}, \mathbf{y} \rangle - \mathbf{z} - 2N$ (which is always negative), even when restricted to arbitrary large combinatorial rectangles.

To construct $\mathbf{z}$, we first define another independent random variable $\mathbf{k}$ and then let $\mathbf{z} = \langle \mathbf{x}, \mathbf{y} \rangle + \mathbf{k}$, or $\mathbf{z} = \langle \mathbf{x}, \mathbf{y} \rangle + \mathbf{k} - 2N$, with equal probabilities. We choose $\mathbf{k} = \mathbf{k}_1 + \mathbf{k}_2$ for $\mathbf{k}_1, \mathbf{k}_2$ independent uniform elements from $[N]$ so that $\mathbf{k}$ is smooth enough for the analysis to go through.

We bound the discrepancy $\mathrm{Disc}_\nu(f)$ as follows. Fix a combinatorial rectangle $A \times B \subset ([N]^2 \times [-3N^2, 3N^2])$. We want to bound the correlation of f with $1_A 1_B$ under the distribution ν. This boils down to showing that after conditioning on the input being in $A \times B$, the distribution $(\langle \mathbf{x}, \mathbf{y} \rangle - \mathbf{z})|_{A,B}$ has small total variation distance with its translation by $2N$. We prove a stronger statement, and show that in fact this is true even if we fix $\mathbf{x} = x$ to a typical x (and therefore choosing $A \subset \{x\} \times [-3N^2, 3N^2]$), namely, after conditioning $\mathbf{x} = x$, and $\mathbf{y} \in B$, the distribution of $(\langle x, \mathbf{y} \rangle - \mathbf{z})|_{\mathbf{y} \in B}$ has small total variation distance with its translation by $2N$. To prove the claim we appeal to Fourier analysis and estimate the Fourier coefficients of the random variable, and verify that the only potentially large Fourier coefficients correspond to Fourier characters that are almost invariant under translations by $2N$. Computing these Fourier coefficients involves computing some partial exponential sums whose details may be seen in Lemma 4.3 and Lemma 4.4.

Paper organization. We give preliminary definitions need for the proof in Section 2. We discuss the structure of matrices of sign-rank 1 and 2 in Section 3. We prove Theorem 1.1 in Section 4.

2 Preliminaries

Notations. To simplify the presentation, we often use $\lesssim$ or $\approx$ instead of the big-O notation. That is, $x \lesssim y$ means $x = O(y)$, and $x \approx y$ means $x = \Theta(y)$. For integers $N \leq M$ we denote $[N, M] = \{N, \ldots, M\}$, and we shorthand $[N] = [1, N]$.

Discrepancy. Let $\mathcal{X}, \mathcal{Y}$ be finite sets, and ν be a probability distribution on $\mathcal{X} \times \mathcal{Y}$. The discrepancy of a function $f : \mathcal{X} \times \mathcal{Y} \to \{-1, 1\}$ with respect to ν and a combinatorial rectangle $A \times B \subseteq \mathcal{X} \times \mathcal{Y}$ is defined as

$$\mathrm{Disc}_\nu^{A \times B}(f) = \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \nu} \left[f(\mathbf{x}, \mathbf{y}) 1_A(\mathbf{x}) 1_B(\mathbf{y}) \right].$$

The discrepancy of f with respect to ν is defined as

$$\mathrm{Disc}_\nu(f) = \max_{A,B} \mathrm{Disc}_\nu^{A \times B}(f),$$

and finally the discrepancy of f is defined as

$$\mathrm{Disc}(f) = \min_\nu \mathrm{Disc}_\nu(f).$$

Probability. We denote random variables with bold letters. Given a random variable $\mathbf{r}$, let $\mu = \mu_{\mathbf{r}}$ denote its distribution. The conditional distribution of $\mathbf{r}$, conditioned on $\mathbf{r} \in S$ for some set S, is denoted by $\mu|_S$. Given a finite set S, we denote the uniform measure on S by μ_S. If $\mathbf{r}$ is uniformly sampled from S, we denote it by $\mathbf{r} \sim S$.

Fourier analysis. The proof of Theorem 1.2 is based on Fourier analysis over cyclic groups. We introduce the relevant notation in the following. Let p be a prime. For $f, g : \mathbb{Z}_p \to \mathbb{C}$ define

$$\langle f, g \rangle = \frac{1}{p} \sum_{x \in \mathbb{Z}_p} f(x) \overline{g}(x),$$

and

$$f * g(z) = \frac{1}{p} \sum_{x \in \mathbb{Z}_p} f(x) g(z - x).$$

Let $\mathrm{e}_p : \mathbb{Z}_p \to \mathbb{C}$ denote the function $\mathrm{e}_p : x \mapsto e^{2\pi i x/p}$. For $a \in \mathbb{Z}_p$ define the character $\chi_a : x \mapsto \mathrm{e}_p(-ax)$. The Fourier expansion of $f : \mathbb{Z}_p \to \mathbb{C}$ is the sum

$$f(x) = \sum_{x \in \mathbb{Z}_p} \widehat{f}(a) \chi_a(x),$$

where $\widehat{f}(a) = \langle f, \chi_a \rangle$. Note that we our definition,

$$\widehat{f}(a) = \frac{1}{p} \sum_{x \in \mathbb{Z}_p} f(x) \mathrm{e}_p(ax).$$

It follows from the properties of the characters that

$$f * g(z) = \sum_{a \in \mathbb{Z}_p} \widehat{f}(a) \widehat{g}(a) \chi_a(z),$$

showing that $\widehat{f * g}(a) = \widehat{f}(a)\widehat{g}(a)$. In particular, if $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are independent random variables taking values in $\mathbb{Z}_p$, and if $\mathbf{x} = \mathbf{x}_1 + \ldots + \mathbf{x}_k$, then

$$\widehat{\mu_{\mathbf{x}}}(a) = p^{k-1} \prod_{i=1}^{k} \widehat{\mu_{\mathbf{x}_i}}.$$

Number theory estimates. Given an integer x, we denote the distance of x to the closest multiple of p (and abusing standard notation) by

$$\|x\|_p = \min\{|x - zp| : z \in \mathbb{Z}\}.$$

We will often use the estimate

$$|\mathrm{e}_p(x) - 1| \approx \frac{\|x\|_p}{p},$$

which follows from the easy estimate that $4|x| \le |e^{2\pi i x} - 1| \le 8|x|$ for $x \in [-1/2, 1/2]$.

3 Sign-rank 1 and 2

In this section we demonstrate that Boolean matrices with sign-rank 1 or 2 are very simple combinatorially. Let A be an $N \times N$ Boolean matrix for $N = 2^n$. If A has sign-rank 1, then there exist nonzero numbers $a_1, \ldots, a_N, b_1, \ldots, b_N \in \mathbb{R}$ such that $A_{i,j} = \mathrm{sign}(a_i b_j)$. In particular, if we partition the a_i and the b_j to the positive and negative numbers, we see that A can be partitioned into 4 monochromatic sub-matrices. This implies that $\mathrm{Disc}(A) \geq \Omega(1)$.

Assume next that A has sign-rank 2. Then there exist vectors $u_1, \ldots, u_N, v_1, \ldots, v_N \in \mathbb{R}^2$ such that $A_{i,j} = \mathrm{sign}(\langle u_i, v_j \rangle)$. By applying a rotation to the vectors, we may assume that their coordinates are all nonzero. Next, by scaling the vectors, we may assume that $u_i = (\pm 1, a_i)$ and $v_j = (b_j, \pm 1)$ for all i, j. Next, partition the a_i and the b_j to the positive and negative numbers. Consider without loss of generality the sub-matrix in which $u_i = (1, a_i)$ and $v_j = (b_j, -1)$ for all i, j (the other three cases are equivalent). In this sub-matrix, $A_{i,j} = \mathrm{sign}(a_i - b_j)$. By removing repeated rows and columns, we get that the sub-matrix is an upper triangular matrix, with 1 on or above the diagonal and -1 below the diagonal. That is, the sub-matrix is equivalent to the matrix corresponding to the Greater-Than Boolean function on at most n bits. Nisan [Nis93] showed that the bounded-error communication complexity of this matrix is $O(\log n)$, which in particular implies that the discrepancy is at least $n^{-O(1)}$. This implies that also $\mathrm{Disc}(A) \geq n^{-O(1)}$.

4 Sign-rank 3 vs. discrepancy

We now turn to prove Theorem 1.2. To recall, Alice's input is the pair $[x, z]$ with $x \in [N]^2, z \in [-3N^2, 3N^2]$, and Bob's input is $y \in [N]^2$. The hard distribution ν is defined as follows. First, sample $\mathbf{x}, \mathbf{y}$ uniformly and independently from $[N]^2$. Next, sample $\mathbf{k}_1, \mathbf{k}_2 \in [N]$ uniformly and independently, and let $\mathbf{k} = \mathbf{k}_1 + \mathbf{k}_2$. Define $\mathbf{z}$ as follows: choose $\mathbf{z} = \langle \mathbf{x}, \mathbf{y} \rangle + \mathbf{k}$ or $\mathbf{z} = \langle \mathbf{x}, \mathbf{y} \rangle + \mathbf{k} - 2N$, each with probability 1/2. Observe that in the former case $\langle \mathbf{x}, \mathbf{y} \rangle < \mathbf{z}$ and hence $f([\mathbf{x}, \mathbf{z}], \mathbf{y}) = 1$; and in the latter case $\langle \mathbf{x}, \mathbf{y} \rangle \geq \mathbf{z}$ and hence $f([\mathbf{x}, \mathbf{z}], \mathbf{y}) = -1$. Thus f is balanced:

$$\Pr[f([\mathbf{x}, \mathbf{z}], \mathbf{y}) = 1] = \Pr[f([\mathbf{x}, \mathbf{z}], \mathbf{y}) = -1] = 1/2.$$

In order to prove the theorem, we bound the correlation of f with a rectangle $A \times B$, where $A \subseteq [N]^2 \times [-3N^2, 3N^2]$ and $B \subseteq [N]^2$. For $x \in [N]^2$, let

$$A_x = \{z : [x, z] \in A\}.$$

We have

$$\begin{aligned}
\mathrm{Disc}_\nu^{A \times B}(f) &= \mathbb{E}_{([\mathbf{x},\mathbf{z}],\mathbf{y}) \sim \nu} \left[f([\mathbf{x}, \mathbf{z}], \mathbf{y}) 1_A(\mathbf{x}, \mathbf{z}) 1_B(\mathbf{y}) \right] \\
&= \mathbb{E}_{\mathbf{x},\mathbf{y} \sim [N]^2} 1_B(\mathbf{y}) \mathbb{E}_{\mathbf{z}|\mathbf{x},\mathbf{y}} \left[f([\mathbf{x}, \mathbf{z}], \mathbf{y}) 1_{A_\mathbf{x}}(\mathbf{z}) \right].
\end{aligned}$$

Recall the definition of f and that $\mathbf{z} = \langle \mathbf{x}, \mathbf{y} \rangle + \mathbf{k}$ or $\mathbf{z} = \langle \mathbf{x}, \mathbf{y} \rangle + \mathbf{k} - 2N$ with equal probabilities.

In the former case f evaluates to 1, and it the latter case it evaluates to -1. We thus have

$$\begin{aligned}\mathrm{Disc}_{\nu}^{A\times B}(f) &= \frac{1}{2}\mathbb{E}_{\mathbf{x},\mathbf{y},\mathbf{k}}\left[f([\mathbf{x},\langle\mathbf{x},\mathbf{y}\rangle+\mathbf{k}],\mathbf{y})1_B(\mathbf{y})1_{A_{\mathbf{x}}}(\langle\mathbf{x},\mathbf{y}\rangle+\mathbf{k})\right] \\ &\quad+\frac{1}{2}\mathbb{E}_{\mathbf{x},\mathbf{y},\mathbf{k}}\left[f([\mathbf{x},\langle\mathbf{x},\mathbf{y}\rangle+\mathbf{k}-2N],\mathbf{y}))1_B(\mathbf{y})1_{A_{\mathbf{x}}}(\langle\mathbf{x},\mathbf{y}\rangle+\mathbf{k}-2N)\right] \\ &= \frac{1}{2}\mathbb{E}_{\mathbf{x},\mathbf{y},\mathbf{k}}\left[1_B(\mathbf{y})1_{A_{\mathbf{x}}}(\langle\mathbf{x},\mathbf{y}\rangle+\mathbf{k})-1_B(\mathbf{y})1_{A_{\mathbf{x}}}(\langle\mathbf{x},\mathbf{y}\rangle+\mathbf{k}-2N)\right] \\ &= \frac{|B|}{2N^2}\mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{y}\sim B}\mathbb{E}_{\mathbf{k}}\left[1_{A_{\mathbf{x}}}(\langle\mathbf{x},\mathbf{y}\rangle+\mathbf{k})-1_{A_{\mathbf{x}}}(\langle\mathbf{x},\mathbf{y}\rangle+\mathbf{k}-2N)\right].\end{aligned}$$

For $x\in[N]^2$ let ν_x^B denote the distribution of $\langle\mathbf{x},\mathbf{y}\rangle+\mathbf{k}$ conditioned on $\mathbf{x}=x,\mathbf{y}\in B$. With this notation,

$$\begin{aligned}\mathrm{Disc}_{\nu}^{A\times B}(f) &= \frac{|B|}{2N^2}\mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{w}\sim\nu_{\mathbf{x}}^B}\left[1_{A_{\mathbf{x}}}(\mathbf{w})-1_{A_{\mathbf{x}}}(\mathbf{w}-2N)\right] \\ &= \frac{|B|}{2N^2}\mathbb{E}_{\mathbf{x}}\sum_{w\in\mathbb{Z}}1_{A_{\mathbf{x}}}(w)\nu_{\mathbf{x}}^B(w)-1_{A_{\mathbf{x}}}(w-2N)\nu_{\mathbf{x}}^B(w) \\ &= \frac{|B|}{2N^2}\mathbb{E}_{\mathbf{x}}\sum_{w\in\mathbb{Z}}1_{A_{\mathbf{x}}}(w)\nu_{\mathbf{x}}^B(w)-1_{A_{\mathbf{x}}}(w)\nu_{\mathbf{x}}^B(w+2N) \\ &\le \frac{|B|}{2N^2}\mathbb{E}_{\mathbf{x}}\sum_{w\in\mathbb{Z}}\left|\nu_{\mathbf{x}}^B(w)-\nu_{\mathbf{x}}^B(w+2N)\right|,\end{aligned}$$

which no longer depends on A. The random variable $\langle\mathbf{x},\mathbf{y}\rangle+\mathbf{k}$ is in the range $[-3N^2,3N^2]$ so we embed $[-3N^2,3N^2]$ into $\mathbb{Z}_p$ for some prime $p\in[6N^2+1,12N^2]$. We consider ν_x^B as a distribution over $\mathbb{Z}_p$, and thus we can rewrite

$$\begin{aligned}\mathrm{Disc}_{\nu}^{A\times B}(f) &\le \frac{p|B|}{2N^2}\mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{w}\sim\mathbb{Z}_p}|\nu_{\mathbf{x}}^B(\mathbf{w})-\nu_{\mathbf{x}}^B(\mathbf{w}+2N)| \\ &\lesssim |B|\cdot\mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{w}\sim\mathbb{Z}_p}|\nu_{\mathbf{x}}^B(\mathbf{w})-\nu_{\mathbf{x}}^B(\mathbf{w}+2N)|.\end{aligned}$$

The following lemma, whose proof is deferred to the next section, completes the proof.

Lemma 4.1. *Let $\tilde{N}\approx N$. Then $\mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{w}\sim\mathbb{Z}_p}|\nu_{\mathbf{x}}^B(\mathbf{w})-\nu_{\mathbf{x}}^B(\mathbf{w}+\tilde{N})|\lesssim\frac{\log N}{\sqrt{|B|N^3}}$.*

By invoking Lemma 4.1 for $\tilde{N}=2N$ we obtain

$$\mathrm{Disc}(f)\le\mathrm{Disc}_{\nu}^{A\times B}(f)\lesssim|B|\frac{\log N}{\sqrt{|B|N^3}}\le\sqrt{\frac{|B|}{N^3}}\log N\le N^{-\frac{1}{2}}\log N\lesssim n2^{-n/8}.$$

4.1 Invariance of $\nu_{\mathbf{x}}^B$ under translation

The goal of this section is to prove Lemma 4.1. We will prove that for a typical x, the measure ν_x^B is almost invariant under the translations by $\tilde{N}\approx N$. First we compute the Fourier expansion of this measure.

Lemma 4.2. *For all $x \in [N]^2$ and $a \in \mathbb{Z}_p$, we have*

$$\widehat{\nu_x^B}(a) = \frac{1}{p}\mathrm{e}_p(2a)\left(\frac{1}{N}\frac{\mathrm{e}_p(Na)-1}{\mathrm{e}_p(a)-1}\right)^2 \mathbb{E}_{\mathbf{y}\sim B}\left[\mathrm{e}_p(a\langle x,\mathbf{y}\rangle)\right].$$

Proof. Recall that ν_x^B is the distribution of $\langle x,\mathbf{y}\rangle+\mathbf{k}_1+\mathbf{k}_2$ where $\mathbf{y}\sim B$ and $\mathbf{k}_1,\mathbf{k}_2\sim[N]$. Therefore for all $a\in\mathbb{Z}_p$,

$$\widehat{\nu_x^B}(a) = p^2\widehat{\mu_{\langle x,\mathbf{y}\rangle}}(a)\widehat{\mu_{\mathbf{k}_1}}(a)\widehat{\mu_{\mathbf{k}_2}}(a) = p^2\widehat{\mu_{\langle x,\mathbf{y}\rangle}}(a)\widehat{\mu_{[N]}}(a)^2,$$

where to recall $\mu_{[N]}$ is the uniform distribution on $[N]$. First, we compute the Fourier coefficients of $\mu_{\langle x,\mathbf{y}\rangle}$:

$$\widehat{\mu_{\langle x,\mathbf{y}\rangle}}(a) = \frac{1}{p}\sum_{t\in\mathbb{Z}_p}\mu_{\langle x,\mathbf{y}\rangle}(t)\mathrm{e}_p(at) = \frac{1}{p}\mathbb{E}_{\mathbf{y}\sim B}\left[\mathrm{e}_p(a\langle x,y\rangle)\right].$$

Next, we compute the Fourier coefficients of $\mu_{[N]}$:

$$\widehat{\mu_{[N]}}(a) = \frac{1}{p}\sum_{t=1}^{N}\frac{1}{N}\mathrm{e}_p(at) = \frac{\mathrm{e}_p(a)}{pN}\cdot\frac{\mathrm{e}_p(Na)-1}{\mathrm{e}_p(a)-1},$$

where we have computed the partial sum of the geometric series $\{\mathrm{e}_p(at)\}_{t=1,\ldots,N}$. The lemma follows. □

With the Fourier coefficients $\widehat{\nu_x^B}(a)$ computed in Lemma 4.2, we can analyze the distance of $\nu_{\mathbf{x}}^B$ from its translation by $\tilde{N}\approx N$.

Proof of Lemma 4.1. Let $\mathbf{w}\sim\mathbb{Z}_p$. Recall that $\mathbf{x}\sim[N]^2$ and that $\tilde{N}\approx N$. Using the Fourier expansion of $\nu_{\mathbf{x}}^B$ we can write

$$s := \mathbb{E}_{\mathbf{x},\mathbf{w}}|\nu_{\mathbf{x}}^B(\mathbf{w})-\nu_{\mathbf{x}}^B(\mathbf{w}+\tilde{N})| = \mathbb{E}_{\mathbf{x},\mathbf{w}}\left|\sum_{a\in\mathbb{Z}_p}\widehat{\nu_{\mathbf{x}}^B}(a)\left(\chi_a(\mathbf{w})-\chi_a(\mathbf{w}+\tilde{N})\right)\right|.$$

We may now use Lemma 4.2 and substitute the Fourier coefficient $\widehat{\nu_{\mathbf{x}}^B}(a)$,

$$s = \frac{1}{p}\mathbb{E}_{\mathbf{x},\mathbf{w}}\left|\sum_{a\in\mathbb{Z}_p}e_p(2a)\left(\frac{1}{N}\frac{\mathrm{e}_p(Na)-1}{\mathrm{e}_p(a)-1}\right)^2\mathbb{E}_{\mathbf{y}\sim B}\left[\mathrm{e}_p(a\langle\mathbf{x},\mathbf{y}\rangle)\right](1-\mathrm{e}_p(-\tilde{N}a))\chi_a(\mathbf{w})\right|.$$

Squaring both sides, and applying Cauchy-Schwarz and then Parseval's identity, we get

$$\begin{aligned}
s^2p^2 &\leq \mathbb{E}_{\mathbf{x},\mathbf{w}}\left|\sum_{a\in\mathbb{Z}_p}\mathrm{e}_p(2a)\mathbb{E}_{\mathbf{y}\sim B}\left[\mathrm{e}_p(a\langle\mathbf{x},\mathbf{y}\rangle)\right]\left(\frac{1}{N}\frac{\mathrm{e}_p(Na)-1}{\mathrm{e}_p(a)-1}\right)^2(1-\mathrm{e}_p(-\tilde{N}a))\chi_a(\mathbf{w})\right|^2 \\
&= \mathbb{E}_{\mathbf{x}}\sum_{a\in\mathbb{Z}_p}\left|\mathbb{E}_{\mathbf{y}\sim B}\left[\mathrm{e}_p(a\langle\mathbf{x},\mathbf{y}\rangle)\right]\right|^2\left|\frac{1}{N}\frac{\mathrm{e}_p(Na)-1}{\mathrm{e}_p(a)-1}\right|^4|1-\mathrm{e}_p(-\tilde{N}a)|^2 \\
&= \sum_{a\in\mathbb{Z}_p}\left(\mathbb{E}_{\mathbf{x}}\left|\mathbb{E}_{\mathbf{y}\sim B}\left[\mathrm{e}_p(a\langle\mathbf{x},\mathbf{y}\rangle)\right]\right|^2\right)\left|\frac{1}{N}\frac{\mathrm{e}_p(Na)-1}{\mathrm{e}_p(a)-1}\right|^4|1-\mathrm{e}_p(\tilde{N}a)|^2.
\end{aligned}$$

Recalling that $p \approx N^2$, note that for $a \neq 0$ it holds that

$$\left|\frac{1}{N}\frac{\mathrm{e}_p(Na)-1}{\mathrm{e}_p(a)-1}\right| \approx \frac{\|Na\|_p}{N\,\|a\|_p} \lesssim \min\left(1, \frac{N}{\|a\|_p}\right)$$

and

$$|\mathrm{e}_p(\tilde{N}a)-1| \approx \frac{\left\|\tilde{N}a\right\|_p}{p} \lesssim \min\left(1, \frac{\|a\|_p}{N}\right).$$

Let us denote $E_a(B) := \mathbb{E}_{\mathbf{x}}\left|\mathbb{E}_{\mathbf{y}\sim B}\left[\mathrm{e}_p(a\langle \mathbf{x},\mathbf{y}\rangle)\right]\right|^2$. We break the sum into two parts and for each part use a different estimate for $E_a(B)$ using Lemma 4.3 below.

$$\begin{aligned}
s^2 &\lesssim \frac{1}{p^2}\sum_{\|a\|_p<N} E_a(B)|\mathrm{e}_p(\tilde{N}a)-1|^2 + \frac{1}{p^2}\sum_{\|a\|_p\geq N} E_a(B)\left|\frac{1}{N}\frac{\mathrm{e}_p(Na)-1}{\mathrm{e}_p(a)-1}\right|^4 \\
&\lesssim \frac{1}{p^2}\sum_{\|a\|_p<N} E_a(B)\left(\frac{\|a\|_p}{N}\right)^2 + \frac{1}{p^2}\sum_{\|a\|_p\geq N} E_a(B)\left(\frac{N}{\|a\|_p}\right)^4 \\
&\lesssim \frac{1}{p^2}\sum_{\|a\|_p<N} \frac{N^2}{\|a\|_p^2}\cdot\frac{\log^2 N}{|B|}\left(\frac{\|a\|_p}{N}\right)^2 + \frac{1}{p^2}\sum_{\|a\|_p\geq N} \frac{\|a\|_p^2}{N^2}\cdot\frac{\log^2 N}{|B|}\left(\frac{N}{\|a\|_p}\right)^4 \\
&\lesssim \frac{1}{N^2|B|}\left(\sum_{\|a\|_p<N}\frac{\log^2 N}{N^2} + \sum_{\|a\|_p\geq N}\frac{\log^2 N}{\|a\|_p^2}\right) \\
&\lesssim \frac{1}{N^2|B|}\left(N\cdot\frac{\log^2 N}{N^2} + \sum_{t\geq N}\frac{\log^2 N}{t^2}\right) \\
&\lesssim \frac{1}{N^2|B|}\frac{\log^2 N}{N} = \frac{\log^2 N}{|B|N^3}.
\end{aligned}$$

□

4.2 Uniformity of product sets over $\mathbb{Z}_p$

Recall that $E_a(B) := \mathbb{E}_{\mathbf{x}\sim[N]^2}\left|\mathbb{E}_{\mathbf{y}\sim B}\left[\chi_a(\langle \mathbf{x},\mathbf{y}\rangle)\right]\right|^2$. The following lemma provides estimates for it.

Lemma 4.3. $E_a(B) \lesssim \max\left(\frac{\|a\|_p^2}{N^2}, \frac{N^2}{\|a\|_p^2}\right)\cdot\frac{\log^2 N}{|B|}$.

Proof. We have

$$\begin{aligned}
E_a(B) &= \frac{1}{|B|^2}\mathbb{E}_{\mathbf{x}\sim[N]^2}\left|\sum_{y\in B}\chi_a(\langle \mathbf{x},y\rangle)\right|^2 \\
&= \frac{1}{|B|^2}\sum_{y',y''\in B}\mathbb{E}_{\mathbf{x}\sim[N]^2}\chi_a(\langle \mathbf{x},y'-y''\rangle) \\
&\leq \frac{1}{|B|^2}\sum_{y',y''\in B}\left|\mathbb{E}_{\mathbf{x}\sim[N]^2}\chi_a(\langle \mathbf{x},y'-y''\rangle)\right|.
\end{aligned}$$

Let $B - B = \{y' - y'' : y', y'' \in B\} \subset \mathbb{Z}_p^2$. Any element $y \in B - B$ can be expressed as $y = y' - y''$ for $y', y'' \in B$ in at most $|B|$ ways. Thus we can bound

$$E_a(B) \leq \frac{1}{|B|} \sum_{y \in B-B} \left| \mathbb{E}_{\mathbf{x} \sim [N]^2} \chi_a(\langle \mathbf{x}, y \rangle) \right|.$$

Since $B - B \subseteq [N]^2 - [N]^2 \subseteq [-N, N]^2$, we can simplify the above to

$$\begin{aligned} E_a(B) &\leq \frac{1}{N^2|B|} \sum_{y \in [-N,N]^2} \left| \sum_{x \in [N]^2} \chi_a(\langle x, y \rangle) \right| \\ &= \frac{1}{N^2|B|} \sum_{y_1, y_2 \in [-N,N]} \left| \sum_{x_1, x_2 \in [N]} \chi_a(x_1 y_1) \cdot \chi_a(x_2 y_2) \right| \\ &= \frac{1}{N^2|B|} \sum_{y_1, y_2 \in [-N,N]} \left| \sum_{x_1 \in [N]} \chi_a(x_1 y_1) \right| \left| \sum_{x_2 \in [N]} \chi_a(x_2 y_2) \right| \\ &= \frac{1}{N^2|B|} \left(\sum_{y \in [-N,N]} \left| \sum_{x \in [N]} \chi_a(xy) \right| \right)^2 \\ &\lesssim \frac{1}{N^2|B|} \left(\sum_{y \in [0,N]} \left| \sum_{x \in [N]} \chi_a(xy) \right| \right)^2. \end{aligned}$$

Note that for a fixed $y \neq 0$, $\sum_{x \in [N]} \chi_a(xy)$ is a sum of a geometric series which satisfies $\left| \sum_{x \in [N]} \chi_a(xy) \right| = \left| \frac{\mathrm{e}_p(Nay)-1}{\mathrm{e}_p(ay)-1} \right|$, and hence

$$\sum_{y \in [0,N]} \left| \sum_{x \in [N]} \chi_a(xy) \right| \leq N + \sum_{y \in [N]} \left| \frac{\mathrm{e}_p(Nay) - 1}{\mathrm{e}_p(ay) - 1} \right| \lesssim N + \sum_{y \in [N]} \frac{\|Nay\|_p}{\|ay\|_p}.$$

Invoking Lemma 4.4 below finishes the proof. □

Lemma 4.4. *Let $p \geq N^2$ be prime and let $a \in \mathbb{Z}_p \setminus \{0\}$. Then*

$$\sum_{y \in [N]} \frac{\|Nay\|_p}{\|ay\|_p} \lesssim \max\left(\|a\|_p + \frac{p}{N}, \frac{p}{\|a\|_p} \right) \cdot \log p.$$

We need the following simple claim in the proof of Lemma 4.4.

Claim 4.5. *Let $\mathbf{r}$ be a random variable which takes values in $[K]$. Let $g : [K] \to \mathbb{R}$. Then*

$$\mathbb{E}_{\mathbf{r}} g(\mathbf{r}) = g(K) + \sum_{i=1}^{K-1} (g(i) - g(i+1)) \Pr[\mathbf{r} \leq i].$$

Proof.

$$\begin{aligned}
\mathbb{E}_{\mathbf{r}} g(\mathbf{r}) &= \sum_{i=1}^{K} g(i)\Pr[\mathbf{r} = i] \\
&= \sum_{i=1}^{K} g(i)\left(\Pr[\mathbf{r} \leq i] - \Pr[\mathbf{r} \leq i-1]\right) \\
&= g(K) + \sum_{i=1}^{K-1} \left(g(i) - g(i+1)\right)\Pr[\mathbf{r} \leq i].
\end{aligned}$$

□

Proof of Lemma 4.4. We separate the proof to two cases of $\|a\|_p < N$ and $\|a\|_p \geq N$. Consider an integer k with $\|a\|_p \leq k \leq p$. We start by estimating the size of the set

$$S_k = \{y \in [N] : \|ya\|_p \leq k\}.$$

Note that if $y \in S_k$, then $ya \in ph + [-k, k]$ for some integer $h \geq 0$. Since $y \in [N]$, we have $h \leq \frac{N\|a\|_p + k}{p}$, and hence there are at most $\frac{N\|a\|_p}{p} + 1$ such values of h. Fixing h, we have $y \in \frac{ph}{\|a\|_p} + [-k/\|a\|_p, k/\|a\|_p]$, and there are at most $\frac{2k}{\|a\|_p} + 1 \leq \frac{3k}{\|a\|_p}$ such values of y. We conclude that

$$|S_k| \leq \left(\frac{N\|a\|_p}{p} + 1\right) \times \frac{3k}{\|a\|_p} \leq \frac{3Nk}{p} + \frac{3k}{\|a\|_p} \lesssim \frac{k}{N} + \frac{k}{\|a\|_p}.$$

Note that this bound obviously holds also for $k \geq p$.

Now to compute $\sum_{y\in[N]} \frac{\|Nay\|_p}{\|ay\|_p}$ we separate to two cases depending on whether $\|a\|_p \geq N$ or not, and then use Claim 4.5.

The case $\|a\|_p \geq N$: First, note that in this case we can bound $|S_k| \lesssim \frac{k}{N}$. Also to bound $\frac{\|Nay\|_p}{\|ay\|_p}$, for $y \in S_{\|a\|_p}$, we use the bound $\frac{\|Nay\|_p}{\|ay\|_p} \leq N$, otherwise we use the bound $\|Nay\|_p \leq p$. We get

$$\sum_{y\in[N]} \frac{\|Nay\|_p}{\|ay\|_p} \leq \sum_{y\in S_{\|a\|_p}} N + p\sum_{y\in[N]} \frac{1}{\|ay\|_p}.$$

To compute $\sum_{y\in[N]} \frac{1}{\|ay\|_p}$ we use Claim 4.5. Let $\mathbf{u} \sim [N]$ be uniformly chosen, and set the random variable $\mathbf{r}$ to be $\mathbf{r} = \|a\mathbf{u}\|_p$. Set $g(x) = \frac{1}{x}$. Then we have

$$\begin{aligned}
\frac{1}{N}\sum_{y\in[N]} \frac{1}{\|ay\|_p} &= \mathbb{E}_{\mathbf{r}} g(\mathbf{r}) \\
&= g(p) + \sum_{i=1}^{p-1} \left(g(i) - g(i+1)\right) \Pr[\mathbf{r} \le i] \\
&= \frac{1}{p} + \sum_{i=1}^{p-1} \left(\frac{1}{i} - \frac{1}{i+1}\right) \frac{|S_i|}{N} \\
&\lesssim \frac{1}{p} + \sum_{i=1}^{p-1} \frac{1}{i^2} \cdot \frac{i}{N^2} \\
&\lesssim \frac{\log p}{N^2}.
\end{aligned}$$

Overall we get

$$\sum_{y\in[N]} \frac{\|Nay\|_p}{\|ay\|_p} \le \sum_{y\in S_{\|a\|_p}} N + p \sum_{y\in[N]} \frac{1}{\|ay\|_p} \lesssim \|a\|_p + \frac{p\log p}{N}.$$

The case $\|a\|_p < N$: Here we use the estimate $|S_k| \lesssim \frac{k}{\|a\|_p}$. Also similar to the previous case, for $y \in S_N$ we use the bound $\frac{\|Nay\|_p}{\|ay\|_p} \le N$, otherwise we use the bound $\frac{\|Nay\|_p}{\|ay\|_p} \le \frac{p}{\|ay\|_p}$. Similar to the previous case, we have

$$\begin{aligned}
\frac{1}{N}\sum_{y\in[N]} \frac{1}{\|ay\|_p} &= g(p) + \sum_{i=1}^{p-1} \left(g(i) - g(i+1)\right) \Pr[\mathbf{r} \le i] \\
&= \frac{1}{p} + \sum_{i=1}^{p-1} \left(\frac{1}{i} - \frac{1}{i+1}\right) \frac{|S_i|}{N} \\
&\lesssim \frac{1}{p} + \sum_{i=1}^{p-1} \frac{1}{i^2} \cdot \frac{i}{\|a\|_p N} \\
&\lesssim \frac{\log p}{\|a\|_p N}.
\end{aligned}$$

So we have

$$\begin{aligned}
\sum_{y\in[N]} \frac{\|Nay\|_p}{\|ay\|_p} &\le \sum_{y\in S_N} N + p \sum_{y\in[N]} \frac{1}{\|ay\|_p} \\
&\lesssim \frac{N^2}{\|a\|_p} + \frac{p\log p}{\|a\|_p} \\
&\lesssim \frac{p\log p}{\|a\|_p}.
\end{aligned}$$

The lemma follows.

□

We remark that the following more general statement regarding uniformity of product sets follows by a similar proof to Lemma 4.3 which we record here as it may be of independent interest.

Lemma 4.6. *Let $p \geq N^2$ be prime, and let $B \subseteq [N]^d$ for some positive integer d. Then*

$$\mathbb{E}_{\mathbf{x}\sim[N]^d}|\mathbb{E}_{\mathbf{y}\sim B}\chi_a(\langle \mathbf{x},\mathbf{y}\rangle)|^2 \lesssim \max\left(\|a\|_p^d, \frac{p^d}{\|a\|_p^d}\right)\cdot\frac{\log^d p}{|B|N^d}.$$

References

[AFR85] Noga Alon, Peter Frankl, and V Rodl. Geometrical realization of set systems and probabilistic communication complexity. In *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)*, pages 277–280. IEEE, 1985.

[AMY16] Noga Alon, Shay Moran, and Amir Yehudayoff. Sign rank versus vc dimension. In *Conference on Learning Theory*, pages 47–80, 2016.

[BFG+09] Ronen Basri, Pedro F Felzenszwalb, Ross B Girshick, David W Jacobs, and Caroline J Klivans. Visibility constraints on features of 3d objects. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1231–1238. IEEE, 2009.

[BFS86] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory. In *27th Annual Symposium on Foundations of Computer Science (sfcs 1986)*, pages 337–347. IEEE, 1986.

[BK15] Amey Bhangale and Swastik Kopparty. The complexity of computing the minimum rank of a sign pattern matrix. *arXiv preprint arXiv:1503.04486*, 2015.

[BT16] Mark Bun and Justin Thaler. Improved bounds on the sign-rank of AC^0. In *43rd International Colloquium on Automata, Languages, and Programming*, volume 55 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 37, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016.

[BVdW07] Harry Buhrman, Nikolay Vereshchagin, and Ronald de Wolf. On computation and communication with small bias. In *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)*, pages 24–32. IEEE, 2007.

[CG88] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988.

[For02] Jürgen Forster. A linear lower bound on the unbounded error probabilistic communication complexity. *Journal of Computer and System Sciences*, 65(4):612–625, 2002.

[Kla01] Hartmut Klauck. Lower bounds for quantum communication complexity. In *Proceedings 2001 IEEE International Conference on Cluster Computing*, pages 288–297. IEEE, 2001.

[KS04] Adam R. Klivans and Rocco A. Servedio. Learning DNF in time $2^{O(n^{1/3})}$. *J. Comput. System Sci.*, 68(2):303–318, 2004.

[LMSS07] Nati Linial, Shahar Mendelson, Gideon Schechtman, and Adi Shraibman. Complexity measures of sign matrices. *Combinatorica*, 27(4):439–463, 2007.

[LS09a] Troy Lee and Adi Shraibman. An approximation algorithm for approximation rank. In *2009 24th Annual IEEE Conference on Computational Complexity*, pages 351–357. IEEE, 2009.

[LS09b] Nati Linial and Adi Shraibman. Learning complexity vs. communication complexity. *Combin. Probab. Comput.*, 18(1-2):227–245, 2009.

[LS09c] Nati Linial and Adi Shraibman. Lower bounds in communication complexity based on factorization norms. *Random Structures & Algorithms*, 34(3):368–394, 2009.

[Nis93] Noam Nisan. The communication complexity of threshold gates. *Combinatorics, Paul Erdos is Eighty*, 1:301–315, 1993.

[PS86] Ramamohan Paturi and Janos Simon. Probabilistic communication complexity. *Journal of Computer and System Sciences*, 33(1):106–123, 1986.

[RS10] Alexander A. Razborov and Alexander A. Sherstov. The sign-rank of AC^0. *SIAM J. Comput.*, 39(5):1833–1855, 2010.

[She08] Alexander A Sherstov. Halfspace matrices. *Computational Complexity*, 17(2):149–178, 2008.

[She11] Alexander A Sherstov. The pattern matrix method. *SIAM Journal on Computing*, 40(6):1969–2000, 2011.

[She13] Alexander A Sherstov. Optimal bounds for sign-representing the intersection of two halfspaces by polynomials. *Combinatorica*, 33(1):73–96, 2013.

[She19] Alexander A. Sherstov. The hardest halfspace. *CoRR*, abs/1902.01765, 2019.

[Tha16] Justin Thaler. Lower bounds for the approximate degree of block-composed functions. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

[War68] Hugh E. Warren. Lower bounds for approximation by nonlinear manifolds. *Trans. Amer. Math. Soc.*, 133:167–178, 1968.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il