Electronic Colloquium on Computational Complexity, Report No. 79 (2022)

Lower Bound Methods for Sign-rank and their Limitations

Hamed Hatami [*] Pooya Hatami [†] William Pires [‡] Ran Tao [§] Rosie Zhao [¶]

May 25, 2022

Abstract

The *sign-rank* of a matrix A with ± 1 entries is the smallest rank of a real matrix with the same sign pattern as A. To the best of our knowledge, there are only three known methods for proving lower bounds on the sign-rank of explicit matrices: (i) Sign-rank is at least the VC-dimension; (ii) Forster's method, which states that sign-rank is at least the inverse of the largest possible average margin among the representations of the matrix by points and half-spaces; (iii) Sign-rank is at least a logarithmic function of the density of the largest monochromatic rectangle.

We prove several results regarding the limitations of these methods.

- We prove that, qualitatively, the monochromatic rectangle density is the strongest of these three lower bounds. If it fails to provide a super-constant lower bound for the sign-rank of a matrix, then the other two methods will fail as well.
- We show that there exist $n \times n$ matrices with sign-rank $n^{\Omega(1)}$ for which none of these methods can provide a *super-constant* lower bound.
- We show that sign-rank is at most an exponential function of the deterministic communication complexity with access to an equality oracle. We combine this result with Green and Sanders' quantitative version of Cohen's idempotent theorem to show that for a large class of sign matrices (e.g., XOR-lifts), sign-rank is at most an exponential function of the γ_2 norm of the matrix. We conjecture that this holds for all sign matrices.
- Towards answering a question of Linial, Mendelson, Schechtman, and Shraibman regarding the relation between sign-rank and discrepancy, we conjecture that sign-ranks of the ± 1 adjacency matrices of hypercube graphs can be arbitrarily large. We prove that none of the three lower bound techniques can resolve this conjecture in the affirmative.

1 Introduction

A *sign matrix* is a matrix with ± 1 entries. The *sign-rank* of a sign matrix $A_{m\times n}$ 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. This fundamental notion arises naturally in areas as diverse as

[*]School of Computer Science, McGill University. hatami@cs.mcgill.ca. Supported by an NSERC grant.

[†]Department of Computer Science and Engineering, The Ohio State University. pooyahat@gmail.com. Supported by NSF grant CCF-1947546

[‡]School of Computer Science, McGill University. william.pires@mail.mcgill.ca.

[§]Department of Mathematics and Statistics, McGill University. ran.tao6@mail.mcgill.ca.

[¶]School of Computer Science, McGill University. rosie.zhao@mail.mcgill.ca.

ISSN 1433-8092

learning theory [BDES02, KS07, She08a, SS05, Fel17, FGV21], discrete geometry and geometric graphs [AFR85, FGL$^+$12, FPS$^+$17, Suk16, EMRPS14], communication complexity [PS86, CM18, She08b, HHL20], circuit complexity [RS10, BT16, SW19], and the theory of Banach spaces [Mat96, Nao18].

The notion of sign-rank was formally defined in 1986 in connection with randomized communication complexity in the unbounded-error model [PS86]. After almost four decades of research, sign-rank remains one of the most elusive matrix parameters in discrete analysis. To the best of our knowledge, there are only three known methods for proving lower bounds on the sign-rank of an explicit matrix: VC-dimension, size of the largest monochromatic rectangle, and Forster's method, and among those, only Forster's method can imply super-logarithmic lower bounds.

The results presented in this paper arose from our attempts to solve two fundamental open problems about sign-rank, presented as Question 1.4 and Question 1.11 below. Attempting to give negative answers to these questions, we proved that none of the known techniques could yield adequate sign-rank lower bounds for these purposes. Of course, this observation does not necessarily imply that the techniques are inherently weak, as there is a possibility that the correct answer to both questions is positive. As a natural next step, we examined the limitations of these techniques more carefully and, among other things, proved the existence of $n \times n$ matrices with sign-rank $n^{\Omega(1)}$, for which none of these methods could provide a *super-constant* lower bound.

We start by reviewing and reformulating the results that are relevant to this article.

Counting argument: Shortly after the introduction of sign-rank in [PS86], Alon, Frankl, and Rödl [AFR85] used results of [Mil64, Tho65, War68] on the number of connected components of real algebraic varieties and obtained a linear lower bound on the sign-rank of random sign matrices. This argument was later refined in [AMY16, Lemma 24] to the following bound on the number of low sign-rank matrices.

Lemma 1.1 (See [AMY16, Lemma 24]). *For $d \leq \frac{n}{2}$, the number of $n \times n$ sign matrices of sign-rank at most d does not exceed $(O(n/d))^{2dn} \leq 2^{O(dn\log(n))}$.*

It follows from Lemma 1.1 that most $n \times n$ sign matrices have sign-rank $\Omega(n)$.

The VC-dimension lower bound: The *Vapnik-Chervonenkis* (VC) dimension of a sign matrix A is the largest k such that A contains a submatrix with k columns and 2^k *distinct* rows. To state the relation between the VC dimension and sign-rank, we discuss a geometric definition of sign-rank.

A real matrix $B_{\mathcal{X}\times\mathcal{Y}}$ has rank d iff the entries of B can be represented as $B_{xy} = \langle u_x, v_y \rangle$ for vectors $u_x, v_y \in \mathbb{R}^d$. Since the normalization of these vectors does not affect the signs of $\langle u_x, v_y \rangle$, we can restate the definition of sign-rank as follows.

Definition 1.2 (Sign-rank). *The sign-rank of a sign matrix $A_{\mathcal{X}\times\mathcal{Y}}$, denoted by $\mathrm{rank}_\pm(A)$, is the smallest d such that there exist* unit *vectors $u_x, v_y \in \mathbb{R}^d$ with $A_{xy} = \mathrm{sgn}(\langle u_x, v_y \rangle)$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.*

The vectors in Definition 1.2 represent A as points and half-spaces in the d-dimensional space: $A_{xy} = 1$ iff the point u_x belongs to the half-space $\{z : \langle z, v_y \rangle > 0\}$. Since the VC dimension of any such configuration of points and half-spaces in $\mathbb{R}^d$ is at most d, we have

$$\mathrm{rank}_\pm(A) \geq \mathrm{VC}(A). \tag{1}$$

This lower bound was already implicit in the paper of Paturi and Simon [PS86, Theorem 4]. Since the VC dimension of every $n \times n$ matrix is at most $\log n$, this method cannot prove super-logarithmic lower bounds on sign-rank. In addition, Alon, Moran, and Yehudayoff [AMY16] established strong separations between the two parameters. For example, they showed that there are $n \times n$ sign matrices of VC dimension 3 that have sign-rank $\Omega\left(\frac{\sqrt{n}}{\log n}\right)$.

Margin and Discrepancy: There is another natural parameter that is associated with the representations of a sign matrix as points and half-spaces. The quantity $\min_{x,y} |\langle u_x, v_y \rangle|$ is called the *margin* of such a representation; it measures the smallest distance between the points u_x and the hyperplanes defined by v_y.

Definition 1.3 (Margin)**.** *The* margin *of a sign matrix $A_{\mathcal{X}\times\mathcal{Y}}$ is*

$$\mathrm{m}(A) \coloneqq \sup \min_{x,y} |\langle u_x, v_y \rangle| ,$$

where the supremum is over all $d \in \mathbb{N}$ and unit vectors $u_x, v_y \in \mathbb{R}^d$ with $A_{xy} = \mathrm{sgn}(\langle u_x, v_y \rangle)$.

Linial and Shraibman [LS09a] proved that margin essentially coincides with the well-studied parameter of *discrepancy* in communication complexity, defined as

$$\mathrm{disc}(A) \coloneqq \inf_{\mu} \max_{\substack{S \subseteq \mathcal{X} \\ T \subseteq \mathcal{Y}}} |\mathbb{E}_{xy\sim\mu}[A_{xy}\mathbf{1}_S(x)\mathbf{1}_T(y)]| , \tag{2}$$

where the infimum is over all probability distributions μ on $\mathcal{X} \times \mathcal{Y}$. They proved

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

The notion of discrepancy is a well-understood parameter, and many lower bounds in communication complexity are established by proving that the discrepancy of the corresponding matrix is small. Such proofs often entail finding a "hard" distribution μ such that the maximum in Equation (2) is small. We shall discuss this more later in the context of Forster's lower bound method.

The problem of understanding the relation between sign-rank and margin is an important one because these notions optimize two fundamental attributes of the geometric representations of the matrix. Sign-rank minimizes the dimension while allowing the margin to be arbitrarily small. Margin maximizes the margin of the representation while allowing the dimension to be arbitrarily large.

Hatami, Hosseini, and Lovett [HHL20] constructed $n \times n$ sign matrices that have a very small margin (equivalently discrepancy) of $O\left(\frac{\log(n)}{n^{1/8}}\right)$ while their sign-rank is only 3. The converse direction regarding the question of margin vs sign-rank remains open. Does large margin imply small sign-rank?

Question 1.4. *Is there a function η such that for every sign matrix A, we have $\mathrm{rank}_{\pm}(A) \leq \eta(\mathrm{m}(A)^{-1})$?*

Question 1.4 is essentially due to [LMSS07, Corollary 3.2, Lemma 4.2, and Section 8], where they proved the inequality $\mathrm{rank}_{\pm}(A) \leq \mathrm{m}(A)^{-2} \cdot \log(n)$, and asked whether the log factor in this inequality is necessary.

It is known that margin, discrepancy, public-coin randomized communication complexity, and approximate γ_2 norms are all related, in the sense that each can be used to provide an upper bound on any other (see Section 3.3 for more details). Therefore, one can *equivalently* restate Question 1.4, with $\mathrm{m}(A)^{-1}$ replaced with any of the mentioned parameters. We propose the following conjecture that would imply a negative answer to Question 1.4, as $\mathrm{m}(Q_d)^{-1} = O(1)$ (see Proposition 3.4).

Conjecture 1.5 (Sign-rank of hypercube graphs)**.** *Let Q_d be the $2^{d-1} \times 2^{d-1}$ sign matrix whose rows and columns are indexed with, respectively, odd-parity and even-parity elements of $\{0,1\}^d$, and $Q_d(x,y) = -1$ iff x and y differ in exactly one coordinate. Then*

$$\lim_{d\to\infty} \mathrm{rank}_\pm(Q_d) = \infty.$$

Forster's sign-rank lower bound: For *explicit* matrices, the VC-dimension lower bound remained state of the art for almost two decades until the breakthrough work of Forster [For02]. Forster used a convex geometric approach to prove a *linear* lower bound on the sign-rank of Hadamard matrices, establishing the first super-logarithmic lower bound on the sign-rank of an explicit matrix.

Forster's proof first transforms the vectors v_y to be in isotropic position, and then uses the anti-concentration of measure in low dimensions to show that the average $\mathbb{E}_y |\langle u_x, v_y\rangle|$ is relatively large for every vector u_x. In other words, the "average margin" of such a representation is large. We will consider a slight generalization of Forster's approach that allows arbitrary distributions on $\mathcal{Y}$.

Definition 1.6 (Average margin)**.** *The* average margin *of a sign matrix $A_{\mathcal{X}\times\mathcal{Y}}$* with respect to a probability distribution *ν on $\mathcal{Y}$ is defined as*

$$\mathrm{m}^{\mathrm{avg}}_\nu(A) = \sup \min_x \mathbb{E}_{y\sim\nu} |\langle u_x, v_y\rangle|,$$

where the supremum is over all sign-representations of A using unit vectors $u_x, v_y \in \mathbb{R}^d$ for any d. The average margin *of A is defined as $\mathrm{m}^{\mathrm{avg}}(A) = \inf_\nu \mathrm{m}^{\mathrm{avg}}_\nu(A)$.*

Note that $\mathrm{m}^{\mathrm{avg}}(A) \geq \mathrm{m}(A)$ since $\mathbb{E}_{y\sim\nu} |\langle u_x, v_y\rangle| \geq \min_y |\langle u_x, v_y\rangle|$. A slightly different notion of average margin is studied by Kallweit and Simon in [KS11], however, since $\mathrm{m}^{\mathrm{avg}}(A)$ is always smaller than Kallweit and Simon's notion of average margin, it provides a stronger lower bound on sign-rank in Theorem 1.7 below. We summarize Forster's approach as the following theorem.

Theorem 1.7 (Forster [For02])**.** *For every sign-matrix A, we have*

$$\mathrm{rank}_\pm(A) \geq \mathrm{m}^{\mathrm{avg}}(A)^{-1}.$$

We will provide a proof of Theorem 1.7 in Section 2.1 since our formulation of Forster's approach is slightly more general than the ones appearing in the literature.

Forster's original paper [For02] applies the average margin method to show that sign-rank is large when the spectral norm is small. Subsequent works [FKL+01, FS06, RS10] showed that, more generally, this method can extend discrepancy bounds to lower bounds on sign-rank if the witnessing hard distribution μ in Equation (2) is well-spread on most of the entries. This is intuitive considering that discrepancy is equivalent to margin, and the lower bound in Theorem 1.7 is based on average margin.

It is straightforward (see Proposition 2.5) to prove that every sign matrix satisfies $\mathrm{m}^{\mathrm{avg}}(A)^{-1} \geq \sqrt{\mathrm{VC}(A)}$, which demonstrates that VC dimension is essentially a weaker lower bound technique than Forster's method.

Monochromatic rectangles and sign-rank: A submatrix of a matrix A is called a *monochromatic rectangle* if all the entries in this submatrix have the same value. In addition to VC-dimension and Forster's method, there is a third known approach for proving super-constant lower bounds on sign-ranks of explicit matrices, which is based on the size of the largest monochromatic rectangle.

We define the following parameter based on the size of monochromatic rectangles.

Definition 1.8 (Monochromatic rectangle ratio)**.** *For every sign-matrix $A_{\mathcal{X}\times\mathcal{Y}}$, define*

$$\mathrm{rect}(A) = \inf_{\mu\times\nu} \max_{R} \mu \times \nu(R),$$

where the infimum is over all product *probability measures $\mu \times \nu$ on $\mathcal{X} \times \mathcal{Y}$, and the maximum is over all monochromatic rectangles in A.*

Alon, Pach, Pinchasi, Radoičić and Sharir [APP$^+$05] proved that every $m \times n$ sign matrix of sign-rank d contains an $\frac{m}{2^{d+1}} \times \frac{n}{2^{d+1}}$ monochromatic rectangle. Similar bounds are obtained by Fox, Pach, and Suk [FPS16] using the cutting lemma of Chazelle [Cha93]. We provide a different proof in Proposition 2.6. While Proposition 2.6 follows from the result of [FPS16], we believe our short and simple proof could provide some geometric intuition for why matrices of low sign-rank contain large monochromatic rectangles.

The following relation between sign-rank and monochromatic rectangle ratio follows from the bound of [APP$^+$05, Theorem 1.3].

Theorem 1.9 (See [APP$^+$05, Theorem 1.3])**.** *For every sign-matrix A, we have*

$$\mathrm{rank}_{\pm}(A) \geq \frac{\log_2\left(\mathrm{rect}(A)^{-1}\right)}{2} - 1. \tag{3}$$

Remark 1.10. The result of [APP$^+$05, Theorem 1.3] says that every $m \times n$ sign matrix of sign-rank d contains an $\frac{m}{2^{d+1}} \times \frac{n}{2^{d+1}}$ monochromatic rectangle. To deduce Equation (3), given a product probability measure $\mu \times \nu$ on $\mathcal{X} \times \mathcal{Y}$, pick a large K and duplicate every row x for $\lfloor K\mu(x) \rfloor$ many times. Similarly duplicate every column y for $\lfloor K\nu(y) \rfloor$ many times. The resulting $M \times N$ matrix has the same sign-rank as A and thus contains an $\frac{M}{2^{d+1}} \times \frac{N}{2^{d+1}}$ monochromatic rectangle, which translates to a monochromatic rectangle of $\mu \times \nu$-measure approximately $\frac{1}{2^{d+1}} \times \frac{1}{2^{d+1}}$ in A. Taking the limit as K grows to infinity yields Equation (3).

Note that similar to the VC-dimension, Theorem 1.9 cannot imply super-logarithmic lower bounds on sign-rank, since every $n \times n$ sign matrix satisfies $\mathrm{rect}(A) \geq \frac{1}{2n}$. To see the latter claim, note that for every probability distribution μ on the rows, there is always a row x with measure $\geq \frac{1}{n}$, and any probability distribution ν over the columns has measure at least $\frac{1}{2}$ on either the 1's or the -1's of this row.

Sign-rank of semi-algebraic matrices, an open problem: A real semi-algebraic set in $\mathbb{R}^d$ is the set of all points that satisfy a given finite Boolean combination of polynomial inequalities in the d coordinates. We say that such a set has *description complexity* t if in some representation, the number of inequalities and the degrees of the corresponding polynomials are all bounded from above by t.

Every collection of points $u_1, \dots, u_m \in \mathbb{R}^d$ and semi-algebraic sets $K_1, \dots, K_n \subseteq \mathbb{R}^d$ define a sign matrix $A_{m\times n}$ where $A_{ij} = 1$ iff $u_i \in K_j$. We say that A has a representation in $\mathbb{R}^d$ with *description complexity* t if every K_i has description complexity t.

We call a class of sign matrices *semi-algebraic* if there exists $d, t \in \mathbb{N}$ such that every matrix in this class has a representation in $\mathbb{R}^d$ of description complexity at most t. Semi-algebraic classes of sign matrices capture natural geometric constructions of graphs on finite dimensional real spaces, such as interval graphs, incidence graphs, disc graphs, and more generally, all graph classes where vertices are points in a real Euclidean space and the edges are defined by a semi-algebraic relation of constant complexity.

An affirmative answer to the following question would imply that semi-algebraic classes of sign matrices coincide with bounded sign-rank classes.

Question 1.11 (Sign-rank of semi-algebraic matrices)**.** *Is there a function $\eta : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that every sign matrix with a d-dimensional representation of description complexity t has sign-rank at most $\eta(d, t)$?*

For the converse direction, note that if $\mathrm{rank}_{\pm}(A) = \eta$, then the corresponding sign-representation of A using vectors $u_i, v_j \in \mathbb{R}^\eta$ is a representation with description complexity 1: We have $A_{ij} = 1$ iff $u_i \in \left\{x \in \mathbb{R}^d \ : \ \langle v_j, x\rangle > 0\right\}$, and note that $\langle v_j, x\rangle$ is a polynomial of degree 1 in the coordinates of x.

Let $\Gamma : \{-1, 1\}^t \to \{-1, 1\}$ be a predicate and let $A_1, \ldots, A_t$ be $m \times n$ sign matrices. Let $\Gamma(A_1, \ldots, A_t)$ denote the $m \times n$ sign matrix with ij-entries $\Gamma(A_1(i, j), \ldots, A_t(i, j))$. As we will discuss in Section 3.5, a simple linearization trick shows that Question 1.11 can be reformulated as the following question.

Question 1.12 (First reformulation of Question 1.11)**.** *Is there a function $\eta : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that for every predicate $\Gamma : \{-1, 1\}^t \to \{-1, 1\}$ and every set of $m \times n$ sign matrices $A_1, \ldots, A_t$ with sign-ranks at most d, we have*

$$\mathrm{rank}_{\pm}(\Gamma(A_1, \ldots, A_t)) \leq \eta(d, t)?$$

The formulation in Question 1.12 is interesting from the perspective of learning theory: Consider a binary data set encoded as a sign matrix Γ. The entry Γ_{ij} is called the label of the data-point j according to the concept i. Suppose that these labels are determined by a few other binary labels. For example, whether a person i is likely to watch a movie j may be determined by whether j is the genre of movie that they like, whether j features some of their favorite actors, and whether j is available at a theater near them. Now suppose that each of these latter binary data sets has a low-dimensional representation. Does this mean that our data set has a low-dimensional representation?

The formulation in Question 1.12 is also interesting from the perspective of communication complexity: since the logarithm of sign-rank is equivalent to the unbounded-error communication complexity (see Equation (10) below), Question 1.12 asks whether a matrix constructed by the entrywise application of a logical predicate to matrices $A_1, \ldots, A_t$, each with a small unbounded-error communication complexity, must have a small unbounded-error communication complexity. It is straightforward to show that a similar statement is indeed true in the *bounded-error* case.

Question 1.12 can be further simplified to a fascinating simple-to-state question. Let $A \wedge B$ be the matrix whose ij-th entries are the point-wise minimums of the entries of A and B, corresponding to the Boolean AND operator. Let $\neg A := -A$. Recall that $\{\wedge, \neg\}$ is a complete basis, i.e., it is a functionally complete set in the logical sense. Hence the function Γ in Question 1.12 can be implemented using the two operations $\wedge$ and $\neg$, and since for every sign matrix A, we have $\mathrm{rank}_{\pm}(A) = \mathrm{rank}_{\pm}(\neg A)$, Question 1.12 is equivalent to the following.

Question 1.13 (Second reformulation of Question 1.11)**.** *Is there a function* $\eta : \mathbb{N} \to \mathbb{N}$ *such that for every two sign matrices* A *and* B *with sign-ranks at most* d*, we have* $\mathrm{rank}_{\pm}(A \wedge B) \le \eta(d)$*?*

In comparison, let us consider the Hadamard product $A \circ B$ of two matrices A and B, which corresponds to entrywise $\oplus$ operator in the Boolean setting. It is well-known that $\mathrm{rank}(A \circ B) \le \mathrm{rank}(A) \cdot \mathrm{rank}(B)$, which implies that for every two $m \times n$ sign matrices A and B, we have

$$\mathrm{rank}_{\pm}(A \circ B) \le \mathrm{rank}_{\pm}(A) \cdot \mathrm{rank}_{\pm}(B).$$

However, this cannot be used in a similar argument as the AND case above to answer Question 1.11, as $\{\oplus, \neg\}$ is not a complete basis.

Contributions and organization: For the following discussion, recall the three aforementioned lower bound techniques for sign-rank:

$$\mathrm{VC}(A) \le \mathrm{rank}_{\pm}(A), \qquad \mathrm{m}^{\mathrm{avg}}(A)^{-1} \le \mathrm{rank}_{\pm}(A), \qquad \frac{\log_2\left(\mathrm{rect}(A)^{-1}\right)}{2} - 1 \le \mathrm{rank}_{\pm}(A),$$

and note that all these lower bounds are non-increasing when restricting to submatrices: For every submatrix M of A, we have

$$\mathrm{VC}(M) \le \mathrm{VC}(A), \qquad \mathrm{m}^{\mathrm{avg}}(M)^{-1} \le \mathrm{m}^{\mathrm{avg}}(A)^{-1}, \qquad \mathrm{rect}(M)^{-1} \le \mathrm{rect}(A)^{-1}.$$

- In Section 3.1, we study the relation between the average margin and the rectangle ratio. In Theorem 3.1, we prove that
$$\mathrm{m}^{\mathrm{avg}}(A)^{-1} \le \mathrm{rect}(A)^{-1},$$
which, combined with Proposition 2.5, shows
$$\sqrt{\mathrm{VC}(A)} \le \mathrm{m}^{\mathrm{avg}}(A)^{-1} \le \mathrm{rect}(A)^{-1}. \tag{4}$$
These inequalities demonstrate that if the monochromatic rectangle ratio cannot provide a super-constant lower bound for the sign-rank of a matrix, then the other two methods will fail as well.

  The significance of Theorem 3.1 is that proving an upper bound on $\mathrm{rect}(A)^{-1}$ is often much easier than directly analyzing the average margin. This allows us to demonstrate some limitations of Forster's method.

- In Section 3.2, we combine Theorem 3.1 with a counting argument to prove our main separation result: In Theorem 3.2, we show the existence of $n \times n$ sign matrices A that have sign-rank $n^{\Omega(1)}$ but $\mathrm{VC}(A)$, $\mathrm{m}(A)^{-1}$ and $\mathrm{rect}(A)^{-1}$ are all $O(1)$. In other words, there exists matrices of very large sign-rank such that none of the known lower bound techniques can provide a lower bound that is larger than $O(1)$.

- In Section 3.3, we discuss the limitation of sign-rank lower bounds in answering Question 1.4 and Conjecture 1.5. In particular, in Proposition 3.4 we observe that $\mathrm{rect}(Q_d)^{-1} = O(1)$, and thus none of the known lower bound methods can prove Conjecture 1.5.

- In Section 3.4, we study a question that is closely related to the relation between margin and sign-rank (i.e., Question 1.4). As discussed above, one can equivalently rephrase Question 1.4 in terms of upper-bounding sign-rank by a function of the approximate γ_2 norm (see Definition 2.1). As stated in Conjecture 1.5, we believe the answer to be negative. However, one can strengthen the assumption and ask whether the sign-rank can be upper-bounded by a function of the γ_2 norm instead:

  Conjecture 1.14. *There exists a function η such that for every sign matrix A, we have $\mathrm{rank}_\pm(A) \leq \eta(\|A\|_{\gamma_2})$.*

  Towards proving Conjecture 1.14, in Theorem 3.8, we show that

  $$\mathrm{rank}_\pm(A) \leq 4^{\mathrm{D}^{eq}(A)}, \tag{5}$$

  where $\mathrm{D}^{eq}(A)$ denotes the deterministic communication complexity of the matrix A with access to an equality oracle. In Corollary 3.9, we combine this with Green and Sanders' [GS08a, GS08b, San19, San11, San20a] quantitative versions of Cohen's idempotent theorem and a theorem of [HHH21] to verify Conjecture 1.14 for a broad class of sign-matrices: We prove there exists a function η such that if $f : G \to \{-1, 1\}$ for a finite group G, and $A_{G\times G}$ is the sign matrix with entries $A(x, y) = f(xy^{-1})$, then

  $$\mathrm{rank}_\pm(A) \leq \eta(\|A\|_{\gamma_2}).$$

  In the case of abelian G, we have

  $$\mathrm{rank}_\pm(A) \leq \exp\big(\exp\big(C\|A\|_{\gamma_2}^4\big)\big),$$

  where C is a universal constant. Note that taking $G = \mathbb{Z}_2^n$ corresponds to the class of XOR-lifts.

  Equation (5) is also interesting from the point of view of communication complexity. It implies

  $$\mathrm{U}(A) \leq 2\,\mathrm{D}^{eq}(A) + O(1).$$

  where $\mathrm{U}(A)$ denotes the unbounded-error randomized communication complexity of A, formally defined in Equation (10).

- In Section 3.5, we study the sign-rank of semi-algebraic sign matrices. In Corollary 3.10, we prove that if A and B are two sign matrices of sign-rank at most d, then

  $$\mathrm{VC}(A \wedge B) \leq 20d \qquad \text{and} \qquad \mathrm{m}^{\mathrm{avg}}(A \wedge B)^{-1} \leq \mathrm{rect}(A \wedge B)^{-1} \leq 2^{-2d-2}.$$

  These demonstrate the inability of the known lower bound techniques to give a negative answer to Question 1.11 by providing a super-constant lower bound on the sign-rank of semi-algebraic matrices.

- In Section 3.6 we prove that sign matrices of sign-rank d have small communication complexity in the average communication model over any product distribution.

2 Notation, Background, and Basic Observations

Much of the notation we will use is implicit in the introduction, but it may be helpful to clarify things here.

We will use the standard computer science asymptotic notations [CLRS01] of $O(\cdot)$, $\Omega(\cdot)$, $\Theta(\cdot)$, $o(\cdot)$, and $\omega(\cdot)$. We denote the indicator function of a set S by $\mathbf{1}_S$, that is, $\mathbf{1}_S(x) \coloneqq 1$ if $x \in S$, and $\mathbf{1}_S(x) \coloneqq 0$ otherwise. For $i = 1, \ldots, d$, we denote the i-th standard vector by $\mathbf{e}_i \in \mathbb{R}^d$. For a vector $u \in \mathbb{R}^d$, we denote the Euclidean norm of u by $\|u\|$.

For a real matrix $B_{\mathcal{X}\times\mathcal{Y}}$, we denote by $\operatorname{sgn}(B)$ the sign matrix corresponding to the signs of the entries of B. We say that the unit vectors $u_x, v_y \in \mathbb{R}^d$ *sign-represent* $A_{\mathcal{X}\times\mathcal{Y}}$ if $A_{xy} = \operatorname{sgn}(\langle u_x, v_y\rangle)$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

A finite set of vectors $v_1, \ldots, v_m \in \mathbb{R}^d$ are in *isotropic position* if for every unit vector $u \in \mathbb{R}^d$, we have

$$\frac{1}{m}\sum_{i=1}^{m} |\langle u, v_i\rangle|^2 = \frac{1}{d}.$$

All matrices in this article are real and finite, and all normed spaces are defined over the reals. The spectral norm of a matrix $A_{\mathcal{X}\times\mathcal{Y}}$ is defined as

$$\|A\| = \max_{x\in\mathbb{R}^{\mathcal{Y}}:\|x\|=1} \|Ax\|,$$

and its trace norm is defined as

$$\|A\|_{\mathrm{tr}} = \operatorname{tr}\left(\sqrt{A^t A}\right) = \sum_{i=1}^{\min(|\mathcal{X}|,|\mathcal{Y}|)} \sigma_i,$$

where σ_i are the singular values of A. Next, we define the γ_2 norm of a matrix, which is an important tool for proving lower and upper bounds in discrepancy theory and communication complexity [LS09b].

Definition 2.1 (γ_2 norm)**.** *The γ_2 norm of a matrix $A_{\mathcal{X}\times\mathcal{Y}}$, denoted by $\|A\|_{\gamma_2}$, is the smallest $c \geq 0$ such that there exists $d \in \mathbb{N}$ and vectors $u_x, v_y \in \mathbb{R}^d$ with* $\max_{x,y} \|u_x\| \cdot \|v_y\| \leq c$ *and $A_{xy} = \langle u_x, v_y\rangle$ for all x, y.*

For $\epsilon \in [0, 1)$, the *approximate γ_2 norm* of $A_{\mathcal{X}\times\mathcal{Y}}$ with error parameter ϵ is defined as

$$\|A\|_{\gamma_2,\epsilon} = \inf_B \|B\|_{\gamma_2},$$

where the infimum is over all real matrices $B_{\mathcal{X}\times\mathcal{Y}}$ with $\max_{x,y} |A_{xy} - B_{xy}| \leq \epsilon$. Note that despite what the notation might suggest, $\|\cdot\|_{\gamma_2,\epsilon}$ is not a norm.

By definition, a matrix $B_{\mathcal{X}\times\mathcal{Y}}$ satisfies $\|B\|_{\gamma_2} = 1$ if and only if for some $d \in \mathbb{N}$, there exist unit vectors $u_x, v_y \in \mathbb{R}^d$ with $B_{xy} = \langle u_x, v_y\rangle$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Hence, one can reformulate Definition 1.3 and Definition 1.6 in terms of the γ_2 norm as

$$\mathrm{m}(A) = \sup_{\substack{B:\|B\|_{\gamma_2}=1\\ \operatorname{sgn}(B)=A}} \min_{x,y} |B_{xy}|,$$

and

$$\mathrm{m}_\nu^{\mathrm{avg}}(A) = \sup_{\substack{B:\|B\|_{\gamma_2}=1\\ \mathrm{sgn}(B)=A}} \min_x \mathbb{E}_{y\sim\nu} |B_{xy}|.$$

Finally, note that the dual of the γ_2 norm is

$$\|A\|_{\gamma_2^*} := \sup_{B:\|B\|_{\gamma_2}=1} \mathrm{tr}\big(AB^t\big) = \sup_{B:\|B\|_{\gamma_2}=1} \sum_{x,y\in\mathcal{X}\times\mathcal{Y}} A_{xy}B_{xy}, \tag{6}$$

where both A and B are $\mathcal{X}\times\mathcal{Y}$ matrices.

2.1 Forster's lower bound

Forster's lower bound is based on the geometric fact that every set of vectors in general position is transformable to be in isotropic position by applying an invertible linear transformation. This powerful fact in convex geometry was first established by Barthe [Bar98] as a key step in his proof of a reverse form of the Brascamp-Lieb inequality. It seems that Forster was unaware of Barthe's paper, and he gave a different proof in his paper [For02]. The following variation of this fact from [HKLM20] is proved using Barthe's theorem.

Theorem 2.2 (Isotropic position [HKLM20, Thereom A.2]). *Consider a probability distribution ν over a finite set $\mathcal{Y}$ and non-zero vectors $v_y \in \mathbb{R}^d$ for all $y\in\mathcal{Y}$. There is an invertible linear transformation T such that*

$$\mathbb{E}_{y\sim\nu}\langle Tv_y, u\rangle^2 = \frac{1}{d}$$

for every unit vector $u\in\mathbb{R}^d$ iff the following holds. For every $0<k<d$ and every k-dimensional subspace V, either

- *$\nu(\{y : v_y \in V\}) < \frac{k}{d}$, or*
- *$\nu(\{y : v_y \in V\}) = \frac{k}{d}$ and the remaining mass lies in a $(d-k)$-dimensional subspace.*

We will summarize Forster's approach for proving lower bounds on sign-rank as the following theorem.

Theorem 2.3 (Theorem 1.7 restated). *For every sign-matrix $A_{\mathcal{X}\times\mathcal{Y}}$, we have*

$$\mathrm{rank}_\pm(A) \geq \mathrm{m}^{\mathrm{avg}}(A)^{-1}.$$

Proof. Let $d = \mathrm{rank}_\pm(A)$ and ν be a probability distribution on $\mathcal{Y}$. We consider two cases.

Case I: There exists $y_0\in\mathcal{Y}$ such that $\nu(y_0)\geq\frac{1}{d}$. In this case, we show that regardless of the value of $\mathrm{rank}_\pm(A)$, we always have $\mathrm{m}_\nu^{\mathrm{avg}}(A)\geq\frac{1}{d}$. Pick any sign representation of A with unit vectors $u_x, v_y\in\mathbb{R}^k$ for a $k\in\mathbb{N}$. Consider a small $\delta>0$ and define the unit vectors

$$u'_x = (\delta u_x, \sqrt{1-\delta^2}\,\mathrm{sgn}(A_{xy_0})) \in \mathbb{R}^{k+1} \qquad \text{for all } x\in\mathcal{X},$$

and

$$v'_y = \begin{cases} \mathbf{e}_{k+1} & y = y_0 \\ (v_y, 0) & y\neq y_0 \end{cases}$$

for all $y \in \mathcal{Y}$. These vectors provide a sign representation of A by unit vectors in $\mathbb{R}^{k+1}$. Moreover, for every $x \in \mathcal{X}$,

$$\mathbb{E}_{y\sim\nu} \left|\langle u'_x, v'_y \rangle\right| \geq \frac{1}{d} \left|\langle u'_x, v'_{y_0} \rangle\right| \geq \frac{\sqrt{1-\delta^2}}{d}.$$

Therefore, $\mathrm{m}_\nu^{\mathrm{avg}}(A) \geq \frac{\sqrt{1-\delta^2}}{d}$ for every $\delta > 0$. Taking the limit as δ tends to 0 shows $\mathrm{m}_\nu^{\mathrm{avg}}(A) \geq \frac{1}{d}$.

Case II: For every $y \in \mathcal{Y}$, we have $\nu(y) < \frac{1}{d}$. Let $u_x, v_y \in \mathbb{R}^d$ be unit vectors that sign-represent A. By applying a small perturbation to the vectors u_x and v_y, we can assume without loss of generality that they are in general position.

Since the vectors v_y are in general position, for every $0 < k < d$ and every k-dimensional subspace V, we have $|\{y : v_y \in V\}| \leq k$, and thus $\nu(\{y : v_y \in V\}) < \frac{k}{d}$. Hence, by Theorem 2.2, there exists an invertible linear transformation T such that for every unit vector $u \in \mathbb{R}^d$, we have

$$\mathbb{E}_{y\sim\nu} \langle Tv_y, u\rangle^2 = \frac{1}{d}.$$

Since $\langle T^{-1}u_x, Tv_y\rangle = \langle u_x, v_y\rangle$, by replacing u_x and v_y with unit vectors $\frac{T^{-1}u_x}{\|T^{-1}u_x\|}$ and $\frac{Tv_y}{\|Tv_y\|}$, respectively, we can assume without loss of generality that

$$\mathbb{E}_{y\sim\nu} \langle v_y, u\rangle^2 = \frac{1}{d},$$

for every unit vector $u \in \mathbb{R}^d$. Since $|\langle u_x, v_y\rangle| \leq 1$, it follows that

$$\min_{x\in\mathcal{X}} \mathbb{E}_{y\sim\nu} \left|\langle u_x, v_y\rangle\right| \geq \min_{x} \mathbb{E}_{y\sim\nu} \langle u_x, v_y\rangle^2 = \frac{1}{d}.$$

Therefore, $\mathrm{m}_\nu^{\mathrm{avg}}(A) \geq \frac{1}{d}$. □

Note that for any matrix $A_{\mathcal{X}\times\mathcal{Y}}$ and unit vectors $u_x, v_y \in \mathbb{R}^d$, we have

$$\begin{aligned}
\sum_{x,y} A_{xy}\langle u_x, v_y\rangle &= \sum_{i=1}^{d} \sum_{x,y} A_{xy} u_x(i) v_y(i) \\
&\leq \sum_{i=1}^{d} \|A\| \left(\sum_x |u_x(i)|^2\right)^{1/2} \left(\sum_y |v_y(i)|^2\right)^{1/2} \\
&\leq \|A\| \left(\sum_{i=1}^{d}\sum_x |u_x(i)|^2\right)^{1/2} \left(\sum_{i=1}^{d}\sum_y |v_y(i)|^2\right)^{1/2} \\
&\leq \|A\|\sqrt{|\mathcal{X}||\mathcal{Y}|}.
\end{aligned}$$

Therefore, by Equation (6), we have

$$\|A\|_{\gamma_2^*} \leq \|A\|\sqrt{|\mathcal{X}||\mathcal{Y}|}. \tag{7}$$

Forster's original paper [For02] shows

$$\mathrm{rank}_{\pm}(A) \geq \frac{\sqrt{|\mathcal{X}||\mathcal{Y}|}}{\|A\|},$$

and [LS09b] later improved this bound to

$$\operatorname{rank}_{\pm}(A) \geq \frac{|\mathcal{X}||\mathcal{Y}|}{\|A\|_{\gamma_2}^*}.$$

The following proposition, which is based on [LS09b, For02], recovers these bounds, as $\operatorname{rank}_{\pm}(A) \geq \mathrm{m}^{\mathrm{avg}}(A)^{-1}$ by Theorem 1.7.

Proposition 2.4. *For every sign-matrix $A_{\mathcal{X}\times\mathcal{Y}}$, we have*

$$\mathrm{m}^{\mathrm{avg}}(A)^{-1} \geq \frac{|\mathcal{X}||\mathcal{Y}|}{\|A\|_{\gamma_2}^*} \geq \frac{\sqrt{|\mathcal{X}||\mathcal{Y}|}}{\|A\|}.$$

Proof. Let ν be the uniform distribution on $\mathcal{Y}$, and consider a sign representation of A with unit vectors $u_x, v_y \in \mathbb{R}^d$ with $\mathrm{m}_\nu^{\mathrm{avg}}(A) = \min_x \mathbb{E}_y |\langle u_x, v_y\rangle|$. We have

$$\mathrm{m}^{\mathrm{avg}}(A) \leq \mathrm{m}_\nu^{\mathrm{avg}}(A) = \min_x \mathbb{E}_y |\langle u_x, v_y\rangle| \leq \mathbb{E}_{x,y} |\langle u_x, v_y\rangle| = \mathbb{E}_{x,y} \langle u_x, v_y\rangle A_{xy} \leq \frac{\|A\|_{\gamma_2}^*}{|\mathcal{X}||\mathcal{Y}|},$$

which combined with Equation (7) finishes the proof. □

2.2 VC dimension and average margin

Here we record a simple argument that shows that VC dimension is a weaker lower bound technique than Forster's method.

Proposition 2.5. *For every sign matrix A, we have $\mathrm{m}^{\mathrm{avg}}(A)^{-1} \geq \sqrt{\mathrm{VC}(A)}$.*

Proof. Suppose $\mathrm{VC}(A) = k$. By the definition of the VC dimension, A contains a $2^k \times k$ submatrix U_k with all the possible different rows. Note that

$$U_k^T U_k = 2^k \mathbf{I}_k.$$

In particular, we have $\|U_k\| = 2^{k/2}$, which combined with Proposition 2.4, gives

$$\mathrm{m}^{\mathrm{avg}}(A) \leq \mathrm{m}^{\mathrm{avg}}(U_k) \leq \frac{2^{k/2}}{\sqrt{k2^k}} = \frac{1}{\sqrt{k}}.$$ □

2.3 Small sign-rank implies large monochromatic rectangles

In this section, we provide a short and robust geometric argument for the fact that sign matrices of small sign-rank contain large monochromatic rectangles. Our proof is quite different from the proof of [FPS16], which is based on the divide-and-conquer cutting lemma of Chazelle [Cha93]. However, we note that our bound is slightly weaker than the $\frac{n}{2^{O(d\log d)}} \times \frac{n}{O(1)}$ bound of [FPS16].

Proposition 2.6. *There exists a constant $c > 0$ such that the following holds. Every sign matrix $A_{n\times n}$ with sign-rank d contains a monochromatic rectangle of size*

$$\frac{n}{2^{cd\log d}} \times \frac{n}{4d}.$$

Proof. Let S^{d-1} denote the unit sphere in $\mathbb{R}^d$. Consider a sign representation of A with unit vectors $u_i, v_j \in \mathbb{R}^d$. Without loss of generality, we can assume that the v_j's are in isotropic position.

For every $u \in S^{d-1}$, consider the spherical cap of height $\alpha := \frac{1}{\sqrt{2d}}$, defined as

$$C_u = \left\{ x \in S^{d-1} \ : \ \langle u, x\rangle \geq \sqrt{1-\alpha^2} \right\},$$

and the equator region

$$E_u = \left\{ x \in S^{d-1} \ : \ |\langle u, x\rangle| \leq \alpha \right\}.$$

Note that the sets

$$R_u^+ := \{i : u_i \in C_u\} \times \{j : v_j \notin E_u, \langle v_j, u\rangle > 0\}$$

and

$$R_u^- := \{i : u_i \in C_u\} \times \{j : v_j \notin E_u, \langle v_j, u\rangle < 0\}$$

correspond, respectively, to a $(+1)$-monochromatic and a (-1)-monochromatic rectangle in A.

Since the v_j's are in isotropic position, for every $u \in S^{d-1}$, we have

$$\sum_{j=1}^{n} \langle u, v_j\rangle^2 = \frac{n}{d}.$$

On the other hand

$$\sum_{j : v_j \in E_u} \langle u, v_j\rangle^2 \leq n\alpha^2 = \frac{n}{2d},$$

which shows

$$|\{j : v_j \notin E_u\}| \geq \frac{n}{d} - \frac{n}{2d} = \frac{n}{2d}.$$

In particular

$$|\{j : v_j \notin E_u, \langle v_j, u\rangle > 0\}| \geq \frac{n}{4d} \qquad \text{or} \qquad |\{j : v_j \notin E_u, \langle v_j, u\rangle < 0\}| \geq \frac{n}{4d}.$$

To estimate the surface area of C_u, recall that the surface area of the d-dimensional sphere of radius r is given by

$$A_d(r) := \frac{2\pi^{d/2}}{\Gamma(d/2)} r^{d-1} = \int_{-1}^{1} A_{d-1}\left(\sqrt{1-h^2}\right) dh = \frac{2\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})} \int_{-1}^{1} \left(\sqrt{1-h^2}\right)^{d-2} dh.$$

Hence the ratio between the surface area of C_u and the whole sphere S^{d-1} can be estimated as

$$\begin{aligned} \frac{|C_u|}{A_d(1)} &= \frac{\int_{\sqrt{1-\alpha^2}}^{1} \left(\sqrt{1-h^2}\right)^{d-2} dh}{\int_{-1}^{1} \left(\sqrt{1-h^2}\right)^{d-2} dh} \geq \frac{\int_{\sqrt{1-\alpha^2}}^{\sqrt{1-\frac{\alpha^2}{4}}} \left(\sqrt{1-h^2}\right)^{d-2} dh}{\int_{-1}^{1} 1 dh} \\ &\geq \frac{\sqrt{1-\frac{\alpha^2}{4}} - \sqrt{1-\alpha^2}}{2} \times (\alpha/2)^{d-2} = 2^{-O(d\log d)}. \end{aligned}$$

Picking a $u \in S^{d-1}$ uniformly at random, with positive probability, one of the rectangles R_u^+ or R_u^- satisfies the assertion of the theorem. $\square$

3 Main Results

3.1 Monochromatic rectangle ratio vs average margin

Our first theorem relates the monochromatic rectangle ratio of a sign matrix to its average margin.

Theorem 3.1. *For every sign matrix A, we have*

$$\mathrm{m}^{\mathrm{avg}}(A)^{-1} \le \mathrm{rect}(A)^{-1}.$$

Proof. Suppose $A_{\mathcal{X}\times\mathcal{Y}}$ is a sign matrix, and let ν be a probability distribution on $\mathcal{Y}$. Consider the following zero-sum game, where the first player chooses $x \in \mathcal{X}$ and the second player chooses a monochromatic rectangle R of A. Define the payoff of the game to be

$$\mathbb{E}_{y\sim\nu}\left[\mathbf{1}_R(x,y)\right] = \Pr_{y\sim\nu}\left[(x,y)\in R\right].$$

By the minimax theorem, we have

$$\mathrm{rect}(A) \le \min_{\mu}\max_{R}\mathbb{E}_{x\sim\mu}\,\mathbb{E}_{y\sim\nu}\left[\mathbf{1}_R(x,y)\right] = \max_{\eta}\min_{x}\mathbb{E}_{R\sim\eta}\,\mathbb{E}_{y\sim\nu}\left[\mathbf{1}_R(x,y)\right], \tag{8}$$

where μ ranges over all probability distributions over $\mathcal{X}$ and η ranges over all probability distributions on the set of monochromatic rectangles of A. Take η to maximize this quantity. Denote by $R_i = S_i \times T_i$ the i-th monochromatic rectangle in the support of η, and define $a_i \in \{-1,1\}$ to be the value of A on the rectangle R_i, and let $\eta_i = \Pr_{R\sim\eta}[R = R_i]$. Now, consider the vectors u_x and v_y defined with coordinates

$$u_x(i) = \sqrt{\eta_i}\cdot\mathbf{1}_{S_i}(x), \qquad v_y(i) = a_i\sqrt{\eta_i}\cdot\mathbf{1}_{T_i}(y).$$

Note that $\|u_x\|^2 = \Pr_{R=S\times T\sim\eta}[x\in S] \le 1$ and similarly $\|v_y\|^2 = \Pr_{R=S\times T\sim\eta}[y\in T]\le 1$. Let S be the matrix with entries $S_{xy} = \langle u_x, v_y\rangle$. For $(x,y)\in\mathcal{X}\times\mathcal{Y}$, it is clear that

$$S_{xy} = \sum_i \eta_i\cdot a_i\cdot\mathbf{1}_{R_i}(x,y) = A_{xy}\cdot\Pr_{R\sim\eta}\left[(x,y)\in R\right],$$

and thus $\mathrm{sgn}(S_{xy}) = A_{xy}$. Therefore, by Equation (8),

$$\mathrm{m}^{\mathrm{avg}}_{\nu}(A) \ge \min_x \mathbb{E}_{y\sim\nu}[|S_{xy}|] = \min_x \mathbb{E}_{y\sim\nu}\,\mathbb{E}_{R\sim\eta}[\mathbf{1}_R(x,y)] \ge \mathrm{rect}(A),$$

where the first inequality follows from the definition of average margin since S sign-represents A and the u_x and v_y's have Euclidean norm at most 1. $\square$

3.2 Sign-rank vs. current lower bound methods

Our next theorem shows a significant limitation for the three discussed lower bound methods. It shows that there are matrices with polynomially large sign-rank, while neither of the known methods can yield super constant bounds.

Theorem 3.2 (Main Theorem)**.** *There exists $n\times n$ sign matrices A with sign-rank $\Omega(\frac{n^{1/3}}{\log(n)})$ that satisfy*

$$\mathrm{VC}(A) \le 2 \quad \textit{and} \quad \mathrm{m}^{\mathrm{avg}}(A)^{-1} \le \mathrm{rect}(A)^{-1} \le 2^{15}.$$

Proof. The idea is to construct a large collection of sign matrices, each with a large monochromatic rectangle ratio. The statement then would follow from the upper bound on the number of matrices of small sign-rank, presented in Lemma 1.1.

Let N be a positive integer, and consider the sets

$$\mathcal{P} = \{(x, y) \in \mathbb{Z}^2 : 1 \leq x \leq N, 1 \leq y \leq 2N^2\}$$

and

$$\mathcal{L} = \{(a, b) \in \mathbb{Z}^2 : 1 \leq a \leq N, 1 \leq b \leq 2N^2\}.$$

We think of the elements $\ell = (a, b) \in \mathcal{L}$ as lines $y = ax + b$ in $\mathbb{R}^2$, and we consider $(x, y) \in \mathcal{P}$ as points in $\mathbb{R}^2$.

Define the sign matrix $F_{\mathcal{L},\mathcal{P}}$ by point line incidences:

$$F_{\ell,p} = \begin{cases} -1 & p \in \ell \\ 1 & p \notin \ell \end{cases}.$$

Set $n = N^3$ and note that F is a $2n \times 2n$ matrix, and for every $\ell = (a, b)$ and $p = (x, y)$, we have

$$\begin{aligned} F_{\ell,p} &= \operatorname{sgn}\left((ax + b - y)^2 - \frac{1}{2}\right) \\ &= \operatorname{sgn}\left(a^2x^2 - 2axy + y^2 + 2abx - 2by + \left(b^2 - \frac{1}{2}\right)\right). \end{aligned}$$

Since each term in the last line corresponds to a rank 1 matrix, we have

$$\operatorname{rank}_{\pm}(F) \leq 6.$$

Additionally, F has the following useful properties:

1. Since any two distinct lines have at most one point in common, F does not contain any 2×2 (-1)-monochromatic subrectangles.

2. Each line $\ell = (a, b)$ with $b \leq N^2$ goes through $N = n^{1/3}$ points from $\mathcal{P}$. Consequently, F contains at least $n^{\frac{4}{3}}$ negative entries.

Consider all $2n \times 2n$ sign matrices A that can be obtained from F by changing the sign of a subset of the negative entries to positive. There are at least $2^{n^{4/3}}$ such matrices. By Lemma 1.1, most such matrices A have sign-rank $\Omega(n^{1/3}/\log n)$. Let A be any such matrix obtained from F, so that $\operatorname{rank}_{\pm}(A) = \Omega(n^{1/3}/\log n)$.

Since A is obtained from a submatrix of F by only changing its -1 entries, A also satisfies the first property above. That is, A does not contain any 2×2 (-1)-monochromatic subrectangle, and consequently $\mathrm{VC}(A) \leq 2$ as desired.

We proceed to bounding the rectangle ratio and hence also the average margin of A. Let $\mu \times \nu$ be any product distribution on $\mathcal{L} \times \mathcal{P}$. Since $\operatorname{rank}_{\pm}(F) \leq 6$, by Theorem 1.9, there exists a monochromatic rectangle R of F with

$$\mu \times \nu(R) \geq 2^{-14}.$$

If R is a 1-monochromatic rectangle in F, then it is also a 1-monochromatic rectangle in A. On the other hand, if R is a (-1)-monochromatic rectangle in F, then by the first property above, it is either a $1 \times k$ or a $k \times 1$ rectangle for some k. In both cases R contains a subrectangle $R' \subseteq R$ that is monochromatic in A and satisfies

$$\mu \times \nu(R') \geq \frac{\mu \times \nu(R)}{2} \geq 2^{-15}.$$

We conclude that

$$\mathrm{rect}(A) \geq 2^{-15}.$$

Finally, by Theorem 3.1, we have $\mathrm{m}^{\mathrm{avg}}(A)^{-1} \leq \mathrm{rect}(A)^{-1} \leq 2^{15}$. $\square$

3.3 Does large margin imply small sign-rank?

Next, we discuss the relation between sign-rank and margin, namely Question 1.4 and Conjecture 1.5. We start with a short discussion of the equivalence of margin and several other complexity and analytic parameters associated with sign matrices. We have already mentioned the result of Linial and Shraibman [LS09a] stating

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

Let $\mathrm{R}_\epsilon(A)$ denote the public-coin randomized communication complexity of the matrix A with two-sided error ϵ. We refer the reader to [KN97] for a formal definition of this complexity measure. The following folklore proposition shows that for any fixed $\epsilon \in (0, \frac{1}{2})$, the gap between $\mathrm{disc}(A)^{-1}$ and $\mathrm{R}_\epsilon(A)$ is at most exponential.

Proposition 3.3 (folklore)**.** *For every $\epsilon \in (0, \frac{1}{2})$ and every sign-matrix A, we have*

$$\log\left((1-2\epsilon)\cdot \mathrm{disc}(A)^{-1}\right) \leq \mathrm{R}_\epsilon(A) \leq O\left(\log\left(\frac{1}{\epsilon}\right)\mathrm{disc}(A)^{-2}\right). \tag{9}$$

Proof. For a proof of the lower bound in Equation (9), we refer the reader to [LS07, Theorem 4.9]. We provide a proof of the upper bound.

Suppose $A_{\mathcal{X}\times\mathcal{Y}}$ is a sign matrix with $\mathrm{disc}(A) = \delta$. We will show that $\mathrm{R}_{\frac{1-\delta}{2}}(A) \leq 2$, and the upper bound of Equation (9) then follows by applying a standard error reduction procedure.

Recall the definition $\mathrm{disc}(A) = \min_\mu \max_R |\mathbb{E}_{xy\sim\mu}[A_{xy}\mathbf{1}_R(x,y)]|$, where R ranges over all combinatorial rectangles of $\mathcal{X}\times\mathcal{Y}$. Consider the zero-sum game, where the first player chooses (R, b) where R is a rectangle and $b \in \{-1, 1\}$ and the second player chooses $(x, y) \in \mathcal{X}\times\mathcal{Y}$. If $(x, y) \notin R$, then the payoff is zero, otherwise, the payoff is 1 if $b = A_{xy}$ and -1 if $b \neq A_{xy}$. Denote the mixed strategies of the first and the second players by ν and μ, respectively. By the minimax principle, we have

$$\mathrm{disc}(A) = \min_\mu \max_{R,b} \mathbb{E}_{xy\sim\mu}\left[bA_{xy}\mathbf{1}_R(x,y)\right] = \max_\nu \min_{x,y} \mathbb{E}_{(R,b)\sim\nu}\left[bA_{xy}\mathbf{1}_R(x,y)\right].$$

Therefore, there is a distribution ν over the set of all (R, b) such that for every (x, y), we have

$$\mathbb{E}_{(R,b)\sim\nu}\left[bA_{xy}\mathbf{1}_R(x,y)\right] \geq \delta.$$

Consider the two-bit randomized communication protocol π for A, where on inputs (x, y), the players use the public randomness to sample $(R, b) \sim \nu$, and output b if $(x, y) \in R$, and otherwise, output uniformly at random a ± 1 bit. Let $\pi(x, y)$ denote the output of the protocol on input (x, y). For every $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we have

$$\Pr[\pi(x,y) = A_{xy}] = \frac{1}{2} + \frac{1}{2}\mathbb{E}[\pi(x,y)\cdot A_{xy}] = \frac{1}{2} + \frac{1}{2}\Pr_R[(x,y)\in R]\,\mathbb{E}[b\cdot A_{xy}|(x,y)\in R] \geq \frac{1}{2} + \frac{\delta}{2},$$

where the second equality is due to the fact that $\mathbb{E}[\pi(x,y)A_{xy}] = 0$ when $(x,y) \notin R$.

We conclude that $\mathrm{R}_{\frac{1-\delta}{2}}(A) \leq 2$. Finally, by a standard error-reduction (see [RY20, Chapter 3] for example), we have

$$\mathrm{R}_\epsilon(A) \leq \mathrm{R}_{\frac{1-\delta}{2}}(A) \cdot O\left(\log\left(\frac{1}{\epsilon}\right) \cdot \delta^{-2}\right). \qquad \square$$

By Proposition 3.3, one can equivalently consider $\mathrm{R}_\epsilon(A)$ instead of $\mathrm{m}(A)^{-1}$ in Question 1.4 and Conjecture 1.5. This is particularly interesting in light of the equivalence of the logarithm of sign-rank and the unbounded-error communication complexity $\mathrm{U}(A)$, due to Paturi and Simon [PS86]:

$$\mathrm{U}(A) \coloneqq \lim_{\epsilon \nearrow \frac{1}{2}} \mathrm{R}^{\mathrm{prv}}_\epsilon(A) = \log(\mathrm{rank}_\pm(A)) + O(1). \tag{10}$$

We refer the reader to [KN97] for the definition of the private-coin randomized communication complexity $\mathrm{R}^{\mathrm{prv}}_\epsilon(A)$.

Finally, let us discuss the equivalence to approximate γ_2 norms. The following relationship with public-coin randomized communication complexity is known

$$\log\|A\|_{\gamma_2,\epsilon} \leq \mathrm{R}_{\frac{\epsilon}{2}}(A) \leq O\left(\frac{\log(1/\epsilon)}{(1-\epsilon)^2}\|A\|^2_{\gamma_2,\epsilon}\right), \tag{11}$$

where A is a sign matrix and $\epsilon \in (0,1)$. The lower bound is from [LS09b] and the upper bound is proven in [HHH21, Corollary 2.8 (c)]. However, since those papers use a different notation, for the convenience of the reader, we provide a proof in Proposition A.2.

To summarize, for every fixed $\epsilon \in (0, \frac{1}{2})$, we have

$$\mathrm{m}(A)^{-1} \approx \mathrm{disc}(A)^{-1} \approx \|A\|_{\gamma_2,\epsilon} \approx \mathrm{R}_\epsilon(A), \tag{12}$$

where the equivalence notation $\approx$ means that each parameter can be bounded by applying a universal function (that could depend on ϵ) to the other parameter.

The following proposition shows that a positive answer to Conjecture 1.5 is beyond the reach of the current known lower bound techniques.

Proposition 3.4 (Barrier to Conjecture 1.5)**.** *Let Q_d be the sign matrix defined in Conjecture 1.5. There exists a constant c such that for every $d \in \mathbb{N}$, we have*

$$\mathrm{m}(Q_d)^{-1} \leq c,$$

and

$$\mathrm{VC}(Q_d), \mathrm{m}^{\mathrm{avg}}(Q_d)^{-1}, \mathrm{rect}(Q_d)^{-1} \leq c.$$

Proof. The bound $\mathrm{m}(Q_d)^{-1} = O(1)$ follows from the equivalence of the margin and the randomized communication complexity discussed above, and the fact that $\mathrm{R}_{1/3}(Q_d) = O(1)$, due to [ZS09]. Since $\sqrt{\mathrm{VC}(A)} \leq \mathrm{m}^{\mathrm{avg}}(Q_d)^{-1} \leq \mathrm{m}(Q_d)^{-1}$, it only remains to show $\mathrm{rect}(Q_d)^{-1} = O(1)$.

Next, we will show how to bound $\mathrm{rect}(Q_d)^{-1}$. Let $\mathcal{X}$ and $\mathcal{Y}$ be the set of odd-parity and even-parity elements of $\{0,1\}^d$ corresponding, respectively, to the rows and columns of Q_d. Let μ and ν be distributions, respectively, over $\mathcal{X}$ and $\mathcal{Y}$. Recall that $Q_d(x,y) = -1$ iff $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ differ in exactly one coordinate. We will use the fact that Q_d does not contain any 2×3 or 3×2 (-1)-monochromatic rectangles, which also directly implies $\mathrm{VC}(Q_d) \leq 3$. We will consider two cases.

Case 1. Suppose

$$\Pr_{x\sim\mu, y\sim\nu}[Q_d(x,y) = -1] \geq c := 1/2.$$

Applying Jensen's inequality twice, we have

$$\begin{aligned} c^6 \leq \left(\mathbb{E}_{x\sim\mu, y\sim\nu}[\mathbf{1}_{Q_d(x,y)=-1}]\right)^6 &\leq \left(\mathbb{E}_{x\sim\mu}(\mathbb{E}_{y\sim\nu}[\mathbf{1}_{Q_d(x,y)=-1}])^3\right)^2 \\ &= \left(\mathbb{E}_{x\sim\mu}\,\mathbb{E}_{y_1,y_2,y_3\sim\nu}\left[\prod_j \mathbf{1}_{Q_d(x,y_j)=-1}\right]\right)^2 \\ &\leq \mathbb{E}_{y_1,y_2,y_3\sim\nu}\left(\mathbb{E}_{x\sim\mu}\left[\prod_{j=1}^3 \mathbf{1}_{Q_d(x,y_j)=-1}\right]\right)^2 \\ &= \mathbb{E}_{x_1,x_2\sim\mu, y_1,y_2,y_3\sim\nu}\left[\prod_{i,j} \mathbf{1}_{Q_d(x_i,y_j)=-1}\right]. \end{aligned}$$

The last term is the probability that the random rectangle $\{x_1,x_2\} \times \{y_1,y_2,y_3\}$ is a (-1)-monochromatic rectangle of Q_d. Since Q_d does not contain any 2×3 (-1)-monochromatic rectangle, we must have

$$\Pr[x_1 = x_2 \ \vee\ |\{y_1,y_2,y_3\}| \leq 2] \geq c^6.$$

Therefore, one of the two distributions μ or ν has noticeable collision probability. Specifically, either $\Pr_{x,x'\sim\mu}\Pr[x = x'] \geq c^6/4$ or $\Pr_{y,y'\sim\nu}[y = y'] \geq \frac{1}{3}\Pr[|\{y_1,y_2,y_3\}| \leq 2] \geq c^6/4$. Without loss of generality, assume that the former is true. In this case

$$\Pr\big[x = x'\big] = \sum_{a\in\mathcal{X}} \Pr[x = a]^2 \leq \max_{a\in\mathcal{X}} \Pr[x = a].$$

Therefore, there is an $a \in \mathcal{X}$ such that $\Pr[x = a] \geq c^6/8$. Now, note that the a'th row of Q_d either has a $\mu \times \nu$-measure of at least $c^6/16$ on its (-1)'s or on its 1's.

Case 2. If Case 1 does not hold, then

$$\Pr_{x\sim\mu, y\sim\nu}[|x-y|_1 \geq 3] \geq 1/2, \tag{13}$$

where $|x-y|_1$ denotes the Hamming distance between x and y. For a subset $S \subseteq [d]$, let $\phi_S : \{0,1\}^d \to \{0,1,2,3\}$ be defined as $\phi_S(x) = \sum_{i\in S} x_i \mod 4$.

For $x, y \in \{0,1\}^d$ satisfying $|x-y|_1 \geq 3$, let j_1, j_2, j_3 be distinct indices where they differ. Pick $S \subseteq [d]$ uniformly at random by first picking a random subset $S_1 \subseteq [d] \setminus \{j_1, j_2, j_3\}$ and then taking its union with a random $S_2 \subseteq \{j_1, j_2, j_3\}$. For every choice of S_1, there exists at least one choice of S_2 such that $|\phi_S(x) - \phi_S(y)| = 2$. Therefore,

$$\Pr_S\left[|\phi_S(x) - \phi_S(y)| = 2\right] \geq 1/8.$$

Combining with Equation (13), we have

$$\Pr_{\substack{S \\ x\sim\mu, y\sim\nu}}\left[|\phi_S(x) - \phi_S(y)| = 2\right] \geq \Pr_S\left[|\phi_S(x) - \phi_S(y)| = 2 \mid |x-y|_1 \geq 3\right] \Pr_{x\sim\mu, y\sim\nu}\left[|x-y|_1 \geq 3\right] \geq 1/16.$$

Hence, there is a choice of $S \subseteq [d]$ such that

$$\Pr_{x\sim\mu, y\sim\nu}\left[|\phi_S(x) - \phi_S(y)| = 2\right] \geq 1/16.$$

Hence, there exist $r, t \in \{0,1,2,3\}$ with $|r-t| = 2$ such that

$$\Pr_{x\sim\mu, y\sim\nu}\left[\phi_S(x) = r \text{ and } \phi_S(y) = t\right] \geq 2^{-8}.$$

In this case, the set $\{x|\phi_S(x) = r\} \times \{y|\phi_S(y) = t\}$ is a 1-monochromatic rectangle of measure at least 2^{-8}. □

In light of Proposition 3.4 it might seem worthwhile to seek a different candidate sign matrix for establishing a negative answer to Question 1.4. By Proposition 2.5 and the definition of average margin, for every sign matrix A, we have

$$\sqrt{\mathrm{VC}(A)} \leq \mathrm{m}^{\mathrm{avg}}(A)^{-1} \leq \mathrm{m}(A)^{-1}, \tag{14}$$

and thus Forster's method and the VC dimension method cannot imply a negative answer to Question 1.4. Therefore, $\mathrm{rect}(A)^{-1}$ remains the only known approach.

The following conjecture of Chattopadhyay, Lovett, and Vinyals [CLV19, Problem 6.1] (see also [HHH21, Conjecture I]), if true, would imply that $\mathrm{rect}(A)^{-1}$ is also small if $\mathrm{m}(A)^{-1}$ is small.

Conjecture 3.5 (Chattopadhyay, Lovett, Vinyals [CLV19])**.** *There exists a function η such that every sign matrix $A_{\mathcal{X}\times\mathcal{Y}}$ contains an $\frac{|\mathcal{X}|}{k} \times \frac{|\mathcal{Y}|}{k}$ monochromatic rectangle for $k \leq \eta(\mathrm{m}(A)^{-1})$.*

By Remark 1.10, Conjecture 3.5 is equivalent to the existence of a function η such that every sign matrix $A_{\mathcal{X}\times\mathcal{Y}}$, we have

$$\mathrm{rect}(A)^{-1} \leq \eta(\mathrm{m}(A)^{-1}).$$

In particular, assuming Conjecture 3.5, even $\mathrm{rect}(A)^{-1}$ cannot be used towards giving a negative answer to Question 1.4.

3.4 Communication Complexity with Equality Oracle

In Section 3.3, we showed that Question 1.4 can be formulated in terms of the approximate γ_2 norm: Is it true that for sign matrices, $\|A\|_{\gamma_2,\epsilon} = O(1)$ implies $\mathrm{rank}_\pm(A) = O(1)$? As we mentioned in Conjecture 1.5, we believe that the answer to this question is negative. However, it seems plausible that such a statement could hold if we strengthen the assumption by replacing the approximate γ_2 norm with the γ_2 norm:

Conjecture 3.6 (Conjecture 1.14 restated)**.** *There exists a function η such that for every sign matrix A, we have $\mathrm{rank}_\pm(A) \leq \eta(\|A\|_{\gamma_2})$.*

Zero-one valued matrices that satisfy $\|A\|_{\gamma_2} = O(1)$ are important in operator theory as they correspond to the bounded idempotents of the algebra of Schur multipliers. Inspired by Cohen's idempotent theorem in harmonic analysis, a characterization of these matrices was conjectured in [HHH21]. To state this conjecture, we need to introduce the notion of a blocky matrix. We call a zero-one valued matrix $M_{\mathcal{X}\times\mathcal{Y}}$ *blocky* if

$$\{(x,y) \mid M_{xy} = 1\} = \bigcup_i \mathcal{X}_i \times \mathcal{Y}_i,$$

for disjoint sets $\mathcal{X}_i \subseteq \mathcal{X}$ and disjoint sets $\mathcal{Y}_i \subseteq \mathcal{Y}$. A simple example of a blocky matrix is the *identity matrix*. Note that every blocky matrix can be obtained from the identity matrix by duplicating rows and columns and adding all zero rows and columns. Since the γ_2 norm is invariant under these operations, every non-zero blocky matrix M satisfies $\|M\|_{\gamma_2} = 1$. It is shown in [Liv95] that blocky matrices are precisely the set of Boolean matrices with $\|M\|_{\gamma_2} \leq 1$.

Blocky matrices are related to deterministic communication complexity with access to an equality oracle. In this model, a protocol for a sign matrix A corresponds to a binary tree. Each non-leaf node v in the tree corresponds to a query to $eq(a_v(x), b_v(y))$, where $eq(a,b) = 1$ if $a = b$ and -1 otherwise. Note that $a_v(x)$ and $b_v(y)$ can be computed, respectively, by the first and the second party in the communication protocol. Every input (x,y) naturally corresponds to a path from the root of the tree to a leaf, and it is required that the leaf is labeled with the correct value A_{xy}. The cost of the protocol is the depth of the tree. The deterministic communication complexity of the matrix A with access to an equality oracle, denoted by $\mathrm{D}^{eq}(A)$, is the smallest depth of such a protocol for A.

Note that for any two functions $a(x)$ and $b(y)$, the function $(x,y) \mapsto eq(a(x), b(y))$ corresponds to an $\mathcal{X}\times\mathcal{Y}$ blocky matrix as its 1's consist of a union of row-disjoint and column-disjoint rectangles.

Conjecture 3.7 ([HHH21, Conjecture III])**.** *For every sign-matrix A, if $\|A\|_{\gamma_2} = O(1)$, then A can be expressed as a ± 1-linear combination of $O(1)$ blocky matrices, equivalently $\mathrm{D}^{eq}(A) = O(1)$.*

The following theorem shows that if Conjecture 3.7 is true, then the answer to Conjecture 1.14 is positive.

Theorem 3.8. *For every sign matrix $A_{\mathcal{X}\times\mathcal{Y}}$, we have*

$$\mathrm{rank}_{\pm 1}(A) \leq 4^{\mathrm{D}^{eq}(A)}.$$

Proof. We proceed by induction on $d := \mathrm{D}^{eq}(A)$. When $d = 1$, A corresponds to a blocky matrix, which in fact has $\mathrm{rank}_\pm(A) \leq 3$. For larger d, consider a cost d protocol for a sign matrix A and suppose the equality query at the root of tree is $eq(a(x), b(y))$. We may assume without loss of generality that $a(x), b(y) \in \mathbb{N}$. Let $S_{\mathcal{X}\times\mathcal{Y}}$ be the matrix with entries $S_{xy} = \mathbf{1}_{a(x)=b(y)}$. We branch according to the output of the first query either to the left or the right subtree of the root, each corresponding to a protocol of cost at most $d-1$. Let the corresponding matrices for these protocols be Π_1 and Π_2, and note that

$$A = S \circ \Pi_1 + (\mathbf{J} - S) \circ \Pi_2,$$

where $\mathbf{J} \coloneqq \mathbf{J}_{\mathcal{X}\times\mathcal{Y}}$ is the all-ones matrix. By the induction hypothesis, Π_1 and Π_2 have sign-rank at most $\leq 4^{d-1}$. Let $\widetilde{\Pi}_1$ and $\widetilde{\Pi}_2$ be real matrices with rank at most 4^{d-1} that satisfy $\mathrm{sgn}(\widetilde{\Pi}_1) = \Pi_1$ and $\mathrm{sgn}(\widetilde{\Pi}_2) = \Pi_2$. Let $E_{\mathcal{X}\times\mathcal{Y}}$ be the rank-3 matrix with entries $E_{xy} = (a(x) - b(y))^2$. Note that for a sufficiently large k, we have

$$A = \mathrm{sgn}(\widetilde{\Pi}_1 + kE \circ \widetilde{\Pi}_2).$$

Finally, we have

$$\mathrm{rank}(\widetilde{\Pi}_1 + k\widetilde{\Pi}_2 \circ E) \leq \mathrm{rank}(\widetilde{\Pi}_1) + \mathrm{rank}(\widetilde{\Pi}_2) \cdot \mathrm{rank}(E) \leq 4^{d-1} + 3 \cdot 4^{d-1} = 4^d. \qquad \square$$

Conjecture 3.7 is inspired by quantitative versions of Cohen's seminal idempotent theorem in harmonic analysis, developed by Green and Sanders [GS08a, GS08b] and Sanders [San19, San11, San20a]. As it is noticed in [HHH21], these theorems verify Conjecture 3.7 for a large natural class of matrices: sign matrices $A_{G\times G}$ where G is a finite group and the entries are defined as $A_{xy} = f(xy^{-1})$ for some $f : G \to \{-1, 1\}$. Note that taking $G = \mathbb{Z}_2^n$ corresponds to the class of XOR-lifts, which is a well studied class of functions in communication complexity. In the following corollary, we combine these results with Theorem 3.8 to verify Conjecture 1.14 for this class of matrices.

Corollary 3.9. *There exists a function η such that the following holds. If $f : G \to \{-1, 1\}$ for a finite group G, and $A_{G\times G}$ is the sign matrix with entries $A(x, y) = f(xy^{-1})$, then*

$$\mathrm{rank}_{\pm}(A) \leq \eta(\|A\|_{\gamma_2}).$$

In the case of abelian G, we have

$$\mathrm{rank}_{\pm}(A) \leq \exp\big(\exp\big(C\|A\|_{\gamma_2}^4\big)\big),$$

Proof. By applying results of Davidson and Donsig [DD07] and Mathias [Mat93], it is shown in [HHH21, Corollary 3.13] that every matrix $A_{G\times G}$ with $A_{xy} = f(xy^{-1})$ for a function $f : G \to \mathbb{R}$ satisfies

$$\|A\|_{\gamma_2} = \frac{1}{|G|}\|A\|_{\mathrm{tr}} = \|f\|_{A(G)},$$

where $\|f\|_{A(G)}$ denotes the Fourier algebra norm of f.

In particular, for the matrix A in the statement of Corollary 3.9, we have $\|A\|_{\gamma_2} = \|f\|_{A(G)}$, and $f : G \to \{-1, 1\}$. This puts us in the setting of idempotent theorems in harmonic analysis: [San11, Theorem 1.2] states that there is a constant $\ell = L(\|f\|_{A(G)})$, subgroups $H_1, \ldots, H_\ell \subseteq G$, elements $a_1, \ldots, a_\ell \in G$, and signs $\sigma_1, \ldots, \sigma_\ell \in \{-1, 1\}$ such that

$$f = \sum_{i=1}^{\ell} \sigma_i \mathbf{1}_{H_i a_i}.$$

In particular

$$A_{xy} = \sum_{i=1}^{\ell} \sigma_i \mathbf{1}_{H_i a_i}(xy^{-1}).$$

This implies $\mathrm{D}^{eq}(A) \leq \ell$ as $\mathbf{1}_{H_i a_i}(xy^{-1})$ can be evaluated using a single equality oracle query to check that x and $a_i y$ belong to the same right coset of H_i.

When G is an abelian group, Sanders [San20b] proved that it is possible to take $\ell \leq \exp\big(C\|f\|_A^4\big)$, which combined with Theorem 3.8 gives the desired double exponential bound. $\square$

3.5 Sign-rank of Semi-algebraic matrices, an open problem.

We start by discussing why Question 1.11, Question 1.12, and Question 1.13 are equivalent. Recall that a d-dimensional *semi-algebraic* set of description complexity t is of the form

$$\left\{ y \in \mathbb{R}^d \ : \ \Gamma(\mathbf{1}_{p_1(y)\geq 0}, \ldots, \mathbf{1}_{p_t(y)\geq 0}) = 1 \right\}.$$

for a predicate $\Gamma : \{0,1\}^t \to \{0,1\}$ and polynomials $p_1, \ldots, p_t$ on d variables.

Proof of Equivalence of Question 1.11 and Question 1.12. Clearly, Question 1.12 is a special case of Question 1.11. In order to prove the nontrivial direction of this equivalence, consider a semi-algebraic sign-matrix A defined by points $u_1, \ldots, u_m \in \mathbb{R}^d$ and semi-algebraic sets $K_1, \ldots, K_n \subseteq \mathbb{R}^d$, each with description complexity t. Note that there are only 2^{2^t} different possible predicates $\{0,1\}^t \to \{0,1\}$, and hence in Question 1.11, we can assume without loss of generality that all the sets K_i are defined using the same predicate $\Gamma : \{0,1\}^t \to \{0,1\}$.

Let $p \in \mathbb{R}[x_1, \ldots, x_d]$ be a polynomial of degree t. Let $I_{d,t}$ denote the set of all $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{Z}_{\geq 0}^d$ with $\sum_{i=1}^d \alpha_i \leq t$. The monomials of degree at most t in variables $x_1, \ldots, x_d$ are indexed by $\alpha \in I_{d,t}$ with the correspondence $x^\alpha = x_1^{\alpha_1} \ldots x_d^{\alpha_d}$. Note that every polynomial $p(x) = \sum_{\alpha \in I_{d,t}} a_\alpha x^\alpha$ of degree at most t corresponds to an inner product

$$p(x) = \langle \Psi_t(p), \Phi_t(x) \rangle,$$

where $\Psi_t(p) \coloneqq (a_\alpha)_{\alpha \in I_{d,t}} \in \mathbb{R}^{|I_{d,t}|}$ and $\Phi_t(x) \coloneqq (x^\alpha)_{\alpha \in I_{d,t}} \in \mathbb{R}^{|I_{d,t}|}$. Applying this linearization idea to all the defining polynomials of the semi-algebraic sets allows us to view the matrix A as a single predicate applied to a collection of sign matrices, each of sign-rank at most $|I_{d,t}|$ each. $\square$

Hence, Question 1.11, Question 1.12, and Question 1.13 are all equivalent. Question 1.13, in particular, has a simple statement. Regarding this formulation, Bun, Mande, and Thaler [BMT21] used Forster's method to show the existence of matrices A and B of sign-rank d such that $\mathrm{rank}_{\pm}(A \wedge B) \geq \Omega(d^2)$. However, the following corollary of Theorem 3.1 shows that neither of the known methods can imply a negative answer to Question 1.13.

Corollary 3.10 (Corollary to Theorem 3.1). *If A and B are two $m \times n$ sign matrices of sign-rank at most d, then*

$$\mathrm{VC}(A \wedge B) \leq 20d \qquad \textit{and} \qquad \mathrm{m}^{\mathrm{avg}}(A \wedge B)^{-1} \leq \mathrm{rect}(A \wedge B)^{-1} \leq 2^{4d+4}.$$

Proof. By Equation (1), we have $\mathrm{VC}(A), \mathrm{VC}(B) \leq d$. By Theorem 1.9, we have

$$\mathrm{rect}(A), \mathrm{rect}(B) \geq 2^{-2(d+1)},$$

which immediately implies

$$\mathrm{rect}(A \wedge B) \geq \mathrm{rect}(A)\,\mathrm{rect}(B) \geq 2^{-4(d+1)}.$$

Hence, by Theorem 3.1, we have

$$\mathrm{m}^{\mathrm{avg}}(A \wedge B)^{-1} \leq \mathrm{rect}(A \wedge B)^{-1} \leq 2^{4d+4}.$$

We can combine this with Proposition 2.5 to upper bound $\mathrm{VC}(A \wedge B)$. However, as it is shown in [BEHW89], a direct proof yields a stronger upper bound:

Consider a set of k columns. By the Sauer-Shelah lemma, the corresponding $m \times k$ submatrices of A and B have at most $\sum_{i=0}^{d} \binom{k}{i}$ distinct rows. It follows that the corresponding submatrix in $A \wedge B$ has at most $\left(\sum_{i=0}^{d} \binom{k}{i}\right)^2$ distinct rows. For $k = 20d$, we have

$$\left(\sum_{i=0}^{d} \binom{k}{i}\right)^2 < (d+1)^2 \left(\frac{ke}{d}\right)^{2d} \leq (d+1)^2 \,(20e)^{2d} \leq (d+1)^2 2^{12d} < 2^{20d},$$

and thus no set of $20d$ columns is shattered in $A \wedge B$. □

Intersections of Half-spaces. The problem of bounding the sign-rank of $A \wedge B$ is closely related to bounding the sign-rank of the matrices that are defined by points and intersections of pairs of half-spaces. For *distinct* $y, y' \in \mathbb{R}^d$, let $I_{y,y'} = \{z \mid \langle y, z\rangle > 0\} \cap \{z \mid \langle y', z\rangle > 0\} \subset \mathbb{R}^d$ denote the intersection of the two half-spaces defined by y and y', respectively. We refer to these sets as half-space intersections. Given a finite set of points $\mathcal{X} \subseteq \mathbb{R}^d$ and a finite set of half-space intersections $\mathcal{I}$ in $\mathbb{R}^d$, define the matrix $F_{\mathcal{X} \times \mathcal{I}}$ as

$$F_{x,I} = \begin{cases} 1 & x \in I \\ -1 & x \notin I \end{cases}.$$

Is sign-rank of F bounded by a constant c_d? Note that for $x \in \mathcal{X}$ and $I_{y,y'} \in \mathcal{I}$, we have $F_{x,I} = \mathrm{sgn}\,\langle x, y\rangle \wedge \mathrm{sgn}\,\langle x, y'\rangle$, and thus F can be expressed as the $\wedge$ of two sign matrices of sign-rank d. Consequently, such a constant c_d exists if the answer to Question 1.13 regarding the sign-rank of $A \wedge B$ is positive.

It turns out that the opposite direction is also true, but with a slight increase in the value of d.

Claim 3.11. *If the constant c_{2d-1} exists, then for sign matrices A and B with $\mathrm{rank}_\pm(A), \mathrm{rank}_\pm(B) \leq d$, we have $\mathrm{rank}_\pm(A \wedge B) = O(c_{2d-1})$.*

Proof. Consider two sign matrices $A_{m\times n}$ and $B_{m\times n}$ of sign-rank at most d. There are vectors $u_i, v_j, u'_i, v'_j \in \mathbb{R}^d$ such that $A_{ij} = \mathrm{sgn}\langle u_i, v_j\rangle$ and $B_{ij} = \mathrm{sgn}\langle u'_i, v'_j\rangle$ for $i \in [m]$ and $j \in [n]$. By adding a small independent noise to the vectors, we may assume that all the coordinates of these vectors are non-zero. We then normalize these vectors according to the value of their last coordinate and assume that the their last coordinates are ± 1.

We partition the set of rows $i \in [m]$ into 2^{2d} many parts according to the value of $(\mathrm{sgn}(u_i), \mathrm{sgn}(u'_i)) \in \{-1,1\}^d \times \{-1,1\}^d$. Similarly, we partition the columns into 2^{2d} parts according to the sign patterns of v_j and v'_j. These two partitions divide $A \wedge B$ into $2^{2d} \times 2^{2d}$ blocks, and to prove the claim, it suffices to bound the sign-rank of each block. Hence, without loss of generality, we assume that $(\mathrm{sgn}(u_i), \mathrm{sgn}(u'_i), \mathrm{sgn}(v_j), \mathrm{sgn}(v'_j))$ is a fixed vector in $\{-1,1\}^{4d}$ that does not depend on i or j.

Since $\langle u_i, v_j\rangle = \langle -u_i, -v_j\rangle$, we can assume that the last coordinate of every u_i and u'_i is positive and thus it is equal to 1. Let $a, a' \in \{-1,1\}$ denote the last coordinates of v_j's and v'_j's, respectively. For every vector $w \in \mathbb{R}^d$, let $\widetilde{w} \in \mathbb{R}^{d-1}$ denote the restriction of w to the first $d-1$ coordinates. Note that

$$\langle u_i, v_j\rangle = \left\langle (\widetilde{u_i}, \widetilde{u'_i}, 1), (\widetilde{v_j}, 0^{d-1}, a)\right\rangle$$

and

$$\langle u'_i, v'_j\rangle = \left\langle (\widetilde{u_i}, \widetilde{u'_i}, 1), (0^{d-1}, \widetilde{v'_j}, a')\right\rangle,$$

thus $\mathrm{sgn}\,\langle u_i, v_j\rangle \wedge \mathrm{sgn}\left\langle u'_i, v'_j\right\rangle = 1$ iff $(\widetilde{u_i}, \widetilde{u'_i}, 1)$ belongs to the intersection of the two half-spaces defined by $(\widetilde{v_j}, 0^{d-1}, a)$ and $(0^{d-1}, \widetilde{v'_j}, a')$. $\square$

It is communicated to us by Shay Moran that it is known that the matrix F defined by half-space intersections in $\mathbb{R}^3$ has bounded sign-rank.

Proposition 3.12 (Communicated by Shay Moran)**.** *There exists a constant c_3 such that given a finite set $\mathcal{X}$ of points $x \in \mathbb{R}^3$ and a finite set $\mathcal{I}$ of half-space intersections $I_{y,y'}$ in $\mathbb{R}^3$, the matrix $F_{\mathcal{X}\times\mathcal{I}}$ with entries*

$$F_{x,I} = \begin{cases} 1 & x \in I \\ -1 & x \notin I \end{cases}$$

satisfies $\mathrm{rank}_{\pm}(F) \leq c_3$.

Proof. Recall that for $x \in \mathcal{X}$ and $I \coloneqq I_{y,y'} \in \mathcal{I}$, we have

$$F_{x,I} = \mathrm{sgn}\,\langle x, y\rangle \wedge \mathrm{sgn}\,\langle x, y'\rangle.$$

Adding a small independent noise to the vectors x, y, y' allows us to assume throughout the proof that these vectors have several non-degeneracy properties. For example, we may assume that all the coordinates are non-zero. Similar to the proof of Claim 3.11, it suffices to consider the case where

- The sign pattern of x, y, y' is fixed. More precisely, there exists $\Gamma \in (\{-1,1\}^3)^3$ such that $(\mathrm{sgn}(x), \mathrm{sgn}(y), \mathrm{sgn}(y')) = \Gamma$ for all $x \in \mathcal{X}$ and $I_{y,y'} \in \mathcal{I}$.
- The third coordinate of every $x \in \mathcal{X}$ is 1.

Every $x \in \mathcal{X}$ is of the form $x = (x_1, x_2, 1)$, corresponding to the point $(x_1, x_2) \in \mathbb{R}^2$. For every $y \in \mathbb{R}^3$, define $\ell_y : \mathbb{R}^2 \to \mathbb{R}$ as $\ell_y(x) \coloneqq \langle x, y\rangle = x_1y_1 + x_2y_2 + y_3$. For $I \coloneqq I_{y,y'} \in \mathcal{I}$, consider the two lines on $\mathbb{R}^2$ defined by $\ell_y(x) = 0$ and $\ell_{y'}(x) = 0$, and let $q^I = (q_1^I, q_2^I)$ be their intersection, i.e., the unique solution to $\ell_y(q) = \ell_{y'}(q) = 0$. By the discussion in the first paragraph of the proof, we may assume that there exists an $\epsilon > 0$ such that

- The two lines $\ell_y(x) = 0$ and $\ell_{y'}(x) = 0$ are not parallel, and thus the point q^I is well-defined.

- $|q_1^I - x_1| > \epsilon$ for all $x = (x_1, x_2, 1) \in \mathcal{X}$ and $I \in \mathcal{I}$.
- For every $I_{y,y'} \in \mathcal{I}$ and $x \in \mathcal{X}$, we have $|\ell_y(x)|, |\ell_{y'}(x)| > \epsilon$.

Note that the lines $\ell_y(x) = 0$ and $\ell_{y'}(x) = 0$ divide $\mathbb{R}^+$ into four open regions such that the signs of $\ell_y(x)$ and $\ell_{y'}(x)$ are fixed in each region. Let P_I be the unique region that is entirely contained in the half-space $\{(x_1, x_2) \ : \ x_1 > q_1^I\}$. Let $(\sigma_I, \sigma'_I) \in \{-1, +1\}^2$ be such that $\operatorname{sgn}(\ell_y(x)) = \sigma_I$ and $\operatorname{sgn}(\ell_{y'}(x)) = \sigma'_I$ for every $x \in P_I$.

We will construct a real matrix $B_{\mathcal{X}\times\mathcal{I}}$ of low rank such that for all $x \in \mathcal{X}$ and $I \coloneqq I_{y,y'} \in \mathcal{I}$, we have

$$\operatorname{sgn}(B_{x,I}) = \operatorname{sgn}\langle x, y\rangle \wedge \operatorname{sgn}\langle x, y'\rangle = F_{x,I}.$$

To define the entries of B, we consider the four possible cases for (σ_I, σ'_I).

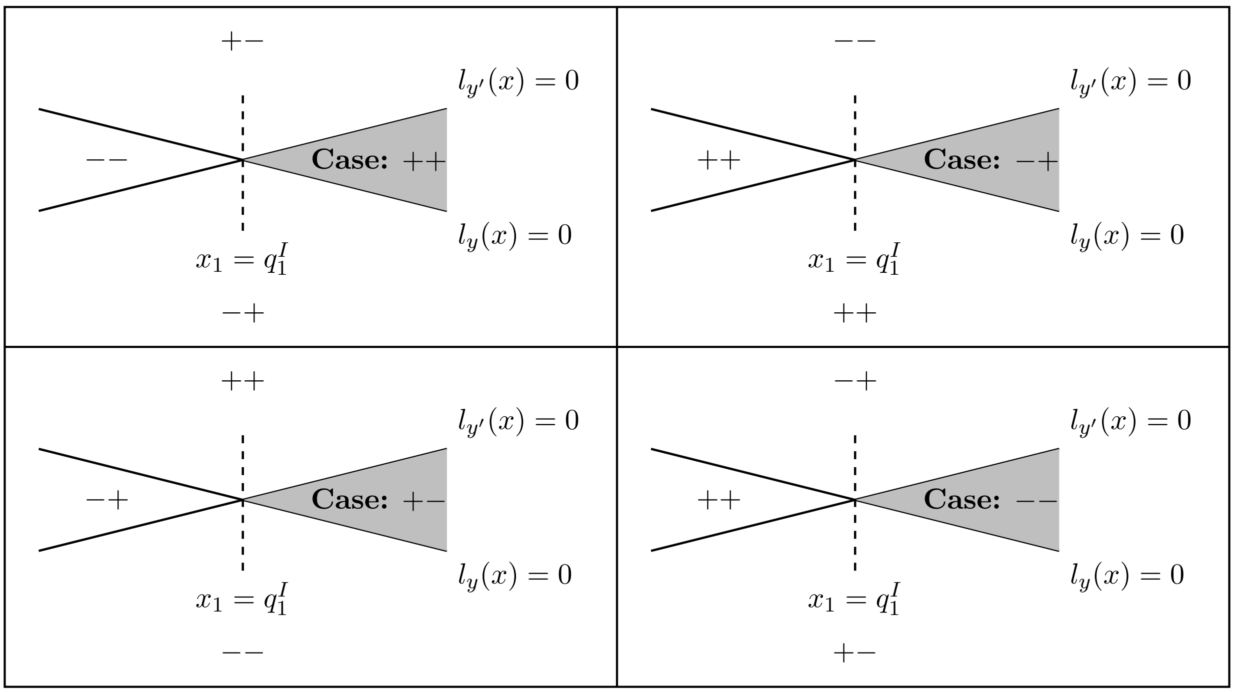

Figure 1: The grey area is P_I. The pairs of signs correspond to $\operatorname{sgn}(\ell_y), \operatorname{sgn}(\ell_{y'})$ on each region.

- **Case** $(\sigma_I, \sigma'_I) = (+1, +1)$**:** Let K be a large constant, and for every $x \in \mathcal{X}$, define

$$B_{x,I} \coloneqq e^{K(x_1 - q_1^I)}\ell_y(x)\ell_{y'}(x) - 1.$$

By choosing K to be sufficiently large, for every $x \in \mathcal{X}$, we have

$$\operatorname{sgn}(B_{x,I}) = \begin{cases} -1 & x_1 < q_1^I \\ \operatorname{sgn}(\ell_y(x)\ell_{y'}(x)) & x_1 > q_1^I \end{cases}.$$

Note that as a function of x, we have

$$B_{x,I} \in \operatorname{span}\left\{e^{Kx_1}, e^{Kx_1}x_1, e^{Kx_1}x_2, e^{Kx_1}x_1^2, e^{Kx_1}x_2^2, e^{Kx_1}x_1x_2, 1\right\},$$

and thus the restriction of B to the points $I_{y,y'}$ that satisfy the conditions of this case has rank at most 7.

- **Case** $(\sigma_I, \sigma'_I) = (-1, +1)$**:** In this case, define

$$B_{x,I} := e^{K(x_1 - q_1^I)}\ell_y(x) + \ell_{y'}(x),$$

and note that for every $x \in \mathcal{X}$, we have

$$\operatorname{sgn}(B_{x,I}) = \begin{cases} \operatorname{sgn}(\ell_{y'}(x)) & x_1 < q_1^I \\ \operatorname{sgn}(\ell_y(x)) & x_1 > q_1^I \end{cases}.$$

In this case, as a function of x, we have

$$B_{x,I} \in \operatorname{span}\left\{e^{Kx_1}, e^{Kx_1}x_1, e^{Kx_1}x_2, x_1, x_2, 1\right\},$$

and thus the restriction of B to these columns has rank at most 6.

- **Case** $(\sigma_I, \sigma'_I) = (+1, -1)$**:** In this case, define

$$B_{x,(y,y')} := e^{K(x_1 - q_1^I)}\ell_{y'}(x) + \ell_y(x),$$

and note that for every $x \in \mathcal{X}$, we have

$$\operatorname{sgn}(B_{x,I}) = \begin{cases} \operatorname{sgn}(\ell_y(x)) & x_1 < q_1^I \\ \operatorname{sgn}(\ell_{y'}(x)) & x_1 > q_1^I \end{cases}.$$

Similar to the previous case, the restriction of B to these columns has rank at most 6.

- **Case** $(\sigma_I, \sigma'_I) = (-1, -1)$**:** In this case, define

$$B_{x,I} := e^{K(q_1^I - x_1)}\ell_y(x)\ell_{y'}(x) - 1,$$

and note that for every $x \in \mathcal{X}$, we have

$$\operatorname{sgn}(B_{x,I}) = \begin{cases} -1 & x_1 > q_1^I \\ \operatorname{sgn}(\ell_y(x)\ell_{y'}(x)) & x_1 < q_1^I \end{cases}.$$

In this case, as a function of x, we have

$$B_{x,I} \in \operatorname{span}\left\{e^{-Kx_1}, e^{-Kx_1}x_1, e^{-Kx_1}x_2, e^{-Kx_1}x_1^2, e^{-Kx_1}x_2^2, e^{-Kx_1}x_1x_2, 1\right\},$$

which shows that the rank is at most 7.

These cases are illustrated in Figure 1. In all cases, for every $x \in \mathcal{X}$, we have

$$\operatorname{sgn}(B_{x,I}) = \operatorname{sgn}\langle x, y\rangle \wedge \operatorname{sgn}\langle x, y'\rangle = F_{x,I}.$$

□

Proposition 3.12 combined with Claim 3.11 implies the following special case of Question 1.13 for sign matrices of sign-rank at most 2.

Corollary 3.13. *There is a constant $c > 0$ such that for every two sign matrices $B_{m\times n}$ and $C_{m\times n}$ with sign-rank at most 2, we have* $\operatorname{rank}_{\pm}(B \wedge C) \le c$.

3.6 Average Communication Complexity

In this section, we observe a simple connection between sign-rank and another model of communication complexity, average communication complexity. For any distribution μ over $\mathcal{X} \times \mathcal{Y}$, let $\mathrm{CC}^{\mathrm{avg}}_{\mu}(A)$ be the smallest expected communication complexity of a deterministic protocol that computes A correctly on all inputs. Moreover, define

$$\mathrm{CC}^{\mathrm{avg}}(A) = \sup_{\mu} \mathrm{CC}^{\mathrm{avg}}_{\mu}(A),$$

where μ ranges over all *product* distributions over $\mathcal{X} \times \mathcal{Y}$.

Proposition 3.14. *For every sign-matrix $A_{\mathcal{X}\times\mathcal{Y}}$, we have*

$$\mathrm{CC}^{\mathrm{avg}}(A) \leq 2\,\mathrm{rect}(A)^{-1}.$$

Proof. Let μ be any distribution on $\mathcal{X} \times \mathcal{Y}$ and let $\delta = \mathrm{rect}(A)$. By definition, A has a monochromatic rectangle $R = S \times T$ such that $\mu(R) \geq \delta$. The two parties recursively proceed as follows. Given x and y as inputs, after communicating the two bits $\mathbf{1}_{x\in S}$ and $\mathbf{1}_{y\in T}$, they can agree on whether $(x, y) \in R$. At which point, they have reduced their search to one of the four matrices $A_{S\times T}$, $A_{S^c\times T}$, $A_{S\times T^c}$, and $A_{S^c\times T^c}$. Note that in the first case, both parties know the answer and can conclude the protocol. In all the other three cases, the μ-measure of the search-space has been reduced to at most $1 - \delta$, and they can recurse on the resulting submatrix according to the same protocol applied to μ conditioned on the submatrix.

For a distribution μ, let c_μ denote the average cost of the above protocol, and let μ be the maximizer for c_μ. Let μ_1, μ_2, μ_3 denote μ conditioned on $S \times T^c$, $S^c \times T^c$, and $S^c \times T$ respectively. We have

$$c_\mu \leq 2\Pr[(x, y) \in R] + \Pr[(x, y) \notin R] \cdot (2 + \max_i c_{\mu_i}) = 2 + \Pr[(x, y) \notin R] \max_i c_{\mu_i} \leq 2 + (1 - \delta)c_\mu.$$

Therefore, $c_\mu \leq 2/\delta$ as claimed. □

Combined with Theorem 1.9, we get the following bound in terms of sign-rank.

Corollary 3.15. *For every sign matrix A we have*

$$\mathrm{CC}^{\mathrm{avg}}(A) \leq 2^{2\mathrm{rank}_{\pm}(A)+3}.$$

Theorem 3.2 shows that there is no converse to Corollary 3.15. In particular, there are $n\times n$ sign matrices A with sign-rank $n^{\Omega(1)}$ and $\mathrm{rect}(A) = O(1)$. By Proposition 3.14, we have $\mathrm{CC}^{\mathrm{avg}}(A) = O(1)$, and thus there is a strong separation between sign-rank and $\mathrm{CC}^{\mathrm{avg}}(A)$.

4 Concluding remarks

In light of the results in the present paper, the following open problem captures the limitation of the currently known lower bound techniques for sign-rank.

Problem 4.1. *Construct an explicit sequence of matrices A_n such that $\mathrm{rect}(A_n)^{-1} = O(1)$ and*

$$\lim_{n\to\infty} \mathrm{rank}_{\pm}(A_n) = \infty.$$

By Theorem 3.2, we know such sequences of matrices exist. On the other hand, by Theorem 3.1 and Proposition 2.5, we have

$$\sqrt{\mathrm{VC}(A)} \leq \mathrm{m}^{\mathrm{avg}}(A)^{-1} \leq \mathrm{rect}^{-1}(A),$$

and thus none of the known lower bound techniques are capable of solving Problem 4.1. Note that a positive answer to Conjecture 1.5 would solve Problem 4.1.

Finally, let us mention that it is unclear whether the proof of Proposition 3.12 can be generalized to infinite matrices, which raises the following intriguing question.

Question 4.2. *Is the sign-rank of an infinite sign matrix $A_{\mathbb{N}\times\mathbb{N}}$ finite if the sign-rank of every finite submatrix of $A_{\mathbb{N}\times\mathbb{N}}$ is bounded by a fixed constant d?*

Acknowledgement We wish to thank Shachar Lovett and Shay Moran for helpful discussions. We would like to thank Shay Moran for sharing the proof of Proposition 3.12 with us.

References

[AFR85] Noga Alon, Peter Frankl, and Vojtech Rödl. Geometrical realization of set systems and probabilistic communication complexity. In *26th Annual Symposium on Foundations of Computer Science, FOCS 1985*, pages 277–280. IEEE Computer Society, 1985.

[AMY16] Noga Alon, Shay Moran, and Amir Yehudayoff. Sign rank versus VC dimension. In *Proceedings of the 29th Conference on Learning Theory, COLT 2016*, volume 49, pages 47–80, 2016.

[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.

[Bar98] Franck Barthe. On a reverse form of the Brascamp-Lieb inequality. *Invent. Math.*, 134(2):335–361, 1998.

[BDES02] Shai Ben-David, Nadav Eiron, and Hans Ulrich Simon. Limitations of learning via embeddings in Euclidean half spaces. *J. Mach. Learn. Res.*, 3(Spec. Issue Comput. Learn. Theory):441–461, 2002.

[BEHW89] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *J. Assoc. Comput. Mach.*, 36(4):929–965, 1989.

[BMT21] Mark Bun, Nikhil S. Mande, and Justin Thaler. Sign-rank can increase under intersection. *ACM Trans. Comput. Theory*, 13(4):Art. 24, 17, 2021.

[BT16] Mark Bun and Justin Thaler. Improved Bounds on the Sign-Rank of AC0. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, volume 55 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 37:1–37:14, 2016.

[Cha93] Bernard Chazelle. Cutting hyperplanes for divide-and-conquer. *Discrete Comput. Geom.*, 9(2):145–158, 1993.

[CLRS01] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to algorithms.* MIT Press, Cambridge, MA; McGraw-Hill Book Co., Boston, MA, second edition, 2001.

[CLV19] Arkadev Chattopadhyay, Shachar Lovett, and Marc Vinyals. Equality alone does not simulate randomness. In *34th Computational Complexity Conference (CCC 2019)*, 2019.

[CM18] Arkadev Chattopadhyay and Nikhil S. Mande. Separation of unbounded-error models in multi-party communication complexity. *Theory Comput.*, 14(1):1–23, 2018.

[DD07] Kenneth R. Davidson and Allan P. Donsig. Norms of Schur multipliers. *Illinois J. Math.*, 51(3):743–766, 2007.

[EMRPS14] Marek Eliáš, Jiří Matoušek, Edgardo Roldán-Pensado, and Zuzana Safernová. Lower bounds on geometric Ramsey functions. *SIAM J. Discrete Math.*, 28(4):1960–1970, 2014.

[Fel17] Vitaly Feldman. A general characterization of the statistical query complexity. In *Proceedings of the 30th Conference on Learning Theory, COLT 2017*, volume 65 of *Proceedings of Machine Learning Research*, pages 785–830, 2017.

[FGL+12] Jacob Fox, Mikhail Gromov, Vincent Lafforgue, Assaf Naor, and János Pach. Overlap properties of geometric expanders. *J. Reine Angew. Math.*, 671:49–83, 2012.

[FGV21] Vitaly Feldman, Cristóbal Guzmán, and Santosh Vempala. Statistical query algorithms for mean vector estimation and stochastic convex optimization. *Math. Oper. Res.*, 46(3):912–945, 2021.

[FKL+01] Jürgen Forster, Matthias Krause, Satyanarayana V. Lokam, Rustam Mubarakzjanov, Niels Schmitt, and Hans Ulrich Simon. Relations between communication complexity, linear arrangements, and computational complexity. In *FST TCS 2001: Foundations of software technology and theoretical computer science*, volume 2245, pages 171–182. 2001.

[For02] Jürgen Forster. A linear lower bound on the unbounded error probabilistic communication complexity. volume 65, pages 612–625. 2002. Special issue on complexity, 2001 (Chicago, IL).

[FPS16] Jacob Fox, János Pach, and Andrew Suk. A polynomial regularity lemma for semi-algebraic hypergraphs and its applications in geometry and property testing. *SIAM Journal on Computing*, 45(6):2199–2223, 2016.

[FPS+17] Jacob Fox, János Pach, Adam Sheffer, Andrew Suk, and Joshua Zahl. A semi-algebraic version of Zarankiewicz's problem. *J. Eur. Math. Soc. (JEMS)*, 19(6):1785–1810, 2017.

[FS06] Jürgen Forster and Hans Ulrich Simon. On the smallest possible dimension and the largest possible margin of linear arrangements representing given concept classes. *Theoret. Comput. Sci.*, 350(1):40–48, 2006.

[GS08a] Ben Green and Tom Sanders. Boolean functions with small spectral norm. *Geometric and Functional Analysis*, 18(1):144–162, 2008.

[GS08b] Ben Green and Tom Sanders. A quantitative version of the idempotent theorem in harmonic analysis. *Ann. of Math. (2)*, 168(3):1025–1054, 2008.

[HHH21] Lianna Hambardzumyan, Hamed Hatami, and Pooya Hatami. Dimension-free bounds and structural results in communication complexity. *Israel J. Math.*, 2021. To appear.

[HHL20] Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Sign rank vs discrepancy. In *35th Computational Complexity Conference*, volume 169 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 18, 14. 2020.

[HKLM20] Max Hopkins, Daniel Kane, Shachar Lovett, and Gaurav Mahajan. Point location and active learning: Learning halfspaces almost optimally. In *In 61st Annual Symposium on Foundations of Computer Science (FOCS 2020)*, pages 1034–1044. IEEE Computer Society, 2020.

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

[KS07] Adam R. Klivans and Alexander A. Sherstov. Unconditional lower bounds for learning intersections of halfspaces. *Mach. Learn.*, 69(2-3):97–114, 2007.

[KS11] Michael Kallweit and Hans Ulrich Simon. A close look to margin complexity and related parameters. In Sham M. Kakade and Ulrike von Luxburg, editors, *COLT 2011 - The 24th Annual Conference on Learning Theory*, volume 19 of *JMLR Proceedings*, pages 437–456. JMLR.org, 2011.

[Liv95] Leo Livshits. A note on 0-1 Schur multipliers. *Linear Algebra Appl.*, 222:15–22, 1995.

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

[LS07] Troy Lee and Adi Shraibman. Lower bounds in communication complexity. *Found. Trends Theor. Comput. Sci.*, 3(4):front matter, 263–399 (2009), 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.

[Mat93] Roy Mathias. The Hadamard operator norm of a circulant and applications. *SIAM J. Matrix Anal. Appl.*, 14(4):1152–1167, 1993.

[Mat96] Jiří Matoušek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel J. Math.*, 93:333–344, 1996.

[Mil64] J. Milnor. On the Betti numbers of real varieties. *Proc. Amer. Math. Soc.*, 15:275–280, 1964.

[Nao18] Assaf Naor. Metric dimension reduction: a snapshot of the Ribe program. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. I. Plenary lectures*, pages 759–837. World Sci. Publ., Hackensack, NJ, 2018.

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

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

[RY20] Anup Rao and Amir Yehudayoff. *Communication Complexity: and Applications*. Cambridge University Press, 2020.

[San11] Tom Sanders. A quantitative version of the non-abelian idempotent theorem. *Geom. Funct. Anal.*, 21(1):141–221, 2011.

[San19] Tom Sanders. Boolean functions with small spectral norm, revisited. *Math. Proc. Cambridge Philos. Soc.*, 167(2):335–344, 2019.

[San20a] Tom Sanders. Bounds in Cohen's idempotent theorem. *J. Fourier Anal. Appl.*, 26(2):Paper No. 25, 64, 2020.

[San20b] Tom Sanders. Bounds in cohen's idempotent theorem. *Journal of Fourier Analysis and Applications*, 26(2):1–64, 2020.

[She08a] Alexander A. Sherstov. Halfspace matrices. *Comput. Complexity*, 17(2):149–178, 2008.

[She08b] Alexander A. Sherstov. The unbounded-error communication complexity of symmetric functions. In *Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '08, page 384–393, 2008.

[SS05] Nathan Srebro and Adi Shraibman. Rank, trace-norm and max-norm. In *Learning theory*, volume 3559 of *Lecture Notes in Comput. Sci.*, pages 545–560. Springer, Berlin, 2005.

[Suk16] Andrew Suk. Semi-algebraic Ramsey numbers. *J. Combin. Theory Ser. B*, 116:465–483, 2016.

[SW19] Alexander A. Sherstov and Pei Wu. Near-optimal lower bounds on the threshold degree and sign-rank of AC0. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, page 401–412, New York, NY, USA, 2019. Association for Computing Machinery.

[Tho65] René Thom. Sur l'homologie des variétés algébriques réelles. In *Differential and Combinatorial Topology (A Symposium in Honor of Marston Morse)*, pages 255–265. Princeton Univ. Press, Princeton, N.J., 1965.

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

[ZS09] Zhiqiang Zhang and Yaoyun Shi. Communication complexities of symmetric XOR functions. *Quantum Inf. Comput.*, 9(3-4):255–263, 2009.

A Appendix

Recall the following well-known inequality.

Lemma A.1 (Hoeffding's inequality)**.** *For $i = 1, \ldots, n$, let X_i be independent random variables taking values from range $[a_i, b_i]$ and let $X = \sum_{i=1}^n X_i$. Then,*

$$\Pr[|X - \mathbb{E}[X]| \geq t] < 2\exp\left(-\frac{2t^2}{\sum_i (b_i - a_i)^2}\right).$$

The next proposition proves the equivalence of the approximate γ_2 norm and the randomized communication complexity.

Proposition A.2 ([LS09b, HHH21])**.** *For every sign matrix $A_{\mathcal{X}\times\mathcal{Y}}$ and every $\epsilon \in (0,1)$, we have*

$$\log \|A\|_{\gamma_2,\epsilon} \leq \mathrm{R}_{\frac{\epsilon}{2}}(A) \leq O\left(\frac{\log(1/\epsilon)}{(1-\epsilon)^2}\|A\|_{\gamma_2,\epsilon}^2\right).$$

Proof. **Lower-bound:** Consider a randomized protocol π_R of cost $c = \mathrm{R}_{\frac{\epsilon}{2}}(A)$ that computes $A_{\mathcal{X}\times\mathcal{Y}}$ with two-sided error at most $\frac{\epsilon}{2}$. In this notation, the subscript R denotes the random variable that corresponds to the randomness in the protocol, and any fixation of R to a value r corresponds to a deterministic protocol π_r of communication cost at most c. Let Π_r denote the matrix that corresponds to the output of the deterministic protocol π_r. A deterministic communication protocol π_r of cost c provides a partition of $\mathcal{X} \times \mathcal{Y}$ into at most 2^c rectangles, and thus Π_r can be written as a sum of at most 2^c rank-1 sign matrices. Since the γ_2 norm of a non-zero rank-1 sign matrix is 1, we have $\|\Pi_r\|_{\gamma_2} \leq 2^c$. By convexity

$$\|\mathbb{E}_R[\Pi_R]\|_{\gamma_2} \leq \mathbb{E}_R[\|\Pi_R\|_{\gamma_2}] \leq \max_r \|\Pi_r\|_{\gamma_2} \leq 2^c.$$

Since π_R has error at most $\epsilon/2$, we have

$$|A_{xy} - \mathbb{E}_R[\pi_R(x,y)]| = 2 \cdot \Pr[A_{xy} \neq \pi_R(x,y)] \leq \epsilon,$$

which implies $\|A\|_{\gamma_2,\epsilon} \leq 2^c$ as desired.

Upper-bound: The approximate norm $\|A\|_{\gamma_2,\epsilon}$ is defined as the *infimum* of $\|B\|_{\gamma_2}$ such that $\|A - B\| \leq \epsilon$. Hence, for every $\eta > 0$, there exists a real matrix B with $\|B\|_{\gamma_2} \leq \|A\|_{\gamma_2,\epsilon}$ and $\|A - B\|_\infty \leq \epsilon + \eta$. Pick a small positive $\eta < \frac{1-\epsilon}{2}$, and consider such a B.

As it is stated in [LS09a, Equation (2.3)], it follows from Grothendieck's inequality that the γ_2 norm is equivalent to the so-called ν-norm. In particular, there exist rank-1 sign matrices $B_1, \ldots, B_m$ and real numbers $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ with $L := \sum_{i=1}^m |\lambda_i| \leq \frac{\pi}{2\ln(1+\sqrt{2})}\|B\|_{\gamma_2}$ such that

$$B = \sum_{i=1}^m \lambda_i B_i.$$

We will convert this to a randomized protocol. Pick D randomly from $\{B_1, \ldots, B_m\}$ according to the probability distribution

$$\Pr[D = B_i] = \frac{|\lambda_i|}{\sum_{i=1}^k |\lambda_i|}.$$

Note that for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we have $\mathbb{E}[D_{xy}] = B_{xy}/L$ and $|D_{xy}| = 1$. Let $\delta = \frac{1-\epsilon}{2}$ and $N = 2\delta^{-2}L^2 \log(4/\epsilon) = \frac{8L^2 \log(4/\epsilon)}{(1-\epsilon)^2}$. Let $D_1, \ldots, D_N$ be i.i.d. copies of D and define $\widetilde{D} = \frac{L}{N}\sum_{i=1}^N D_i$.

Note that for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$, by applying Hoeffding's inequality (Lemma A.1), we have

$$\Pr\left[|\widetilde{D}_{xy} - B_{xy}| \geq \delta\right] < 2\exp\left(-\frac{2\delta^2}{4N \cdot (L/N)^2}\right) \leq \frac{\epsilon}{2},$$

where the last inequality is by the choice of N.

Let E be the ± 1 rounding of $\widetilde{D}$, that is $E_{xy} = 1$ iff $\widetilde{D}_{xy} \geq 0$. Since $\|B - A\|_\infty \leq \epsilon + \eta$, for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we have

$$\begin{aligned} \Pr[E_{xy} \neq A_{xy}] &\leq \Pr\left[|\tilde{D}_{xy} - B_{xy}| \geq 1 - \epsilon - \eta\right] \leq \Pr\left[|\tilde{D}_{xy} - B_{xy}| \geq \frac{1-\epsilon}{2}\right] \\ &\leq \Pr\left[|\widetilde{D}_{xy} - B_{xy}| \geq \delta\right] \leq \frac{\epsilon}{2}. \end{aligned}$$

Each D_i can be computed with communication cost at most 2. Since $\tilde{D}_{xy}$ can be computed by rounding a linear combination of N such D_i's, it can be computed with communication cost at most $2N = O\left(\frac{\log(1/\epsilon)}{(1-\epsilon)^2}\|A\|^2_{\gamma_2,\epsilon}\right)$. This concludes the statement. $\square$

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il