Electronic Colloquium on Computational Complexity, Report No. 50 (2023)

Communication complexity of half-plane membership

Manasseh Ahmed* Tsun-Ming Cheung† Hamed Hatami ‡ Kusha Sareen §

April 18, 2023

Abstract

We study the randomized communication complexity of the following problem. Alice receives the *integer* coordinates of a point in the plane, and Bob receives the *integer* parameters of a half-plane, and their goal is to determine whether Alice's point belongs to Bob's half-plane.

This communication task corresponds to determining whether $x_1y_1 + y_2 \geq x_2$, where the first player knows $(x_1, x_2) \in [n]^2$ and the second player knows $(y_1, y_2) \in [n]^2$. We prove that its randomized communication complexity is $\Omega(\log n)$.

Our lower bound extends a recent result of Hatami, Hosseini, and Lovett (CCC '20 and ToC '22) regarding the largest possible gap between sign-rank and randomized communication complexity.

1 Introduction

We study the randomized communication complexity of the following communication task. Let $\mathcal{P}$ be a finite set of points in the plane, and let $\mathcal{H}$ be a finite set of half-planes. Alice receives a point in $\mathcal{P}$, and Bob receives a half-plane in $\mathcal{H}$, and their goal is to determine whether Alice's point belongs to Bob's half-plane. We refer to this communication problem as the *half-plane membership problem.*

We represent every point in $\mathcal{P}$ by its coordinates $(x_1, x_2) \in \mathbb{R}^2$. Similarly, we represent every half-plane in $\mathcal{H}$ by a pair $(y_1, y_2) \in \mathbb{R}^2$, corresponding to the half-plane

$$H_{y_1,y_2} \coloneqq \{(z_1, z_2) \in \mathbb{R}^2 \ : \ y_1z_1 + y_2 \geq z_2\}.$$

We show that the randomized communication complexity of the half-plane membership problem is large, even if the points and half-planes are chosen from $[n]^2$, where $[n] \coloneqq \{1, \dots, n\}$.

Theorem 1.1. *The randomized communication complexity of the half-plane membership problem is* $\Omega(\log n)$ *when*

$$\mathcal{P} \coloneqq \{(x_1, x_2) \ : (x_1, x_2) \in [n]^2\} \qquad \text{and} \qquad \mathcal{H} \coloneqq \{H_{y_1,y_2} \ : \ (y_1, y_2) \in [n]^2\}. \tag{1}$$

Note that the lower bound of Theorem 1.1 matches the trivial upper bound of $O(\log n)$, which is witnessed by the (deterministic) protocol where Alice sends her input to Bob, and Bob replies with the output.

*Marianopolis College. Email: `manassehahmed@gmail.com`.
†School of Computer Science, McGill University. Email: `tsun.ming.cheung@mail.mcgill.ca`.
‡School of Computer Science, McGill University. Email: `hatami@cs.mcgill.ca`. Supported by an NSERC grant.
§School of Computer Science, McGill University. Email: `kushagra.sareen@mail.mcgill.ca`.

ISSN 1433-8092

1.1 Connection to Hatami, Hosseini, and Lovett [HHL22]

A recent work by Hatami, Hosseini, and Lovett [HHL22] considers the following communication problem based on points and half-spaces in dimension *three*: Alice receives $(x_1, x_2, x_3) \in [n]^3$ and Bob receives $(y_1, y_2) \in [n]^2$, and their goal is to determine whether $x_1y_1 + x_2y_2 \geq x_3$. They prove that the randomized communication complexity of this problem is $\Omega(\log n)$.

We can translate the above problem into a half-plane membership problem as follows: $x_1y_1 + x_2y_2 \geq x_3$ iff the point $p = (x_1/x_2, x_3/x_2)$ belongs to the half-plane H_{y_1,y_2}. Therefore, the result of [HHL22] says that the randomized communication complexity of the half-plane membership problem is large when

$$\mathcal{P} = \{(x_1/x_2, x_3/x_2) \ : (x_1, x_2, x_3) \in [n]^3\} \qquad \text{and} \qquad \mathcal{H} = \{H_{y_1,y_2} \ : \ (y_1, y_2) \in [n]^2\}. \tag{2}$$

Theorem 1.1 extends this lower bound to the more natural setting where the points and half-planes are chosen from the integer lattice. A few remarks are in order.

- The half-plane membership problem of Theorem 1.1 corresponds to determining whether $x_1y_1 + y_2 \geq x_2$, where Alice knows $(x_1, x_2) \in [n]^2$ and Bob knows $(y_1, y_2) \in [n]^2$.

  Theorem 1.1 is an extension of the result of [HHL22] as the set of points and half-planes in Theorem 1.1 are subsets of those in Eq. (2). Indeed the half-plane membership problem of Eq. (1) is obtained by restricting to $x_2 = 1$ in $x_1y_1 + x_2y_2 \geq x_3$.

- The proof of Theorem 1.1 follows the general proof strategy of [HHL22]. Both proofs use Fourier analysis of the cyclic group and various estimates of partial exponential sums. However, a key step of bounding the discrepancy of Q in [HHL22] relies crucially on the mixing property of the function $x_1y_1 + x_2y_2$. For the matrix P, the corresponding function $x_1y_1 + y_2$ lacks those desirable properties, and this key step fails when applied to our problem.

  The differences between the mixing properties of $x_1y_1 + x_2y_2$ and $x_1y_1 + y_2$ initially seemed a serious barrier to extending the proof of [HHL22] to Theorem 1.1, and raised some doubts among the authors that perhaps the randomized communication complexity of the half-plane membership problem of Eq. (1) is small. Eventually, we circumvented the broken step in the proof of [HHL22] by an averaging argument based on the fact that the L_1 sum of the Fourier coefficients of the convolution of two Boolean functions is always at most 1.

- Finally, we simplify some parts of the proof that are common to both Theorem 1.1 and [HHL22]. In this sense, Theorem 1.1 not only strengthens the result of [HHL22] but also provides a shorter and simpler proof. We explain the differences between the two proofs in more detail in Section 4.

1.2 Discrepancy

We prove the lower bound of Theorem 1.1 by employing the *discrepancy* method, one of the most commonly used lower bound methods in communication complexity theory.

A *sign matrix* is a matrix with ± 1 entries. The discrepancy of a sign matrix measures how balanced its submatrices are. Formally, the *discrepancy* of a sign matrix $F_{\mathcal{X}\times\mathcal{Y}}$ with respect to a probability distribution μ on $\mathcal{X} \times \mathcal{Y}$ is

$$\mathrm{Disc}_\mu(F) \coloneqq \max_{\substack{A\subseteq\mathcal{X}\\ B\subseteq\mathcal{Y}}} \mathrm{Disc}_\mu^{A\times B}(F), \tag{3}$$

where

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

The *discrepancy* of F, denoted by $\mathrm{Disc}(F)$, is the minimum of $\mathrm{Disc}_{\mu}(F)$ over all probability distributions μ.

The combinatorial parameter of discrepancy is closely related to the complexity of randomized communication protocols. Chor and Goldreich [CG88] proved that for every $0 < \epsilon < 1/2$,

$$\mathrm{R}_{\epsilon}(F) \geq \log \frac{1-2\epsilon}{\mathrm{Disc}(F)}, \tag{4}$$

where $\mathrm{R}_{\epsilon}(F)$ denotes the randomized communication complexity of F with error ϵ in the shared randomness model (See [KN97, Section 3] for the precise definition).

Every $n \times n$ sign matrix F satisfies $\mathrm{R}_{\epsilon}(F) \leq 1 + \log n$ and $\mathrm{Disc}(F) \geq \Omega(1/\sqrt{n})$. The first inequality follows from the trivial (deterministic) protocol where Alice sends her input to Bob, and Bob replies with the output. We refer readers to [LS09, Observation 1.1] for the second inequality.

The following theorem, which immediately implies Theorem 1.1, shows the half-plane membership problem of Theorem 1.1 essentially matches these worst-case bounds.

Theorem 1.2 (Main theorem)**.** *Let $n = m^3$ be positive integers and consider the matrix $P_{n\times n}$, whose rows and columns are indexed by $[m] \times [m^2]$, and*

$$P([x_1, x_2], [y_1, y_2]) = \begin{cases} 1 & \text{if } x_1y_1 + y_2 \geq x_2 \\ -1 & \text{otherwise} \end{cases}. \tag{5}$$

We have

$$\mathrm{Disc}(P) = O(n^{-1/6}\log^{3/2} n) \qquad \text{and} \qquad \mathrm{R}_{1/3}(P) = \Theta(\log n).$$

In view of the equivalence of discrepancy and margin, proved by Linial and Shraibman [LS09], Theorem 1.2 has a geometric interpretation: while the matrix P is representable in dimension two as points and half-planes, the normalized margin of the point-halfspace representation of P in any dimension is small. We refer the reader to [HHL22, Section 1.1] and [LS09] for the definition of margin and more details on this interpretation.

We remark that it is essential to have the half-planes in $\mathcal{H}$ not limited to *homogeneous half-planes*, which are the half-planes defined by lines that pass through the origin. Indeed, limiting to homogeneous half-planes results in the communication problem $x_1y_1 \geq x_2$, which is equivalent to $y_1 > x_2/x_1$. Since Alice has full information of x_2/x_1 and Bob has full information of y_1, this reduces to an instance of the so-called Greater-than communication problem. Nisan [Nis93] showed that the randomized communication complexity of the $n\times n$ Greater-than problem is $O(\log\log n)$. Moreover, Braverman and Weinstein [BW16] proved that the discrepancy of this matrix is $\Omega(1/\sqrt{\log n})$.

1.3 Sign-rank versus Discrepancy

The *sign-rank* of a sign matrix $A_{m\times n}$, denoted by $\mathrm{rank}_{\pm}(A)$, is the smallest rank of a real matrix $B_{m\times n}$ such that the entries of B are nonzero and have the same signs as their corresponding entries in A. The notion of sign-rank was introduced in 1986 in connection with randomized communication complexity in the unbounded-error model of Paturi and Simon [PS86]. This fundamental notion arises naturally in areas as diverse as learning theory, discrete geometry and geometric graphs, communication complexity, circuit complexity, and the theory of Banach spaces (see [HHP$^+$22] and the references therein).

The pioneering paper of Babai, Frankl, and Simon [BFS86], which introduced communication complexity classes, initiated a line of research investigating the gap between two fundamental notions in communication complexity, namely sign-rank and discrepancy. This separation question was posed in [BFS86] in the equivalent form of separating the two communication complexity classes $\mathbf{PP}^{\mathrm{cc}}$ and $\mathbf{UPP}^{\mathrm{cc}}$, i.e., *weakly-unbounded-error* and *unbounded-error* communication complexity classes. We will not define the complexity classes and the related measures here, and we refer the reader to [HHL22] for a more comprehensive discussion of these connections.

The question of Babai, Frankl and Simon [BFS86] remained unanswered for over two decades. Finally, Buhrman et al. [BVdW07] and independently Sherstov [She08b] showed that there are $n \times n$ sign matrices with $\mathbf{rk}_{\pm}(F) = O(\log n)$ but $\mathrm{Disc}(F) = 2^{-\log^{\Omega(1)}(n)}$. This separation was enhanced along a subsequent line of works [She11, She13, Tha16, She19] to $\mathbf{rk}_{\pm}(F) = O(\log n)$ and $\mathrm{Disc}(F) = n^{-\Omega(1)}$ of [She19].

Recently, [HHL22] improved the separation to $\mathbf{rk}_{\pm}(F) = 3$ and $\mathrm{Disc}(F) = O(n^{-1/8}\log n)$. The sign-rank 3 of this separation is tight since every sign matrix of sign-rank 2 consists of a few copies of the Greater-Than matrix, and thus, by the result of Braverman and Weinstein [BW16], has discrepancy $\Omega(1/\sqrt{\log n})$.

Notice that the matrix P in Theorem 1.2 also has sign-rank 3 and it provides a slightly stronger upper bound on the discrepancy.

1.4 Discrepancy with respect to product measures

Sign matrices with *sub-logarithmic* sign-rank inherit interesting structural properties from low dimensional geometry. For example, Alon, Pach, Pinchasi, and Sharir [APP$^+$05, Theorem 1.3] proved that if $F_{n\times n}$ is a matrix with sign-rank d, then F contains a large monochromatic rectangle. It follows that for such a matrix, for every *product measure* $\lambda \times \nu$ (where λ and ν are probability measures over rows and columns, respectively), we have

$$\mathrm{Disc}_{\lambda\times\nu}(F) \geq \frac{1}{2^{2d+2}}.$$

This is a meaningful lower bound when $d = o(\log n)$. It is particularly interesting to contrast this result with Theorem 1.2. As the matrix P of Theorem 1.2 has sign-rank 3, it satisfies that

$$\inf_{\lambda\times\nu} \mathrm{Disc}_{\lambda\times\nu}(P) \geq 2^{-8},$$

while Theorem 1.2 shows if we allow the infimum to include non-product measures, then

$$\inf_{\mu} \mathrm{Disc}_{\mu}(P) \leq O(n^{-1/6}\log^{3/2} n).$$

From the communication complexity perspective, the above observations lead to another example that separates (general) distributional complexity and product distributional complexity.

For a distribution μ, the *μ-distributional complexity* of F, denoted by $\mathrm{D}^{\mu}_{\epsilon}(F)$, is the least cost of a deterministic protocol that computes F on input sampled from μ with error probability at most ϵ. Yao's minimax principle [Yao83] states that the randomized communication complexity is exactly the maximum distributional complexity. Therefore, by Theorem 1.2, one has

$$\max_{\mu} \mathrm{D}^{\mu}_{1/3}(P) = \Theta(\log n).$$

On the other hand, for any sign matrix F and product distribution $\lambda \times \nu$, [KNR95] proved that

$$\mathrm{D}_{\epsilon}^{\lambda\times\nu}(F) = O\left(\frac{1}{\epsilon}\,\mathrm{VC}(F)\log\frac{1}{\epsilon}\right),$$

where $\mathrm{VC}(F)$ denotes the Vapnik-Chervonenkis (VC) dimension of F. It is well known that the sign-rank upper bounds the VC dimension (see [HHP$^+$22]). Therefore, in the case of the constant sign-rank matrix P, one can deduce that

$$\max_{\lambda\times\nu} \mathrm{D}_{1/3(P)}^{\lambda\times\nu} = O(1).$$

Consequently, Theorem 1.1 recovers the $O(1)$-versus-$\Omega(\log n)$ separation between general distributional complexity and product distributional complexity proven by Sherstov [She08a].

2 Preliminaries

Notations. To simplify the presentation, we often use $\lesssim$ or $\approx$ instead of the big-O notation whenever the constants are unimportant. That is, $x \lesssim y$ means $x = O(y)$, and $x \approx y$ means $x = \Theta(y)$. For integers $s < t$, we denote $[s, t] = \{s, \dots, t\}$, and we shorthand $[s] = [1, s]$.

For a random variable r, we denote $\mu = \mu_r$ the distribution of r. For a finite set S, we write $r \sim S$ to indicate that r is uniformly sampled from S.

Fourier analysis. We introduce the relevant notations and fundamental results in Fourier analysis over cyclic groups, the primary tool for the proof of our main result. Let p be a prime. For $f, g : \mathbb{Z}_p \to \mathbb{C}$, define the inner product by

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

Let $\mathbf{e}_p : \mathbb{Z}_p \to \mathbb{C}$ denote the exponentiation by a p-th root of unity, that is $\mathbf{e}_p : x \mapsto e^{2\pi i x/p}$. For $a \in \mathbb{Z}_p$, define the character function $\chi_a : x \mapsto \mathbf{e}_p(-ax)$. Note that $\{\chi_a : \ a \in \mathbb{Z}_p\}$ forms an orthonormal basis with respect to the inner product defined above.

The Fourier expansion of $f : \mathbb{Z}_p \to \mathbb{C}$ is given by

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

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

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

A fundamental identity of Fourier analysis is Parseval's identity:

$$\sum_{a\in\mathbb{Z}_p} |\widehat{f}(a)|^2 = \mathop{\mathbb{E}}_{x\in\mathbb{Z}_p} |f(x)|^2.$$

The convolution of two functions $f, g : \mathbb{Z}_p \to \mathbb{C}$ is defined to be

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

From the orthonormality of characters, it follows that

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

in other words, $\widehat{f * g}(a) = \widehat{f}(a)\widehat{g}(a)$. In particular, if $x_1, \ldots, x_k$ are independent random variables taking values in $\mathbb{Z}_p$, and then the Fourier coefficient of the distribution of the random variable $x \coloneqq x_1 + \ldots + x_k$ is

$$\widehat{\mu_x}(a) = p^{k-1} \prod_{i=1}^{k} \widehat{\mu_{x_i}}(a).$$

Number theory estimates. Fix a prime p. For $x \in \mathbb{Z}$, denote by $|x|_p$ the minimum distance of x to a multiple of p, that is

$$|x|_p = \min\{|x - pk| : \ k \in \mathbb{Z}\}.$$

We will often use the estimate

$$\frac{4|x|_p}{p} \leq |\mathbf{e}_p(x) - 1| \leq \frac{8|x|_p}{p},$$

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

3 Proof of Theorem 1.2

Let m be sufficiently large and set $\mathcal{X} = [m] \times [m^2]$. The matrix P is an $\mathcal{X} \times \mathcal{X}$ matrix.

Construction of hard distribution. We introduce a distribution μ on $\mathcal{X} \times \mathcal{X}$ by sampling $(x_1, x_2, y_1, y_2) \in \mathcal{X} \times \mathcal{X}$ as follows.

- Select $x_1, y_1 \sim [m/2]$, $y_2 \sim \left[m^2/4, m^2/2\right]$ uniformly and independently.
- Let $t = \lfloor 10 \log m \rfloor$. Select $k_1, \ldots, k_t \sim [20m]$ uniformly and independently and set $k = k_1 + \cdots + k_t$. Set $x_2 = x_1 y_1 + y_2 + k$ or $x_2 = x_1 y_1 + y_2 + k - 20mt$, each with probability $1/2$.

Assuming m is sufficiently large, we have $0 < x_2 \leq m^2$ and thus μ is indeed supported on $\mathcal{X} \times \mathcal{X}$.

To make the presentation cleaner, instead of analyzing μ directly, we work with a similar measure on the extended domain $\mathbb{Z}^2 \times \mathbb{Z}^2$. We also extend the definition of P in Eq. (5) to $\mathbb{Z} \times \mathbb{Z}$.

We introduce a distribution ν on $\mathbb{Z}^2 \times \mathbb{Z}^2$ by sampling (x_1, x_2, y_1, y_2) as follows:

- Select $x_1, y_1 \sim [m]$, $y_2 \sim [m^2]$ uniformly and independently.
- Select $k_1, \ldots, k_t \sim [20m]$ uniformly and independently and set $k = k_1 + \ldots + k_t$. Set $x_2 = x_1 y_1 + y_2 + k$ or $x_2 = x_1 y_1 + y_2 + k - 20mt$, each with probability $1/2$. Note that in the former case, $x_1 y_1 + y_2 < x_2$ and in the latter case, $x_1 y_1 + y_2 \geq x_2$.

Let $(x_1, x_2, y_1, y_2) \sim \nu$ and consider the event

$$\mathcal{S} \coloneqq \left\{(x_1, x_2, y_1, y_2) \mid x_1, y_1 \in [m/2] \text{ and } y_2 \in \left[m^2/4, m^2/2\right]\right\}.$$

The distribution μ, defined earlier, is ν conditioned on $\mathcal{S}$.

Consider $A, B \subseteq \mathcal{X}$, and let A' and B' be A and B restricted to $\mathcal{S}$, that is

$$A' = \{(x_1, x_2) \in A \mid x_1 \le m/2\} \subseteq A,$$

and

$$B' = \left\{(y_1, y_2) \in B \mid y_1 \le m/2 \text{ and } y_2 \in \left[m^2/4, m^2/2\right]\right\} \subseteq B.$$

We shorthand $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$. By the definition of μ, we have

$$\begin{aligned}
\mathrm{Disc}_\mu^{A\times B}(P) &= |\,\mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mu}[P(\mathbf{x},\mathbf{y})\mathbf{1}_{A'}(\mathbf{x})\mathbf{1}_{B'}(\mathbf{y})]| = \frac{1}{\Pr_\nu[\mathcal{S}]}|\,\mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\nu}[P(\mathbf{x},\mathbf{y})\mathbf{1}_{A'}(\mathbf{x})\mathbf{1}_{B'}(\mathbf{y})]| \\
&= 16|\,\mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\nu}[P(\mathbf{x},\mathbf{y})\mathbf{1}_{A'}(\mathbf{x})\mathbf{1}_{B'}(\mathbf{y})]| = 16\,\mathrm{Disc}_\nu^{A'\times B'}(P).
\end{aligned}$$

Therefore, it suffices to show that for every $A, B \subseteq \mathcal{X}$, we have

$$\mathrm{Disc}_\nu^{A\times B}(P) = O(m^{-1/2}\log^{3/2} m).$$

The rest of the proof of Theorem 1.2 is dedicated to proving this bound.

Invariance under shift. For every $x_1 \in [m]$, define $A_{x_1} = \{x_2 : (x_1, x_2) \in A\}$. We have

$$\begin{aligned}
\mathrm{Disc}_\nu^{A\times B}(P) &= \left|\mathbb{E}_{x_1\sim[m]}\,\mathbb{E}_{\mathbf{y}\sim[m]\times[m^2]}\left[\mathbf{1}_B(\mathbf{y})\,\mathbb{E}_{x_2|x_1,\mathbf{y}}[P(\mathbf{x},\mathbf{y})\mathbf{1}_{A_{x_1}}(x_2)]\right]\right| \\
&= \frac{|B|}{m^3}\left|\mathbb{E}_{x_1\sim[m]}\,\mathbb{E}_{\mathbf{y}\sim B}\,\mathbb{E}_{x_2|x_1,\mathbf{y}}[P(\mathbf{x},\mathbf{y})\mathbf{1}_{A_{x_1}}(x_2)]\right| \\
&= \frac{|B|}{2m^3}\left|\mathbb{E}_{x_1\sim[m],\mathbf{y}\sim B,k}[\mathbf{1}_{A_{x_1}}(x_1y_1+y_2+k) - \mathbf{1}_{A_{x_1}}(x_1y_1+y_2+k-20mt)]\right|.
\end{aligned}$$

Here, the last line follows from the definition of x_2 and ν.

Let $\nu_{x_1}^B$ denote the distribution of $x_1y_1 + y_2 + k$ conditioned on the value of x_1 and the event $(y_1, y_2) \in B$. Note that $\nu_{x_1}^B$ is supported on $[0, 3m^2]$. We embed this distribution into $\mathbb{Z}_p$ for some prime $p \in [4m^2, 5m^2]$. With this notation, we can rewrite

$$\begin{aligned}
\mathrm{Disc}_\nu^{A\times B}(P) &= \frac{|B|}{2m^3}\left|\mathbb{E}_{x_1}\,\mathbb{E}_{w\sim\nu_{x_1}^B}[\mathbf{1}_{A_{x_1}}(w) - \mathbf{1}_{A_{x_1}}(w-20mt)]\right| \\
&= \frac{|B|}{2m^3}\left|\mathbb{E}_{x_1}\sum_{w\in\mathbb{Z}}[\mathbf{1}_{A_{x_1}}(w)\nu_{x_1}^B(w) - \mathbf{1}_{A_{x_1}}(w-20mt)\nu_{x_1}^B(w)]\right| \\
&= \frac{|B|}{2m^3}\left|\mathbb{E}_{x_1}\sum_{w\in\mathbb{Z}}[\mathbf{1}_{A_{x_1}}(w)\nu_{x_1}^B(w) - \mathbf{1}_{A_{x_1}}(w)\nu_{x_1}^B(w+20mt)]\right| \\
&\le \frac{|B|}{2m^3}\,\mathbb{E}_{x_1}\sum_{w\in\mathbb{Z}}\left|\nu_{x_1}^B(w) - \nu_{x_1}^B(w+20mt)\right| \\
&= \frac{|B|}{2m^3}\,\mathbb{E}_{x_1}\sum_{w\in\mathbb{Z}_p}\left|\nu_{x_1}^B(w) - \nu_{x_1}^B(w+20mt)\right| \\
&\lesssim \frac{|B|}{m}\,\mathbb{E}_{x_1}\,\mathbb{E}_{w\sim\mathbb{Z}_p}\left|\nu_{x_1}^B(w) - \nu_{x_1}^B(w+20mt)\right|.
\end{aligned}$$

The above analysis shows that in order to prove that $\mathrm{Disc}_\nu^{A\times B}(P)$ is small, we need to show that typically $\nu_{x_1}^B$ is almost invariant under a shift of $20mt$.

Fourier Expansion of $\nu_{x_1}^B$. In order to analyze the shift-invariance of $\nu_{x_1}^B$, we examine the Fourier expansion of $\nu_{x_1}^B(w)$ as a function on $\mathbb{Z}_p$.

Lemma 3.1. *For a fixed x_1, for every $a \in \mathbb{Z}_p \setminus \{0\}$,*

$$\widehat{\nu_{x_1}^B}(a) = \frac{1}{p}\mathbf{e}_p(ta)\left(\frac{1}{20m}\frac{\mathbf{e}_p(20ma)-1}{\mathbf{e}_p(a)-1}\right)^t \mathbb{E}_{\mathbf{y}\sim B}[\mathbf{e}_p(x_1y_1+y_2)].$$

Proof. For the fixed x_1, denote by η the distribution of $x_1y_1+y_2$ for random $\mathbf{y}\sim B$. For $j\in[t]$, denote by μ_j the distribution of k_j. Note that

$$\widehat{\eta}(a) = \frac{1}{p}\sum_{u\in\mathbb{Z}_p}\eta(u)\mathbf{e}_p(au) = \frac{1}{p}\mathbb{E}_{\mathbf{y}\sim B}[\mathbf{e}_p(a(x_1y_1+y_2)],$$

and for every j, by the partial sum formula of a geometric series,

$$\widehat{\mu_j}(a) = \frac{1}{p}\sum_{u=1}^{20m}\frac{1}{20m}\mathbf{e}_p(au) = \frac{\mathbf{e}_p(a)}{20mp}\cdot\frac{\mathbf{e}_p(20ma)-1}{\mathbf{e}_p(a)-1}.$$

Since $\nu_{x_1}^B = x_1y_1+y_2+k_1+\ldots+k_t$, we have $\widehat{\nu_{x_1}^B}(a) = p^t\widehat{\eta}(a)\widehat{\mu_1}(a)\ldots\widehat{\mu_t}(a)$, and the result follows. □

Invariance via Fourier expansion. Our earlier upper bound on $\mathrm{Disc}_\nu^{A\times B}(P)$ translates to

$$\begin{aligned}\mathrm{Disc}_\nu^{A\times B}(P) &\lesssim \frac{|B|}{m}\,\mathbb{E}_{x_1,w}\,|\nu_{x_1}^B(w)-\nu_{x_1}^B(w+20mt)|\\ &= \frac{|B|}{m}\,\mathbb{E}_{x_1,w}\left|\sum_{a\in\mathbb{Z}_p}\widehat{\nu_{x_1}^B}(a)(\chi_a(w)-\chi_a(w+20mt))\right|\\ &= \frac{|B|}{m}\,\mathbb{E}_{x_1,w}\left|\sum_{a\in\mathbb{Z}_p}\widehat{\nu_{x_1}^B}(a)(1-\mathbf{e}_p(-20mta))\chi_a(w)\right|.\end{aligned}$$

We now square both sides and apply Cauchy-Schwarz, then Parseval's identity, to obtain

$$\mathrm{Disc}_\nu^{A\times B}(P)^2 \lesssim \left(\frac{|B|}{m}\right)^2\mathbb{E}_{x_1}\sum_{a\in\mathbb{Z}_p}|\widehat{\nu_{x_1}^B}(a)|^2|1-\mathbf{e}_p(-20mta)|^2.$$

Substituting $\widehat{\nu_{x_1}^B}(a)$ for its value from Lemma 3.1 yields

$$\mathrm{Disc}_\nu^{A\times B}(P)^2 \lesssim \left(\frac{|B|}{pm}\right)^2\sum_{a\in\mathbb{Z}_p}\mathop{\mathbb{E}}_{x_1}\left|\mathop{\mathbb{E}}_{\mathbf{y}\sim B}\mathbf{e}_p(a(x_1y_1+y_2))\right|^2\left|\frac{1}{20m}\frac{\mathbf{e}_p(20ma)-1}{\mathbf{e}_p(a)-1}\right|^{2t}|1-\mathbf{e}_p(-20mta)|^2. \tag{6}$$

Since $4m^2 \le p \le 5m^2$, for $a \neq 0$, it follows from the trivial bound $|ma|_p \le m|a|_p$ that

$$|\mathbf{e}_p(20mta) - 1| \approx \frac{|20mta|_p}{p} \lesssim \min\left\{1, \frac{mt|a|_p}{p}\right\} \lesssim \min\left\{1, \frac{t|a|_p}{m}\right\},$$

and

$$\left|\frac{1}{20m}\frac{\mathbf{e}_p(20ma) - 1}{\mathbf{e}_p(a) - 1}\right| \le \min\left\{1, \frac{1}{20m} \times \frac{8|20ma|_p}{4|a|_p}\right\} \le \min\left\{1, \frac{p}{10m|a|_p}\right\} \le \min\left\{1, \frac{m}{2|a|_p}\right\}.$$

Denote $\mathcal{E}_a(B) := \mathbb{E}_{x_1} \left|\mathbb{E}_{\mathbf{y}\sim B}\, \mathbf{e}_p(a(x_1y_1 + y_2))\right|^2$, and note that $\mathcal{E}_a(B) \le 1$. We can split our sum in Eq. (6) as

$$\begin{aligned}
\mathrm{Disc}_\nu^{A\times B}(P)^2 &\lesssim \left(\frac{|B|}{pm}\right)^2 \left(\sum_{|a|_p \ge m} \mathcal{E}_a(B) \left|\frac{1}{20m}\frac{\mathbf{e}_p(20ma) - 1}{\mathbf{e}_p(a) - 1}\right|^{2t} + \sum_{|a|_p < m} \mathcal{E}_a(B)\, |1 - \mathbf{e}_p(-20mta)|^2\right) \\
&\lesssim \left(\frac{|B|}{pm}\right)^2 \sum_{|a|_p \ge m} \mathcal{E}_a(B) \left(\frac{m}{2|a|_p}\right)^{2t} + \left(\frac{|B|}{pm}\right)^2 \sum_{|a|_p < m} \mathcal{E}_a(B) \left(\frac{t|a|_p}{m}\right)^2 \\
&\le \frac{p}{2^t} + \left(\frac{|B|}{pm}\right)^2 \sum_{|a|_p < m} \mathcal{E}_a(B) \left(\frac{t|a|_p}{m}\right)^2. \qquad (7)
\end{aligned}$$

Here in the last line, we use $|B| \le pm$ and the fact that there are at most p terms in the sum.

Key estimates, analyzing $\mathcal{E}_a(B)$: The only mysterious term in (7) is $\mathcal{E}_a(B)$. In this part of the proof, we obtain the required upper bounds on this quantity.

Lemma 3.2. *Let $0 < L < U < m$. Then*

$$\sum_{a \in [L,U]} \mathcal{E}_a(B) \lesssim \frac{p^2 m^2 \log m}{|B|^2 L}.$$

Proof. For $y_1 \in [m]$, define $B_{y_1} : \mathbb{Z}_p \to \{0,1\}$ as $B_{y_1}(y) = 1$ iff $(y_1, y) \in B$. Considering the Fourier expansion of B_{y_1}, for each y, we have

$$B_{y_1}(y) = \sum_{b \in \mathbb{Z}_p} \widehat{B_{y_1}}(b)\mathbf{e}_p(by).$$

Now we can rewrite the sum of $\mathcal{E}_a(B)$:

$$\begin{aligned}
\sum_{a\in[L,U]} \mathcal{E}_a(B) &= \sum_{a\in[L,U]} \mathbb{E}_{x_1\sim[m]} \left| \mathbb{E}_{\mathbf{y}\sim B}\, \mathbf{e}_p(ax_1y_1+ay_2)\right|^2 \\
&= \left(\frac{pm}{|B|}\right)^2 \sum_{a\in[L,U]} \mathop{\mathbb{E}}_{x_1\sim[m]} \left| \mathop{\mathbb{E}}_{y_1\sim[m]} \mathop{\mathbb{E}}_{y_2\sim\mathbb{Z}_p} B_{y_1}(y_2)\mathbf{e}_p(ax_1y_1+ay_2)\right|^2 \\
&= \left(\frac{pm}{|B|}\right)^2 \sum_{a\in[L,U]} \mathop{\mathbb{E}}_{x_1\sim[m]} \mathop{\mathbb{E}}_{y_1,y_1'\sim[m]} \mathop{\mathbb{E}}_{y_2,y_2'\sim\mathbb{Z}_p} B_{y_1}(y_2)B_{y_1'}(y_2')\mathbf{e}_p(ax_1(y_1-y_1')+a(y_2-y_2')) \\
&= \left(\frac{pm}{|B|}\right)^2 \sum_{a\in[L,U]} \mathop{\mathbb{E}}_{y_1,y_1'\sim[m]} \left(\mathop{\mathbb{E}}_{x_1\sim[m]} \mathbf{e}_p(ax_1(y_1-y_1'))\right) \mathop{\mathbb{E}}_{y_2,y_2'\sim\mathbb{Z}_p} B_{y_1}(y_2)B_{y_1'}(y_2')\mathbf{e}_p(a(y_2-y_2')) \\
&= \left(\frac{pm}{|B|}\right)^2 \sum_{a\in[L,U]} \mathop{\mathbb{E}}_{y_1,y_1'\sim[m]} \left(\mathop{\mathbb{E}}_{x_1\sim[m]} \mathbf{e}_p(ax_1(y_1-y_1'))\right) \widehat{B_{y_1}}(-a)\widehat{B_{y_1'}}(a).
\end{aligned}$$

By the Cauchy-Schwarz inequality and Parseval's identity, one has

$$\begin{aligned}
\sum_{a\in[L,U]} |\widehat{B_{y_1}}(-a)\widehat{B_{y_1'}}(a)| &\le \left(\sum_{a\in[L,U]} |\widehat{B_{y_1}}(-a)|^2\right)^{1/2} \left(\sum_{a\in[L,U]} |\widehat{B_{y_1'}}(a)|^2\right)^{1/2} \\
&\le \left(\sum_{a\in\mathbb{Z}_p} |\widehat{B_{y_1}}(-a)|^2\right)^{1/2} \left(\sum_{a\in\mathbb{Z}_p} |\widehat{B_{y_1'}}(a)|^2\right)^{1/2} \\
&= |\,\mathbb{E}_y\, B_{y_1}(y)|^{1/2} |\,\mathbb{E}_y\, B_{y_1'}(y)|^{1/2} \le 1.
\end{aligned}$$

Combining this fact with the previous calculations, we obtain

$$\sum_{a\in[L,U]} \mathcal{E}_a(B) \le \left(\frac{pm}{|B|}\right)^2 \mathop{\mathbb{E}}_{y_1,y_1'\sim[m]} \max_{a\in[L,U]} \left| \mathop{\mathbb{E}}_{x_1\sim[m]} \mathbf{e}_p(ax_1(y_1-y_1'))\right|.$$

Observe that for any $y_1, y_1' \in [m]$, we have $y_1 - y_1' \in [-m, m]$, and moreover, for every $y \in [-m, m]$, we have $\Pr_{y_1,y_1'\sim[m]}[y_1 - y_1' = y] \le \frac{1}{m}$. Therefore,

$$\sum_{a\in[L,U]} \mathcal{E}_a(B) \le \frac{p^2m}{|B|^2} \sum_{y=-m}^{m} \max_{a\in[L,U]} \left| \mathop{\mathbb{E}}_{x_1\sim[m]} \mathbf{e}_p(ax_1y)\right| = \frac{p^2m}{|B|^2}\left(1 + 2\sum_{y\in[m]} \max_{a\in[L,U]} \left| \mathop{\mathbb{E}}_{x_1\sim[m]} \mathbf{e}_p(ax_1y)\right|\right).$$

Substituting

$$\left| \mathop{\mathbb{E}}_{x_1\sim[m]} \mathbf{e}_p(ax_1y)\right| = \left|\frac{1}{m}\frac{\mathbf{e}_p(may)-1}{\mathbf{e}_p(ay)-1}\right| \lesssim \frac{|may|_p}{m|ay|_p} \lesssim \frac{p}{m|ay|_p} \lesssim \frac{m}{|ay|_p},$$

we obtain

$$\sum_{a\in[L,U]} \mathcal{E}_a(B) \lesssim \frac{p^2m}{|B|^2}\left(1 + \sum_{y\in[m]} \max_{a\in[L,U]} \frac{m}{|ay|_p}\right).$$

Since $|x|_p = x$ for $x \in [0, p/2]$, together with the assumptions of $L < m$ and $p > 2m^2$, we have

$$\sum_{a \in [L,U]} \mathcal{E}_a(B) \lesssim \frac{p^2 m}{|B|^2} \left(1 + \sum_{y \in [m]} \frac{m}{Ly} \right) \lesssim \frac{p^2 m^2 \log m}{|B|^2 L}.$$

□

With Lemma 3.2, we can bound the sum in Eq. (7) as

$$\begin{aligned}
\left(\frac{|B|}{pm}\right)^2 \sum_{|a|_p < m} \mathcal{E}_a(B) \left(\frac{t|a|_p}{m}\right)^2 &\approx \left(\frac{|B|}{pm}\right)^2 \frac{t^2}{m^2} \sum_{c=1}^{\log m} \sum_{|a|_p \in [2^{c-1}, 2^c]} |a|_p^2 \mathcal{E}_a(B) \\
&\lesssim \left(\frac{|B|}{pm}\right)^2 \frac{t^2}{m^2} \sum_{c=1}^{\log m} 2^{2c} \cdot \frac{p^2 m^2 \log m}{|B|^2 2^{c-1}} \\
&\approx \frac{t^2}{m^2} \log m \sum_{c=1}^{\log m} 2^c \\
&\approx \frac{t^2 \log m}{m}.
\end{aligned}$$

Since $t \geq 10 \log m$, we have $2^{-t} \leq m^{-10}$ and hence

$$\mathrm{Disc}_\nu^{A \times B}(P) \lesssim \sqrt{\max\left\{\frac{p}{2^t}, \frac{t^2 \log m}{m}\right\}} \approx \sqrt{\frac{\log^3 m}{m}} = m^{-1/2} \log^{3/2} m.$$

4 Concluding remarks

A key step of the proof of [HHL22] relies on the mixing properties of $x_1 y_1 + x_2 y_2$, thus resulting in a strong upper bound on

$$\mathbb{E}_{(x_1,x_2) \sim [m]^2} \left| \mathbb{E}_{(y_1,y_2) \sim B} \mathbf{e}_p(a(x_1 y_1 + x_2 y_2)) \right|^2,$$

for every $|a|_p < m$ and every $B \subseteq [m]^2$. However, the analogous quantity

$$\mathcal{E}_a(B) = \mathbb{E}_{x_1 \sim [m]} \left| \mathbb{E}_{\mathbf{y} \sim B} \mathbf{e}_p(a(x_1 y_1 + y_2)) \right|^2$$

that arises in the proof of Theorem 1.2 can generally be large even when $|a|_p < m$. This seemingly presented a serious obstacle to extending the proof of [HHL22] to Theorem 1.2 at first. Ultimately, we bypassed this issue in Lemma 3.2, by using the fact that the L_1 sum of the Fourier coefficients of the convolution of two Boolean functions is always at most 1. This allowed us to show that while individual $\mathcal{E}_a(B)$ can be large, their average over the interval $[L, U]$ is small (when L and U are small). In this sense, Lemma 3.2 is the major novel component of the proof that allowed us to extend the result of [HHL22].

Another key technical difference with [HHL22] is the choice of the random variable k in constructing the hard distribution. In this work, we choose k as a sum of $\Theta(\log m)$ independent uniform random variables in setting x_2 in the hard distribution μ. By taking k as a sum of a super-constant number of uniform elements, we remove the need for a strong bound on $\mathcal{E}_a(B)$ when $|a|_p \geq m$ and hence simplify and shorten the proof in [HHL22].

Finally, we mention an open problem regarding the sharpness of the bound of Theorem 1.2. Recall that every sign matrix $A_{n\times n}$ satisfies $\mathrm{Disc}(A) \geq \Omega(1/\sqrt{n})$. Can a matrix of sign-rank 3 match this bound?

Question 4.1. *Are there sign matrices $A_{n\times n}$ with sign-rank 3 and*

$$\mathrm{Disc}(A) \leq n^{-\frac{1}{2}+o(1)}?$$

References

[APP+05] Noga Alon, János Pach, Rom Pinchasi, Radoš Radoičić, and Micha Sharir. Crossing patterns of semi-algebraic sets. *Journal of Combinatorial Theory, Series A*, 111(2):310–326, 2005.

[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.

[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.

[BW16] Mark Braverman and Omri Weinstein. A discrepancy lower bound for information complexity. *Algorithmica*, 76(3):846–864, 2016.

[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.

[HHL22] Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Sign-rank vs. discrepancy. *Theory Comput.*, 18:Paper No. 19, 22, 2022.

[HHP+22] Hamed Hatami, Pooya Hatami, William Pires, Ran Tao, and Rosie Zhao. Lower bound methods for sign-rank and their limitations. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2022)*, volume 245 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 22:1–22:24, 2022.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication complexity.* Cambridge University Press, Cambridge, 1997.

[KNR95] Ilan Kremer, Noam Nisan, and Dana Ron. On randomized one-round communication complexity. In *Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '95, page 596–605, New York, NY, USA, 1995. Association for Computing Machinery.

[LS09] Nati Linial and Adi Shraibman. Learning complexity vs. communication complexity. *Combin. Probab. Comput.*, 18(1-2):227–245, 2009.

[Nis93] N. Nisan. The communication complexity of threshold gates. In *Combinatorics, Paul Erdős is eighty, Vol. 1*, Bolyai Soc. Math. Stud., pages 301–315. János Bolyai Math. Soc., Budapest, 1993.

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

[She08a] Alexander A. Sherstov. Communication complexity under product and nonproduct distributions. In *2008 23rd Annual IEEE Conference on Computational Complexity*, pages 64–70, 2008.

[She08b] 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.

[Yao83] Andrew C. Yao. Lower bounds by probabilistic arguments. In *24th Annual Symposium on Foundations of Computer Science (sfcs 1983)*, pages 420–428, 1983.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il