An approximation algorithm for approximation rank

Troy Lee
Department of Computer Science
Columbia University *

Adi Shraibman
Department of Mathematics
Weizmann Institute of Science †

Abstract

One of the strongest techniques available for showing lower bounds on bounded-error communication complexity is the logarithm of the approximation rank of the communication matrix—the minimum rank of a matrix which is close to the communication matrix in ℓ_∞ norm. Krause [Kra96] showed that the logarithm of approximation rank is a lower bound in the randomized case, and later Buhrman and de Wolf [BW01] showed it could also be used for quantum communication complexity. As a lower bound technique, approximation rank has two main drawbacks: it is difficult to compute, and it is not known to lower bound the model of quantum communication complexity with entanglement.

Linial and Shraibman [LS07] recently introduced a quantity, called γ_2^α, to quantum communication complexity, showing that it can be used to lower bound communication in the model with shared entanglement. Here α is a measure of approximation which is related to the allowable error probability of the protocol. This quantity can be written as a semidefinite program and gives bounds at least as large as many techniques in the literature, although it is smaller than the corresponding α-approximation rank, rk_α. We show that in fact $\log \gamma_2^\alpha(A)$ and $\log \mathrm{rk}_\alpha(A)$ agree up to small factors. As corollaries we obtain a constant factor polynomial time approximation algorithm to the logarithm of approximation rank, and that the logarithm of approximation rank is a lower bound for quantum communication complexity with entanglement.

1 Introduction

Often when trying to show that a problem is computationally hard we ourselves face a computationally hard problem. The minimum cost algorithm for a problem is naturally phrased as an optimization problem, and frequently techniques to lower bound this cost are also hard combinatorial optimization problems.

When taking such an computational view of lower bounds, it is natural to borrow ideas from approximation algorithms which have had a good deal of success in dealing with NP-hardness.

*Work conducted in part while at Rutgers University, supported by a NSF Mathematical Sciences Postdoctoral Fellowship. Email: troyjlee@gmail.com

†Email: adi.shraibman@weizmann.ac.il

Beginning with the seminal approximation algorithm for MAX CUT of Goemans and Williamson [GW95], a now common approach to hard combinatorial optimization problems is to look at a semidefinite relaxation of the problem with the hope of showing that such a relaxation provides a good approximation to the original problem.

We take this approach in dealing with approximation rank, an optimization problem that arises in communication complexity. In communication complexity, introduced by Yao [Yao79], two parties Alice and Bob wish to compute a function $f : X \times Y \to \{-1, +1\}$, where Alice receives $x \in X$ and Bob receives $y \in Y$. The question is how much they have to communicate to evaluate $f(x, y)$ for the most difficult pair (x, y). Associate to f a $|X|$-by-$|Y|$ *communication matrix* M_f where $M_f[x, y] = f(x, y)$. A well-known lower bound on the deterministic communication complexity of f due to Melhorn and Schmidt [MS82] is $\log \mathrm{rk}(M_f)$. This lower bound has many nice features—rank is easy to compute, at least from a theoretical perspective, and the famous log rank conjecture of Lovász and Saks [LS88] asserts that this bound is nearly tight in the sense that there is a universal constant c such that $(\log \mathrm{rk}(M_f))^c$ is an upper bound on the deterministic communication complexity of f, for every function f.

When we look at bounded-error randomized communication complexity, where Alice and Bob are allowed to flip coins and answer incorrectly with some small probability, the relevant quantity is no longer rank but *approximation rank*. For a sign matrix A, the α-approximation rank, denoted $\mathrm{rk}_\alpha(A)$, is the minimum rank of a matrix B which has the same sign pattern as A and whose entries have magnitude between 1 and α. When used to lower bound randomized communication complexity, the approximation factor α is related to the allowable error probability of the protocol. In the limit as $\alpha \to \infty$ we obtain the sign rank, denoted $\mathrm{rk}_\infty(A)$, the minimum rank of a matrix with the same sign pattern as A. Paturi and Simon [PS86] showed that the $\log \mathrm{rk}_\infty(M_f)$ exactly characterizes the unbounded error complexity of f, where Alice and Bob only have to get the correct answer on every input with probability strictly larger than $1/2$. Krause [Kra96] extended this to the bounded-error case by showing that $\log \mathrm{rk}_\alpha(M_f)$ is a lower bound on the $\frac{\alpha-1}{2\alpha}$-error randomized communication complexity of f. Later, Buhrman and de Wolf [BW01] showed that one-half this quantity is also a lower bound on the bounded-error *quantum* communication complexity of f, when the players do not share entanglement. Approximation rank is one of the strongest lower bound techniques available for either of these bounded-error models. In view of the log rank conjecture it is natural to conjecture as well that a polynomial in the logarithm of approximation rank is an upper bound on randomized communication complexity.

As a lower bound technique, however, approximation rank suffers from two deficiencies. The first is that it is quite difficult to compute in practice. Although we do not know if it is NP-hard to compute, the class of problems minimizing rank subject to linear constraints does contain NP-hard instances (see, for example, Section 7.3 in the survey of Vandenberghe and Boyd [VB96]). The second drawback is that it is not known to lower bound quantum communication complexity with entanglement.

We address both of these problems. We make use of a quantity γ_2^α which was introduced in the context of communication complexity by Linial et al. [LMSS07]. This quantity can naturally be viewed as a semidefinite relaxation of rank, and it is not hard to show that $(\frac{1}{\alpha}\gamma_2^\alpha(A))^2 \leq \mathrm{rk}_\alpha(A)$. We show that this lower bound is in fact reasonably tight.

Theorem 1 *Let $1 < \alpha < \infty$. Then for any m-by-n sign matrix A*

$$\frac{1}{\alpha^2}\,\gamma_2^\alpha(A)^2 \leq \mathrm{rk}_\alpha(A) = O_\alpha\left(\ln(mn)\gamma_2^\alpha(A)^2\right)^3$$

The quantity $\gamma_2^\alpha(A)$ can be written as a semidefinite program and so can be computed up to additive error ϵ in time polynomial in the size of A and $\log(1/\epsilon)$ by the ellipsoid method (see, for example, the textbook [GLS88]). Thus Theorem 1 gives a constant factor polynomial time approximation algorithm to compute $\log \mathrm{rk}_\alpha(A)$. Moreover, the proof of this theorem gives a method to find a near optimal low rank approximation to A in randomized polynomial time.

Linial and Shraibman [LS07] have shown that $\log \gamma_2^\alpha(A)$ is a lower bound on the $\frac{\alpha-1}{2\alpha}$-error quantum communication complexity of the sign matrix A with entanglement, thus we also obtain the following corollary.

Corollary 2 *Let $0 < \epsilon < 1/2$. Let $Q_\epsilon^*(A)$ be the quantum communication complexity of a m-by-n sign matrix A with entanglement. Then*

$$Q_\epsilon^*(A) \geq \frac{1}{6}\log \mathrm{rk}_{\alpha_\epsilon}(A) - \frac{1}{2}\log\log(mn) - \log\frac{\alpha_\epsilon^2}{\alpha_\epsilon - 1} - O(1),$$

where $\alpha_\epsilon = \frac{1}{1-2\epsilon}$.

The log log factor is necessary as the n-bit equality function with communication matrix of size 2^n-by-2^n has approximation rank $\Omega(\log n)$ [Alo08], but can be solved by a bounded-error quantum protocol with entanglement—or randomized protocol with public coins—with $O(1)$ bits of communication. This corollary means that approximation rank cannot be used to show a large gap between the models of quantum communication complexity with and without entanglement, if indeed such a gap exists.

Our proof works roughly as follows. Note that the rank of a m-by-n matrix A is the smallest k such that A can be factored as $A = XY^T$ where X is a m-by-k matrix and Y is a n-by-k matrix. The factorization norm $\gamma_2(A)$ can be defined as $\min_{X,Y:XY^T=A} r(X)r(Y)$ where $r(X)$ is the largest ℓ_2 norm of a row of X. Let X_0, Y_0 be an optimal solution to this program so that all rows of X_0, Y_0 have squared ℓ_2 norm at most $\gamma_2(A)$. The problem is that, although the rows of X_0, Y_0 have small ℓ_2 norm, they might still have large *dimension*. Intuitively, however, if the rows of X_0 have small ℓ_2 norm but X_0 has many columns, then one would think that many of the columns are rather sparse and one could somehow compress the matrix without causing too much damage. The Johnson-Lindenstrauss dimension reduction lemma [JL84] can be used to make this intuition precise. We randomly project X_0 and Y_0 to matrices X_1, Y_1 with column space of dimension roughly $\ln(mn)\gamma_2^\alpha(A)^2$. One can argue that with high probability after such a projection $X_1Y_1^T$ still provides a decent approximation to A. In the second step of the proof, we do an error reduction step to show that one can then improve this approximation without increasing the rank of $X_1Y_1^T$ by too much.

Ben-David, Eiron, and Simon [BES02] have previously used this dimension reduction technique to show that $\mathrm{rk}_\infty(A) = O(\ln(mn)\gamma_2^\infty(A)^2)$ for a sign matrix A. In this limiting case, however, $\gamma_2^\infty(A)$ fails to be a lower bound on $\mathrm{rk}_\infty(A)$. Buhrman, Vereshchagin, and de Wolf [BVW07], and independently Sherstov [She08], have given an example of a sign matrix A where $\gamma_2^\infty(A)$ is exponentially larger than $\mathrm{rk}_\infty(A)$.

2 Preliminaries

We will use the following form of Hoeffding's inequality [Hoe63].

Lemma 3 ([Hoe63]) *Let* $a_1, \ldots, a_n \in \mathbb{R}$*, and* $\delta_1, \ldots, \delta_n$ *be random variables with* $\Pr[\delta_i = 1] = \Pr[\delta_i = -1] = 1/2$*. Then for any* $t \geq 0$

$$\Pr\left[\left|\sum_{i=1}^{n} a_i \delta_i\right| > t\right] \leq 2\exp\left(\frac{-t^2}{2\sum_i a_i^2}\right)$$

2.1 Matrix notation

We will work with real matrices and vectors throughout this paper. For a vector u, we use $\|u\|$ for the ℓ_2 norm of u, and $\|u\|_\infty$ for the ℓ_∞ norm of u. For a matrix A let A^T denote the transpose of A. We let $A \circ B$ denote the entrywise product of A and B. For a positive semidefinite matrix M let $\lambda_1(M) \geq \cdots \geq \lambda_n(M) \geq 0$ be the eigenvalues of M. We define the i^{th} singular value of A, denoted $\sigma_i(A)$, as $\sigma_i(A) = \sqrt{\lambda_i(AA^T)}$. The rank of A, denoted $\mathrm{rk}(A)$ is the number of nonzero singular values of A. We will use several matrix norms.

- Spectral or operator norm: $\|A\| = \sigma_1(A)$.
- Trace norm: $\|A\|_{tr} = \sum_i \sigma_i(A)$.
- Frobenius norm: $\|A\|_F = \sqrt{\sum_i \sigma_i(A)^2}$.

One can alternatively see that $\|A\|_F^2 = \mathrm{Tr}(AA^T) = \sum_{i,j} A[i,j]^2$.

Our main tool will be the factorization norm γ_2 [TJ89], introduced in the context of complexity measures of matrices by Linial et al. [LMSS07]. This norm can naturally be viewed as a semidefinite programming relaxation of rank as we now explain. We take the following as our primary definition of γ_2:

Definition 4 ([TJ89],[LMSS07]) *Let* A *be a matrix. Then*

$$\gamma_2(A) = \min_{X,Y: XY^T = A} r(X)r(Y),$$

where $r(X)$ *is the largest* ℓ_2 *norm of a row of* X*.*

The quantity γ_2 can equivalently be written as the optimum of a maximization problem known as the Schur product operator norm: $\gamma_2(A) = \max_{X:\|X\|=1} \|A \circ X\|$. The book of Bhatia (Thm. 3.4.3 [Bha07]) contains a nice discussion of this equivalence and attributes it to an unpublished manuscript of Haagerup. An alternative proof can be obtained by writing the optimization problem defining γ_2 as a semidefinite programming problem and taking its dual [LSŠ08].

More convenient for our purposes will be a formulation of γ_2 in terms of the trace norm. One can see that this next formulation is equivalent to the Schur product operator norm formulation using the fact that $\|A\|_{tr} = \max_{B:\|B\|\leq 1} \mathrm{Tr}(AB^T)$.

Proposition 5 (cf. [LSŠ08]) *Let A be matrix. Then*

$$\gamma_2(A) = \max_{\substack{u,v \\ \|u\|=\|v\|=1}} \|A \circ vu^T\|_{tr}$$

From this formulation we can easily see the connection of γ_2 to matrix rank. This connection is well known in Banach spaces theory, where it is proved in a more general setting, but the following proof is more elementary.

Proposition 6 ([TJ89], [LSŠ08]) *Let A be a matrix. Then*

$$\mathrm{rk}(A) \geq \frac{\gamma_2(A)^2}{\|A\|_\infty^2}.$$

Proof: Let u, v be unit vectors such that $\gamma_2(A) = \|A \circ vu^T\|_{tr}$. As the rank of A is equal to the number of nonzero singular values of A, we see by the Cauchy-Schwarz inequality that

$$\mathrm{rk}(A) \geq \frac{\|A\|_{tr}^2}{\|A\|_F^2}.$$

As $\mathrm{rk}(A \circ vu^T) \leq \mathrm{rk}(A)$ we obtain

$$\begin{aligned} \mathrm{rk}(A) &\geq \frac{\|A \circ vu^T\|_{tr}^2}{\|A \circ vu^T\|_F^2} \\ &\geq \frac{\gamma_2(A)^2}{\|A\|_\infty^2} \end{aligned}$$

□

Finally, we define the approximate version of the γ_2 norm.

Definition 7 ([LS07]) *Let A be a sign matrix, and let $\alpha \geq 1$.*

$$\begin{aligned} \gamma_2^\alpha(A) &= \min_{B: 1 \leq A[i,j]B[i,j] \leq \alpha} \gamma_2(B) \\ \gamma_2^\infty(A) &= \min_{B: 1 \leq A[i,j]B[i,j]} \gamma_2(B) \end{aligned}$$

Similarly, we define can define approximation rank.

Definition 8 (approximation rank) *Let A be a sign matrix, and let $\alpha \geq 1$.*

$$\begin{aligned} \mathrm{rk}_\alpha(A) &= \min_{B: 1 \leq A[i,j]B[i,j] \leq \alpha} \mathrm{rk}(B) \\ \mathrm{rk}_\infty(A) &= \min_{B: 1 \leq A[i,j]B[i,j]} \mathrm{rk}(B) \end{aligned}$$

As corollary of Proposition 6 we get

Corollary 9 *Let A be a sign matrix and $\alpha \geq 1$.*

$$\mathrm{rk}_\alpha(A) \geq \frac{1}{\alpha^2}\,\gamma_2^\alpha(A)^2$$

We will also use a related norm ν known as the nuclear norm.

Definition 10 ([Jam87]) *Let A be a m-by$n$ matrix.*

$$\nu(A) = \min_{\alpha_i} \left\{ \sum |\alpha_i| : A = \sum_i \alpha_i x_i y_i^T \right\}$$

where $x_i \in \{-1,+1\}^m$ and $y_i \in \{-1,+1\}^n$.

It follows from Grothendieck's inequality that $\nu(A)$ and $\gamma_2(A)$ agree up to a constant multiplicative factor. For details see [LS07].

Proposition 11 (Grothendieck's inequality) *Let A be a matrix.*

$$\gamma_2(A) \leq \nu(A) \leq K_G\,\gamma_2(A),$$

where $1.67 \leq K_G \leq 1.78\ldots$ is Grothendieck's constant.

The best lower bound on Grothendieck's constant can be found in a paper of Reeds [Ree91], and the best upper bound is due to Krivine [Kri79].

3 Main Result

In this section we present our main result relating $\gamma_2^\alpha(A)$ and $\mathrm{rk}_\alpha(A)$. We show this in two steps: first we upper bound $\mathrm{rk}_{2\alpha-1}(A)$ in terms of $\gamma_2^\alpha(A)$ using dimension reduction. The second step of error reduction shows that $\mathrm{rk}_{2\alpha-1}(A)$ and $\mathrm{rk}_\alpha(A)$ are fairly closely related.

3.1 Dimension reduction

Theorem 12 *Let A be a m-by-n sign matrix and $\alpha > 1$.*

$$\mathrm{rk}_{2\alpha-1}(A) \leq \frac{16\alpha^2}{(\alpha-1)^2}\ln(4mn)\gamma_2^\alpha(A)^2$$

Proof: Instead of γ_2^α we will work with ν^α which makes the analysis easier and is only larger by at most a multiplicative factor of 2.

Let B be such that $\nu(B) = \nu^\alpha(A)$ and $J \leq A \circ B \leq \alpha J$, where J is the all ones matrix. By definition of ν, there is a decomposition

$$B = \sum_i \beta_i x_i y_i^T$$

where x_i, y_i are sign vectors and $\sum_i |\beta_i| = \nu(B)$. We can alternatively view this decomposition as a factorization of B. Let X be a matrix whose i^{th} column is $\sqrt{\beta_i} x_i$ and Y the matrix whose i^{th} column is $\sqrt{\beta_i} y_i$. The above equation gives that $XY^T = B$. The matrices X and Y also have the nice property that all entries in column i have the same magnitude $|\sqrt{\beta_i}|$.

Say that X is a m-by-k matrix and Y is a n-by-k matrix. Notice that while we know the ℓ_2 norm of the rows of X, Y is at most $\nu(B)$, the *number* of columns k could be much larger. The idea will be to randomly map X, Y down to matrices X_1, Y_1 with k' many columns for $k' \approx \nu(B)^2$ and show that the entries of $X_1 Y_1^T$ are close to those of XY^T with high probability.

To this end, let $\delta_{ij} \in \{-1, +1\}$ be independent identically distributed random variables taking on -1 and $+1$ with equal probability, and let R be a k-by-k' matrix where $R[i,j] = \frac{1}{\sqrt{k'}}\delta_{ij}$. Define $X_1 = XR, Y_1 = YR$. Then $\mathrm{rk}(X_1 Y_1^T) \leq k'$. We now bound how closely $X_1 Y_1^T$ approximates XY^T entrywise.

$$\begin{aligned}
X_1 Y_1^T[i,j] &= \sum_{\ell=1}^{k'} X_1[i,\ell] Y_1[j,\ell] \\
&= \sum_{\ell=1}^{k'} \left(\sum_{r=1}^{k} X[i,r] R[r,\ell] \right) \left(\sum_{r=1}^{k} Y[j,r] R[r,\ell] \right) \\
&= \sum_{r=1}^{k} X[i,r] Y[j,r] + \sum_{\ell=1}^{k'} \sum_{r \neq r'} X[i,r] R[r,\ell] Y[j,r'] R[r',\ell].
\end{aligned}$$

The first term is simply $B[i,j]$ thus it suffices to bound the magnitude of the second term.

$$\begin{aligned}
\Pr_R \left[\left| \sum_{\ell=1}^{k'} \sum_{r \neq r'} X[i,r] R[r,\ell] Y[j,r'] R[r',\ell] \right| > t \right] &= \Pr_{\{\delta_{r,\ell}\}} \left[\left| \sum_{\ell=1}^{k'} \sum_{r \neq r'} \frac{\delta_{r,\ell} \delta_{r',\ell} \sqrt{\beta_r \beta_{r'}}}{k} \right| > t \right] \\
&\leq 2 \exp \left(\frac{-t^2 k'}{2\nu(B)^2} \right)
\end{aligned}$$

by Hoeffding's inequality Lemma 3.

By taking $k' = 2\nu(B)^2 \ln(4mn)/t^2$ we can make this probability less than $1/2mn$. Then by a union bound there exists an R such that

$$\left| (XY^t)[i,j] - (X_1 Y_1^t)[i,j] \right| \leq t$$

for all i, j. Taking $t = \frac{\alpha - 1}{2\alpha}$ and rescaling $X_1 Y_1^T$ appropriately gives the theorem. □

3.2 Error-reduction

In this section, we will see that given a matrix A' which is an $\alpha' > 1$ approximation to the sign matrix A, we can obtain a matrix which is a better approximation to A and whose rank is not too much larger than that of A' by applying a low-degree polynomial approximation of the sign function to the entries of A'. This technique has been used several times before, for example [Alo08, KS07].

Let $p(x) = a_0 + a_1 x + \ldots + a_d x^d$ be a degree d polynomial. For a matrix A, we define $p(A)$ to be the matrix $a_0 J + a_1 A + \ldots + a_d A^{\circ d}$ where $A^{\circ s}$ is the matrix whose (i,j) entry is $A[i,j]^s$, and J is the all ones matrix.

Lemma 13 *Let A be a matrix and p be a degree d polynomial. Then* $\mathrm{rk}(p(A)) \leq (d+1)\mathrm{rk}(A)^d$

Proof: The result follows using subadditivity of rank and that $\mathrm{rk}(A^{\circ s}) \leq \mathrm{rk}(A^{\otimes s}) = \mathrm{rk}(A)^s$ since $A^{\circ s}$ is a submatrix of $A^{\otimes s}$. □

In general for any constants $1 < \beta \leq \alpha < \infty$ one can show that there is a constant c such that $\mathrm{rk}_\beta(A) \leq \mathrm{rk}_\alpha(A)^c$ by looking at low degree approximations of the sign function (see Corollary 1 of [KS07] for such a statement). As we are interested in the special case where α, β are quite close, we give an explicit construction in an attempt to keep the exponent as small as possible.

Proposition 14 *Fix $\epsilon > 0$. Let $a_3 = 1/(2 + 6\epsilon + 4\epsilon^2)$, and $a_1 = 1 + a_3$. Then the polynomial*

$$p(x) = a_1 x - a_3 x^3$$

maps $[1, 1+2\epsilon]$ into $[1, 1+\epsilon]$ and $[-1-2\epsilon, -1]$ into $[-1-\epsilon, -1]$.

Proof: As p is an odd polynomial, we only need to check that it maps $[1, 1+2\epsilon]$ into $[1, 1+\epsilon]$. With our choice of a_1, a_3, we see that $p(1) = p(1+2\epsilon) = 1$. Furthermore, $p(x) \geq 1$ for all $x \in [1, 1+2\epsilon]$, thus we just need to check that the maximum value of $p(x)$ in this interval does not exceed $1+\epsilon$.

Calculus shows that the maximum value of $p(x)$ is attained at $x = (\frac{1+a_3}{3a_3})^{1/2}$. Plugging this into the expression for $p(x)$, we see that the maximum value is

$$\max_{x \in [1, 1+2\epsilon]} p(x) = \frac{2}{3\sqrt{3}} \frac{(1+a_3)^{3/2}}{\sqrt{a_3}}.$$

We want to show that this is at most $1+\epsilon$, or equivalently that

$$\frac{2}{3\sqrt{3}} \frac{\sqrt{2+6\epsilon+4\epsilon^2}}{1+\epsilon} \left(\frac{3+6\epsilon+4\epsilon^2}{2+6\epsilon+4\epsilon^2} \right)^{3/2} \leq 1.$$

One can verify that this inequality is true for all $\epsilon \geq 0$. □

3.3 Putting everything together

Now we are ready to put everything together.

Theorem 15 *Fix $\alpha > 1$ and let A be a m-by-n sign matrix. Then*

$$\frac{1}{\alpha^2}\,\gamma_2^\alpha(A)^2 \leq \mathrm{rk}_\alpha(A) \leq \frac{2^{13}\alpha^6}{(\alpha-1)^6}\ln^3(4mn)\gamma_2^\alpha(A)^6.$$

Proof: By Theorem 12

$$\mathrm{rk}_{2\alpha-1}(A) \leq \frac{16\alpha^2}{(\alpha-1)^2}\ln(4mn)\gamma_2^\alpha(A)^2.$$

Now we can use the polynomial constructed in Proposition 14 and Lemma 13 to obtain

$$\mathrm{rk}_\alpha(A) \leq 2\mathrm{rk}_{2\alpha-1}(A)^3 \leq \frac{2^{13}\alpha^6}{(\alpha-1)^6}\ln^3(4mn)\gamma_2^\alpha(A)^6.$$

□

4 Discussion and open problems

One of the fundamental questions of quantum information is the power of entanglement. If we believe that there can be a large gap between the communication complexity of a function with and without entanglement then we must develop techniques to lower bound quantum communication complexity without entanglement that do not also work for communication complexity with entanglement. We have eliminated one of these possibilities in approximation rank.

As can be seen in Theorem 15, the relationship between $\gamma_2^\alpha(A)$ and $\mathrm{rk}_\alpha(A)$ weakens as $\alpha \to \infty$ because the *lower bound* becomes worse. Indeed, Buhrman, Vereshchagin, and de Wolf [BVW07], and independently Sherstov [She08], have given examples where $\gamma_2^\infty(A)$ is exponentially larger than $\mathrm{rk}_\infty(A)$. It is an interesting open problem to find a polynomial time approximation algorithm for sign rank $\mathrm{rk}_\infty(A)$. As far as we are aware, it is also an open question if sign rank is NP-hard to compute.

Acknowledgments

We would like to thank Ronald de Wolf for helpful comments on an earlier version of this manuscript and Gideon Schechtman for helpful conversations.

References

[Alo08] N. Alon. Perturbed identity matrices have high rank: proof and applications. *Combinatorics, Probability, and Computing*, 2008.

[BES02] S. Ben-David, N. Eiron, and H. Simon. Limitations of learning via embeddings in Euclidean half spaces. *Journal of Machine Learning Research*, 3:441–461, 2002.

[Bha07] R. Bhatia. *Positive definite matrices*. Princeton University Press, 2007.

[BVW07] H. Buhrman, N. Vereshchagin, and R. de Wolf. On computation and communication with small bias. In *Proceedings of the 22nd IEEE Conference on Computational Complexity*, pages 24–32. IEEE, 2007.

[BW01] H. Buhrman and R. de Wolf. Communication complexity lower bounds by polynomials. In *Proceedings of the 16th IEEE Conference on Computational Complexity*, pages 120–130, 2001.

[GLS88] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer-Verlag, 1988.

[GW95] M. Goemans and D. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42:1115–1145, 1995.

[Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[Jam87] G. J. O. Jameson. *Summing and nuclear norms in Banach space theory*. London mathematical society student texts. Cambridge university press, 1987.

[JL84] W. Johnson and J. Lindenstrauss. Extensions of Lipschitz mapping into Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

[Kra96] M. Krause. Geometric arguments yield better bounds for threshold circuits and distributed computing. *Theoretical Computer Science*, 156:99–117, 1996.

[Kri79] J. Krivine. Constantes de Grothendieck et fonctions de type positif sur les sphères. *Adv. Math.*, 31:16–30, 1979.

[KS07] A. Klivans and A. Sherstov. A lower bound for agnostically learning disjunctions. In *Proceedings of the 20th Conference on Learning Theory*, 2007.

[LMSS07] N. Linial, S. Mendelson, G. Schechtman, and A. Shraibman. Complexity measures of sign matrices. *Combinatorica*, 27(4):439–463, 2007.

[LS88] L. Lovász and M. Saks. Möbius functions and communication complexity. In *Proceedings of the 29th IEEE Symposium on Foundations of Computer Science*, pages 81–90. IEEE, 1988.

[LS07] N. Linial and A. Shraibman. Lower bounds in communication complexity based on factorization norms. In *Proceedings of the 39th ACM Symposium on the Theory of Computing*. ACM, 2007.

[LSŠ08] T. Lee, A. Shraibman, and R. Špalek. A direct product theorem for discrepancy. In *Proceedings of the 23rd IEEE Conference on Computational Complexity*, pages 71–80. IEEE, 2008.

[MS82] K. Melhorn and E. Schmidt. Las Vegas is better than determinism in VLSI and distributed computing. In *Proceedings of the 14th ACM Symposium on the Theory of Computing*, pages 330–337. ACM, 1982.

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

[Ree91] J. Reeds. A new lower bound on the real Grothendieck constant. Available at `http://www.dtc.umn.edu/~reedsj`, 1991.

[She08] A. Sherstov. Halfspace matrices. *Computational Complexity*, 17(2):149–178, 2008.

[TJ89] N. Tomczak-Jaegermann. *Banach-Mazur distances and finite-dimensional operator ideals*, volume 38 of *Pitman Monographs and Surveys in Pure and Applied Mathematics*. Longman Scientific & Technical, Harlow, 1989.

[VB96] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38:49–95, 1996.

[Yao79] A. Yao. Some complexity questions related to distributive computing. In *Proceedings of the 11th ACM Symposium on the Theory of Computing*, pages 209–213. ACM, 1979.