

An approximation algorithm for approximation rank

Troy Lee

Department of Computer Science
Rutgers University *

Adi Shraibman

Department of Mathematics
Weizmann Institute of Science †

Abstract

One of the strongest techniques available for showing lower bounds on quantum 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. This technique has two main drawbacks: it is difficult to compute, and it is not known to lower bound quantum communication complexity with entanglement.

Linial and Shraibman recently introduced a norm, called γ_2^α , to quantum communication complexity, showing that it can be used to lower bound communication with entanglement. Here α is a measure of approximation which is related to the allowable error probability of the protocol. This bound 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 $\log \text{rk}_\alpha(A)$ agree up to small factors. As corollaries we obtain a constant factor polynomial time approximation algorithm to the logarithm of approximate 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 a computational view of lower bounds, a natural idea is to borrow ideas from approximation algorithms which have had a good deal of success in dealing with NP-hardness. 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.

*Supported by a NSF Mathematical Sciences Postdoctoral Fellowship. Email: troylee@gmail.com

†Email: adi.shraibman@weizmann.ac.il

We take this approach in dealing with approximation rank, an optimization problem that arises in communication complexity. In communication complexity, two parties Alice and Bob wish to compute a function $f : X \times Y \rightarrow \{-1, +1\}$, where Alice receives $x \in X$ and Bob receives $y \in Y$. The question is how much they have to communicate to evaluate $f(x, y)$. We can 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 is $\log \text{rk}(M_f)$; the long standing log rank conjecture asserts that this bound is nearly tight in the sense that there is a universal constant c such that $(\log \text{rk}(M_f))^c$ is an upper bound on the communication complexity of f .

When we turn our attention to bounded-error communication complexity, where Alice and Bob are allowed to get the wrong answer with some small probability, the relevant quantity is no longer rank but *approximation rank*. For a sign matrix A , the approximation rank, denoted $\text{rk}_\alpha(A)$, is the minimum rank of a matrix A' such that $J \leq A \circ A' \leq \alpha J$, where J denotes the all ones matrix and \circ is entrywise product. Here α is a measure of approximation and is related to the allowable error probability of the protocol. It is known that $\Omega(\log \text{rk}_2(M_f))$ is a lower bound on the bounded-error randomized communication complexity of f and the quantum communication complexity of f when the players do not share entanglement. In addition, approximation rank is one of the strongest lower bound techniques available for these models. In view of the log rank conjecture it is also natural to conjecture that a polynomial in the logarithm of approximation rank is an upper bound on the communication complexity of these models.

As a lower bound technique, however, approximation rank suffers from two drawbacks. The first is that it is quite difficult to compute in practice. Although we do not know if it is NP-hard, many similar rank minimization problems are in fact NP-hard [VB96]. The second drawback is that it is not known to lower bound quantum communication complexity with entanglement.

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

$$\frac{1}{\alpha^2} \gamma_2^\alpha(A)^2 \leq \text{rk}_\alpha(A) \leq \frac{8192\alpha^6}{(\alpha - 1)^6} \ln^3(4mn) \gamma_2^\alpha(A)^6$$

$\gamma_2^\alpha(A)$ 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 [GLS88]). Thus Theorem 1 gives a constant factor polynomial time approximation algorithm to compute $\log \text{rk}_\alpha(A)$. Moreover, the proof of this theorem gives a method to find a near optimal low rank approximation in randomized polynomial time.

Linial and Shraibman [LS07] have shown that $\log \gamma_2^\alpha(M_f)$ is a lower bound on quantum communication complexity 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*

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

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

The log log factor is necessary as the communication matrix of 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.

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 $A = X^T Y$ where X is a k -by- m matrix and Y is a k -by- n matrix. The factorization norm $\gamma_2(A)$ can be defined as $\min_{X,Y:X^T Y=A} c(X)c(Y)$ where $c(X)$ is the smallest ℓ_2 norm of a column of X . Let X_0, Y_0 be an optimal solution to this program so that all columns of X_0, Y_0 have squared ℓ_2 norm at most $\gamma_2(A)$. The problem is that, although the columns of X_0, Y_0 have small ℓ_2 norm, they might still have large *dimension*. Intuitively, however, if the columns of X_0 have small ℓ_2 norm but X_0 has many rows, then one would think that many of the rows 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 row space of dimension roughly $\ln(mn)\gamma_2^\alpha(A)^2$. One can argue that with high probability after such a projection $X_1^T Y_1$ 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 $X_1^T Y_1$ by too much.

2 Preliminaries

We will need some basic notation. For a vector u , we use $\|u\|$ for the ℓ_2 norm of u . For a matrix A , we use $\|A\|$ for the spectral or operator norm and $\|A\|_{tr}$ for the trace norm. The notation $A \otimes B$ denotes tensor product, and $A \circ B$ denotes the entrywise product of A and B .

We will use the following form of Hoeffding’s inequality [Hoe63]:

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

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

Our main tool will be the factorization norm γ_2 [TJ89], introduced in the context of complexity measures of matrices by [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*

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

where $c(X)$ is the largest ℓ_2 norm of a column 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]. Using duality of the operator norm and trace norm it is then straightforward to see that the Schur product operator norm is equivalent to the expression given in the next proposition, which is most convenient for our purposes. A proof of this last equivalence can be found in [Mat93].

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*

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

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

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

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

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

□

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

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

$$\gamma_2^\alpha(A) = \min_{A': J \leq A \circ A' \leq \alpha J} \gamma_2(A'),$$

where J is the all ones matrix.

As corollary of Proposition 6 we get

Corollary 8 *Let A be a sign matrix and $\alpha \geq 1$.*

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

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

Definition 9 ([Jam87]) *Let A be a matrix.*

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

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

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

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

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

3 Main Result

In this section we present our main result relating $\gamma_2^\alpha(A)$ and $\text{rk}_\alpha(A)$. We show this in two steps: first we upper bound $\text{rk}_{\alpha'}(A)$ in terms of $\gamma_2^\alpha(A)$ where α' is slightly larger than α . In the second step, we show that this approximation can be tightened to α without increasing the rank too much.

3.1 Dimension reduction

Theorem 11 *Let A be a m -by- n sign matrix and $\alpha \geq 1$. Then for any $0 < t < 1$*

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

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

Thus let $X^T Y$ be a factorization realizing $\nu^\alpha(A)$ —in other words, $J \leq X^T Y \circ A \leq \alpha J$, and all the columns of X, Y have squared ℓ_2 norm at most $\nu^\alpha(A)$. Moreover, X, Y are layered in the sense that for every row, all elements in that row have the same magnitude. Indeed, balancing as necessary, we may assume that $|X[i, j]| = |Y[i, j]| = \sqrt{\beta_i}$ for all j , and that $\nu^\alpha(A) = \sum_i \beta_i$.

Say that X is a k -by- m matrix and Y is a k -by- n matrix. Let $\delta_{i,j}$ be independent identically distributed variables chosen to be $+1$ with probability $1/2$ and -1 with probability $1/2$. For an integer k' to be fixed later, we let R be a k -by- k' random matrix where $R[i, j] = \frac{1}{\sqrt{k'}}\delta_{i,j}$.

Let $X_1 = R^T X$ and $Y_1 = R^T Y$. Then clearly $\text{rk}(X_1^T Y_1) \leq k'$. We now estimate the probability over choice of R that $X_1^T Y_1$ gives a good approximation to A .

Consider

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

Thus to bound the difference between $X^T Y[i, j]$ and $X_1^T Y_1[i, j]$, it suffices to bound the magnitude of the second term.

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

by Hoeffding's inequality (Lemma 3).

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

$$|X^T Y[i, j] - X_1^T Y_1[i, j]| \leq t$$

for all i, j . Rescaling the matrix $X_1^T Y_1$ gives the theorem. \square

3.2 Error-reduction

In this section, we will see that given a matrix A' which is an $\alpha' > 1$ approximation to 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 [Al08, KS07].

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

Lemma 12 *Let A be a matrix and p be a degree d polynomial. Then $\text{rk}(p(A)) \leq (d+1)\text{rk}(A)^d$*

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

In general for any constants $1 < \beta \leq \alpha < \infty$ one can show that there is a constant c such that $\text{rk}_\beta(A) \leq \text{rk}_\alpha(A)^c$ 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 13 *Fix $\epsilon > 0$. Let $a_3 = 1/(2 + 6\epsilon + 4\epsilon^2)$, and $a_1 = 1 + a_3$. Then the polynomial*

$$p(x) = a_1x - a_3x^3$$

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 $\epsilon \geq 0$. \square

3.3 Putting everything together

Now we are ready to put everything together.

Theorem 14 *Fix $\alpha > 1$ and let A be a m -by- n sign matrix. Then*

$$\frac{1}{\alpha^2} \gamma_2^\alpha(A)^2 \leq \text{rk}_\alpha(A) \leq \frac{8192\alpha^6}{(\alpha - 1)^6} \ln^3(4mn) \gamma_2^\alpha(A)^6.$$

Proof: We first apply Theorem 11 with $t = \frac{\alpha-1}{2\alpha}$. In this way, we have $\frac{\alpha+t}{1-t} \leq 2\alpha - 1$. Thus we obtain

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

Now we can use the polynomial constructed in Proposition 13 and Lemma 12 to obtain

$$\text{rk}_\alpha(A) \leq 2\text{rk}_{2\alpha-1}(A)^3 \leq \frac{8192\alpha^6}{(\alpha-1)^6} \ln^3(4mn)\gamma_2^\alpha(A)^6.$$

□

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.

The strong relation between $\gamma_2^\alpha(A)$ and $\text{rk}_\alpha(A)$ for constant α might be surprising in light of the behavior we encounter when $\alpha = 1$ or when α goes to infinity. When $\alpha = 1$ the n dimensional identity matrix provides an example where γ_2 is constant while the rank is equal to n . When $\alpha \rightarrow \infty$ the quantity $\text{rk}_\infty(A)$ is the well-studied *sign rank*. Using a similar random projection technique, Ben-David et al. [BES02] show that $\text{rk}_\infty(A) \leq O(\log(mn)\gamma_2^\infty(A)^2)$ for a m -by- n matrix A . In this case, however, the lower bound fails as there are examples of an exponential gap between sign rank and γ_2^∞ [BVW07, She08].

Back to the computational point of view, it remains an interesting open problem to find an efficient approximation algorithm for sign rank. More generally, one can ask about other kinds of rank minimization problems. We have just dealt with a very specific case where one has a constraint of the form $A \circ X \geq 0$. One could look more generally at linear equality constraints of the form $A_1 \bullet X = b_1, \dots, A_k \bullet X = b_k$. Such problems arise in many areas and are known to be NP-hard.

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.
- [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.
- [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.
- [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*. IEEE, 2008.
- [Mat93] R. Mathias. The Hadamard operator norm of a circulant and applications. *SIAM journal on matrix analysis and applications*, 14(4):1152–1167, 1993.
- [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.