Electronic Colloquium on Computational Complexity, Report No. 130 (2022)

A Borsuk-Ulam lower bound for sign-rank and its applications

Hamed Hatami * Kaave Hosseini † Xiang Meng‡

September 15, 2022

Abstract

We introduce a new topological argument based on the Borsuk-Ulam theorem to prove a lower bound on sign-rank.

- This result implies the strongest possible separation between randomized and unbounded-error communication complexity. More precisely, we show that for a particular range of parameters, the randomized communication complexity of the Gap Hamming Distance problem is $O(1)$ while its unbounded-error communication complexity is $\Omega(\log(n))$.

  Previously, it was unknown whether the unbounded-error communication complexity could be asymptotically larger than the randomized communication complexity.
- In connection to learning theory, we prove that, despite its learnability properties, the class of large margin half-spaces in $\mathbb{R}^d$ is genuinely high-dimensional, i.e., it cannot be embedded in $\mathbb{R}^{d-1}$. This result is closely related to a recent conjecture of Alon, Hanneke, Holzman, and Moran (FOCS 2021) about the VC dimension of this class.
- Our final application is to the theory of dimension reductions. The Johnson-Lindenstrauss theorem implies that any set of N unit vectors is embeddable in dimension $O(\gamma^{-2}\log N)$ without altering the signs of those pairwise inner products that have absolute values at least $\gamma > 0$. Our result establishes the tightness of this bound, which answers a question of Linial, Mendelson, Schechtman, and Shraibman (Combinatorica, 27(2007)) in the case of partial functions.

1 Introduction

Our main result, which is a geometric fact about the two notions of *margin* and *dimension*, has different formulations and consequences in communication complexity, learning theory, and the theory of metric dimension reductions.

In communication complexity, we study the relation between the *(shared randomness) bounded-error randomized communication complexity*, denoted by $\mathrm{R}(\cdot)$, and the *unbounded-error communication complexity*, denoted by $\mathrm{U}(\cdot)$.

The unbounded-error model allows the error probability to be arbitrarily close to $\frac{1}{2}$, and this relaxed requirement makes it more powerful than any of the usual communication models. Therefore, lower bounds against this model are highly desirable as they apply to many other models and

*School of Computer Science, McGill University. Email: hatami@cs.mcgill.ca. Supported by an NSERC grant.
†Department of Computer Science, University of Rochester. Email: kaave.hosseini@rochester.edu.
‡School of Computer Science, McGill University. Email: xiang.meng2@mail.mcgill.ca.

ISSN 1433-8092

problems. However, this additional power naturally means that proving such lower bounds is often difficult and requires richer mathematics.

In comparison, the bounded-error model is generally much weaker even though it has access to shared randomness. An extensive collection of tools for proving lower bounds against this model exist, and in fact, all the known lower bound techniques against $\mathrm{U}(M)$, such as the counting argument of [AFR85], the Vapnik-Chervonenkis (VC) dimension lower bound of [PS86], or Forster's method [For02] and its extensions, also imply strong lower bounds against $\mathrm{R}(M)$.

These considerations raise an interesting question: *whether there exists any communication problem that is easier for the bounded-error model than for the unbounded-error model.*

Naturally, proving a positive answer to this question requires a lower bound technique against $\mathrm{U}(\cdot)$ that would not apply to $\mathrm{R}(\cdot)$. In the present paper, we use a novel topological argument based on the Borsuk-Ulam theorem to achieve this goal.

There is a (partial) Boolean matrix $M_{2^n \times 2^n}$ with $\mathrm{R}(M) = O(1)$, but $\mathrm{U}(M) = \Omega(\log n)$.

It follows from Newman's theorem [New91] that this is the best possible separation.

Next, we discuss the connection to learning theory. Binary classification is one of the most widely studied problems in machine learning. Its goal is to learn a model that can distinguish between positive and negative examples. The algorithm receives training data, a collection of fully labelled samples, and it must produce a model that can make low-error predictions for the labels of the unseen data points.

The geometric representation of concepts as half-spaces is central to binary classification. Indeed, many standard approaches, such as support vector machines, embed complex concept classes in half-spaces and afterward apply efficient learning algorithms for half-spaces.

Dimension is arguably the most critical attribute of these representations as it corresponds to the number of parameters in the model. The smallest possible dimension where such a representation is possible is called the *dimension complexity* of the concept class. The dimension complexity is equal to the so-called *sign-rank* of the matrix representing the concept class (see Definition 1.1).

The *curse of dimensionality* refers to a commonly observed phenomenon in which the performances of many algorithms deteriorate exponentially as the number of parameters increases. However, a widely used class of learning algorithms, called *support vector machines*, can overcome high dimensionality challenges by requiring extra assumptions about the data. Regardless of the dimension, these algorithms have strong success guarantees when we confine the learning task to the data points separated from the decision boundary by a non-negligible margin $\gamma > 0$.

To be more rigorous, let the unit sphere $\mathbf{S}^{d-1} \subset \mathbb{R}^d$ represent the set of data points. Fix a parameter $\gamma \in (0,1)$, and define $\mathbb{G}_\gamma^d$ to be the class of all partial concepts $h_y : \mathbf{S}^{d-1} \to \{-1, 1, *\}$ for $y \in \mathbf{S}^{d-1}$ where

$$h_y : x \mapsto \begin{cases} 1 & \langle x, y \rangle > \gamma \\ -1 & \langle x, y \rangle < -\gamma \\ * & \text{otherwise} \end{cases} . \tag{1}$$

The success of support vector machines in dealing with high-dimensional data is due to the fact that their sample complexity in learning $\mathbb{G}_\gamma^d$ does not depend on the dimension d.

The fact that $\mathbb{G}_\gamma^d$ evades some aspects of the curse of dimensionality raises the question of whether it is genuinely high-dimensional. Perhaps surprisingly, in Theorem 1.5, we prove that the answer is *yes* in the following sense.

For every $\gamma \in (0,1)$, the dimension complexity of $\mathbb{G}^d_\gamma$ is exactly d.

This result is also closely related to a recent conjecture of Alon, Hanneke, Holzman, and Moran [AHHM22] about the VC dimension of $\mathbb{G}^d_\gamma$. We will elaborate on these connections in Section 1.4.1.

Finally, we discuss the connection to dimension reductions. The Johnson–Lindenstrauss lemma states that every finite set of points in any Euclidean space can be embedded into a Euclidean space of low dimension in such a way that distances between the points are nearly preserved. In particular, every set of N unit vectors is embeddable into $\mathbb{R}^{O(\log(N))}$ in such a way that the pairwise inner products between the points are approximately preserved. Such an approximation will not alter the signs of those inner products that are bounded away from zero by a large margin. It follows that [AV06, LMSS07] any set of N unit vectors is embeddable in dimension $O(\gamma^{-2}\log N)$ without altering the signs of those pairwise inner products that have magnitude at least $\gamma > 0$. Linial, Mendelson, Schechtman, and Shraibman [LMSS07] asked whether the dependency on N is necessary[1]. In Theorem 1.11, we essentially resolve their question by proving a lower bound of $\Omega(\gamma^{-2}\log N)$.

In the remaining sections of the introduction, we formally state our results and elaborate on their applications.

1.1 Notation

We will use the standard computer science asymptotic notations [CLRS01] of $O(\cdot)$, $\Omega(\cdot)$, $\Theta(\cdot)$, $o(\cdot)$, and $\omega(\cdot)$. The inner product of $x = (x_1, \cdots, x_d) \in \mathbb{R}^d$ and $y = (y_1, \cdots, y_d) \in \mathbb{R}^d$ is $\langle x, y\rangle \coloneqq \sum_{i=1}^d x_i \cdot y_i$. This defines the Euclidean norm $\|x\| \coloneqq \sqrt{\langle x, x\rangle} = \sqrt{\sum_{i=1}^d x_i^2}$.

We denote the $(d-1)$-dimensional unit sphere by $\mathbf{S}^{d-1} = \{x \in \mathbb{R}^d \ : \ \|x\| = 1\}$. Given a parameter $\delta \in (0,1)$, a δ-net for $\mathbf{S}^{d-1}$ is a finite subset $T \subset \mathbf{S}^{d-1}$ such that for all $x \in \mathbf{S}^{d-1}$ there is a $t \in T$ with $\|x - t\| \le \delta$.

Throughout the article, we sometimes identify a matrix $A_{\mathcal{X}\times\mathcal{Y}}$ with the corresponding function on $\mathcal{X} \times \mathcal{Y}$, defined as $(x, y) \mapsto A_{x,y}$.

1.2 Background

A *sign matrix* is a matrix with ± 1 entries. A *partial sign matrix* is a matrix with $\{-1, 1, *\}$ entries where $*$'s represent "invalid" entries. When there is ambiguity, we will use the term *total sign matrix* to differentiate sign matrices from partial sign matrices.

The *sign-rank* of a sign matrix $A_{\mathcal{X}\times\mathcal{Y}}$, denoted by $\mathbf{rk}_\pm(A)$, is the smallest rank of a real matrix $B_{\mathcal{X}\times\mathcal{Y}}$ such that the entries of B are nonzero and have the same signs as their corresponding entries in A. Note that the definition of sign-rank naturally extends to partial matrices, where for invalid entries of A, the corresponding entry in B could be any real number.

Sign-rank is a fundamental notion in learning theory [BDES02, KS07, She08a, SS05, Fel17, FGV21], communication complexity [PS86, CM18, She08b, HHL22], and discrete geometry [AFR85, FGL+12, FPS+17, Suk16, EMRPS14]. It also arises naturally as a lower bound tool in circuit complexity [RS10, BT16, SW19] and the theory of metric dimension reductions [Mat96, Nao18].

[1] To be more precise, [LMSS07] assumes that *all* the initial pairwise inner products have magnitude at least $\gamma > 0$.

Geometrically, sign-rank is the smallest dimension in which the matrix has a realization as points and half-spaces, and for this reason, sign-rank is sometimes called the *dimension complexity.* For the record, we will state the definition of sign-rank in this terminology.

Definition 1.1 (Sign-rank)**.** *The sign-rank of a partial sign matrix $A_{\mathcal{X}\times\mathcal{Y}}$ is the smallest d such that there exist unit vectors $u_x, v_y \in \mathbb{R}^d$ with $A_{xy} = \operatorname{sgn}(\langle u_x, v_y\rangle)$ for all valid $(x, y) \in \mathcal{X} \times \mathcal{Y}$.*

The unit vectors in Definition 1.1 represent A as points and half-spaces in the d-dimensional space: $A_{xy} = 1$ if and only if the point u_x belongs to the half-space $\{z : \langle z, v_y\rangle > 0\}$.

There is a second geometric 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 point u_x and the hyperplanes defined by v_y.

Definition 1.2 (Margin)**.** *The* margin *of a partial 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 minimum is over all valid $(x, y) \in \mathcal{X} \times \mathcal{Y}$, and the supremum is over all $d \in \mathbb{N}$ and unit vectors $u_x, v_y \in \mathbb{R}^d$ with $A_{xy} = \operatorname{sgn}(\langle u_x, v_y\rangle)$ for all valid $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

Linial and Shraibman [LS09] proved that for total matrices, the margin is equivalent to the well-studied notion of discrepancy in communication complexity:

$$\operatorname{disc}(A) \le \mathrm{m}(A) \le 8 \operatorname{disc}(A).$$

1.3 Communication Complexity

Every sign matrix $A_{\mathcal{X}\times\mathcal{Y}}$ corresponds to a communication problem in the standard two-party model, where one player receives $x \in \mathcal{X}$, and the other receives $y \in \mathcal{Y}$. Their objective is to determine $A_{x,y}$ by exchanging as few bits of communication as possible. In communication complexity, it is common to assume that the rows and columns are indexed by n-bit strings, i.e. $\mathcal{X} = \mathcal{Y} \coloneqq \{-1, 1\}^n$. We refer to those problems as *n-bit communication problems.*

Many natural problems in communication complexity are *promise problems* where the input is promised to belong to a particular subset of all possible inputs. Equivalently, one can allow invalid inputs but assume that -1 and 1 are both acceptable outputs in case the players receive an invalid input. We represent these problems by partial sign matrices.

The *randomized communication complexity* of A, denoted by $\mathrm{R}(A)$, is the smallest cost of a probabilistic communication protocol that computes A in the *shared randomness model* with error probability at most $1/3$ on every valid input [KN97, Definition 3.12]. The particular choice of $1/3$ is unimportant as long as one is concerned with a fixed error probability in $(0, 1/2)$. In this case, the error can be reduced to any desired constant by running the protocol a few times and outputting the majority answer.

In the *unbounded-error* model of communication, the players have access to their *private* sources of randomness. Their objective is only to outperform a random guess: for every valid input $(x, y) \in \mathcal{X}\times\mathcal{Y}$, they must correctly produce the value of $A_{x,y}$ with a probability strictly larger than $1/2$. The unbounded-error communication complexity of A, denoted by $\mathrm{U}(A)$, is the smallest protocol cost that computes A in this model. We emphasize that the unbounded-error communication complexity must be defined with *private randomness* since in the shared randomness model, achieving the error probability less than $1/2$ is always possible using 2 bits of communication.

Paturi and Simon [PS86] proved that sign-rank provides an elegant characterization of the unbounded-error communication complexity:

$$\log \mathbf{rk}_{\pm}(A) \leq \mathrm{U}(A) \leq 2 + \log \mathbf{rk}_{\pm}(A). \tag{2}$$

It is straightforward to verify that this characterization is true for *partial* sign matrices as well.

The unbounded-error model is one of the most powerful models of communication, and the unbounded-error communication complexity provides lower bounds for many other notions of complexity. Nevertheless, we prove a separation that shows that the unbounded-error communication complexity could be asymptotically larger than the randomized communication complexity. The example giving this separation is the well-known Gap Hamming distance problem.

Definition 1.3. *For $\varepsilon \in (0,1)$ and $n \in \mathbb{N}$, the gap Hamming distance is defined on $\{-1,1\}^n \times \{-1,1\}^n$ as*

$$\mathrm{GHD}^n_{1-\varepsilon}(x,y) = \begin{cases} 1 & \langle x,y\rangle > (1-\varepsilon)n \\ -1 & \langle x,y\rangle < -(1-\varepsilon)n \\ * & \textit{otherwise} \end{cases}.$$

The communication complexity of the gap Hamming distance problem is well-studied in the regime of $\varepsilon = 1 - \Theta(\frac{1}{\sqrt{n}})$, where Chakrabarti and Regev [CR11] (see also [She12]) showed that the bounded-error randomized communication complexity is $\Theta(n)$. However, here we are interested in a different regime, namely when ε is bounded away from 1.

Theorem 1.4 (Communication complexity of Gap Hamming Distance)**.** *For $\varepsilon \in (0, \frac{1}{9})$, and $n \in \mathbb{N}$, we have*

(i) $\mathrm{R}(\mathrm{GHD}^n_{1-\varepsilon}) = 2$.

(ii) $\mathrm{U}(\mathrm{GHD}^n_{1-\varepsilon}) = \log(n) - O\left(\log\left(\varepsilon^{-1}\right)\right)$.

The upper bound in Theorem 1.4 (i) is easy. A formal proof is given in Section 4. The difficulty in proving the theorem lies in establishing (ii), which requires proving a strong lower bound on the sign-rank of the gap Hamming distance problem. This is achieved in Theorem 1.11.

The tightness of the separation of Theorem 1.4 follows from Newman's theorem [New91], which implies that for every (total or partial) n-bit communication problem A, we have

$$\mathrm{U}(A) \leq \mathrm{R}(A) + O(\log(n)). \tag{3}$$

Theorem 1.4 shows that, at least for partial matrices, the logarithmic additive term in Eq. (3) is necessary.

1.4 Learning theory

In learning theory, sign matrices represent binary concept classes, which correspond to statistical binary classification problems. Every column $y \in \mathcal{Y}$ of a sign matrix $A_{\mathcal{X}\times\mathcal{Y}}$ corresponds to an object in some domain. A classification algorithm receives, as input, a training set consisting of labelled samples $(y_1, A_{x,y_1}), \ldots, (y_m, A_{x,y_m})$ for an unknown $x \in \mathcal{X}$, and its task is to predict $A_{x,y}$ for other values of y.

Characterizing learnability is the primary subject matter in learning theory. The fundamental theorem of statistical learning states that a binary (total) concept class is learnable if and only if its VC dimension is finite. However, most known learning algorithms do not rely on explicit upper bounds on the VC dimension; instead, they use 'low-complexity' geometric representations of the concept classes. The data points are represented by real-valued feature vectors and the concepts are modelled by real-valued parameters. The learning algorithm uses the training data to navigate this geometric space to find a setting of the parameters that represents a good hypothesis. This broad description encompasses many standard learning algorithms.

Dimension, perhaps, is the most natural way to measure the complexity of these geometric representations. It corresponds to the number of features describing the data points and the number of parameters the algorithm uses to model hypotheses.

Margin is another notion that arises in the context of many common learning algorithms. It corresponds to the distance between the data points and the *decision boundaries*, which are the hypersurfaces in the classification model that partition the underlying space into positive and negative points.

While the notions of dimension and margin, in these general terms, appear in the context of many different learning algorithms, e.g., neural networks, in this article, we are interested in the setting of linear classifiers. In this setting, points and half-spaces represent concept classes as in Definition 1.1. The minimum possible choice for the dimension is called the *dimension complexity* of the concept class, and it is equal to the *sign-rank* of the corresponding matrix.

The notion of margin, introduced in Definition 1.2, quantifies how well the large margin classifiers such as *support vector machines* can learn the class. To be more precise, every finite partial sign matrix A with $\mathrm{m}(A) = \gamma$ is a submatrix of $\mathbb{G}_\gamma^d$ in some dimension d. Moreover, the sample complexity of the support vector machines in learning $\mathbb{G}_\gamma^d$ depends only on γ^{-1} and is independent of the dimension [SSBD14, Theorem 15.4, Theorem 26.13]. In fact, even the classical perceptron algorithm of Rosenblatt [Ros58] is guaranteed to make at most γ^{-2} mistakes, regardless of the value of d [SSBD14, Theorem 9.1].

It is a curious fact that the partial concept class $\mathbb{G}_\gamma^d$, despite its high dimensional representation, is easy to learn. This might raise the suspicion that $\mathbb{G}_\gamma^d$ has a low dimensional linear representation. But in the following theorem, we show that this is not true, and in fact, $\mathbb{G}_\gamma^d$ cannot be realized even in a $(d-1)$-dimensional Euclidean space.

Theorem 1.5 (Sign-rank of Gap Inner Product)**.** *For every $\gamma \in (0,1)$ and every $d \in \mathbb{N}$, the sign-rank of the $\mathbf{S}^{d-1} \times \mathbf{S}^{d-1}$* gap inner product

$$\mathbb{G}_\gamma^d(x,y) = \begin{cases} 1 & \langle x,y\rangle > \gamma \\ -1 & \langle x,y\rangle < -\gamma \\ * & \textit{otherwise} \end{cases}.$$

is exactly d. For $\gamma = 1$, the sign-rank is 1*.*

We present the proof of Theorem 1.5 in Section 2. The proof uses the following version of the Borsuk-Ulam theorem.

Theorem 1.6 (Borsuk-Ulam)**.** *For every continuous $f : \mathbf{S}^d \to \mathbb{R}^d$, there is $x \in \mathbf{S}^d$ such that $f(x) = f(-x)$.*

We refer the reader to the excellent book of Matoušek [Mat03] for more applications of this fascinating theorem.

1.4.1 The VC theory

Recall that the *VC dimension* of a total sign matrix A is the largest d such that A contains a $2^d \times d$ submatrix with distinct rows. The fundamental theorem of PAC learning asserts that a total concept class is PAC-learnable if and only if its VC dimension is bounded.

In a recent paper, Alon, Hanneke, Holzman, and Moran [AHHM22] initiated a theory of PAC learning for partial concept classes. One of the examples that motivated their study is the learnability of $\mathbb{G}_\gamma^d$. They proposed the following conjecture [AHHM22, Question 21] to show that the learnability of $\mathbb{G}_\gamma^d$ does not follow from the standard VC dimension considerations.

Conjecture 1.7 ([AHHM22])**.** *The smallest possible VC dimension of any completion of $\mathbb{G}_\gamma^d$ into an $\mathbf{S}^{d-1} \times \mathbf{S}^{d-1}$* total *concept class tends to infinity as d grows.*

Remark 1.8. It follows from the analysis of the Perceptron algorithm that the VC dimension of every *total* submatrix of $\mathbb{G}_\gamma^d$ is at most $1/\gamma^2$, which does not depend on d.

Conjecture 1.7 is the VC dimension analogue of Theorem 1.5. In fact, the two statements are more concretely related since the smallest possible VC dimension of a completion of $\mathbb{G}_\gamma^d$ into a total concept class is a lower bound for the sign-rank of $\mathbb{G}_\gamma^d$.

In the following theorem, we show that, unlike sign-rank, the VC dimension of completions of $\mathbb{G}_\gamma^d$ can be much smaller than d.

Theorem 1.9. *There is a completion of $\mathbb{G}_\gamma^d$ to a total matrix G that has at most $O(d \log d \cdot \gamma^{-d})$ distinct rows, therefore,*

$$\mathrm{VC}(G) \leq O(d \cdot \log(\gamma^{-1}) + \log d).$$

Consequently, for $\gamma = 1 - \frac{1}{d}$,

$$\mathrm{VC}(G) \leq O(\log d).$$

Theorem 1.9 shows that Theorem 1.5 (i.e., $\mathbf{rk}_\pm(\mathbb{G}_\gamma^d) = d$) does not follow from a lower bound on the VC dimensions of *arbitrary* completions of $\mathbb{G}_\gamma^d$. However, it still could be the case that all "geometric" completions of $\mathbb{G}_\gamma^d$ have VC dimension d. That is, completions that are of the form $M(x, y) = \mathrm{sgn}\langle\phi(x), \psi(y)\rangle$ for some arbitrary choice of maps $\phi, \psi : \mathbf{S}^{d-1} \to \mathbb{R}^n$ and $n \geq d$. We prove this statement when ϕ and ψ are continuous.

Theorem 1.10. *Fix $d \in \mathbb{N}$ and $\gamma \in (0, 1)$, and suppose that the two maps $\phi, \psi : \mathbf{S}^{d-1} \to \mathbf{S}^{d-1}$ are continuous and satisfy*

$$\mathrm{sgn}\langle\phi(x), \psi(y)\rangle = \begin{cases} 1 & \langle x, y\rangle > \gamma \\ -1 & \langle x, y\rangle < -\gamma \end{cases}. \tag{4}$$

Then the matrix $M_{\mathbf{S}^{d-1} \times \mathbf{S}^{d-1}}$, with entries $M(x, y) = \mathrm{sgn}\langle\phi(x), \psi(y)\rangle$, has VC dimension d.

Note that the matrix M in Theorem 1.10 is a completion of $\mathbb{G}_\gamma^d$ into a total sign matrix. Also observe that if one proves a generalization of Theorem 1.10 for arbitrary maps $\phi, \psi : \mathbf{S}^{d-1} \to \mathbf{S}^{d-1}$, then it will immediately imply Theorem 1.5

1.5 Johnson–Lindenstrauss in bounded margin regime

The Johnson–Lindenstrauss lemma implies that (as it was shown in [LMSS07] and [AV06]) that every $N \times N$ sign matrix A with margin $\Omega(1)$ has a realization in dimension $O(\log(N))$. More precisely,

$$\mathbf{rk}_{\pm}(A) \leq O(\mathrm{m}(A)^{-2} \cdot \log(N)). \tag{5}$$

Linial, Mendelson, Schechtman, and Shraibman [LMSS07] asked whether this dependency on N is necessary. We prove the following theorem in Section 3, which answers this question in the case of partial matrices, and shows that Eq. (5) is tight.

Theorem 1.11 (Sign-rank of Gap Hamming Distance)**.** *For $\varepsilon \in (0, \frac{1}{2})$ and $n \in \mathbb{N}$,*

(i) The margin of $\mathrm{GHD}^n_{1-\varepsilon}$ is $1 - \varepsilon$.

(ii) For the sign-rank of $\mathrm{GHD}^n_{1-\varepsilon}$, we have

$$\Omega\left(\frac{\varepsilon n}{\log(\varepsilon^{-1})}\right) \leq \mathbf{rk}_{\pm}(\mathrm{GHD}^n_{1-\varepsilon}) \leq \lceil \varepsilon n \rceil.$$

Remark 1.12. The case of $\varepsilon = 0$ corresponds to the Equality function, which has sign-rank 3.

Remark 1.13. Theorem 1.11 provides an example of a partial $N \times N$ sign matrix with margin $\Theta(1)$ and sign-rank $\Theta(\log(N))$. In the converse direction, [HHL22] uses sum-product type ideas to construct a total sign matrix $A_{N\times N}$ with sign-rank 3 and margin $\Theta(N^{-\Omega(1)})$. In particular, this matrix corresponds to an n-bit communication problem with $\mathrm{U}(A) = O(1)$ and $\mathrm{R}(A) = \Theta(n)$.

2 Proof of Theorem 1.5

We recall Theorem 1.5.

Theorem 1.5 (Sign-rank of Gap Inner Product)**.** *For every $\gamma \in (0, 1)$ and every $d \in \mathbb{N}$, the sign-rank of the $\mathbf{S}^{d-1} \times \mathbf{S}^{d-1}$* gap inner product

$$\mathbb{G}^d_\gamma(x, y) = \begin{cases} 1 & \langle x, y\rangle > \gamma \\ -1 & \langle x, y\rangle < -\gamma \\ * & \textit{otherwise} \end{cases}.$$

is exactly d. For $\gamma = 1$, the sign-rank is 1.

Proof. First, we show that for $\gamma = 1$, the sign-rank of $\mathbb{G}^d_\gamma$ is 1. We need to find $\phi, \psi : \mathbf{S}^{d-1} \to \{-1, 1\}$ such that

$$\mathrm{sgn}\langle \phi(x), \psi(y)\rangle = \begin{cases} 1 & x = y \\ -1 & x = -y \end{cases}.$$

This is easily achieved by defining $\phi(x) = \psi(x) \in \{-1, 1\}$ to be the sign of the first non-zero coordinate of x.

Next, we turn to the case $\gamma \in (0,1)$. Suppose, contrary to the statement of the theorem, that there are $\phi, \psi : \mathbf{S}^{d-1} \to \mathbf{S}^{d-2}$ such that for all $x, y \in \mathbf{S}^{d-1}$,

$$\operatorname{sgn}\langle \phi(x), \psi(y) \rangle = \begin{cases} 1 & \langle x, y \rangle > \gamma \\ -1 & \langle x, y \rangle < -\gamma \end{cases}. \tag{6}$$

First, assume that at least one of the maps, say ϕ, is continuous. Later on, we will remove the continuity assumption.

Continuous setting. By the Borsuk-Ulam theorem, there exists $x \in \mathbf{S}^{d-1}$ such that $\phi(x) = \phi(-x)$. Pick any $y \in \mathbf{S}^{d-1}$ such that $\langle x, y \rangle > \gamma$. By Eq. (6), we have

$$0 < \langle \phi(x), \psi(y) \rangle.$$

Then $\langle \phi(-x), \psi(y) \rangle = \langle \phi(x), \psi(y) \rangle > 0$ while $\langle -x, y \rangle < -\gamma$, which shows that Eq. (6) is not satisfied for $-x$ and y.

General setting. Suppose ϕ is an arbitrary map and not necessarily continuous. We modify ϕ into a continuous map at the cost of slightly increasing γ. Let $\delta > 0$ be a small constant to be determined later. Let T be a finite δ-net for $\mathbf{S}^{d-1}$, which means that for every $x \in \mathbf{S}^{d-1}$, there exists $t \in T$ such that $\|x - t\| < \delta$. Let $\rho : \mathbb{R} \to [0,1]$ be the continuous function

$$\rho(z) = \begin{cases} 1 & z \leq \delta \\ 2 - \frac{z}{\delta} & z \in (\delta, 2\delta) \\ 0 & z \geq 2\delta \end{cases},$$

and define $\widetilde{\phi} : \mathbf{S}^{d-1} \to \mathbb{R}^{d-1}$ as

$$\widetilde{\phi} : x \mapsto \sum_{t \in T} \rho(\|x - t\|) \phi(t).$$

It follows from the continuity of ρ that $\widetilde{\phi}$ is continuous. Here, $\widetilde{\phi}(x)$ is a weighted sum of $\phi(t)$ over all $t \in T$ that assigns a zero weight to the points that are in distance at least 2δ from x. We will show that for all $x, y \in \mathbf{S}^{d-1}$,

$$\operatorname{sgn}\langle \widetilde{\phi}(x), \psi(y) \rangle = \begin{cases} 1 & \langle x, y \rangle > \gamma + 2\delta \\ -1 & \langle x, y \rangle < -(\gamma + 2\delta) \end{cases}.$$

Consider x, y with $\langle x, y \rangle \geq \gamma + 2\delta$. Every $t \in T$ with $\|x - t\| < 2\delta$ satisfies

$$\langle t, y \rangle \geq \langle x, y \rangle - \|x - t\| > \langle x, y \rangle - 2\delta \geq \gamma,$$

and thus by Eq. (6), it must satisfy $\langle \phi(t), \psi(y) \rangle > 0$. Consequently,

$$\langle \widetilde{\phi}(x), \psi(y) \rangle = \sum_{t \in T} \rho(\|x - t\|) \langle \phi(t), \psi(y) \rangle > 0.$$

Similarly, one can show that if $\langle x, y \rangle < -(\gamma + 2\delta)$, then $\langle \widetilde{\phi}(x), \psi(y) \rangle < 0$.

We can pick any $\delta > 0$ small enough so that $\gamma + 2\delta < 1$ and let $\gamma' := \gamma + 2\delta$. To get a contradiction, we apply the continuous case to the pair $\widetilde{\phi}, \psi$ with parameter γ' instead of γ.

□

3 Proof of Theorem 1.11

We first recall Theorem 1.11.

Theorem 1.11 (Sign-rank of Gap Hamming Distance)**.** *For $\varepsilon \in (0, \frac{1}{2})$ and $n \in \mathbb{N}$,*

(i) The margin of $\mathrm{GHD}^n_{1-\varepsilon}$ is $1 - \varepsilon$.

(ii) For the sign-rank of $\mathrm{GHD}^n_{1-\varepsilon}$, we have

$$\Omega\left(\frac{\varepsilon n}{\log(\varepsilon^{-1})}\right) \leq \mathbf{rk}_{\pm}(\mathrm{GHD}^n_{1-\varepsilon}) \leq \lceil \varepsilon n \rceil.$$

3.1 Upper bound

Claim 3.1. *Let $\varepsilon \in [0,1]$. Then $\mathbf{rk}_{\pm}(\mathrm{GHD}^n_{1-\varepsilon}) \leq \lceil \varepsilon n \rceil$.*

Proof. Let $k := \lceil \varepsilon n \rceil$ and let $\phi : \{-1,1\}^n \to \{-1,1\}^k$ be the projection to the first k coordinates. For every $x, y \in \{-1,1\}^n$,

$$\langle x, y\rangle > (1-\varepsilon)n \implies \langle \phi(x), \phi(y)\rangle \geq \langle x, y\rangle - (n-k) > 0$$

and

$$\langle x, y\rangle \leq -(1-\varepsilon)n \implies \langle \phi(x), \phi(y)\rangle < \langle x, y\rangle + (n-k) < 0,$$

which shows that $\mathbf{rk}_{\pm}(\mathrm{GHD}^n_{1-\varepsilon}) \leq k$ as desired. $\square$

3.2 Lower bound

First, we prove the following lemma that embeds the Gap Inner Product function of Theorem 1.5 in the Gap Hamming Distance function of Theorem 1.11.

Lemma 3.2. *Let $\delta \in (0, \frac{1}{2})$ be a parameter. There exist $n = O\left(\frac{d}{\sqrt{\delta}}\log(1/\delta)\right)$ and a map $\xi : \mathbf{S}^{d-1} \to \{-1,1\}^n$ such that for all $u, v \in \mathbf{S}^{d-1}$, we have*

$$\langle u, v\rangle > 1 - \delta \implies \langle \xi(u), \xi(v)\rangle > n(1 - 6\sqrt{\delta})$$

and

$$\langle u, v\rangle < -(1-\delta) \implies \langle \xi(u), \xi(v)\rangle < -n(1 - 6\sqrt{\delta}).$$

Proof. Let T be a $\frac{\delta}{2}$-net of size at most $(O(1/\delta))^d = 2^{O(d\log(1/\delta))}$ in $\mathbf{S}^{d-1}$. The upper bound of $(O(1/\delta))^d$ follows from a greedy procedure where one picks as many disjoint balls of radius $\frac{\delta}{4}$ as possible and takes the center of those balls as the $\frac{\delta}{2}$-net. We show the existence of the desired $\xi : \mathbf{S}^{d-1} \to \{-1,1\}^n$ by a probabilistic argument. Let $n \in \mathbb{N}$ to be determined later. Pick $\mathbf{w}_1, \ldots, \mathbf{w}_n \in \mathbf{S}^{d-1}$ independently and uniformly at random, and define a corresponding random function $\boldsymbol{\eta} : T \to \{-1,1\}^n$ as

$$\boldsymbol{\eta} : x \mapsto (\operatorname{sgn}\langle x, \mathbf{w}_i\rangle)_{i=1}^n . \tag{7}$$

Fix $i \in [n]$ and $u, v \in T$ such that $\langle u, v\rangle > 1 - 2\delta$. Since the angle between u and v is $\arccos(\langle u, v\rangle)$, we have

$$\Pr\left[\operatorname{sgn}\langle u, \mathbf{w}_i\rangle \neq \operatorname{sgn}\langle v, \mathbf{w}_i\rangle\right] = \frac{\arccos(\langle u, v\rangle)}{\pi} < \frac{\arccos(1-2\delta)}{\pi} \leq \sqrt{2\delta}.$$

To see the last inequality, note that for $\delta = 0$, we have

$$\frac{\arccos(1-2\delta)}{\pi} = \sqrt{2\delta} = 0,$$

and for every $\delta \in [0, 1/2]$, we have

$$\frac{d}{d\delta}\left(\sqrt{2\delta} - \frac{\arccos(1-2\delta)}{\pi}\right) = \frac{2}{\sqrt{2\delta}}\left(\frac{1}{2} - \frac{1}{\pi\sqrt{2-2\delta}}\right) \geq \frac{2}{\sqrt{2\delta}}\left(\frac{1}{2} - \frac{1}{\pi}\right) \geq 0.$$

For $i = 1, \ldots, n$, define $\mathbf{X}_i \coloneqq \mathbf{X}_i(u,v) \in \{0,1\}$ as $\mathbf{X}_i(u,v) \coloneqq 1$ iff $\operatorname{sgn}\langle u, \mathbf{w}_i\rangle \neq \operatorname{sgn}\langle v, \mathbf{w}_i\rangle$. The variables $\mathbf{X}_i$ are i.i.d. Bernoulli random variables with $\Pr[\mathbf{X}_i = 1] \leq \sqrt{2\delta}$. Hence, by Chernoff bound[2],

$$\Pr\left[\sum_{i=1} \mathbf{X}_i \geq 2n\sqrt{2\delta}\right] \leq e^{-n\sqrt{2\delta}/3}.$$

By Eq. (7), we have

$$\langle \boldsymbol{\eta}(u), \boldsymbol{\eta}(v)\rangle = n - 2\left(\sum_{i=1}^{n} \mathbf{X}_i\right),$$

and thus

$$\Pr\Big[\langle \boldsymbol{\eta}(u), \boldsymbol{\eta}(v)\rangle \leq n(1-6\sqrt{\delta})\Big] \leq \Pr\Big[\langle \boldsymbol{\eta}(u), \boldsymbol{\eta}(v)\rangle \leq n(1-4\sqrt{2\delta})\Big] \leq e^{-n\sqrt{2\delta}/3}.$$

Similarly, for $u, v \in T$ with $\langle u, v\rangle < -(1-2\delta)$, we have

$$\Pr\Big[\langle \boldsymbol{\eta}(u), \boldsymbol{\eta}(v)\rangle \geq -n(1-6\sqrt{\delta})\Big] \leq e^{-n\sqrt{2\delta}/3}.$$

Choose $n \coloneqq O\left(\frac{d}{\sqrt{\delta}}\log(1/\delta)\right)$ so that $|T|^2 e^{-n\sqrt{2\delta}/3} < 1$ and apply the union bound to the above probabilities over all pairs $u, v \in T$. We conclude that there exists $\eta : T \to \{-1,1\}^n$ such that for all $u, v \in T$, we have

$$\langle u, v\rangle > 1 - 2\delta \implies \langle \eta(u), \eta(v)\rangle > n(1-6\sqrt{\delta}),$$

and

$$\langle u, v\rangle < -(1-\delta) \implies \langle \eta(u), \eta(v)\rangle < -n(1-6\sqrt{\delta}).$$

Finally, let $f : \mathbf{S}^{d-1} \to T$ map each point in $\mathbf{S}^{d-1}$ to the closest point in T, breaking the ties arbitrarily, and define $\xi : \mathbf{S}^{d-1} \to \{-1,1\}^n$ as $\xi \coloneqq \eta \circ f$. For $u, v \in \mathbf{S}^{d-1}$ with $\langle u, v\rangle > 1 - \delta$, we have

$$\langle f(u), f(v)\rangle = \langle u, v\rangle + \langle f(u) - u, v\rangle + \langle f(u), f(v) - v\rangle > 1 - \delta - \frac{\delta}{2} - \frac{\delta}{2} \geq 1 - 2\delta.$$

Hence, as desired, we have

$$\langle \xi(u), \xi(v)\rangle = \langle \eta(f(u)), \eta(f(v))\rangle \geq n(1-6\sqrt{\delta}).$$

The case $\langle u, v\rangle < -(1-\delta)$ is similar. $\square$

[2] Chernoff bound implies that the sum of i.i.d. random variables $\mathbf{X}_i$ with expectation $\mu \coloneqq \mathbb{E}[\sum \mathbf{X}_i]$ satisfies $\Pr[\sum \mathbf{X}_i \geq 2\mu] \leq e^{-\mu/3}$.

Finally, we use Lemma 3.2 to complete the proof of Theorem 1.11.

Claim 3.3. *For $\varepsilon \in (0, \frac{1}{2})$,*

$$\mathbf{rk}_{\pm}(\mathrm{GHD}^n_{1-\varepsilon}) = \Omega\left(\frac{\varepsilon n}{\log(\varepsilon^{-1})}\right).$$

Proof. Denote $k \coloneqq \mathbf{rk}_{\pm}(\mathrm{GHD}^n_{1-\varepsilon})$. Then there are maps $\phi, \psi : \{-1,1\}^n \to \mathbb{R}^k$ such that

$$\langle x, y\rangle > (1-\varepsilon)n \implies \langle \phi(x), \psi(y)\rangle > 0$$

and

$$\langle x, y\rangle < -(1-\varepsilon)n \implies \langle \phi(x), \psi(y)\rangle < 0.$$

Let $\delta \coloneqq \left(\frac{\varepsilon}{6}\right)^2$. By applying Lemma 3.2 to parameter δ, we get $d = \Omega\left(\frac{\varepsilon n}{\log(\varepsilon^{-1})}\right)$ and $\xi : \mathbf{S}^{d-1} \to \{-1,1\}^n$ so that

$$\langle u, v\rangle > 1-\delta \implies \langle \xi(u), \xi(v)\rangle > 1-\varepsilon$$

and

$$\langle u, v\rangle < -(1-\delta) \implies \langle \xi(u), \xi(v)\rangle < -(1-\varepsilon),$$

for every $u, v \in \mathbf{S}^{d-1}$. By considering the composed functions $\phi \circ \xi$ and $\psi \circ \xi$, for every $u, v \in \mathbf{S}^{d-1}$, we have

$$\langle u, v\rangle > 1-\delta \implies \langle \xi(u), \xi(v)\rangle > 1-\varepsilon \implies \langle \phi\circ\xi(u), \psi\circ\xi(v)\rangle > 0$$

and

$$\langle u, v\rangle < -(1-\delta) \implies \langle \xi_1(u), \xi_2(v)\rangle < -(1-\varepsilon) \implies \langle \phi\circ\xi_1(u), \psi\circ\xi_2(v)\rangle < 0.$$

Therefore, we must have $\mathbf{rk}_{\pm}(\mathbb{G}^d_{1-\delta}) \leq k$. But by Theorem 1.5, we have $\mathbf{rk}_{\pm}(\mathbb{G}^d_{1-\delta}) \geq d = \Omega\left(\frac{\varepsilon n}{\log(\varepsilon^{-1})}\right)$. ☐

4 Proof of Theorem 1.4

In this section we present the proof of Theorem 1.4.

Theorem 1.4 (Communication complexity of Gap Hamming Distance)**.** *For $\varepsilon \in (0, \frac{1}{9})$, and $n \in \mathbb{N}$, we have*

(i) $\mathrm{R}(\mathrm{GHD}^n_{1-\varepsilon}) = 2$.

(ii) $\mathrm{U}(\mathrm{GHD}^n_{1-\varepsilon}) = \log(n) - O\left(\log\left(\varepsilon^{-1}\right)\right)$.

Proof. We first prove part (i), namely $\mathrm{R}(\mathrm{GHD}^n_{1-\varepsilon}) = 2$ when $\varepsilon \in (0, \frac{1}{9})$.

The two parties use their shared randomness to sample a uniform random $v \in \mathbf{S}^{n-1}$. Then Alice sends the bit $\mathrm{sgn}\langle v, x\rangle$ to Bob. Bob checks if $\mathrm{sgn}\langle v, x\rangle = \mathrm{sgn}\langle v, y\rangle$, in which case the output of the protocol is 1. Otherwise the output is -1.

We show that the error probability of the protocol is at most $\frac{1}{3}$. Suppose that $\mathrm{GHD}^n_{1-\varepsilon}(x, y) = 1$, namely, $\langle x, y\rangle > (1-\varepsilon)n$. The output of the protocol is incorrect if $\mathrm{sgn}\langle v, x\rangle \neq \mathrm{sgn}\langle v, y\rangle$. Similar to the proof of Lemma 3.2, one can see that

$$\Pr_{v\in\mathbf{S}^{n-1}}\left[\mathrm{sgn}\langle v, x\rangle \neq \mathrm{sgn}\langle v, y\rangle\right] = \frac{\arccos\left(\langle \frac{x}{\|x\|}, \frac{y}{\|y\|}\rangle\right)}{\pi} < \frac{\arccos(1-\varepsilon)}{\pi} \leq \sqrt{\varepsilon} \leq \sqrt{\frac{1}{9}} = \frac{1}{3}.$$

The case of $\mathrm{GHD}^n_{1-\varepsilon}(x,y) = -1$ is similar.

Part (ii) immediately follows from the sign-rank bound

$$\Omega\left(\frac{\varepsilon n}{\log(\varepsilon^{-1})}\right) \leq \mathbf{rk}_{\pm}(\mathrm{GHD}^n_{1-\varepsilon}) \leq \lceil \varepsilon n \rceil$$

of Theorem 1.11 and the fact that

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

$\square$

5 VC dimension of $\mathbb{G}^d_\gamma$

In this section, we present the proofs of Theorem 1.9 and Theorem 1.10. We will use the following classical result.

Theorem 5.1 (Rogers [Rog63])**.** *There is an absolute constant c, such that for $\gamma < 1$ and $d \geq 2$, the d-dimensional unit sphere can be covered by less than $cd(\log d)\gamma^{-d}$ balls of radius γ.*

Proof of Theorem 1.9. We define the matrix G based on a suitable partition $\mathcal{P}$ of $\mathbf{S}^{d-1}$ in such a way that $G(x,\cdot) = G(x',\cdot)$ if x and x' belong to the same part in $\mathcal{P}$, i.e.,

$$\forall P \in \mathcal{P}\ \forall x, x' \in P\ \forall y \in \mathbf{S}^{d-1} \qquad G(x,y) = G(x',y). \tag{8}$$

We show how to choose the partition $\mathcal{P}$ and define G. By Theorem 5.1, there exists a *partition* $\mathcal{P}$ of $\mathbf{S}^{d-1}$ into at most $O(d \log d \cdot \gamma^{-d})$ parts such that each part is contained in a ball of radius γ.

For $P \in \mathcal{P}$ and $y \in \mathbf{S}^{d-1}$, if there exists $x \in P$ with $\langle x, y\rangle > \gamma$, then let $G(x', y) \coloneqq 1$ for all $x' \in P$; otherwise let $G(x', y) \coloneqq -1$ for all $x' \in P$.

The matrix G clearly satisfies Eq. (8), and it remains to show that it is a completion of $\mathbb{G}^d_\gamma$. To this end, note that if there is $x \in P$ with $\langle x, y\rangle > \gamma$, then no $x' \in \mathcal{P}$ that can satisfy $\langle x', y\rangle \leq -\gamma$. This is because $\|x - x'\|_2 \leq 2\gamma$ and thus

$$\langle x', y\rangle \geq \langle x, y\rangle - \|x - x'\|_2 > \gamma - 2\gamma = -\gamma. \tag{9}$$

Similarly if $\langle x, y\rangle < -\gamma$, then no $x' \in \mathcal{P}$ satisfies $\langle x', y\rangle \geq \gamma$. Hence, indeed G is a completion of $\mathbb{G}^d_\gamma$ as desired.

Finally, note that G has at most $|\mathcal{P}|$ distinct rows and therefore,

$$\mathrm{VC}(G) \leq \log_2 |\mathcal{P}| \leq O(d \cdot \ln(\gamma^{-1}) + \log d).$$

By taking $\gamma = 1 - \frac{1}{d}$ and using the inequality $\ln x \leq x - 1$ for $x > 0$, we get

$$\mathrm{VC}(G) \leq O(\log d).$$

$\square$

Proof of Theorem 1.10. We first show that both maps ϕ and ψ have to be surjective. Suppose that, say, ϕ is not surjective and there is $a \in \mathbf{S}^{d-1}$ such that $a \notin \phi(\mathbf{S}^{d-1})$. Let P_a be the stereographic projection[3] $P_a : \mathbf{S}^{d-1} \backslash \{a\} \to \mathbb{R}^{d-1}$. Then the map $\phi' = P_a \circ \phi : \mathbf{S}^{d-1} \to \mathbb{R}^{d-1}$ is continuous. By Borsuk-Ulam, Theorem 1.6, there is $x' \in \mathbf{S}^{d-1}$ with $\phi'(x') = \phi'(-x')$. Since P_a is a one-to-one map, there must exist $x \in \mathbf{S}^{d-1}$ with $\phi(x) = \phi(-x)$. Pick an arbitrary $y \in \mathbf{S}^{d-1}$ with $\langle x, y \rangle > \gamma$. Then

$$0 < \langle \phi(x), \psi(y) \rangle = \langle \phi(-x), \psi(y) \rangle < 0,$$

which is a contradiction.

Now suppose both ϕ and ψ are surjective. Let $\{e_1, \cdots, e_d\}$ be the standard basis for $\mathbb{R}^d$. For each $S \subset [d]$, let $x'_S = \frac{1}{\sqrt{d}} \left(\sum_{i \in S} e_i - \sum_{j \notin S} e_j \right)$. By surjectivity of ϕ, there are $x_S \in \mathbf{S}^{d-1}$ with $\phi(x_S) = x'_S$. Similarly by surjectivity of ψ, there are $y_1, \ldots, y_d \in \mathbf{S}^{d-1}$ with $\psi(y_i) = e_i$ for all $i \in [d]$.

Note $\langle \phi(x_S), \psi(y_i) \rangle = \langle x'_S, e_i \rangle$, which is positive if $i \in S$ and negative otherwise. It follows that the matrix M restricted to rows x_S's and columns y_i's has VC dimension d. $\square$

6 Concluding remarks

The main question left open by this work is whether the separations of Theorem 1.4 and Theorem 1.11 are true for total functions.

Question 6.1. *Are there total sign matrices A with $\mathrm{m}(A) = \Omega(1)$ and $\mathbf{rk}_{\pm}(A) = \omega(1)$?*

Remark 6.2. Question 6.1 can be rephrased in terms of $\mathrm{disc}(A)$ or $\mathrm{R}(A)$ since (see [HHP$^+$])

$$\mathrm{m}(A) = \Omega(1) \Leftrightarrow \mathrm{disc}(A) = \Omega(1) \Leftrightarrow \mathrm{R}(A) = O(1).$$

One natural candidate for answering Question 6.1 is the "sign adjacency matrix" of the hypercube as it is known that $\mathrm{R}(Q_n) = O(1)$.

Conjecture 6.3 (Sign-rank of hypercube graphs [HHP$^+$])**.** *Let Q_n be the $\{0,1\}^n \times \{0,1\}^n$ sign matrix with $Q_n(x, y) = -1$ if and only if x and y differ in exactly one coordinate. Then*

$$\lim_{n \to \infty} \mathbf{rk}_{\pm}(Q_n) = \infty.$$

We believe Question 6.1 and Conjecture 6.3 are important questions because, in a sense, they capture an important limitation of *all* known techniques for proving lower bounds for sign-ranks of *explicit* matrices. Indeed, one can summarise the known methods for proving such lower bounds as the following three inequalities.

$$\mathrm{VC}(A) \le \mathbf{rk}_{\pm}(A), \qquad \mathrm{m}^{\mathrm{avg}}(A)^{-1} \le \mathbf{rk}_{\pm}(A), \qquad \frac{\log_2 \left(\mathrm{rect}(A)^{-1}\right)}{2} - 1 \le \mathbf{rk}_{\pm}(A). \tag{10}$$

The first inequality is immediate from the geometric definition of sign-ranks in Definition 1.1. In the second inequality, which captures Forster's method [For02], $\mathrm{m}^{\mathrm{avg}}(A)$ refers to a notion of "average" margin. In the third inequality,

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

[3] Stereographic projection from the north pole $p = (0, \cdots, 0, 1)$ is defined by $P_p : \mathbf{S}^{d-1} \backslash \{p\} \to \mathbb{R}^{d-1} : (x_1, \cdots, x_d) \mapsto \frac{1}{1 - x_d}(x_1, \cdots, x_{d-1})$.

where the infimum is over all *product* probability measures on the entries of A, and the maximum is over all monochromatic rectangles in A. We refer the reader to [HHP+] for more details, where it is also shown that

$$\sqrt{\mathrm{VC}(A)} \leq \mathrm{m}^{\mathrm{avg}}(A)^{-1} \leq \mathrm{rect}(A)^{-1}. \tag{11}$$

It is conjectured [CLV19] that every sign matrix A with $\mathrm{m}(A) = \Omega(1)$ satisfies $\mathrm{rect}(A)^{-1} = O(1)$. If true, then all the lower bounds in Eq. (10) are $O(1)$ for every A with $\mathrm{m}(A) = \Omega(1)$, and thus these methods cannot resolve Question 6.1 in the positive.

It was conjectured in [GKPW19] that there are matrices that satisfy $\mathrm{rect}(A)^{-1} = O(1)$ and $\mathbf{rk}_{\pm}(A) = \omega(1)$. This was resolved in [HHP+] using a counting argument, which showed that there are $N \times N$ sign matrices A with $\mathrm{rect}^{-1}(A) = O(1)$ and $\mathbf{rk}_{\pm}(A) \geq N^{\frac{1}{3}-o(1)}$. However, since for explicit matrices, there are no available tools beyond Eq. (10), no explicit example with $\mathrm{rect}(A)^{-1} = O(1)$ and $\mathbf{rk}_{\pm}(A) = \omega(1)$ is known.

Problem 6.4 ([HHP+])**.** *Construct an explicit sequence of matrices A_n such that* $\mathrm{rect}(A_n)^{-1} = O(1)$ *and*

$$\lim_{n\to\infty} \mathbf{rk}_{\pm}(A_n) = \infty.$$

The above-mentioned limitations of Eq. (10) signify the importance of discovering new methods for proving lower bounds on sign-rank. Since our results are about partial matrices, we cannot concretely compare the strength of our proof method to Eq. (10). Obviously, it will be very interesting to use the Borsuk-Ulam method, or any new approach for that matter, to resolve the above open problems.

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.

[AHHM22] Noga Alon, Steve Hanneke, Ron Holzman, and Shay Moran. A theory of PAC learnability of partial concept classes. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science—FOCS 2021*, pages 658–671. IEEE Computer Soc., Los Alamitos, CA, 2022.

[AV06] Rosa I. Arriaga and Santosh Vempala. An algorithmic theory of learning: Robust concepts and random projection. *Mach. Learn.*, 63(2):161–182, may 2006.

[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.

[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.

[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.

[CR11] Amit Chakrabarti and Oded Regev. An optimal lower bound on the communication complexity of gap-hamming-distance. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 51–60, 2011.

[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.

[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).

[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.

[GKPW19] Mika Göös, Pritish Kamath, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for pnp. *computational complexity*, 28(1):113–144, 2019.

[HHL22] Kaave Hosseini, Hamed Hatami, and Shachar Lovett. Sign-rank vs. discrepancy. *Theory of Computing*, 18(1):1–22, 2022.

[HHP+] Hamed Hatami, Pooya Hatami, William Pires, Ran Tao, and Rosie Zhao. Lower bound methods for sign-rank and their limitations.

[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.

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

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

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

[Mat03] Jiří Matoušek. *Using the Borsuk-Ulam theorem: lectures on topological methods in combinatorics and geometry*, volume 2003. Springer, 2003.

[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.

[New91] Ilan Newman. Private vs. common random bits in communication complexity. *Information processing letters*, 39(2):67–71, 1991.

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

[Rog63] CA Rogers. Covering a sphere with spheres. *Mathematika*, 10(2):157–164, 1963.

[Ros58] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

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

[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.

[She12] Alexander A Sherstov. The communication complexity of gap hamming distance. *Theory of Computing*, 8(1):197–208, 2012.

[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.

[SSBD14] Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, USA, 2014.

[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.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il