THEORY OF COMPUTING, Volume 18 (19), 2022, pp. 1–22
www.theoryofcomputing.org

SPECIAL ISSUE: CCC 2020

Sign-Rank vs. Discrepancy

Hamed Hatami* Kaave Hosseini† Shachar Lovett‡

Received August 11, 2020; Revised May 2, 2022; Published July 9, 2022

Abstract. Sign-rank and discrepancy are two central notions in communication complexity. The seminal paper by Babai, Frankl, and Simon (FOCS'86) 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 $n^{-O(1)}$.

In connection with learning theory, our result implies the existence of Boolean matrices whose entries are represented by points and half-spaces in dimension 3, and yet, the normalized margin of any such representation (angle between the half-spaces and the unit vectors representing the points), even in higher dimensions, is very small.

In the context of communication complexity, 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.

A conference version of this paper appeared in the Proceedings of the 35th Computational Complexity Conference, 2020 (CCC'20) [16].

*Supported by an NSERC grant.

†Supported by NSF grant CCF-1614023

‡Supported by NSF grant CCF-1614023

ACM Classification: F.2.3

AMS Classification: 68Q11

Key words and phrases: communication complexity, sign-rank, discrepancy

 DOI: 10.4086/toc.2022.v018a019

1 Introduction

Sign-rank and discrepancy are arguably the most important analytic notions in the area of communication complexity. Let $A_{\mathcal{X}\times\mathcal{Y}}$ be a matrix with $\{-1,1\}$ entries (we refer to these matrices as boolean matrices in this paper). The *discrepancy* of A is the minimum over all input distributions of the maximum correlation that A has with a rectangle. More formally,

$$\mathrm{Disc}(A) := \min_{\nu} \max_{\substack{S\subseteq\mathcal{X}\\ T\subseteq\mathcal{Y}}} \left|\mathbb{E}_{(x,y)\sim\nu}[A_{x,y}1_S(x)1_T(y)]\right|, \tag{1.1}$$

where the minimum is over all probability distributions ν on $\mathcal{X}\times\mathcal{Y}$.

The classical concept[1] of discrepancy was applied to Hadamard matrices (boolean matrices with orthogonal rows) by Brown and Spencer[2] in 1971 [6] to analyze the Gale—Berlekamp switching game (see [13, Ch. 15]). Adapting their method, Chor and Goldreich [11] applied discrepancy to randomized communication complexity. In deterministic communication complexity, discrepancy was used by Babai, Nisan, and Szegedy [3] to prove lower bounds for the deterministic communication complexity of Generalized Inner Product (GIP) in the number-on-the-forehead model. Nowadays, discrepancy has become one of the most commonly used measures in communication complexity, especially to prove lower bounds for randomized protocols.

The *sign-rank* of A, denoted by $\mathbf{rk}_{\pm}(A)$, 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 [22] in the context of the unbounded-error communication complexity. Since then, its applications have extended beyond communication complexity to areas such as circuit complexity [8, 23], learning theory [18–20], and even connections to algebraic geometry [29].

Boolean matrices in communication complexity correspond to boolean functions: given an n-bits two-player function $f:\{0,1\}^n\times\{0,1\}^n\to\{-1,1\}$, it corresponds to the $2^n\times 2^n$ matrix $A_{x,y}=f(x,y)$. The notions of discrepancy and sign-rank for f are defined as the respective quantities assigned to the corresponding matrix.

Our work is motivated by the following main informal question.

Problem 1.1. Does every function of low sign-rank have an efficient randomized protocol?

If the answer is negative, then the next question is, does it at least have large discrepancy. (Small discrepancy is one technique to prove randomized communication complexity lower bounds, but there are functions showing separations between the two measures, for example set-disjointness [9].)

Problem 1.2. Does every function of low sign-rank have large discrepancy?

In order to build some intuition towards more quantitative questions, let us consider some well-known examples:

[1] Bernard Chazelle writes in the Preface of his book "The Discrepancy Method" [10]: "Discrepancy theory grew out of a question posed by van der Corput [12] in 1935: 'How uniform can an infinite sequence in $[0,1]$ be?' "

[2] Erdős and Spencer credit the key lemma to J. H. Lindsey (John Hathway Lindsey II).

- Greater-than: we interpret x, y as integers in $\{1, \ldots, 2^n\}$ and define $f(x, y) = 1$ if $x \leq y$ and $f(x, y) = -1$ otherwise. This function has sign-rank 2 and requires $\Theta(\log n)$ bits of randomized communication [21]. Moreover, its discrepancy is $n^{-\Theta(1)}$, which proves the communication lower bound.

- Set-disjointness: we interpret x, y as subsets of $[n]$, and define $f(x, y) = 1$ if x, y are disjoint and $f(x, y) = -1$ otherwise. This function has sign-rank $O(n)$ and requires communication complexity of $\Theta(n)$ bits. However, this cannot be shown using discrepancy, as the discrepancy of set-disjointness is $n^{-O(1)}$ [9].

- Sherstov [27] constructed a function with sign-rank $O(n)$ and discrepancy $2^{-\Omega(n)}$.

Thus, it seems that functions with logarithmic sign-rank can already be very complicated, both in terms of their randomized communication complexity and also in terms of their discrepancy. However, the situation is less clear for functions of *constant* sign-rank.

Problem 1.3. Does every function of constant sign-rank have an efficient randomized protocol? In particular, does it have large discrepancy?

Our main result is a resounding no, already for sign-rank 3.

Theorem 1.4 (Main Theorem)**.** *There exists a function $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,1\}$ of sign-rank 3 and discrepancy $O(n \cdot 2^{-n/8}) = 2^{-\Omega(n)}$.*

Note that, it follows from the bound on discrepancy that the function f in Theorem 1.4 has $\Omega(n)$ randomized communication complexity.

The sign-rank 3 in Theorem 1.4 is tight. We show in Section 3 that functions of sign-rank 1 or 2 are very simple combinatorially, and in particular have discrepancy $n^{-O(1)}$ and randomized communication complexity $O(\log n)$.

The function f in Theorem 1.4 is simple to define: the sign on an inner product in dimension 3. Concretely, let $M \approx 2^{n/3}$. Alice gets a vector $\mathbf{a} \in [-M, M]^3$ and Bob gets a vector $\mathbf{b} \in [-M, M]^3$. Define

$$f(\mathbf{a}, \mathbf{b}) = \operatorname{sign}\langle \mathbf{a}, \mathbf{b} \rangle,$$

where $\operatorname{sign} : \mathbb{R} \to \{-1, 1\}$ is the sign function, mapping positive inputs to 1 and zero or negative inputs to -1; and $\langle \cdot, \cdot \rangle$ is inner product over the integers. It is obvious from the definition that f has sign-rank 3. We prove that its discrepancy is exponentially small. The actual function we study is a mild restriction of this function, convenient for the proof. See Theorem 1.9 for details.

1.1 Connections to learning theory

Note that the sign-rank of an $N \times N$ boolean matrix A is the smallest d such that there exist unit vectors $u_i, v_j \in \mathbb{R}^d$ with $A_{i,j} = \operatorname{sign}(\langle u_i, v_j \rangle)$ for all $i, j \in [N]$. These unit vectors u_i, v_j represent A as points and half-spaces in the d-dimensional space: $A_{i,j} = 1$ iff the point u_i belongs to the half-space $\{z : \langle z, v_j \rangle > 0\}$.

The geometric representations of boolean matrices as points and half-spaces is central to the theory of learning. In this context, every column j of A corresponds to an object in some domain. The vectors $v_j \in \mathbb{R}^d$, which are called *feature vectors*, represent each object by d numerical features. A classification algorithm receives as input a sample $(j_1, A_{i,j_1}), \ldots, (j_m, A_{i,j_m})$ for an unknown i, and its task is to predict $A_{i,j}$ for other values of j. For example, linear classifiers, such as support vector machines, aim to produce from the samples a vector u such that $\mathrm{sign}(\langle u, v_j \rangle)$ is a good predictor of $A_{i,j} = \mathrm{sign}(\langle u_i, v_j \rangle)$.

While sign-rank optimizes the dimension (i. e., the number of features), there is a second natural parameter that is associated with such representations. The quantity $\min_{i,j} |\langle u_i, v_j \rangle|$ is called the *margin*; it measures the smallest distance between the points u_i and the hyperplanes defined by v_j.

The *margin* of an $N \times N$ boolean matrix A, denoted by $\mathrm{m}(A)$, is the largest possible margin attainable by such representations. More formally,

$$\mathrm{m}(A) := \sup \min_{i,j} \left| \langle u_i, v_j \rangle \right|,$$

where the supremum is over all $d \in \mathbb{N}$ and unit vectors $u_i, v_j \in \mathbb{R}^d$ with $A_{i,j} = \mathrm{sign}(\langle u_i, v_j \rangle)$.

Dimension and margin are two important parameters that impact the performance of these algorithms. It is desirable to represent the matrix A in a smaller dimension and with a large margin. Therefore, the problem of understanding the relation between sign-rank and margin is an important one. Note that sign-rank minimizes the dimension while allowing the margin to be arbitrarily small, and in contrast, margin maximizes the margin of the representation while allowing the dimension to be arbitrarily large.

Due to the mentioned connections to the theory of learning, the notion of margin had been mainly studied in that context, but Linial and Shraibman [20] proved that margin is essentially equivalent to discrepancy:

$$\mathrm{Disc}(A) \leq \mathrm{m}(A) \leq 8\,\mathrm{Disc}(A).$$

In light of this equivalence, our main result (Theorem 1.4) can be reformulated as the following.

Theorem 1.5 (Reformulation of Theorem 1.4)**.** *There exists $N \times N$ matrices A with sign-rank 3 and margin $O(\log(N)/N^{1/8})$.*

In other words, even though it is possible to represent A in dimension 3, any representation of A in *any dimension* will have a very small margin.

It is worthwhile to contrast Theorem 1.5 with Forster's seminal lower bound for sign-rank [14]. Forster proved that in the representation $A_{i,j} = \mathrm{sign}(\langle u_i, v_j \rangle)$ by unit vectors $u_i, v_j \in \mathbb{R}^d$, it is possible to transform the vectors so that the vectors v_j are in a so-called isotropic position. This in particular implies

$$\mathbb{E}_{i,j \in [N]} |\langle u_i, v_j \rangle| \geq \mathbb{E}_{i,j \in [N]} |\langle u_i, v_j \rangle|^2 = \frac{1}{d}.$$

In other words, when sign-rank is small, there are representations with large "average" margin. In the specific case of the matrix A in Theorem 1.5, there exists a representation of A with unit vectors $u_i, v_j \in \mathbb{R}^3$ such that

$$\mathbb{E}_{i,j\in[N]} |\langle u_i, v_j\rangle| \geq \frac{1}{3},$$

while Theorem 1.5 shows that in any representation (in any dimension), we have

$$\min_{i,j\in[N]} |\langle u_i, v_j\rangle| \leq O(\log(N)/N^{1/8}).$$

Finally, let us mention that regarding the converse direction of the relation between sign-rank and margin, Linial, Mendelson, Schechtman, and Shraibman [19, Corollary 3.2, Lemma 4.2, and Section 8] proved the inequality $\mathbf{rk}_{\pm}(A) \leq \mathrm{m}(A)^{-2} \cdot \log(N)$, and asked whether the log factor in this inequality is necessary. In fact the following question remains open.

Question 1.6. Is there a function η such that for every boolean matrix A, we have $\mathbf{rk}_{\pm}(A) \leq \eta(\mathrm{m}(A)^{-1})$?

1.2 Connections to communication complexity

Theorem 1.4 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

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

   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

$$\mathrm{UPP}(f) = \lim_{\epsilon\to\frac{1}{2}-0} R_\epsilon(f).$$

In their seminal paper, Babai, Frankl and Simon [2] 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 bound 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 $R_{\frac{1}{2}-2^{-\log^c(n)}}(f_n) \leq \log^c(n)$ 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 [22] proved that UPP essentially coincides with the sign-rank of f:

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

Similarly 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 [11], 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 [17] 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 1.4 is the following separation between unbounded-error and weakly bounded-error communication complexity.

Corollary 1.7. *There exists a function* $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,1\}$ *with* $\mathrm{UPP}(f) = O(1)$ *and* $\mathrm{PP}(f) = \Omega(n)$.

1.3 Implications regarding approximate rank

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 1.4.

Corollary 1.8. *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 1.8 follows from Theorem 1.4 and the fact that

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

which is established in [15, Equation (1)].

1.4 Related work

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

When Babai et al. [2] introduced the complexity classes $\mathrm{BPP^{cc}} \subseteq \mathrm{PP^{cc}} \subseteq \mathrm{UPP^{cc}}$, they noticed that the set-disjointness problem separates $\mathrm{BPP^{cc}}$ from $\mathrm{PP^{cc}}$, but they left open the question of separating $\mathrm{UPP^{cc}}$ from $\mathrm{PP^{cc}}$, or equivalently sign-rank from discrepancy. This question remained unanswered for more than two decades until finally Buhrman et al. [7] and independently Sherstov [24] showed that there are n-bit boolean functions 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 work [25–28] with the final recent separation from [27] 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 1.7 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. [4] and independently by Bhangale and Kopparty [5]. 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 [4] and [5], as observed by [1].

1.5 Proof overview

We give an overview of the proof of Theorem 1.4. Let us first slightly modify f in a way that will be convenient for the proof.

Let $N \approx 2^{n/4}$. Alice gets three integers x_1, x_2, z and Bob gets two integers y_1, y_2, where $x_1, x_2, y_1, y_2 \in [N]$ and $z \in [-3N^2, 3N^2]$. We shorthand $x = (x_1; x_2)$ and $y = (y_1; y_2)$, so that Alice's input is $(x; z)$ and Bob's input is y. Note that $x, y \in [N]^2$. Define

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

The following is our main technical result.

Theorem 1.9 (Main result; technical version)**.** *Let f be as above. Then* $\mathrm{Disc}(f) = O(n \cdot 2^{-n/8})$.

We remark that the function f here is a restriction of the function f described after Theorem 1.4, and therefore, Theorem 1.9 implies Theorem 1.4.

To prove Theorem 1.9, it is useful to think about our discrepancy bound in the language of communication complexity. We prove Theorem 1.9 in two steps. Below we denote random variables with bold letters.

Step 1: constructing a hard distribution 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 $\frac{1}{2}$,
- The distribution of $\mathbf{z}$ conditioned on $\langle \mathbf{x}, \mathbf{y} \rangle \geq \mathbf{z}$ is extremely close in total variation distance to the distribution of $\mathbf{z}$ conditioned on $\langle \mathbf{x}, \mathbf{y} \rangle < \mathbf{z}$, even when restricted to arbitrary large combinatorial rectangles. This is formalized in Lemma 4.1 and calculations preceding it. See step 2 bellow for more discussion.

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. Note that the range of $\mathbf{z}$ is really just $[-2N, 2N^2 + 2N]$, and we picked the range of z in the definition of f as $z \in [-3N^2, 3N^2]$ for its simpler shape.

Step 2: translation invariance of $\langle \mathbf{x}, \mathbf{y} \rangle + \mathbf{k}$ We bound the discrepancy $\mathrm{Disc}_\nu(f)$ as follows. Fix a combinatorial rectangle $A \times B \subset ([N]^2 \times [-3N^2, 3N^2]) \times [N]^2$. We bound the correlation of f with $1_A 1_B$ under the distribution ν. In other words, we show under the distribution ν, restricted to the rectangle $A \times B$, we have $\nu(f^{-1}(1)) \approx \nu(f^{-1}(-1))$. This boils down to showing that after conditioning on the input being in $A \times B$, the distribution of $(\langle \mathbf{x}, \mathbf{y} \rangle + \mathbf{k})|_{A,B}$ has small total variation distance with $(\langle \mathbf{x}, \mathbf{y} \rangle + \mathbf{k} - 2N)|_{A,B}$. 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 $(\langle \mathbf{x}, \mathbf{y} \rangle + \mathbf{k})|_{\mathbf{y} \in B}$ 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. At a high level, these boil down to showing that if $\mathbf{x}, \mathbf{y} \in \mathbb{Z}_p^2$ are two random independent variables, uniform over large sets, then their inner product $\langle \mathbf{x}, \mathbf{y} \rangle$ has well-behaved spectral properties.

Paper organization. We give preliminary definitions needed for the proof in Section 2. We discuss the structure of matrices of sign-rank 1 and 2 in Section 3. We prove our main result, Theorem 1.9, in Section 4.

2 Preliminaries

Notation. 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 $N \le M$ 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 1.9 is based on Fourier analysis over cyclic groups. Next we introduce the relevant notation. 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 $\mathbf{e}_p : \mathbb{Z}_p \to \mathbb{C}$ denote the function $\mathbf{e}_p : x \mapsto \mathrm{e}^{2\pi i x/p}$. For $a \in \mathbb{Z}_p$ define the character $\chi_a : x \mapsto \mathbf{e}_p(-ax)$. The Fourier expansion of $f : \mathbb{Z}_p \to \mathbb{C}$ is the sum

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

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

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

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

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

Number theory estimates. Fix a prime p. 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

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

which follows from the easy estimate that $4|y| \le |e^{2\pi i y} - 1| \le 8|y|$ for $y \in [-1/2, 1/2]$, and taking $y = \frac{\operatorname{sign}(x)\|x\|_p}{p}$.

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 $A_{i,j} = \operatorname{sign}(a_i b_j)$. 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 submatrices. This implies that $\operatorname{Disc}(A) = \Omega(1)$.

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 $A_{i,j} = \operatorname{sign}(\langle u_i, v_j \rangle)$. 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 submatrix 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 submatrix, $A_{i,j} = \operatorname{sign}(a_i - b_j)$. By removing repeated rows and columns, we get that the submatrix is an upper triangular matrix, with 1 on or above the diagonal and -1 below the diagonal. That is, the submatrix is equivalent to the matrix corresponding to the Greater-Than boolean function on at most n bits. Nisan [21] 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 proving Theorem 1.9. 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 $f((x;z), \mathbf{y}) = 1$; 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:

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

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. *For all $\tilde{N}$ such that $\tilde{N} \approx N$,*

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

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*

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

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 $\mu_{\langle x,\mathbf{y}\rangle}$:

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

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

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

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

With the Fourier coefficients $\widehat{\nu_x^B}(a)$ computed in Lemma 4.2, we can analyze the distance from $\nu_{\mathbf{x}}^B$ to 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 $\widehat{\nu_{\mathbf{x}}^B}(a)$,

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

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

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

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

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

and

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

both of which follow from trivial upper bounds $\|Na\|_p \leq N\,\|a\|_p$ and $\|x\|_p \leq p \approx N^2$. Let us denote $E_a(B) := \mathbb{E}_{\mathbf{x}}\left|\mathbb{E}_{\mathbf{y}\sim B}\left[\mathbf{e}_p(a\langle \mathbf{x},\mathbf{y}\rangle)\right]\right|^2$. We break the sum into two parts and for each part use a different estimate for $E_a(B)$ using Lemma 4.3 below.

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

□

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

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

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

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

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

$$\begin{aligned}
E_a(B) &\le \frac{1}{N^2|B|}\sum_{y\in[-N,N]^2}\left|\sum_{x\in[N]^2}\chi_a(\langle x,y\rangle)\right| \\
&= \frac{1}{N^2|B|}\sum_{y_1,y_2\in[-N,N]}\left|\sum_{x_1,x_2\in[N]}\chi_a(x_1y_1)\cdot\chi_a(x_2y_2)\right| \\
&= \frac{1}{N^2|B|}\sum_{y_1,y_2\in[-N,N]}\left|\sum_{x_1\in[N]}\chi_a(x_1y_1)\right|\left|\sum_{x_2\in[N]}\chi_a(x_2y_2)\right| \\
&= \frac{1}{N^2|B|}\left(\sum_{y\in[-N,N]}\left|\sum_{x\in[N]}\chi_a(xy)\right|\right)^2 \\
&\lesssim \frac{1}{N^2|B|}\left(\sum_{y\in[0,N]}\left|\sum_{x\in[N]}\chi_a(xy)\right|\right)^2.
\end{aligned}$$

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

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

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*

$$\sum_{y \in [N]} \frac{\|Nay\|_p}{\|ay\|_p} \lesssim \frac{p \log p}{\min(N, \|a\|_p)}.$$

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.

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

□

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 $\|Nay\|_p \leq p$. We get

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

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

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

Overall we get

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

The case $\|a\|_p < N$: Here we use the estimate $|S_k| \lesssim \frac{k}{\|a\|_p}$. Also similar to the previous case, we bound $\frac{\|Nay\|_p}{\|ay\|_p} \leq \frac{p}{\|ay\|_p}$. We get

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

So we have

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

The lemma follows. □

We remark that the following more general statement regarding uniformity of product sets follows by an argument similar to the proof of 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

[1] NOGA ALON, SHAY MORAN, AND AMIR YEHUDAYOFF: Sign rank versus Vapnik–Chervonenkis dimension. *Sbornik Math.*, 208(12):1724–1757, 2017. Preliminary version in COLT'16:PMLR. [doi:10.1070/SM8780] 7

[2] LÁSZLÓ BABAI, PETER FRANKL, AND JANOS SIMON: Complexity classes in communication complexity theory. In *Proc. 27th FOCS*, pp. 337–347. IEEE Comp. Soc., 1986. [doi:10.1109/SFCS.1986.15] 6, 7

[3] LÁSZLÓ BABAI, NOAM NISAN, AND MÁRIÓ SZEGEDY: Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. Comput. System Sci.*, 45(2):204–232, 1992. Preliminary version in STOC'89. 2

[4] RONEN BASRI, PEDRO F. FELZENSZWALB, ROSS B. GIRSHICK, DAVID W. JACOBS, AND CAROLINE J. KLIVANS: Visibility constraints on features of 3D objects. In *Proc. Conf. Comp. Vision and Pattern Recog. (CVPR'09)*, pp. 1231–1238. IEEE Comp. Soc., 2009. [doi:10.1109/CVPR.2009.5206726] 7

[5] AMEY BHANGALE AND SWASTIK KOPPARTY: The complexity of computing the minimum rank of a sign pattern matrix, 2015. [arXiv:1503.04486] 7

[6] THOMAS A. BROWN AND JOEL H. SPENCER: Minimization of ± 1 matrices under line shifts. *Colloquium Mathematicum*, 23:165–171, 1971. [doi:10.4064/cm-23-1-165-171] 2

[7] HARRY BUHRMAN, NIKOLAY VERESHCHAGIN, AND RONALD DE WOLF: On computation and communication with small bias. In *Proc. 22nd IEEE Conf. on Comput. Complexity (CCC'07)*, pp. 24–32. IEEE Comp. Soc., 2007. [doi:10.1109/CCC.2007.18] 7

[8] MARK BUN AND JUSTIN THALER: Improved bounds on the sign-rank of AC^0. In *Proc. 43rd Internat. Colloq. on Automata, Languages, and Programming (ICALP'16)*, pp. 37:1–14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016. [doi:10.4230/LIPIcs.ICALP.2016.37, ECCC:TR16-075] 2

[9] ARKADEV CHATTOPADHYAY AND TONIANN PITASSI: The story of set disjointness. *SIGACT News*, 41(3):59–85, 2010. [doi:10.1145/1855118.1855133] 2, 3

[10] BERNARD CHAZELLE: *The Discrepancy Method: Randomness and Complexity*. Cambridge Univ. Press, 2000. 2

[11] BENNY CHOR AND ODED GOLDREICH: Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.*, 17(2):230–261, 1988. Preliminary version in FOCS'85. [doi:10.1137/0217015] 2, 6

[12] JOHANNES G. VAN DER CORPUT: Verteilungsfunktionen I (Distribution functions, German). *Proc. Akad. Wetenschappen Amsterdam*, 38:813–821, 1935. 2

[13] PAUL ERDŐS AND JOEL H. SPENCER: *Probabilistic Methods in Combinatorics*. Akadémiai Kiadó, Budapest, and Academic Press, 1971. 2

[14] JÜRGEN FORSTER: A linear lower bound on the unbounded error probabilistic communication complexity. *J. Comput. System Sci.*, 65(4):612–625, 2002. Preliminary version in CCC'01. [doi:10.1016/S0022-0000(02)00019-3] 4

[15] ANNA GÁL AND RIDWAN SYED: Upper bounds on communication in terms of approximate rank. In *Proc. Comp. Sci. Symp. in Russia (CSR'21)*, pp. 116–130. Springer, 2021. [doi:10.1007/978-3-030-79416-3_7, ECCC:TR19-006] 7

[16] HAMED HATAMI, KAAVE HOSSEINI, AND SHACHAR LOVETT: Sign rank vs discrepancy. In *Proc. 35th Comput. Complexity Conf. (CCC'20)*, pp. 18:1–14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2020. [doi:10.4230/LIPIcs.CCC.2020.18, ECCC:TR19-067] 1

[17] HARTMUT KLAUCK: Lower bounds for quantum communication complexity. *SIAM J. Comput.*, 37(1):20–46, 2007. Preliminary version in FOCS'01. [doi:10.1137/S0097539702405620] 6

[18] ADAM R. KLIVANS AND ROCCO A. SERVEDIO: Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. *J. Comput. System Sci.*, 68(2):303–318, 2004. Preliminary version in STOC'01. [doi:10.1016/j.jcss.2003.07.007] 2

[19] NATI LINIAL, SHAHAR MENDELSON, GIDEON SCHECHTMAN, AND ADI SHRAIBMAN: Complexity measures of sign matrices. *Combinatorica*, 27(4):439–463, 2007. [doi:10.1007/s00493-007-2160-5] 2, 5

[20] NATI LINIAL AND ADI SHRAIBMAN: Learning complexity vs. communication complexity. *Combin. Probab. Comput.*, 18(1–2):227–245, 2009. Preliminary version in CCC'08. [doi:10.1017/S0963548308009656] 2, 4

[21] NOAM NISAN: The communication complexity of threshold gates. In D. MIKLÓS, T. SZŐNYI, AND V. T. SÓS, editors, *Combinatorics, Paul Erdős is Eighty*, volume 1, pp. 301–315. Bolyai Society, Budapest and North-Holland, 1993. Author's website. 3, 11

[22] RAMAMOHAN PATURI AND JANOS SIMON: Probabilistic communication complexity. *J. Comput. System Sci.*, 33(1):106–123, 1986. Preliminary version in FOCS'84. [doi:10.1016/0022-0000(86)90046-2] 2, 6

[23] ALEXANDER A. RAZBOROV AND ALEXANDER A. SHERSTOV: The sign-rank of AC^0. *SIAM J. Comput.*, 39(5):1833–1855, 2010. Preliminary version in FOCS'08. [doi:10.1137/080744037, ECCC:TR08-016] 2

[24] ALEXANDER A. SHERSTOV: Halfspace matrices. *Comput. Complexity*, 17(2):149–178, 2008. Preliminary version in CCC'07. [doi:10.1007/s00037-008-0242-4] 7

[25] ALEXANDER A. SHERSTOV: The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011. [doi:10.1137/080733644, arXiv:0906.4291] 7

[26] ALEXANDER A. SHERSTOV: Optimal bounds for sign-representing the intersection of two halfspaces by polynomials. *Combinatorica*, 33(1):73–96, 2013. Preliminary version in STOC'10. [doi:10.1007/s00493-013-2759-7, arXiv:0910.4224, ECCC:TR10-025] 7

[27] ALEXANDER A. SHERSTOV: The hardest halfspace. *Comput. Complexity*, 30(2):11, 2021. [doi:10.1007/s00037-021-00211-4, arXiv:1902.01765, ECCC:TR19-016] 3, 7

[28] JUSTIN THALER: Lower bounds for the approximate degree of block-composed functions. In *Proc. 43rd Internat. Colloq. on Automata, Languages, and Programming (ICALP'16)*, pp. 17:1–15. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016. [doi:10.4230/LIPIcs.ICALP.2016.17, ECCC:TR14-150] 7

[29] HUGH E. WARREN: Lower bounds for approximation by nonlinear manifolds. *Trans. AMS*, 133(1):167–178, 1968. [doi:10.1090/S0002-9947-1968-0226281-1] 2

AUTHORS

Hamed Hatami
Associate professor
Department of Computer Science
McGill University
Montreal, Quebec, Canada
hatami@cs.mcgill.edu
https://www.cs.mcgill.ca/~hatami/

Kaave Hosseini
Assistant professor
Department of Computer Science
University of Rochester
Rochester, New York, USA
kaave.hosseini@rochester.edu
https://www.cs.rochester.edu/u/shossei2/

Shachar Lovett
Associate professor
Department of Computer Science
University of California San Diego
San Diego, California, USA
slovett@ucsd.edu
https://cseweb.ucsd.edu/~slovett/home.html

ABOUT THE AUTHORS

HAMED HATAMI received his Ph. D. from the Department of Computer Science at the University of Toronto in 2009 under Michael Molloy and Balázs Szegedy. Afterward, he became a Veblen fellow at the Institute for Advanced Study and the Department of Mathematics at Princeton University until 2010. Since 2010 he has been on the faculty at the School of Computer Science at McGill University. His research focuses on the applications of mathematical analysis to theoretical computer science and combinatorics.

KAAVE HOSSEINI received his Ph. D. from the Department of Computer Science and Engineering at University of California, San Diego in 2019 under the supervision of Shachar Lovett. Then he was a postdoctoral associate in the Department of Mathematical Sciences at Carnegie Mellon University until 2021. Since 2021, he has been on the faculty in the Department of Computer Science at the University of Rochester. His research interests are in additive combinatorics, pseudo-randomness, and their applications in algorithms and computational complexity.

SHACHAR LOVETT received his Ph. D. from the Weizmann Institute of Science in 2010 under the supervision of Omer Reingold and Ran Raz. He was a postdoctoral researcher at the Institute for Advanced Study until 2012. Since 2012, he has been on the faculty at the University of California, San Diego. He is a recipient of an NSF CAREER award and a Sloan fellowship. His research is broadly in theoretical computer science and combinatorics. In particular: computational complexity, randomness and pseudo-randomness, algebraic constructions, coding theory, additive combinatorics and combinatorial aspects of high-dimensional geometry. He is happily married and has four kids.