Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 50 (2023)

Communication Complexity and Discrepancy of Halfplanes

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

March 20, 2024

Abstract

We study the discrepancy of the following communication problem. Alice receives a halfplane, and Bob receives a point in the plane, and their goal is to determine whether Bob's point belongs to Alice's halfplane. This communication task corresponds to determining whether $x_1y_1 + y_2 \geq x_2$, where the first player knows (x_1, x_2) and the second player knows (y_1, y_2).

Denoting $n = m^3$, we show that when the inputs are chosen from $[m] \times [m^2]$, the communication discrepancy of the above problem is $O(n^{-1/6} \log^{3/2} n)$.

On the other hand, through the connections to the notion of hereditary discrepancy by Matoušek, Nikolov, and Tawler (IMRN 2020) and a classical result of Matoušek (Discrete Comput. Geom. 1995), we show that the communication discrepancy of every set of n points and n halfplanes is at least $\Omega(n^{-1/4} \log^{-1} n)$.

1 Introduction

Discrepancy theory is an extensive mathematical area that studies the irregularities of distributions of mathematical objects of all kinds. The various notions of discrepancy usually measure the extent of irregularity a system is bound to exhibit. In a sense, discrepancy theory studies how well discrete objects can be distributed to best approximate continuous ones. The subject of discrepancy theory originated from number-theoretic problems in the 1930s, and since then, it has developed into a broad and diverse research area with close connections to theoretic fields like number theory and combinatorics, as well as newfound motivations from complexity theory, computational geometry and numerical computations [Mat99, Cha00].

In communication complexity, a related notion coincidentally named discrepancy, measures the maximum imbalance over submatrices guaranteed by any probability distribution.

Definition 1.1 (Discrepancy). *For a sign matrix $M \in \{\pm 1\}^{\mathcal{X} \times \mathcal{Y}}$, the discrepancy of M with respect to a probability distribution μ on $\mathcal{X} \times \mathcal{Y}$ is*

$$\mathrm{Disc}_\mu(M) := \max_{\substack{A \subseteq \mathcal{X} \\ B \subseteq \mathcal{Y}}} \mathrm{Disc}_\mu^{A \times B}(M),$$

*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

where

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

The discrepancy *of M, denoted by* $\mathrm{Disc}(M)$*, is the minimum of* $\mathrm{Disc}_\mu(M)$ *over all probability distributions μ on $\mathcal{X}\times\mathcal{Y}$.*

The above notion of communication discrepancy is the main parameter of interest in this work. To forestall possible confusion, we explicitly state that we are referring to the communication-theoretic notion whenever we use the term "discrepancy" alone and denote the other notion as "combinatorial discrepancy".

Formulated as an irregularity measure, the definition of communication discrepancy suggests probable connections with classical discrepancy theory. Matoušek, Nikolov and Talwar [MNT15] noted a relation that linked hereditary discrepancy and communication discrepancy (Eq. (3)), and we illustrate the power of this insight through studying a communication task called the *halfplane membership problem*.

Let $\mathcal{H}$ be a finite set of halfplanes, and let $\mathcal{P}$ be a finite set of points. Alice receives a halfplane in $\mathcal{H}$, and Bob receives a point in $\mathcal{P}$. Their goal is to determine whether Bob's point belongs to Alice's halfplane. We represent every point in $\mathcal{P}$ by its coordinates $(y_1, y_2) \in \mathbb{R}^2$, and every halfplane in $\mathcal{H}$ by a pair $(x_1, x_2) \in \mathbb{R}^2$, corresponding to the halfplane

$$H_{x_1,x_2} := \{(z_1, z_2) \in \mathbb{R}^2 \ : \ x_1 z_1 + z_2 \geq x_2\}.$$

We show that the discrepancy of the halfplane membership problem is small even when the points and halfplanes are chosen from $[n]^2$, where $[n] := \{1, \dots, n\}$. This problem is henceforth denoted as $\mathtt{PH}$. In the context of communication complexity, Chor and Goldreich [CG88] proved that for every $0 < \epsilon < 1/2$,

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

where $\mathrm{R}_\epsilon(M)$ denotes the randomized communication complexity of M with an error of ϵ in the public randomness model (See [KN97, Section 3] for the precise definition). As a consequence, a small discrepancy implies a large randomized communication complexity.

Every $n \times n$ sign matrix M satisfies $\mathrm{R}_\epsilon(M) \leq 1 + \log n$ as illustrated by the trivial protocol where Alice sends her whole input to Bob, and Bob computes the output. Also, $\mathrm{Disc}(M)$ is always at least $\Omega(n^{-1/2})$, which we will discuss in Section 1.2. Our main theorem shows that the halfplane membership problem essentially matches these worst-case bounds.

Theorem 1.2 (Main theorem)**.** *Let $n = m^3$ and consider the matrix $\mathtt{PH} \in \{\pm 1\}^{n\times n}$, whose rows and columns are indexed by $[m]\times[m^2]$, and*

$$\mathtt{PH}([x_1, x_2], [y_1, y_2]) = \begin{cases} 1 & \textit{if } x_1 y_1 + y_2 \geq x_2 \\ -1 & \textit{otherwise} \end{cases}. \tag{2}$$

We have

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

On the other hand, for every point-halfplane membership instance with n points and n halfplanes, the discrepancy is at least $\Omega(n^{-1/4}\log^{-1} n)$.

Remark 1.3. It is essential that the halfplanes in $\mathcal{H}$ are not limited to be *homogeneous*, which are the halfplanes defined by lines passing through the origin. Indeed, limiting to homogeneous halfplanes 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 $\Omega(\log^{-1/2} n)$.

Communication complexity. Theorem 1.2 extends the findings of a recent work by Hatami, Hosseini, and Lovett [HHL22]. Their work considered the following communication problem $\mathtt{PH}'$, which is based on points and half-spaces in dimension *three*. Specifically, for $n = s^4$, Alice receives $(x_1, x_2, x_3) \in [s]^2 \times [-3s^2, 3s^2]$ and Bob receives $(y_1, y_2) \in [s]^2$, the goal is to determine whether $x_1y_1 + x_2y_2 \geq x_3$. Hatami et al. proved that $\mathrm{Disc}(\mathtt{PH}') = O(n^{-1/8}\log n)$ and $\mathrm{R}(\mathtt{PH}') = \Omega(\log n)$. Theorem 1.2 extends the result of [HHL22] since $\mathtt{PH}$ is obtained from $\mathtt{PH}'$ by the restriction $x_2 = 1$. Our result also represents an improvement in the discrepancy upper bound from the magnitude of $n^{-1/8}$ to $n^{-1/6}$.

The discrepancy upper bound for $\mathtt{PH}'$ in [HHL22] crucially depends on the mixing property of the function $x_1y_1 + x_2y_2$, a property absent in the corresponding function $x_1y_2 + y_2$ in $\mathtt{PH}$. To overcome the requirement for a strong mixing property, we employ an averaging argument based in a Fourier-theoretic fact: the L_1 sum of the Fourier coefficients of the convolution of two Boolean functions is always at most 1. This not only circumvents the broken step in the proof of [HHL22], but also significantly simplifies various parts of the proofs shared between Theorem 1.2 and [HHL22]. We elaborate on the differences between the two proofs in more technical detail in Section 4.

Discrepancy theory. Partly motivated by applications to the range searching problem, Chazelle and Lvov [CL00] provided an explicit example of n points and n halfplanes with *hereditary discrepancy* $\Omega(n^{1/4}\log^{-1/2} n)$. Hereditary discrepancy, denoted by $\mathrm{herdisc}(\cdot)$, was introduced by Lovász, Spencer, and Vesztergombi [LSV86] as a robust version of combinatorial discrepancy. It is a well-studied quantity in discrepancy theory with numerous applications in computer science and related fields [Cha00, Mat99, Aga04]. The formal definition of hereditary discrepancy and other related notions will be deferred to subsequent sections.

Matoušek, Nikolov, and Talwar [MNT15] established a connection between hereditary discrepancy and the γ_2 factorization norm, a matrix norm that has found extensive use in communication complexity theory [LMSS07, LS09a, LS09b].

Theorem 1.4 ([MNT15])**.** *There exists a universal constant C such that for any real matrix B with m rows,*

$$\frac{\gamma_2(B)}{C\log m} \leq \mathrm{herdisc}(B) \leq \gamma_2(B) \cdot C\sqrt{\log m}.$$

By means of the notion of *margin*, one can apply Theorem 1.4 to relate communication discrepancy with hereditary discrepancy (see Theorems 2.2 and 2.3). Denoting

$$\mathrm{herdisc}^{\infty}(A) := \min\{\mathrm{herdisc}(B) : \ A_{ij}B_{ij} \geq 1 \ \forall i, j\},$$

Matoušek et al. [MNT15] mentioned as a direct consequence of Theorem 1.4 that the quantities herdisc$^\infty(A)$ and $\mathrm{Disc}(A)^{-1}$ are equivalent up to poly-logarithmic factors:

$$\frac{\mathrm{Disc}(A)^{-1}}{C\log m} \le \mathrm{herdisc}^\infty(A) \le C\,\mathrm{Disc}(A)^{-1}\sqrt{\log m}. \tag{3}$$

By leveraging basic properties of the simple properties of γ_2 norm in combination with Theorem 1.4, we derive the following proposition, allowing for comparisons of *classical discrepancy* results in the boolean setting to *communication discrepancy* results in the sign setting.

Proposition 1.5. *Let $F \in \{0,1\}^{m\times n}$ be a non-zero matrix. If A is the sign matrix obtained by replacing 0 in F with -1, then*

$$\gamma_2(F) \le \gamma_2(A) \le 3\gamma_2(F). \tag{4}$$

In particular,

$$\mathrm{herdisc}(F)\cdot\Omega(\log^{-3/2} m) \le \mathrm{herdisc}(A) \le \mathrm{herdisc}(F)\cdot O(\log^{3/2} m). \tag{5}$$

By Proposition 1.5, Chazelle and Lvov's example corresponds to an $n\times n$ sign matrix P representable by points and halfplanes with

$$\mathrm{herdisc}(P) = \Omega(n^{1/4}\log^{-2} n),$$

whereas Theorem 1.2 provides a polynomial lower bound on the more challenging quantity

$$\mathrm{herdisc}^\infty(\mathtt{PH}) = \Omega(n^{1/6}\log^{-5/2} n).$$

In the other direction, Theorem 1.4 holds greater importance in adapting classical discrepancy results to communication complexity. Theorems 2.2 and 2.3 combined give

$$\mathrm{Disc}(A) = \Omega(\gamma_2(A)^{-1}). \tag{6}$$

Through the connection between γ_2 norm and hereditary discrepancy (Theorem 1.4), as well as the conversion from boolean to sign setting (Proposition 1.5), we can translate classical discrepancy upper bounds on incidence systems into discrepancy lower bounds at an expense of additional poly-logarithmic factors.

Employing this machinery and a celebrated upper bound result by Matoušek, we prove a discrepancy lower bound in terms of the notion of *sign-rank*. The sign-rank of a sign matrix is the minimum rank over real matrices with agreeing sign patterns. We will discuss the geometric interpretation of the sign-rank and its significance in communication complexity theory in Section 1.2.

Theorem 1.6. *Every $n\times n$ sign matrix A with sign-rank d satisfies*

$$\mathrm{Disc}(A) = \Omega\left(n^{-\frac{1}{2}+\frac{1}{2(d-1)}}\log^{-1} n\right) \text{ and } \mathrm{herdisc}(A) = O\left(n^{\frac{1}{2}-\frac{1}{2(d-1)}}\right).$$

We will present the proof in Section 2. The second part of Theorem 1.2 is an immediate corollary of this theorem. This theorem also improves the trivial bound of $\mathrm{Disc}(A) = \Omega(n^{-1/2})$ for any sign matrix A with sign-rank $O\left(\frac{\log n}{\log\log n}\right)$.

1.1 Geometric interpretation of Theorem 1.2

Theorem 1.2 offers an insightful interpretation of the geometry inherent in a point-halfspace representation. A sign matrix can be naturally represented by an incidence system of points and halfspaces. Concretely, a sign representation for a sign matrix A assigns each row with a point u_i and each column with a vector v_j, denoting an oriented halfspace. This representation ensures that $\mathrm{sign}(\langle u_i, v_j\rangle) = A_{ij}$. By its very definition, `PH` is representable by points and (non-homogeneous) halfspaces in dimension two.

Apart from dimension, another parameter of interest is the *margin*, which measures the normalized distance of the points from the boundaries of halfspaces. We defer the formal mathematical definition to Section 1.2. The dimension and margin of sign representations are crucial performance parameters for linear classifiers in learning theory. Indeed, a low-dimension classifier is desirable for its efficiency, while a large-margin classifier is favourable for robustness against perturbations.

In view of the equivalence of discrepancy and margin (Theorem 2.3), Theorem 1.2 illustrates that while the matrix `PH` is representable in dimension two as points and halfplanes, the margin of such a representation of `PH` in any dimension stays small.

1.2 Sign-rank versus discrepancy

The *sign-rank* of a sign matrix $A \in \{\pm 1\}^{m\times n}$, denoted by $\mathrm{rank}_\pm(A)$, is the smallest rank of a real matrix $B \in \mathbb{R}^{m\times n}$ 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 various areas such as learning theory, discrete geometry and geometric graphs, communication complexity, circuit complexity, and Banach spaces theory (see [HHP$^+$22] and the references therein).

Sign-rank can be equivalently defined based on sign representations. Recall that a sign representation of a sign matrix A is a pair of sets of vectors $\{u_i\}, \{v_j\} \subseteq \mathbb{R}^d$ such that $A_{ij} = \mathrm{sign}(\langle u_i, v_j\rangle)$ for all i, j. The sign-rank is the minimum d for such a representation to exist. The margin of the representation is defined as

$$\min_{i,j} \frac{|\langle u_i, v_j\rangle|}{\|u_i\|\|v_j\|},$$

and the margin of the matrix A, denoted by $\mathrm{m}(A)$, is the maximum margin attainable by a sign representation in any dimension. For an $n \times n$ sign matrix A, if we take u_i's to be the rows of A and v_j's to be the columns of the identity matrix, this sign representation achieves the trivial bounds $\mathrm{rank}_\pm(A) \le n$ and $\mathrm{m}(A) \ge 1/\sqrt{n}$. By the equivalence of the discrepancy and the margin (Theorem 2.3), the above bound gives the $\Omega(1/\sqrt{n})$ lower bound for the discrepancy.

The pioneering paper of Babai, Frankl, and Simon [BFS86], which formally introduced the communication complexity classes, initiated a line of research investigating the gap between two fundamental notions in communication complexity – namely, sign-rank and discrepancy. The same paper posed the separation question of these two parameters, in complexity class terms, this is equivalent to asking for a separation between 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 omit the definitions of the complexity classes here, and direct the reader to [HHL22] for a more comprehensive discussion.

The question posed by Babai, Frankl and Simon [BFS86] remained unanswered for over two decades. Finally, it was demonstrated by Buhrman et al. [BVdW07] and independently Sherstov [She08b] that there are $n \times n$ sign matrices with $\mathbf{rk}_{\pm}(F) = O(\log n)$ but $\mathrm{Disc}(F) = 2^{-\log^{\Omega(1)}(n)}$. Subsequent works [She11, She13, Tha16, She19] further refined this separation, and an exponential separation of $\mathbf{rk}_{\pm}(F) = O(\log n)$ and $\mathrm{Disc}(F) = n^{-\Omega(1)}$ was achieved in [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], such a matrix has discrepancy $\Omega(\log^{-1/2} n)$. The matrix $\mathtt{PH}$ in Theorem 1.2 also has sign-rank 3 and it provides a slightly stronger upper bound on the discrepancy.

1.3 Discrepancy with respect to product measures

A sign matrix with a *sub-logarithmic* sign-rank inherit interesting structural properties from low-dimensional geometry. For instance, Alon, Pach, Pinchasi, and Sharir [APP+05, Theorem 1.3] proved that every $n \times n$ sign matrix of sign-rank d contains a monochromatic rectangle of size $\frac{n}{2^{d+1}} \times \frac{n}{2^{d+1}}$. It follows that for such a matrix, and for every *product measure* $\lambda \times \nu$ (where λ and ν are probability measures over rows and columns, respectively), the inequality holds:

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

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

$$\mathrm{Disc}^{\times}(\mathtt{PH}) := \inf_{\lambda \times \nu} \mathrm{Disc}_{\lambda \times \nu}(\mathtt{PH}) \geq 2^{-8},$$

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

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

From the communication complexity perspective, these 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 an input sampled from μ with an error probability of at most ϵ. Yao's minimax principle [Yao83] states that randomized communication complexity is exactly the maximum distributional complexity. Thus by Theorem 1.2, one has

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

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

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

Since the sign-rank upper bounds the VC dimension and $\mathtt{PH}$ has sign-rank 3, we have

$$\mathrm{Disc}^{\times}(\mathtt{PH}) \geq 2^{-8} \qquad \text{and} \qquad \mathrm{D}^{\times}_{1/3}(\mathtt{PH}) := \max_{\lambda \times \nu} \mathrm{D}^{\lambda \times \nu}_{1/3}(\mathtt{PH}) = O(1),$$

while by Theorem 1.2,

$$\mathrm{Disc}(\texttt{PH}) = O(n^{-1/6}\log^{3/2} n) \qquad \text{and} \qquad \mathrm{R}_{1/3}(\texttt{PH}) = \Omega(\log n). \tag{7}$$

Consequently, Theorem 1.2 reinstates the $O(1)$-versus-$\Omega(\log n)$ separation[1] between general distributional complexity and product distributional complexity as proved by Sherstov [She08a].

Theorem 1.7 ([She08a, Theorem 1.2])**.** *Let $\epsilon > 0$ be an arbitrary constant. There exists an $n \times n$ sign matrix F such that*

$$\mathrm{Disc}^{\times}(F) = \Omega(1) \qquad \textit{and} \qquad \mathrm{D}^{\times}_{1/3}(F) = O(1),$$

while

$$\mathrm{Disc}(F) = O\left(n^{-\frac{1}{2}+\epsilon}\right) \qquad \textit{and} \qquad \mathrm{R}_{1/3}(F) = \Omega(\log n).$$

2 Hereditary discrepancy and γ_2 norm

The combinatorial discrepancy of a matrix $A \in \mathbb{R}^{m\times n}$ is defined as $\min_{x\in\{\pm 1\}^n} \|Ax\|_\infty$. The hereditary discrepancy of A is the maximum combinatorial discrepancy among column restrictions of A:

$$\mathrm{herdisc}(A) \coloneqq \max_{S\subseteq[n]} \min_{x\in\{\pm1\}^{|S|}} \|A_S x\|_\infty,$$

where A_S is the restriction of A to columns in S. This concept was introduced by Lovász, Spencer, and Vesztergombi [LSV86] as a more well-behaved notion of combinatorial discrepancy. Indeed, without the hereditary constraint, duplicating a set system of arbitrary combinatorial discrepancy results in a set system with zero combinatorial discrepancy.

The γ_2 factorization norm originates from Banach space theory. In the seminal work by Linial and Shraibman [LS09b], they introduced the norm as a lower-bound technique for randomized and quantum communication complexities. Since then, the γ_2 norm has become a central matrix complexity parameter of interest in communication complexity theory and related areas.

Definition 2.1 (γ_2-factorization norm)**.** *The* γ_2 norm *of a real matrix A is*

$$\gamma_2(A) \coloneqq \min_{X,Y: A=XY} \|X\|_{\mathrm{row}}\|Y\|_{\mathrm{col}},$$

where $\|X\|_{\mathrm{row}}$ and $\|Y\|_{\mathrm{col}}$ denote the largest ℓ_2-norm of a row in X and the largest ℓ_2 norm of a column in Y, respectively.

A notable alternative definition of γ_2 norm is formulated in [LSŠ08]:

$$\gamma_2(A) = \min_{P:\ \|P\|\le 1} \|P \circ A\|. \tag{8}$$

Here $\|\cdot\|$ denotes the spectral norm and $\circ$ denotes Hadamard (entrywise) product. Many properties of γ_2 norm follow easily from Eq. (8).

To establish the connection between communication discrepancy and hereditary discrepancy, we require two important results related to the margin: the margin admits a γ_2 norm formulation, and it is equivalent to the discrepancy up to the constant factor.

[1] We remind readers for the change of notation from [She08a]: Shertov stated the result in terms of a boolean function $f : \{0,1\}^{n'} \times \{0,1\}^{n'}$, whereas our result considers F to be an $n \times n$ matrix.

Theorem 2.2 ([LMSS07])**.** *For every sign matrix A,* $\mathrm{m}(A)^{-1} = \min\{\gamma_2(B) : \ A_{ij}B_{ij} \geq 1 \ \forall i,j\}$*. In particular,* $\mathrm{m}(A)^{-1} \leq \gamma_2(A)$*.*

Theorem 2.3 ([LS09a])**.** *For every sign matrix A,* $\mathrm{Disc}(A) \leq \mathrm{m}(A) \leq 8\,\mathrm{Disc}(A)$*.*

In this work and common to communication complexity theory, it is routine to switch between the boolean setting and sign setting to represent binary values. Eq. (4) of Proposition 1.5 states that the γ_2 norms of a boolean matrix and its sign version are equivalent up to a constant factor. This simple yet crucial fact allows us to bridge communication discrepancy results with classical discrepancy results.

Proof of Proposition 1.5. Note that $A = 2F - J$, where J is the all-one matrix. By triangle inequality and the fact that $\gamma_2(J) = 1$, we have $2\gamma_2(F) - 1 \leq \gamma_2(A) \leq 2\gamma_2(F) + 1$. It can be checked that $\gamma_2(A), \gamma_2(F) \geq 1$, this yields Eq. (4) as required.

Combining Eq. (4) with Theorem 1.4 yields

$$\mathrm{herdisc}(A) \leq \gamma_2(A) \cdot C\sqrt{\log m} \leq \gamma_2(F) \cdot 3C\sqrt{\log m} \leq \mathrm{herdisc}(F) \cdot 3C^2 \log^{3/2} m,$$

and

$$\mathrm{herdisc}(A) \geq \gamma_2(A) \cdot C^{-1} \log^{-1} m \geq \gamma_2(F) \cdot C^{-1} \log^{-1} m \geq \mathrm{herdisc}(F) \cdot C^{-2} \log^{-3/2} m. \qquad \square$$

With the above preparations, we are in a position to prove Theorem 1.6.

Proof of Theorem 1.6. Since A is of sign-rank d, there exist vectors $u_i, v_j \in \mathbb{R}^d$ such that $A_{ij} = \mathrm{sign}(\langle u_i, v_j \rangle)$ for every $i, j \in [n]$, By applying perturbations if necessary, we can assume the d-th coordinate of v_j is non-zero for each j. Furthermore, since we are only concerned with the sign of $\langle u_i, v_j \rangle$, we can rescale v_j's so that $v_j(d) = \pm 1$ for each j. Partition the columns of A into the sets

$$C_+ \coloneqq \{j \in [n] : \ v_j(d) = +1\} \text{ and } C_- \coloneqq \{j \in [n] : v_j(d) = -1\}.$$

The case where all $v_j(d)$'s are equal is a simple special case and the proof easily follows from the general case. We assume $C_+, C_- \neq \emptyset$ thereafter.

By reordering the columns if necessary, we can assume that all columns in C_+ precede the columns in C_-. Consider the matrix A_+ (resp. A_-) obtained by restricting to the columns in C_+ (resp. C_-), so A is the matrix with A_+ placing next to A_-. We show that A_+ and A_- are representable by points and halfspaces in $\mathbb{R}^{d-1}$. For $v \in [n]$, define $\tilde{v}_j = (v_j(1), \ldots, v_j(d-1)) \in \mathbb{R}^{d-1}$. For $i \in [n]$, define

$$H_i^{(+)} = \{w \in \mathbb{R}^{d-1} : \ u_i(1)w(1) + u_i(2)w(2) + \cdots + u_i(d-1)w(d-1) + u_i(d) \geq 0\}.$$

Then, $A_+(i,j) = +1$ iff $\tilde{v}_j \in H_i^{(+)}$ for any $(i,j) \in [n] \times C_+$. As for A_-, for $i \in [n]$ we define

$$H_i^{(-)} = \{w \in \mathbb{R}^{d-1} : \ u_i(1)w(1) + u_i(2)w(2) + \cdots + u_i(d-1)w(d-1) - u_i(d) \geq 0\}.$$

It follows that $A_-(i,j) = +1$ iff $\tilde{v}_j \in H_i^{(-)}$ for any $(i,j) \in [n] \times C_-$.

It was proved in [Mat95] that for a point-halfspace incidence system in $\mathbb{R}^t$ with n points, the combinatorial discrepancy and consequently hereditary discrepancy of the boolean incidence matrix is $O(n^{1/2-1/2t})$. In Proposition A.1 in the appendix, we prove that the same bound also holds for the sign setting.

Observe that for two matrices $P \in \mathbb{R}^{m\times n_1}$ and $Q \in \mathbb{R}^{m\times n_2}$,

$$\operatorname{herdisc}\left(\begin{bmatrix} P & Q \end{bmatrix}\right) \le \operatorname{herdisc}(P) + \operatorname{herdisc}(Q).$$

This concludes that

$$\operatorname{herdisc}(A) \le \operatorname{herdisc}(A_-) + \operatorname{herdisc}(A_+) = O\left(n^{\frac{1}{2}-\frac{1}{2(d-1)}}\right). \tag{9}$$

For the communication discrepancy lower bound, applying Eq. (6) and Theorem 1.4 to Eq. (9) gives the desired bound:

$$\operatorname{Disc}(A) = \Omega(\gamma_2(A)^{-1}) = \Omega((\operatorname{herdisc}(A)\log n)^{-1}) = \Omega\left(n^{-\frac{1}{2}+\frac{1}{2(d-1)}}\log^{-1} n\right). \qquad \square$$

3 Proof of Theorem 1.2

Before presenting the proof of Theorem 1.2, we first provide the technical preludes of the proof.

3.1 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 $[s,t] = \{s,\ldots,t\}$, 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 $\widehat{f * g}(a) = \widehat{f}(a)\widehat{g}(a)$. 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 $x := x_1 + \ldots + x_k$ 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 $|x|_p$ 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

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

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

3.2 Proof

As mentioned in Section 1, the lower bound in Theorem 1.2 follows from Theorem 1.6. For the rest of the section, we focus on the upper bound. Let m be sufficiently large and denote $\mathcal{X} = [m] \times [m^2]$. The matrix $\texttt{PH}$ 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 $0 < x_2 \le m^2$ 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 $\texttt{PH}$ in Eq. (2) 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 $x_2 = x_1y_1 + y_2 + k$ or $x_2 = x_1y_1 + y_2 + k - 20mt$, each with probability $1/2$. Note that in the former case, $x_1y_1 + y_2 < x_2$ and in the latter case, $x_1y_1 + y_2 \geq x_2$.

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

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

and

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

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

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

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

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

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

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

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

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

The above analysis shows that in order to prove that $\mathrm{Disc}_{\nu}^{A\times B}(\texttt{PH})$ 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 $\mathrm{Disc}_\nu^{A\times B}(\mathtt{PH})$ translates to

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

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

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

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

$$\mathrm{Disc}_\nu^{A\times B}(\mathtt{PH})^2 \lesssim \left(\frac{|B|}{pm}\right)^2\sum_{a\in\mathbb{Z}_p}\mathop{\mathbb{E}}_{x_1}\left|\mathop{\mathbb{E}}_{\mathbf{y}\sim B}\mathbf{e}_p(a(x_1y_1+y_2))\right|^2\left|\frac{1}{20m}\frac{\mathbf{e}_p(20ma)-1}{\mathbf{e}_p(a)-1}\right|^{2t}|1-\mathbf{e}_p(-20mta)|^2. \tag{10}$$

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 $\mathcal{E}_a(B)\coloneqq\mathbb{E}_{x_1}\left|\mathbb{E}_{\mathbf{y}\sim B}\,\mathbf{e}_p(a(x_1y_1+y_2))\right|^2$, and note that $\mathcal{E}_a(B)\le 1$. We can split our sum in Eq. (10) as

$$\begin{aligned}\mathrm{Disc}_\nu^{A\times B}(\mathtt{PH})^2 &\lesssim \left(\frac{|B|}{pm}\right)^2\left(\sum_{|a|_p\ge m}\mathcal{E}_a(B)\left|\frac{1}{20m}\frac{\mathbf{e}_p(20ma)-1}{\mathbf{e}_p(a)-1}\right|^{2t}+\sum_{|a|_p<m}\mathcal{E}_a(B)\,|1-\mathbf{e}_p(-20mta)|^2\right)\\ &\lesssim \left(\frac{|B|}{pm}\right)^2\sum_{|a|_p\ge m}\mathcal{E}_a(B)\left(\frac{m}{2|a|_p}\right)^{2t}+\left(\frac{|B|}{pm}\right)^2\sum_{|a|_p<m}\mathcal{E}_a(B)\left(\frac{t|a|_p}{m}\right)^2\\ &\le\frac{p}{2^t}+\left(\frac{|B|}{pm}\right)^2\sum_{|a|_p<m}\mathcal{E}_a(B)\left(\frac{t|a|_p}{m}\right)^2.\end{aligned} \tag{11}$$

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 (11) 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 $\mathcal{E}_a(B)$:

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

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 $\Pr_{y_1, y_1' \sim [m]}[y_1 - y_1' = y] \leq \frac{1}{m}$. Therefore,

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

Substituting

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

we obtain

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

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. (11) 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

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

4 Concluding remarks

We study the halfplane membership problem on the integer lattice and prove an upper bound of $O(n^{-1/6} \log^{3/2} n)$ bound on the discrepancy, widening the largest known gap between sign-rank and discrepancy. It implies a $\Omega(\log n)$ randomized communication complexity lower bound, indicating that the halfplane membership problem belongs to a class of essentially the hardest problems in the randomized communication model. The study of halfplane membership is important since it is arguably one of the simplest known communication problems exhibiting gaps in several complexity parameters.

Through the connection with hereditary discrepancy and an adaptation of the classical combinatorial discrepancy upper bound for point-halfplane systems, we show that the discrepancy of every set of n points and n halfplanes is $\Omega(n^{-1/4}\log^{-1} n)$. More generally, we prove a discrepancy lower bound in terms of sign-rank, and show that the discrepancy of a matrix with bounded sign-rank is set apart from the trivial bound of $\Omega(n^{-1/2})$.

For the technical review, a key step of the proof of [HHL22] relies on the mixing properties of $x_1y_1 + x_2y_2$, thus resulting in a strong upper bound on

$$\mathbb{E}_{(x_1,x_2)\sim[m]^2} \left|\mathbb{E}_{(y_1,y_2)\sim B}\, \mathbf{e}_p(a(x_1y_1 + x_2y_2))\right|^2,$$

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

$$\mathcal{E}_a(B) = \mathbb{E}_{x_1\sim[m]} \left|\mathbb{E}_{\mathbf{y}\sim B}\, \mathbf{e}_p(a(x_1y_1 + y_2))\right|^2$$

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].

References

[Aga04] Pankaj K. Agarwal. Range searching. In Jacob E. Goodman and Joseph O'Rourke, editors, *Handbook of Discrete and Computational Geometry, Second Edition*, pages 809–837. Chapman and Hall/CRC, 2004.

[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.

[Bec81] József Beck. Roth's estimate of the discrepancy of integer sequences is nearly sharp. *Combinatorica*, 1(4):319–325, December 1981.

[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.

[Cha00] Bernard Chazelle. *The Discrepancy Method: Randomness and Complexity*. Cambridge University Press, July 2000.

[CL00] Bernard Chazelle and Alexey Lvov. A trace bound for the hereditary discrepancy. *Discrete & Computational Geometry*, 26:221–231, 2000.

[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.

[Kle66] Daniel J. Kleitman. On a combinatorial conjecture of erdös. *Journal of Combinatorial Theory, Series A*, 1:209–214, 1966.

[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.

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

[LS09a] Nati Linial and Adi Shraibman. Learning complexity vs. communication complexity. *Combin. Probab. Comput.*, 18(1-2):227–245, 2009.

[LS09b] Nati Linial and Adi Shraibman. Lower bounds in communication complexity based on factorization norms. *Random Structures & Algorithms*, 34(3):368–394, 2009.

[LSŠ08] Troy Lee, Adi Shraibman, and Robert Špalek. A direct product theorem for discrepancy. In *2008 23rd Annual IEEE Conference on Computational Complexity*, pages 71–80, 2008.

[LSV86] L. Lovász, J. Spencer, and K. Vesztergombi. Discrepancy of set-systems and matrices. *European Journal of Combinatorics*, 7(2):151–160, April 1986.

[Mat95] Jiří Matoušek. Tight upper bounds for the discrepancy of half-spaces. *Discrete & Computational Geometry*, 13:593–601, 1995.

[Mat99] Jiří Matoušek. *Geometric Discrepancy*. Springer Berlin Heidelberg, 1999.

[MNT15] Jiří Matoušek, Aleksandar Nikolov, and Kunal Talwar. Factorization norms and hereditary discrepancy, 2015.

[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.

[Ska13] Matthew Skala. Hypergeometric tail inequalities: ending the insanity, 2013.

[Spe85] Joel Spencer. Six standard deviations suffice. *Transactions of the American Mathematical Society*, 289(2):679–706, 1985.

[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.

A Combinatorial discrepancy upper bound of sign incidence matrix of halfspaces

In this section, we prove the combinatorial discrepancy upper bound of the sign incidence matrix of halfspaces that matches Matoušek's result [Mat95] in the boolean setting.

Proposition A.1. *Let $H_1, \dots, H_m$ be d-dimensional halfspaces and $X = \{u_1, \dots, u_n\} \subseteq \mathbb{R}^d$. Define the sign matrix $A \in \{\pm 1\}^{m \times n}$ by $A_{ij} = 1$ if $u_j \in H_i$ and $A_{ij} = -1$ otherwise. Then*

$$\min_{\chi \in \{\pm 1\}^n} \|A\chi\|_\infty = O(n^{1/2-1/2d}).$$

Our proof is essentially an emulation of Matoušek's proof, as it closely follows Matoušek's notations and proof strategy. While only the special case of halfspaces is relevant to our work, we note that similar to Matoušek's work, the exact bound and proof hold for the extended case in which the d-dimensional halfspace condition is replaced with the order-d growth condition of the primal shatter function [Mat95, Theorem 1.2].

We first give an overview of Matoušek's proof. To keep the exposition as succinct as possible, we leave out the precise definitions of several notations adopted in [Mat95] and refer readers to the original paper for details.

The original proof starts with a canonical decomposition of each halfspace H_i into sets in some auxiliary set systems $\mathscr{P}_1, \dots, \mathscr{P}_k$, where $k = \lceil \log(n+1) \rceil$. The problem reduces to exhibiting a sign vector χ, which the total combinatorial discrepancy attained by χ over the set systems $\mathscr{P}_1, \dots, \mathscr{P}_k$ is bounded by $O(n^{1/2-1/2d})$. This is done by applying Beck-Spencer's partial colouring method [Bec81, Spe85].

The partial colouring method typically combines the entropy method and Kleitman diametric theorem [Kle66]. First, one shows that over uniform random choice of $\chi \in \{\pm 1\}^n$, the entropy of certain random variables $R(\chi)$ is bounded above by Cn for some constant $C \in (0, 1)$. Consequently, there is a fixed assignment r^* so that at least $2^{(1-C)n}$ χ's are mapped to r^* by R. Kleitman diametric theorem states that among all subsets of n-bit binary strings with pairwise hamming distance at most D (with $D \le n/2$), the Hamming ball of radius $D/2$ has the maximum size. Therefore, one can find χ_1 and χ_2 that are both mapped to r^* and differ in at least αn coordinates for some constant $\alpha = \alpha(C) \in (0, 1)$.

By a suitable choice of $R(\cdot)$, one can establish that $\chi' \coloneqq \frac{\chi_1 - \chi_2}{2} \in \{-1, 0, 1\}^n$ has a small enough combinatorial discrepancy and at most a constant fraction of zeros ("uncoloured"). Now, the paradigm is iterated on the uncoloured points. Since the set of uncoloured points shrinks by a constant factor, this results in a geometric decrease of combinatorial discrepancy of successive partial colourings, so that the final colouring $\chi \coloneqq \chi'_{(1)} + \chi'_{(2)} + \cdots$ satisfies the desired bound.

To modify the argument for the sign setting, notice that the sign incidence matrix A and boolean incidence matrix B are related by $A = 2B - J$ (where J is the all-one matrix). This means $\|A\chi\|_\infty = \|2B\chi - w(\chi) \cdot \mathbf{1}\|_\infty$, which we call $w(\chi) \coloneqq \sum_{u \in U} \chi(u)$ the *weight* of $\chi : U \to \{-1, 0, 1\}$. If we determine a combinatorial discrepancy upper bound of the boolean matrix B under a weight-0 colouring χ, it yields a doubled upper bound for the sign matrix A under the same colouring χ. To this end, we strengthen the key technical lemma ([Mat95, Lemma 2.2]) by adding the fourth item that enforces the resultant partial colouring χ to have weight 0. This additional requirement ensures that the final colouring obtained by iterated applications of this lemma also has weight 0.

Lemma A.2. *There exists a sequence of integers $\Delta_1, \dots, \Delta_k$ and a partial colouring $\chi : X \to \{-1, 0, 1\}$ with the following properties:*

1. $\Delta_1 + \cdots + \Delta_k = O(|X|^{1/2-1/2d})$.
2. $|\chi^{-1}(0)| \le |X|/2$.

3. For every $i \in [k]$ and $S \in \mathscr{P}_i$ the quantity $\chi(S) := \sum_{u \in S} \chi(u)$ satisfies $|\chi(S)| \le \Delta_i$.

4. $w(\chi) = 0$.

The proof of Lemma A.2 is very similar to the original proof of [Mat95, Lemma 2.2]. We only outline the relevant alterations.

Let $\Gamma \subseteq \{\pm 1\}^n$ be the set of weight-0 colourings. Standard calculations give that $|\Gamma| = \binom{n}{n/2} = \Theta(2^n n^{-1/2})$. Let χ be chosen uniformly at random from Γ (instead of a uniform n-bit sign vector), and $\Delta_1, \ldots, \Delta_k$ be as defined in the proof of [Mat95, Lemma 2.2]. For a chosen χ, for $i \in [k]$ and $S \in \mathscr{P}_i$, denote $b_{i,S} = b_{i,S}(\chi)$ the integer nearest to $\chi(S)/\Delta_i$.

Fix i and $S \in \mathscr{P}_i$. For any $j \in \mathbb{Z}$, set

$$p_j := \Pr_{\chi \sim \Gamma}[b_{i,S} = j] = \Pr_{\chi \sim \Gamma}\left[\frac{(2j-1)\Delta_i}{2} \le \chi(S) < \frac{(2j+1)\Delta_i}{2}\right].$$

Notice that the random variable $\chi(S)$ can be modelled by the outcome of a hypergeometric experiment. Denote $\mathcal{H}_{N,M,s}$ the hypergeometric random variable for drawing s items from a population with M successes and $N - M$ failures. Then

$$p_j = \Pr_{\chi}\left[\frac{(2j-1)\Delta_i}{4} \le \mathcal{H}_{n,n/2,|S|} - \frac{|S|}{2} < \frac{(2j+1)\Delta_i}{4}\right].$$

Recall that $\mathbb{E}[\mathcal{H}_{N,M,s}] = sM/N$, so p_j concerns the deviation of the hypergeometric random variable from its expectation. We are in a position to apply the hypergeometric tail bounds, which coincide with the more well-known binomial tail bounds.

Lemma A.3 (see e.g. [Ska13])**.** *Let $\mathcal{H} = \mathcal{H}_{N,M,s}$ be a hypergeometric random variable. For $t \ge 0$,*

$$\Pr[\mathcal{H} - \mathbb{E}[\mathcal{H}] \ge t] \le \exp\left(-\frac{2t^2}{s}\right)$$

and

$$\Pr[\mathcal{H} - \mathbb{E}[\mathcal{H}] \le -t] \le \exp\left(-\frac{2t^2}{s}\right).$$

Applying the tail bounds, we recover the exact bounds found in [Mat95]: for $j \ge 1$,

$$p_j \le \Pr_{\chi}\left[\mathcal{H}_{n,n/2,|S|} - \frac{|S|}{2} \ge \frac{(2j-1)\Delta_i}{4}\right] \le \exp\left(-\frac{(2j-1)^2\Delta_i^2}{8|S|}\right).$$

A similar bound (with $2j+1$ in place of $2j-1$) holds for $j \le -1$. Finally,

$$p_0 \ge 1 - \Pr_{\chi}\left[\left|\mathcal{H}_{n,n/2,|S|} - \frac{|S|}{2}\right| \ge \frac{\Delta_i}{4}\right] \ge 1 - 2\exp\left(-\frac{\Delta_i^2}{8|S|}\right).$$

All the necessary bounds for p_j in the proof of [Mat95, Lemma 2.2] can be established further from the above bounds. Subsequently, the exact entropy estimation for the random variables $\vec{b}$ applies and yields $H(\vec{b}) \le n/10$. Finally, there exists a value $\vec{b}_*$ such that

$$|\{\chi \in \Gamma : \ \vec{b}(\chi) = \vec{b}_*\}| \ge |\Gamma| \cdot 2^{-n/10} = 2^{9n/10 - O(\log n)}.$$

Kleitman diametric theorem guarantees the existence of two weight-0 colourings χ_1 and χ_2 so that $\frac{\chi_1 - \chi_2}{2}$ satisfies all desired properties, this completes the proof of Lemma A.2.

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il