Noname manuscript No.
(will be inserted by the editor)

Some upper and lower bounds on PSD-rank

Troy Lee · Zhaohui Wei · Ronald de Wolf

Received: date / Accepted: date

Abstract Positive semidefinite rank (PSD-rank) is a relatively new complexity measure on matrices, with applications to combinatorial optimization and communication complexity. We first study several basic properties of PSD-rank, and then develop new techniques for showing lower bounds on the PSD-rank. All of these bounds are based on viewing a positive semidefinite factorization of a matrix M as a quantum communication protocol. These lower bounds depend on the entries of the matrix and not only on its support (the zero/nonzero pattern), overcoming a limitation of some previous techniques. We compare these new lower bounds with known bounds, and give examples where the new ones are better. As an application we determine the PSD-rank of (approximations of) some common matrices.

Keywords semidefinite programming · extended formulation · PSD-rank · slack matrix

1 Introduction

1.1 Background

We study the properties of *positive semidefinite factorizations*. Such a factorization (of size r) of a nonnegative m-by-n matrix A is given by r-by-r positive semidefinite matrices $E_1, \ldots, E_m$ and $F_1, \ldots, F_n$ satisfying $A(i,j) = \mathrm{Tr}(E_i F_j)$ for all i, j. The *positive semidefinite rank* (PSD-rank) of A is the smallest r such that A has a positive semidefinite factorization of size r. We denote it by $\mathtt{rank}_{\mathtt{psd}}(A)$. The notion of PSD-rank has been introduced relatively recently because of applications to combinatorial optimization and communication complexity [1,2]. These applications closely parallel those of the *nonnegative rank* of A, which is the minimum number r such that there exists an m-by-r nonnegative matrix B and an r-by-n nonnegative matrix C satisfying $A = BC$.

In the context of combinatorial optimization, a polytope P is associated with a nonnegative matrix known as the *slack matrix* of P. A classic result by Yannakakis shows that the nonnegative rank of the slack matrix of P characterizes the size of a natural way of formulating the optimization of a linear function over P as a linear program [3]. More precisely, the nonnegative rank of the slack matrix of P equals the *linear extended formulation* size of P, which is the minimum number of facets of a (higher-dimensional) polytope Q that projects to P. Analogously, the PSD-rank of the slack matrix of P captures the size of a natural way of optimizing a linear function over P as a *semidefinite* program [1,2]. More precisely, the PSD-rank of the slack matrix of P is equal to the *positive semidefinite extension* size of P, which is the smallest r for which P can be expressed as the projection of an affine slice of the cone of r-dimensional positive semidefinite matrices.

Troy Lee
School of Physics and Mathematical Sciences, Nanyang Technological University and Centre for Quantum Technologies, Singapore
E-mail: troyjlee@gmail.com

Zhaohui Wei
School of Physics and Mathematical Sciences, Nanyang Technological University and Centre for Quantum Technologies, Singapore
E-mail: weizhaohui@gmail.com

Ronald de Wolf
QuSoft, CWI and University of Amsterdam, Amsterdam, The Netherlands
E-mail: rdewolf@cwi.nl

There have recently been great strides in understanding linear extended formulations, showing that the linear extended formulation size for the traveling salesman and matching polytopes is exponentially large in the number of vertices of the underlying graph [2,4]. In a more recent breakthrough, it was similarly proved that the traveling salesman polytope requires superpolynomial *positive semidefinite extension complexity*, and showing this required showing strong lower bounds on the PSD-rank of the corresponding slack matrix [5] (See also [6] for a simple proof for the special case of rank-one positive semidefinite factorizations.)

In communication complexity, nonnegative and PSD-rank arise in the model of computing a function $f : \{0,1\}^m \times \{0,1\}^n \to \mathbb{R}_+$ in expectation. In this model, Alice has an input $x \in \{0,1\}^m$, Bob has an input $y \in \{0,1\}^n$ and their goal is to communicate in order for Bob to output a nonnegative random variable whose expectation is $f(x,y)$. The associated communication matrix for this problem is a 2^m-by-2^n matrix whose (x,y) entry is $f(x,y)$. The nonnegative rank of the communication matrix of f characterizes the amount of classical communication needed to compute f in expectation [7]. Analogously, the PSD-rank of the communication matrix of f characterizes the amount of *quantum* communication needed to compute f in expectation [2]. Alternatively, one can consider the problem where Alice and Bob wish to generate a probability distribution $P(x,y)$ using shared randomness or shared entanglement, but without communication. The number of bits of shared randomness or qubits of shared entanglement are again characterized by the nonnegative rank and PSD-rank, respectively [8,9]. Accordingly, providing lower and upper bounds on the PSD-rank is interesting in the context of communication complexity as well. Among other things, here we will pin down (up to constant factors) the PSD-rank of some common matrices studied in communication complexity like the inner product and non-equality matrices [10].

1.2 Our results

As PSD-rank is a relatively new quantity, even some basic questions about its behavior remain unanswered. We address several properties here. First we show that, unlike the usual rank, PSD-rank is not strictly multiplicative under tensor product: we give an example of a matrix P where $\texttt{rank}_{\texttt{psd}}(P \otimes P) < \texttt{rank}_{\texttt{psd}}(P)^2$. We do this by making a connection between PSD-rank and planar geometry to give a simple sufficient condition for when the PSD-rank is not full.

The second question we address is the dependence of PSD-rank on the underlying field. At the Dagstuhl Seminar 13082 (February 2013), Dirk Oliver Theis raised the question if the PSD-rank where the factorization is by *real* symmetric PSD-matrices is the same as that by *complex* Hermitian PSD-matrices. It is easy to see that the real PSD-rank can be at most a factor of 2 larger than the complex PSD-rank; we give an infinite family of matrices where the real PSD-rank is asymptotically a factor of $\sqrt{2}$ larger than the complex PSD-rank.

Our main goal in this paper is showing lower bounds on the PSD-rank, a task of great importance to both the applications to combinatorial optimization and communication complexity mentioned above. Unfortunately, at this point very few techniques exist to lower bound the PSD-rank. For example, though the technique developed in [5] is very powerful, it is very complicated and not easy to utilize generally.

One lower bound direction is to consider only the *support* of the matrix, that is the pattern of zero/nonzero entries. For the nonnegative rank, this method can show good lower bounds—in particular, support-based arguments sufficed to show exponential lower bounds on the linear extension complexity of the traveling salesman polytope [2]. For the PSD-rank, however, support-based arguments cannot show lower bounds larger than the rank of the matrix [11]. This means that for cases like the traveling salesman polytope, where the positive semidefinite extension complexity is superpolynomial in the rank of the slack matrix, other techniques need to be developed.

We develop three easy-to-compute lower bounds on PSD-rank. All three depend on the values of the matrix and not only on its support structure—in particular, they can show nontrivial lower bounds for matrices with full support, i.e., without zero entries. All three are derived from the viewpoint of PSD-rank of a nonnegative matrix as a quantum communication protocol. We compare these lower bounds with previous techniques and show examples where they are better.

We also give nearly tight bounds on the PSD-rank of (approximations of) the identity matrix and on the PSD-rank of the matrix corresponding to the inner product and nonequality functions.

It should be noted, however, that our new bounds do not take advantage of structural aspects of matrices like their sparsity patterns, and hence will not give tight bounds in many cases. For an example where the technique we develop here can be improved using extra structural information of the problem, see [12].

2 Preliminaries

Let $[n] = \{1, 2, \ldots, n\}$. Let $M = [M(i,j)]$ be an arbitrary m-by-n matrix of rank r with the (i,j)-th entry being $M(i,j)$. The *conjugate transpose* of M is defined as an n-by-m matrix $M^\dagger$ with $M^\dagger(i,j) = \overline{M(j,i)}$, where $\overline{M(j,i)}$ is the complex conjugate of $M(j,i)$.

Let $\sigma_1, \sigma_2, \ldots, \sigma_r$ be the nonzero singular values of M. The *trace norm* of M is defined as $\|M\|_{tr} = \sum_i \sigma_i$, and the *Frobenius norm* of M is defined as $\|M\|_F = (\sum_i \sigma_i^2)^{1/2}$; this equals $(\mathrm{Tr}(M^\dagger M))^{1/2} = (\sum_{i,j} M(i,j)^2)^{1/2}$. Note that $\|M\|_F \leq \|M\|_{tr}$. By the Cauchy-Schwarz inequality we have

$$\mathtt{rank}(M) \geq \left(\frac{\|M\|_{tr}}{\|M\|_F}\right)^2 \tag{1}$$

2.1 PSD-rank

Since it is the central topic of this paper, we repeat the definition of PSD-rank from the introduction:

Definition 1 Let A be a nonnegative m-by-n matrix. A *positive semidefinite factorization* of size r of A is given by r-by-r positive semidefinite matrices $E_1, \ldots, E_m$ and $F_1, \ldots, F_n$ satisfying $A(i,j) = \mathrm{Tr}(E_i F_j)$. The *positive semidefinite rank* (PSD-rank, $\mathtt{rank}_{\mathtt{psd}}(A)$) of A is the smallest integer r such that A has a positive semidefinite factorization of size r.

In the definition of PSD-rank, we allow the matrices of the PSD-factorization to be arbitrary Hermitian PSD matrices, with complex-valued entries. One can also consider the *real* PSD-rank, where the matrices of the factorization are restricted to be real symmetric PSD matrices. For a nonnegative matrix A, we denote its real PSD-rank by $\mathtt{rank}^{\mathbb{R}}_{\mathtt{psd}}(A)$.

Note that for a nonnegative matrix A, the PSD-rank is unchanged when we remove all-zero rows and columns. Also, for nonnegative diagonal matrices D_1, D_2, the PSD-rank of $D_1 A D_2$ is at most that of A. Throughout this paper we will use these facts to achieve a particular normalization for A. In particular, we will frequently assume without loss of generality that each column of A sums to one, i.e., that A is a stochastic matrix.

The following lemma is very useful for giving upper bounds on the PSD-rank.

Lemma 1 *([1, 8]) If A is a nonnegative matrix, then*

$$\mathtt{rank}_{\mathtt{psd}}(A) \leq \min_{M:\ M \circ \overline{M} = A} \mathtt{rank}(M),$$

where $\circ$ is the Hadamard product (entry-wise product) and $\overline{M}$ is the entry-wise complex conjugate of M.

2.2 Quantum background

A *quantum state* ρ is a positive semidefinite matrix with trace $\mathrm{Tr}(\rho) = 1$. If the rank of ρ is 1, it can be written as $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a complex column vector, and $\langle\psi|$ is its conjugate transpose. In this case, we call this a *pure state*, and denote it by $|\psi\rangle$ directly. In order to express an arbitrary r-dimensional pure state, one can choose an orthonormal basis of r unit vectors. A typical choice is the so-called computational basis, $\{|0\rangle, |1\rangle, \ldots, |r-1\rangle\}$, where $|i\rangle$ is the vector that has only one nonzero entry 1, at position $i+1$. If one concatenates two pure states $|x\rangle$ and $|y\rangle$, the state of the joint system is expressed as their tensor product, i.e., $|x\rangle \otimes |y\rangle$, which is sometimes abbreviated to $|x\rangle|y\rangle$ or $|xy\rangle$.

For an r-dimensional quantum system, one can use unitary operations to change its quantum state. A unitary operation can be expressed as an r-by-r matrix U with $UU^\dagger = I$, where I is the identity. As an example, in this paper we will use the Hadamard gate for 2-dimensional quantum states, which can be written as $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

A *POVM* ("Positive Operator Valued Measure") $\mathcal{E} = \{E_i\}$ consists of positive semidefinite matrices E_i that sum to the identity. When measuring a quantum state ρ with this POVM, the outcome is i with probability $p_i = \mathrm{Tr}(\rho E_i)$.

For our purposes, a *(one-way) quantum protocol* between two players Alice (with input x) and Bob (with input y) is the following: Alice sends a quantum state ρ_x to Bob, who measures it with a POVM $\mathcal{E}_y = \{E_i\}$.

Each outcome i of this POVM is associated with a nonnegative value, which is Bob's output. We say the protocol *computes an m-by-n matrix M in expectation* if, for every $x \in [m]$ and $y \in [n]$, the expected value of Bob's output equals $M(x,y)$. Fiorini et al. [2] showed that the minimal dimension of the states ρ_x in such a protocol is either $\mathtt{rank}_{\mathtt{psd}}(M)$ or $\mathtt{rank}_{\mathtt{psd}}(M)+1$, so the minimal number of qubits of communication equals $\lceil \log_2 \mathtt{rank}_{\mathtt{psd}}(M) \rceil$ up to one qubit.

For two quantum states ρ and σ, we definite the *fidelity* between them by

$$F(\rho,\sigma) = \|\sqrt{\sigma}\sqrt{\rho}\|_{tr}.$$

See the excellent book [13, Chapter 9] for additional properties and equivalent formulations of the fidelity. The fidelity between two probability distributions $p=\{p_i\}$ and $q=\{q_i\}$ is $F(p,q)=\sum_i \sqrt{p_i q_i}$.

The following two facts about fidelity will be useful for us.

Fact 1 *If σ, ρ are quantum states, then* $\mathrm{Tr}(\sigma\rho) \le F(\sigma,\rho)^2$.

Proof We have $\mathrm{Tr}(\sigma\rho) = \mathrm{Tr}((\sqrt{\sigma}\sqrt{\rho})(\sqrt{\sigma}\sqrt{\rho})^\dagger) = \|\sqrt{\sigma}\sqrt{\rho}\|_F^2 \le \|\sqrt{\sigma}\sqrt{\rho}\|_{tr}^2 = F(\sigma,\rho)^2$. □

Fact 2 *([13]) If σ, ρ are quantum states, then*

$$F(\sigma,\rho) = \min_{\{E_i\}} F(p,q),$$

where the minimum is over all POVMs $\{E_i\}$, and p and q are the probability distributions when ρ and σ are measured by POVM $\{E_i\}$ respectively, i.e., $p_i = \mathrm{Tr}(\rho E_i)$, and $q_i = \mathrm{Tr}(\sigma E_i)$ for any i.

The (von Neuman) *entropy* of a state ρ is defined as $H(\rho) = -\mathrm{Tr}(\rho \log \rho)$; equivalently, it is the Shannon entropy of the probability distribution given by the eigenvalues of ρ. If the joint state of Alice and Bob is ρ_{AB} (i.e., the state lives on the tensor product of two Hilbert spaces, one for Alice and one for Bob), then we can define the local state of Alice by the partial trace[1]: $\rho_A = \mathrm{Tr}_B(\rho_{AB})$. Similarly Bob's local state is $\rho_B = \mathrm{Tr}_A(\rho_{AB})$, which traces out Alice's part of the state. The *mutual information* between A and B is defined as $H(A:B) = H(\rho_A) + H(\rho_B) - H(\rho_{AB})$.

2.3 Some existing lower bounds

We now review some existing lower bound for the PSD-rank. Firstly, it is well known that the PSD-rank cannot be much smaller than the normal rank $\mathtt{rank}(A)$ of A.

Definition 2 For a nonnegative matrix A, define

$$B_1(A) = \sqrt{\mathtt{rank}(A)} \text{ and } B_1'(A) = \frac{1}{2}\left(\sqrt{1+8\mathtt{rank}(A)}-1\right).$$

Fact 3 *([1])* $\mathtt{rank}_{\mathtt{psd}}(A) \ge B_1(A)$ *and* $\mathtt{rank}^{\mathbb{R}}_{\mathtt{psd}}(A) \ge B_1'(A)$.

This bound does not look very powerful since, as stated in the introduction, usually our goal is to show lower bounds on the PSD-rank that are superpolynomial in the rank. However, this bound can be nearly tight and we give two examples in Section 6 where this is the case.

Jain et al. [9] proved that the amount of quantum communication needed for two separated players to generate a joint probability distribution P is completely characterized by the logarithm of the PSD-rank of P. According to Holevo's bound, if we encode classical information through quantum states and transfer information by sending them, then the amount of classical information that the receiver can retrieve, i.e., the mutual information, is upper bounded by the total number of qubits communicated. For more details on Holevo's bound and mutual information, see [13, Chapter 12]. Combining these two results, a trivial lower bound for PSD-rank is given by mutual information.

Definition 3 Let $P = [P(i,j)]_{i,j}$ be a two-dimensional probability distribution between two players A and B. Define $B_2(P) = 2^{H(A:B)}$, where $H(A:B)$ is the mutual information between the two players.

Fact 4 $\mathtt{rank}_{\mathtt{psd}}(P) \ge B_2(P)$.

[1] $\mathrm{Tr}_B(\rho \otimes \sigma) = \mathrm{Tr}(\sigma)\rho$, which is extended linearly to states that are not tensor products.

As an application of this lower bound, it is easy to see that the PSD-rank of a diagonal nonnegative matrix is the same as its normal rank.

Gouveia et al. [1] introduced a very general result showing that lower bounds on PSD-rank can be asymptotically larger than the rank. More precisely, they show the following.

Fact 5 *([1]) Let $P \subseteq \mathbb{R}^d$ be a polytope with f facets and let S_P be its associated slack matrix. Let $T = \sqrt{\log(f)/d}$. Then*

$$\mathtt{rank}_{\mathtt{psd}}(S_P) = \Omega\left(\frac{T}{\sqrt{\log(T)}}\right)$$

In particular, this shows that the slack matrix of a regular n-gon in $\mathbb{R}^2$, which has n facets and rank 3, has PSD-rank $\Omega(\sqrt{\log n/\log\log n})$. The nonnegative rank of this matrix is known to be $\Theta(\log n)$ [14].

3 Some properties of PSD-rank

The PSD-rank is a relatively new quantity, and even some of its basic properties are still not yet known. In this section we give a simple condition for the PSD-rank of a matrix to not be full. We then use this condition to show that PSD-rank can be strictly sub-multiplicative under tensor product. Finally, we investigate the power of using complex Hermitian over real symmetric matrices in a PSD factorization.

3.1 A sufficient condition for PSD-rank to be less than maximal

We first need a definition and a simple lemma. Let $v \in \mathbb{R}^m$ be a vector. We say that an entry v_k is *dominant* if $|v_k| > \sum_{j \neq k} |v_j|$.

Lemma 2 *Suppose that $v \in \mathbb{R}^m$ is nonnegative and has no dominant entries. Then there exist complex units $e^{i\theta_j}$ such that $\sum_j v_j e^{i\theta_j} = 0$.*

Proof Let $v \in \mathbb{R}^m$. If $m = 1$ then v has a dominant entry and there is nothing to prove. If $m = 2$ and v has no dominant entries, then $v_1 = v_2$ and the lemma holds as $v_1 - v_2 = 0$.

The first interesting case is $m = 3$. That v has no dominant entries means there is a triangle with side lengths v_1, v_2, v_3, as these satisfy the triangle inequality with respect to all permutations. Letting $v_1 e^{i\theta_1}, v_2 e^{i\theta_2}, v_3 e^{i\theta_3}$ be the vectors in the complex plane (oriented head to tail) defining the sides of this triangle gives $v_1 e^{i\theta_1} + v_2 e^{i\theta_2} + v_3 e^{i\theta_3} = 0$ as desired.

We can reduce the case $m > 3$ to the case $m = 3$. Without loss of generality, order v such that $v_1 \geq v_2 \geq \cdots \geq v_m$. Choose the least k such that

$$v_1 + \sum_{j=2}^{k} v_j \geq \sum_{j=k+1}^{m} v_j.$$

Considering the order of v, and the fact that v has no dominant entries, such a $2 \leq k < m$ must exist. The choice of k implies that

$$v_1 + \sum_{j=2}^{k-1} v_j < v_k + \sum_{j=k+1}^{m} v_j,$$

which means that

$$2v_1 + \sum_{j=2}^{k} v_j < 2v_k + v_1 + \sum_{j=k+1}^{m} v_j.$$

Combining with the fact that $v_1 \geq v_k$, this gives that

$$\sum_{j=2}^{k} v_j < v_1 + \sum_{j=k+1}^{m} v_j.$$

Then v_1, $\sum_{j=2}^{k} v_j$, $\sum_{j=k+1}^{m} v_j$ mutually satisfy the triangle inequality and we can repeat the construction from the case $m = 3$ with these lengths. □

Using the construction of Lemma 1, we can give a sufficient condition for A not to have full PSD-rank.

Theorem 6 *Let A be an m-by-n nonnegative matrix, and A' be the entry-wise square root of A (so A' is nonnegative as well). If every column of A' has no dominant entry, then the PSD-rank of A is less than m.*

Proof As each column of A' has no dominant entry, by Lemma 2 there exist complex units $e^{i\theta_{jk}}$ such that $\sum_j A'(j,k)e^{i\theta_{jk}} = 0$ for every k. Define $M(j,k) = A'(j,k)e^{i\theta_{jk}}$. Then $M \circ \overline{M} = A$ and M has rank $< m$: as each column of M sums to zero, the sum of the m rows is the 0-vector so they are linearly dependent. Lemma 1 then completes the proof. □

3.2 The behavior of PSD-rank under tensoring

In this subsection, we discuss how PSD-rank behaves under tensoring. Firstly, we have the following trivial observation on PSD-rank.

Lemma 3 *If P_1 and P_2 are two nonnegative matrices, then it holds that*

$$\texttt{rank}_{\texttt{psd}}(P_1 \otimes P_2) \leq \texttt{rank}_{\texttt{psd}}(P_1)\texttt{rank}_{\texttt{psd}}(P_2).$$

Proof Suppose $\{C_i\}$ and $\{D_j\}$ form a size-optimal PSD-factorization of P_1, and $\{E_k\}$ and $\{F_l\}$ form a size-optimal PSD-factorization of P_2, where the indices are determined by the sizes of P_1 and P_2. Then it can be seen that $\{C_i \otimes E_k\}$ and $\{D_j \otimes F_l\}$ form a PSD-factorization of $P_1 \otimes P_2$. □

We now consider an example. Let x, y be two subsets of $\{1, 2, \dots, n\}$. The disjointness function, $\mathrm{DISJ}_n(x,y)$, is defined to be 1 if $x \cap y = \varnothing$ and 0 otherwise. We denote its corresponding 2^n-by-2^n matrix by D_n, i.e., $D_n(x,y) = \mathrm{DISJ}_n(x,y)$. This function is one of the most important and well-studied in communication complexity. It can be easily checked that for any natural number k, $D_k = D_1^{\otimes k}$. According to the above lemma, we have that $\texttt{rank}_{\texttt{psd}}(D_n) \leq 2^n$, where we used the fact that $\texttt{rank}_{\texttt{psd}}(D_1) = 2$. This upper bound is trivial as the size of D_n is 2^n, but in this case it is tight as we show now.

The following lemma was also found independently by Braun and Pokutta (personal communication).

Lemma 4 *Suppose A is an m-by-n nonnegative matrix, and has the following block expression,*

$$A = \begin{bmatrix} B & C \\ D & 0 \end{bmatrix}.$$

Then $\texttt{rank}_{\texttt{psd}}(A) \geq \texttt{rank}_{\texttt{psd}}(C) + \texttt{rank}_{\texttt{psd}}(D)$.

Proof Let the size of B be k-by-l. Suppose $\{E_1, E_2, \dots, E_m\}$ and $\{F_1, F_2, \dots, F_n\}$ form a size-optimal PSD-factorization of A. Then $\{E_1, E_2, \dots, E_k\}$ and $\{F_{l+1}, F_{l+2}, \dots, F_n\}$ form a PSD-factorization of C, while $\{E_{k+1}, E_{k+2}, \dots, E_m\}$ and $\{F_1, F_2, \dots, F_l\}$ form a PSD-factorization of D.

Let the *support* of a Hermitian operator be the vector space spanned by its eigenvectors with non-zero eigenvalues. We claim that the dimension of the support of $\sum_{i=l+1}^{n} F_i$, denoted by d, will be at least $\texttt{rank}_{\texttt{psd}}(C)$. Suppose this is not the case, i.e., $d < \texttt{rank}_{\texttt{psd}}(C)$. We can find a unitary matrix U such that $U(\sum_{i=l+1}^{n} F_i)U^\dagger$ is diagonal, has rank d, and is zero outside of the upper left d-by-d block.

We claim that each matrix in the set $\{UF_{l+1}U^\dagger, UF_{l+2}U^\dagger, \dots, UF_nU^\dagger\}$ will also be zero outside of the upper left d-by-d block. The (t,t) entry of each $UF_{\ell+i}U^\dagger$ is non-negative as it is positive semidefinite. If $t > d$ then the (t,t) entry of the sum of $UF_{\ell+i}U^\dagger$ for $i = 1, \dots, n-\ell$ is zero. Thus the (t,t) entry of each $UF_{\ell+i}U^\dagger$ must be zero as well. The fact that all entries of $UF_{\ell+i}U^\dagger$ outside of the upper left d-by-d block are zero now follows from the fact that $Z(s,s)Z(t,t) \geq |Z(s,t)|^2$ for a positive semidefinite matrix Z.

Now let X_i be the upper left d-by-d block of $UF_{l+i}U^\dagger$ for $i = 1, \dots, n-\ell$ and similarly let Y_i be the upper left d-by-d block of $UE_iU^\dagger$ for $i = 1, \dots, k$. Then $\{X_i\}, \{Y_i\}$ form a PSD-factorization of C with size d. Since d is smaller than $\texttt{rank}_{\texttt{psd}}(C)$, this is a contradiction. By a similar argument, the dimension of the support of $\sum_{i=k+1}^{m} E_i$ will be at least $\texttt{rank}_{\texttt{psd}}(D)$.

On the other hand, for any $i \in \{k+1, k+2, \dots, m\}$ and $j \in \{l+1, \dots, n\}$, $\mathrm{Tr}(E_iF_j) = 0$, so the support of $\sum_{i=k+1}^{m} E_i$ is orthogonal to that of $\sum_{i=l+1}^{n} F_i$, meaning the kernel of $\sum_{i=k+1}^{m} E_i$ has dimension at least $\texttt{rank}_{\texttt{psd}}(C)$. Hence $\texttt{rank}_{\texttt{psd}}(A) \geq \texttt{rank}_{\texttt{psd}}(C) + \texttt{rank}_{\texttt{psd}}(D)$. □

Then we have that

Theorem 7 $\mathtt{rank}_{\mathtt{psd}}(D_n) = 2^n$.

Proof Note that for any integer k, D_{k+1} can be expressed as the following block matrix.

$$D_{k+1} = \begin{bmatrix} D_k & D_k \\ D_k & 0 \end{bmatrix},$$

Then by Lemma 4 we have that $\mathtt{rank}_{\mathtt{psd}}(D_{k+1}) \geq 2\mathtt{rank}_{\mathtt{psd}}(D_k)$. Since $\mathtt{rank}_{\mathtt{psd}}(D_1) = 2$, it follows that $\mathtt{rank}_{\mathtt{psd}}(D_n) \geq 2^n$. Since $\mathtt{rank}_{\mathtt{psd}}(D_n) \leq 2^n$, this completes the proof, and the PSD-rank of D_n is full. □

Based on this example and by analogy to the normal rank, one might conjecture that generally $\mathtt{rank}_{\mathtt{psd}}(P_1 \otimes P_2) = \mathtt{rank}_{\mathtt{psd}}(P_1)\mathtt{rank}_{\mathtt{psd}}(P_2)$. This is false, however, as shown by the following counterexample.

Example 1 Let $A = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}$ for nonnegative a. Then A has rank 2, and therefore PSD-rank 2, as long as $a \neq 1$. On the other hand,

$$A \otimes A = \begin{bmatrix} 1 & a & a & a^2 \\ a & 1 & a^2 & a \\ a & a^2 & 1 & a \\ a^2 & a & a & 1 \end{bmatrix}$$

satisfies the condition of Theorem 6 for any $a \in [-1+\sqrt{2}, 1+\sqrt{2}]$. Thus for $a \in [-1+\sqrt{2}, 1+\sqrt{2}] \setminus \{1\}$ we have $\mathtt{rank}_{\mathtt{psd}}(A \otimes A) < \mathtt{rank}_{\mathtt{psd}}(A)^2$.

3.3 PSD-rank and real PSD-rank

In the original definition of PSD-rank, the matrices of the PSD-factorization can be arbitrary complex Hermitian PSD matrices. A natural and interesting question is what happens if we restrict these matrices instead to be positive semidefinite *real* matrices. We call this restriction the *real PSD-rank*, and for a nonnegative matrix A we denote it by $\mathtt{rank}^{\mathbb{R}}_{\mathtt{psd}}(A)$. The following observation shows that the multiplicative gap between these notions cannot be too large.

Theorem 8 *If A is a nonnegative matrix, then* $\mathtt{rank}_{\mathtt{psd}}(A) \leq \mathtt{rank}^{\mathbb{R}}_{\mathtt{psd}}(A) \leq 2\mathtt{rank}_{\mathtt{psd}}(A)$.

Proof It is trivial that $\mathtt{rank}_{\mathtt{psd}}(A) \leq \mathtt{rank}^{\mathbb{R}}_{\mathtt{psd}}(A)$, so we only need to prove the second inequality. Suppose $r = \mathtt{rank}_{\mathtt{psd}}(A)$, and $\{E_k\}$ and $\{F_l\}$ are a size-optimal PSD-factorization of A. We now separate all the matrices involved into their real and imaginary parts. Specifically, for any k and l, let $E_k = C_k + i \cdot D_k$, and $F_l = G_l + i \cdot H_l$, where C_k and G_l are real symmetric matrices, and D_k and H_l are real skew-symmetric matrices (i.e., $D_k^T = -D_k$ and $H_l^T = -H_l$). Then it holds that

$$A_{kl} = \mathrm{Tr}(E_k F_l) = (\mathrm{Tr}(C_k G_l) - \mathrm{Tr}(D_k H_l)) + i \cdot (\mathrm{Tr}(D_k G_l) + \mathrm{Tr}(C_k H_l)).$$

Since A_{kl} is real, we in fact have

$$A_{kl} = \mathrm{Tr}(C_k G_l) - \mathrm{Tr}(D_k H_l).$$

Now for any k and l, define new matrices as follows: $S_k = \frac{1}{\sqrt{2}} \begin{bmatrix} C_k & D_k \\ -D_k & C_k \end{bmatrix}$, and $T_l = \frac{1}{\sqrt{2}} \begin{bmatrix} G_l & H_l \\ -H_l & G_l \end{bmatrix}$. Then S_k and T_l are real symmetric matrices, and $\mathrm{Tr}(S_k T_l) = \mathrm{Tr}(C_k G_l) - \mathrm{Tr}(D_k H_l) = A_{kl}$.

It remains to show that the matrices S_k and T_l are positive semidefinite. Suppose $u = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ is a $2r$-dimensional real vector, where v_1 and v_2 are two arbitrary r-dimensional real vectors. Since E_k is positive semidefinite, we have

$$0 \leq (v_2^T - i \cdot v_1^T) E_k (v_2 + i \cdot v_1) = v_1^T C_k v_1 - v_2^T D_k v_1 + v_1^T D_k v_2 + v_2^T C_k v_2 = \sqrt{2} u^T S_k u.$$

Hence S_k is positive semidefinite. Similarly we can show that T_l is positive semidefinite for every l. □

Below in Example 9 we will exhibit a gap between $\mathtt{rank}_{\mathtt{psd}}(A)$ and $\mathtt{rank}^{\mathbb{R}}_{\mathtt{psd}}(A)$ by a factor of $\sqrt{2}$.

4 Three new lower bounds for PSD-rank

In this section we give three new lower bounds on the PSD-rank. All of these bounds are based on the interpretation of PSD-rank in terms of communication complexity.

4.1 A physical explanation of PSD-rank

For a nonnegative $m \times n$ matrix $P = [P(i,j)]_{i,j}$, suppose $\texttt{rank}_{\texttt{psd}}(P) = r$. Then there exist $r \times r$ positive semidefinite matrices E_i, F_j, satisfying that $P(i,j) = \mathrm{Tr}(E_i F_j)$, for every $i \in [m]$ and $j \in [n]$. Fiorini et al. showed how a size-r PSD-factorization of a matrix P induces a one-way quantum communication protocol with $(r+1)$-dimensional messages that computes P in expectation [2]. We will now show that without loss of generality the factors $E_1, \ldots, E_m, F_1, \ldots, F_n$ have a very particular form. Namely, we can assume that $\sum_i E_i = I$ (so they form a POVM) and $\mathrm{Tr}(F_j) = 1$ (so the F_j can be viewed as quantum states). We now give a direct proof of this without increasing the size. This observation will be the key to our lower bounds.

Lemma 5 *Let P be an m-by-n matrix where each column is a probability distribution. If $\texttt{rank}_{\texttt{psd}}(P) = r$, then there exists a PSD-factorization for $P(i,j) = \mathrm{Tr}(E_i F_j)$ such that $\mathrm{Tr}(F_j) = 1$ for each j and*

$$\sum_{i=1}^{m} E_i = I,$$

where I is the r-dimensional identity.

Proof Suppose r-by-r positive semidefinite matrices $C_1, \ldots, C_m$ and $D_1, \ldots, D_n$ form a PSD-factorization for P. Note that for any r-by-r unitary matrix U, it holds that

$$\mathrm{Tr}(C_i D_j) = \mathrm{Tr}((U C_i U^\dagger)(U D_j U^\dagger)).$$

Therefore $U C_i U^\dagger$ and $U D_j U^\dagger$ also form a PSD-factorization for P. In the following, we choose U as the unitary matrix that makes $C' = U C U^\dagger$ diagonal, where $C = \sum_i C_i$.

According to the proof for Lemma 4, the dimension of the support of C cannot be smaller than r. Since the size of C is also r, C must be full-rank. Then C' is also full-rank, and one can always find another full-rank nonnegative diagonal matrix V such that $V C' V^\dagger = I$. Let $E_i = V U C_i U^\dagger V^\dagger$, and $F_j = (V^{-1})^\dagger U D_j U^\dagger V^{-1}$. It is not difficult to verify that E_i and F_j form another PSD-factorization for P with size r, satisfying $\sum_i E_i = I$.

Finally note that $\mathrm{Tr}(F_j) = \mathrm{Tr}(F_j I) = \sum_i \mathrm{Tr}(E_i F_j) = 1$ as each column of P sums to one. □

4.2 A lower bound based on fidelity

Definition 4 For nonnegative stochastic matrix P, define

$$B_3(P) = \max_q \frac{1}{\sum_{i,j} q_i q_j F(P_i, P_j)^2},$$

where P_i is the i^{th} column of P and the max is taken over probability distributions $q = \{q_j\}$.

Theorem 9 $\texttt{rank}_{\texttt{psd}}(P) \geq B_3(P)$.

Proof Let $\{E_i\}, \{\rho_j\}$ be a size-optimal PSD-factorization of P. According to Lemma 5, we may assume that $\sum_i E_i = I$ and $\mathrm{Tr}(\rho_j) = 1$ for each j. For a probability distribution $\{q_j\}$, let $\rho = \sum_j q_j \rho_j$. Notice that the dimension of ρ is $\texttt{rank}_{\texttt{psd}}(P)$, thus the *rank* of ρ will be at most $\texttt{rank}_{\texttt{psd}}(P)$. We use the trace norm bound Eq. (1) to lower bound the rank of ρ giving

$$\texttt{rank}_{\texttt{psd}}(P) \geq \frac{\|\rho\|_{tr}^2}{\|\rho\|_F^2} = \frac{1}{\|\rho\|_F^2}.$$

Let us now proceed to upper bound $\|\rho\|_F^2$. We have

$$\|\rho\|_F^2 = \mathrm{Tr}(\rho^2) = \sum_{i,j} q_i q_j \mathrm{Tr}(\rho_i \rho_j) \leq \sum_{i,j} q_i q_j F(\rho_i, \rho_j)^2,$$

where we used Fact 1. As P_i is obtained from measuring ρ_i with the POVM $\{E_j\}$, according to Fact 2 we have that $F(\rho_i, \rho_j) \leq F(P_i, P_j)$, which gives the bound $\texttt{rank}_{\texttt{psd}}(P) \geq \max_q \frac{1}{\sum_{i,j} q_i q_j F(P_i, P_j)^2}$. □

We can extend the notation $B_3(P)$ to nonnegative matrices P that are not stochastic, by first normalizing the columns of P to make it stochastic and then applying B_3 to the resulting stochastic matrix. As rescaling a nonnegative matrix by multiplying its rows or columns with nonnegative numbers does not increase its PSD-rank, we have the following definition and corollary.

Definition 5 For a nonnegative $m \times n$ matrix $P = [P(i,j)]_{i,j}$, define

$$B'_3(P) = \max_{q,D} \frac{1}{\sum_{i,j} q_i q_j F((DP)_i, (DP)_j)^2},$$

where $q = \{q_j\}$ is a probability distribution, D is a diagonal nonnegative matrix, and $(DP)_i$ is the probability distribution obtained by normalizing the i^{th} column of DP via a constant factor.

Corollary 1 $\texttt{rank}_{\texttt{psd}}(P) \geq B'_3(P)$.

We now see an example where rescaling can improve the bound.

Example 2 Consider the following $n \times n$ nonnegative matrix A, where $n = 10$, and $\epsilon = 0.01$.

$$A = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ \epsilon & 1 & \epsilon & \cdots & \epsilon & \epsilon \\ \epsilon & \epsilon & 1 & \cdots & \epsilon & \epsilon \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \epsilon & \epsilon & \epsilon & \cdots & 1 & \epsilon \\ \epsilon & \epsilon & \epsilon & \cdots & \epsilon & 1 \end{bmatrix}.$$

Let P be the nonnegative stochastic matrix obtained by normalizing the columns of A. P has the same PSD-rank as A. By choosing q as the uniform probability distribution, we can get a lower bound on $B_3(P)$ as follows. Note that for any $i \in [n] \setminus \{1\}$, we have that

$$f_1 := F(P_1, P_i) = \frac{1 + \sqrt{\epsilon} + (n-2)\epsilon}{\sqrt{1 + (n-1)\epsilon} \cdot \sqrt{2 + (n-2)\epsilon}},$$

and for any distinct $i, j \in [n] \setminus \{1\}$, it holds that

$$f_2 := F(P_i, P_j) = \frac{1 + 2\sqrt{\epsilon} + (n-3)\epsilon}{2 + (n-2)\epsilon}.$$

Then we get

$$B_3(A) \geq \frac{n^2}{n + 2(n-1) \cdot {f_1}^2 + (n-2)(n-1) \cdot {f_2}^2} \approx 2.09.$$

We now multiply every row of A by 10 except that the first one is multiplied by 0, i.e., the matrix D in Corollary 1 is a diagonal nonnegative matrix with diagonal $(0, 10, \ldots, 10)$. Then we obtain another nonnegative matrix $\hat{A} = DA$. By a similar calculation as above, it can be verified that $B_3(\hat{A}) \geq 4.88$, hence we have $B'_3(A) \geq 4.88$, which is a better lower bound than $B_3(A)$.

4.3 A lower bound based on the structure of POVMs

Definition 6 For nonnegative stochastic matrix P, define $B_4(P) = \sum_i \max_j P(i,j)$.

Theorem 10 $\texttt{rank}_{\texttt{psd}}(P) \geq B_4(P)$.

Proof Let $\{E_i\}, \{\rho_j\}$ be a size-optimal PSD-factorization of P with $\sum_i E_i = I$ and $\text{Tr}(\rho_j) = 1$ for each j. Note that this condition on the trace of ρ_j implies $I \succeq \rho_j$. Thus

$$\text{Tr}(E_i) = \text{Tr}(E_i \cdot I) \geq \max_j \text{Tr}(E_i \rho_j) = \max_j P(i,j).$$

On the other hand, since $\sum_i E_i = I$, we have

$$\texttt{rank}_{\texttt{psd}}(P) = \sum_i \text{Tr}(E_i) \geq \sum_i \max_j P(i,j),$$

where we used that the size of I is $\texttt{rank}_{\texttt{psd}}(P)$. □

A variant of B_4 involving rescaling can sometimes lead to better bounds:

Definition 7 For a nonnegative $m \times n$ matrix $P = [P(i,j)]_{i,j}$, define

$$B_4'(P) = \max_D \sum_i \max_j ((DP)_j)_i,$$

where D is a diagonal nonnegative matrix, $(DP)_j$ is the probability distribution obtained by normalizing the j^{th} column of DP via a constant factor, and $((DP)_j)_i$ is the i^{th} entry of $(DP)_j$.

Corollary 2 $\texttt{rank}_{\texttt{psd}}(P) \geq B_4'(P)$.

Example 3 We consider the same matrices A and D as in Example 2, and get that

$$B_4(A) = \frac{1}{1+(n-1)\epsilon} + (n-1) \cdot \frac{1}{2+(n-2)\epsilon} \approx 5.24.$$

Similarly, it can be checked that $B_4'(A) \geq 8.33$. The latter indicates that $\texttt{rank}_{\texttt{psd}}(A) \geq 9$, which is better than the bound 4 given by $B_1(A)$ or 6 by $B_2(A)$.

4.4 Another bound that combines B_3 with B_4

Here we will show that B_4 can be strengthened further by combining it with the idea that bounds $\text{Tr}(\sigma^2)$ in B_3, where σ is a quantum state that can be expressed as some linear combination of ρ_i's.

Definition 8 For a nonnegative stochastic matrix $P = [P(i,j)]_{i,j}$, define

$$B_5(P) = \sum_i \max_{q^{(i)}} \frac{\sum_k q_k^{(i)} P(i,k)}{\sqrt{\sum_{s,t} q_s^{(i)} q_t^{(i)} F(P_s, P_t)^2}},$$

where P_s is the s^{th} column of P, and for every i, $q^{(i)} = \{q_k^{(i)}\}$ is a probability distribution.

Theorem 11 $\texttt{rank}_{\texttt{psd}}(P) \geq B_5(P)$.

Proof We define $\{E_i\}$ and $\{\rho_j\}$ as before. For an arbitrary i, we define $\sigma_i = \sum_k q_k^{(i)} \rho_k$. This is a valid quantum state. Since $\text{Tr}(E_i \rho_j) = P(i,j)$, it holds that $\text{Tr}(E_i \sigma_i) = \sum_k q_k^{(i)} P(i,k)$. The Cauchy-Schwarz inequality gives $\text{Tr}^2(E_i \sigma_i) \leq \text{Tr}(E_i^2)\text{Tr}(\sigma_i^2)$. This implies that

$$\left(\sum_k q_k^{(i)} P(i,k)\right)^2 \leq \text{Tr}^2(E_i) \sum_{s,t} q_s^{(i)} q_t^{(i)} F(P_s, P_t)^2,$$

where we used the facts that $\text{Tr}(E_i^2) \leq \text{Tr}^2(E_i)$ and $\text{Tr}(\sigma_i^2) \leq \sum_{s,t} q_s^{(i)} q_t^{(i)} F(P_s, P_t)^2$; the latter has been proved in Theorem 9. Therefore, for any distribution $q^{(i)}$ it holds that

$$\text{Tr}(E_i) \geq \frac{\sum_k q_k^{(i)} P(i,k)}{\sqrt{\sum_{s,t} q_s^{(i)} q_t^{(i)} F(P_s, P_t)^2}}.$$

Substituting this result into the fact that $\sum_i \text{Tr}(E_i) = \texttt{rank}_{\texttt{psd}}(P)$ completes the proof. □

We also have the following corollary that allows rescaling.

Definition 9 For a nonnegative $m \times n$ matrix $P = [P(i,j)]_{i,j}$, define

$$B_5'(P) = \max_D \sum_i \max_{q^{(i)}} \frac{\sum_k q_k^{(i)} ((DP)_k)_i}{\sqrt{\sum_{s,t} q_s^{(i)} q_t^{(i)} F((DP)_s, (DP)_t)^2}},$$

where for every i, $q^{(i)} = \{q_k^{(i)}\}$ is a probability distribution, D is a diagonal nonnegative matrix, $(DP)_k$ is the probability distribution obtained by normalizing the k^{th} column of P via a constant factor, and $((DP)_k)_i$ is the i^{th} entry of $(DP)_k$.

Corollary 3 $\texttt{rank}_{\texttt{psd}}(P) \geq B_5'(P)$.

We now give an example showing that B_5 can be better than B_4.

Example 4 Consider the following $n \times n$ nonnegative matrix A, where $n = 10$, and $\epsilon = 0.01$.

$$A = \begin{bmatrix} 1 & 1 & \epsilon & \cdots & \epsilon & \epsilon \\ \epsilon & 1 & 1 & \cdots & \epsilon & \epsilon \\ \epsilon & \epsilon & 1 & \cdots & \epsilon & \epsilon \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \epsilon & \epsilon & \epsilon & \cdots & 1 & 1 \\ 1 & \epsilon & \epsilon & \cdots & \epsilon & 1 \end{bmatrix}.$$

It can be verified that $B_4(A) \approx 4.81$. In order to provide a lower bound for $B_5(A)$, for any i we choose $q^{(i)}$ as $\{0, \ldots, 0, 1/2, 1/2, 0, \ldots 0\}$, where the positions of $1/2$ are exactly the same as those of 1 in the i^{th} row of A. Straightforward calculation shows that $B_5(A) \geq 5.36$, which is better than $B_4(A)$.

Even B_5 can be quite weak in some cases. For example for the matrix in Example 7 one can show $B_5(A) < 1.1$, which is weaker than $B_1(A) \approx 3.16$.

5 Comparisons between the bounds

In this section we give explicit examples comparing the three new lower bounds on PSD-rank (B_3, B_4 and B_5) and the two that were already known (B_1 and B_2). These examples will show that: (1) for some cases the three new lower bounds are better than B_1 and B_2; (2) the bounds B_3 and B_4 are incomparable.

All our examples will only use positive entries, which trivializes all support-based lower bound methods, i.e., methods that only look at the pattern of zero and non-zero entries in the matrix. Note that most lower bounds on nonnegative rank are in fact support-based (one exception is [15]). Since PSD-rank is always less than or equal to nonnegative rank, the results obtained in the current paper could also serve as new lower bounds for nonnegative rank that apply to arbitrary nonnegative matrices. Serving as lower bounds for nonnegative rank, our bounds are more coarse than the bounds in [15] (this is natural, as we focus on PSD-rank essentially, and the gap between PSD-rank and nonnegative rank can be very large [2]). On the other hand, our bounds are much easier to calculate.

The first example indicates that in some cases B_4 can be at least quadratically better than each of B_1, B_2 and B_3.

Example 5 Consider the following $(n+1) \times (n+1)$ nonnegative matrix A, where $\epsilon = 1/n$.

$$A = \begin{bmatrix} 1 & \epsilon & \epsilon & \cdots & \epsilon & \epsilon \\ \epsilon & 1 & \epsilon & \cdots & \epsilon & \epsilon \\ \epsilon & \epsilon & 1 & \cdots & \epsilon & \epsilon \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \epsilon & \epsilon & \epsilon & \cdots & 1 & \epsilon \\ \epsilon & \epsilon & \epsilon & \cdots & \epsilon & 1 \end{bmatrix}.$$

Theorem 14 (below) shows that $B_4(A) = \frac{n+1}{2}$, and by straightforward calculation one can also get that $B_1(A) = \sqrt{n}$, $B_2(A) = \frac{n+1}{2\sqrt{n}} \approx \frac{\sqrt{n}}{2}$, and numerical calculation indicates that $B_3(A)$ is around 4.

The second example shows that B_3 can also be the best among the four lower bounds B_1, B_2, B_3, B_4, indicating that B_3 and B_4 are incomparable.

Example 6 Consider the following $n \times n$ nonnegative matrix A, where $n = 10$, and $\epsilon = 0.001$.

$$A = \begin{bmatrix} 1 & 1 & \epsilon & \cdots & \epsilon & \epsilon \\ 1 & 1 & 1 & \cdots & \epsilon & \epsilon \\ \epsilon & 1 & 1 & \cdots & \epsilon & \epsilon \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \epsilon & \epsilon & \epsilon & \cdots & 1 & 1 \\ \epsilon & \epsilon & \epsilon & \cdots & 1 & 1 \end{bmatrix}.$$

That is, $A = (1 - \epsilon) \cdot B + \epsilon \cdot J$, where B is the tridiagonal matrix with all nonzero elements being 1, and J is the all-one matrix. By straightforward calculation, we find that $B_1(A) \approx 3.16$, $B_2(A) \approx 3.42$, $B_4(A) \approx 3.99$, and the calculation based on uniform probability distribution q shows that $B_3(A) \geq 4.52$. The result of $B_3(A)$ shows that $\texttt{rank}_{\texttt{psd}}(A) \geq 5$.

Unfortunately, sometimes B_3 and B_4 can be very weak bounds[2], and even the trivial rank-based bound B_1 can be much better than both of them.

Example 7 Consider the following $n \times n$ nonnegative matrix A, where $n = 10$, and $\epsilon = 0.9$.

$$A = \begin{bmatrix} 1 & \epsilon & \epsilon & \cdots & \epsilon & \epsilon \\ \epsilon & 1 & \epsilon & \cdots & \epsilon & \epsilon \\ \epsilon & \epsilon & 1 & \cdots & \epsilon & \epsilon \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \epsilon & \epsilon & \epsilon & \cdots & 1 & \epsilon \\ \epsilon & \epsilon & \epsilon & \cdots & \epsilon & 1 \end{bmatrix}.$$

It can be verified that $B_2(A) \approx 1.0005$, and $B_4(A) \approx 1.099$. For $B_3(A)$, numerical calculation indicates that it is also around 1. However, $B_1(A) = \sqrt{10} \approx 3.16$. Thus, the best lower bound is given by $B_1(A)$, i.e., $\texttt{rank}_{\texttt{psd}}(A) \geq 4$.

Example 8 For slack matrices of regular polygons, the two new bounds B_3 and B_4 are not good either, and in many cases they are at most 3. Moreover, numerical calculations show that rescaling probably cannot improve them by much. Note that the two trivial bounds B_1 and B_2 are also very weak for these cases. As an example, consider the canonical slack matrix of the regular hexagon

$$A = \begin{bmatrix} 0 & 0 & 1 & 2 & 2 & 1 \\ 1 & 0 & 0 & 1 & 2 & 2 \\ 2 & 1 & 0 & 0 & 1 & 2 \\ 2 & 2 & 1 & 0 & 0 & 1 \\ 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 2 & 1 & 0 \end{bmatrix}.$$

It can be verified that $B_1(A) \approx 1.73$, $B_2(A) \approx 1.59$, $B_4(A) = 6 \times \frac{2}{6} = 2$, and choosing q in the definition of $B_3(A)$ as the uniform distribution gives that $B_3(A) > 2.1$. Furthermore, our numerical calculations showed that choosing other distributions or using rescaling could not improve the results much, and never gave lower bounds greater than or equal to 3. Note that the exact PSD-rank of this matrix is 4 [1].

6 PSD-factorizations for specific functions

In this section we show the surprising power of PSD-factorizations by giving nontrivial *upper* bounds on the PSD-ranks of the matrices defined by two important functions in theoretical computer science, i.e., the nonequality and the inner product functions. These bounds are tight up to constant factors.

[2] Even though a nonnegative matrix has the same PSD-rank as its transposition, the bounds given by B_3 (or B_4) can be quite different, for instance for the matrix A of Example 2.

6.1 The nonequality function

The nonequality function defines an n-by-n matrix A_n with entries $A_n(i,i) = 0$ and $A_n(i,j) = 1$ if $i \neq j$. In other words, $A_n = J_n - I_n$ where J_n is the all-ones matrix and I_n is the identity of size n. This is also known as the "derangement matrix." Note that for $n > 1$ it has full rank.

The basic idea of our PSD factorization is the following. We first construct n^2 Hermitian matrices G_{ij} of size n with spectral norm at most 1. Then the matrices $I + G_{ij}$ and $I - G_{ij}$ will be positive semidefinite, and these will form the factorization. Note that

$$\mathrm{Tr}((I + G_{ij})(I - G_{kl})) = \mathrm{Tr}(I) + \mathrm{Tr}(G_{ij}) - \mathrm{Tr}(G_{kl}) - \mathrm{Tr}(G_{ij}G_{kl}).$$

Thus if we can design the G_{ij} such that $\mathrm{Tr}(G_{ij}) = \mathrm{Tr}(G_{kl})$ for all i,j,k,l and $\mathrm{Tr}(G_{ij}G_{kl}) = \delta_{ik}\delta_{jl}n$ (where $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ otherwise), this will give a factorization proportional to the nonequality matrix.

For the case where n is odd, we are able to carry out this plan exactly.

Lemma 6 *Let n be odd. Then there are n^2 Hermitian matrices G_{ij} of size n such that*

- $\mathrm{Tr}(G_{ij}) = \mathrm{Tr}(G_{kl})$ *for all* $i,j,k,l \in \{0,\dots,n-1\}$.
- $\mathrm{Tr}(G_{ij}G_{kl}) = \delta_{ik}\delta_{jl}n$.
- $G_{ij}^2 = I_n$.

Proof We will use two auxiliary matrices in our construction. We will label matrix entries from $\{0,\dots,n-1\}$. Let L be the addition table of $\mathbb{Z}_n$, that is $L(i,j) = i + j \mod n$. Notice that L is a symmetric Latin square[3] with distinct entries along the main diagonal. Let V be the $n \times n$ Vandermonde matrix where $V(k,l) = e^{-2\pi \mathrm{i} kl/n}$ for $k,l \in \{0,\dots,n-1\}$. Note that $VV^\dagger = nI_n$.

We now define the matrices G_{ij} for $i,j \in \{0,\dots,n-1\}$. The matrix G_{ij} will be nonzero only in those (k,l)-entries where $L(k,l) = i$. Thus the zero/nonzero pattern of each G_{ij} forms a permutation matrix with exactly one 1 on the diagonal. These nonzero entries will be filled in from the j-th row of V. We do this in a way to ensure that G_{ij} is Hermitian. Thus $V(j,0) = 1$ will be placed on the diagonal entry of G_{ij}. Now fix an ordering of the $\lfloor n/2 \rfloor$ other pairs $(k,l),(l,k)$ of nonzero entries of G_{ij} (say that each (k,l) is above the diagonal). In the t-th such pair we put the conjugate pair $V(j,t), V(j,n-t)$. In this way, G_{ij} is Hermitian, and as the ordering is the same for all j we have that $\mathrm{Tr}(G_{ij}G_{ik}) = (\langle V_j|, \langle V_k|) = n\delta_{j,k}$, where $\langle V_j|$ is the j-th row of V, and $(\langle V_j|, \langle V_k|)$ is the inner product of the two complex vectors $\langle V_j|$ and $\langle V_k|$.

To finish, we check the other properties. Each G_{ij} has trace one. If $i \neq k$ then $\mathrm{Tr}(G_{ij}G_{kl}) = 0$, because the zero/nonzero patterns are disjoint. Finally, as the zero/nonzero pattern of each G_{ij} is a permutation matrix, and entries are roots of unity, $G_{ij}^2 = I_n$. □

This gives the following theorem for the n^2-by-n^2 nonequality matrix.

Theorem 12 *Suppose n is odd, and let A_{n^2} be the nonequality matrix of size n^2. Then $\mathtt{rank}_{\mathtt{psd}}(A_{n^2}) \leq n$.*

Proof Suppose n^2 Hermitian matrices G_{ij} have been constructed as in Lemma 6. We now define the matrices $X_{ij} = \frac{1}{\sqrt{n}}(I + G_{ij})$ and $Y_{ij} = \frac{1}{\sqrt{n}}(I - G_{ij})$. Note that the spectral norm of each G_{ij} is 1, so X_{ij} and Y_{ij} are PSD. Also, we have

$$\begin{aligned}\mathrm{Tr}(X_{ij}Y_{kl}) &= \frac{1}{n}\left(\mathrm{Tr}(I) + \mathrm{Tr}(G_{ij}) - \mathrm{Tr}(G_{kl}) - \mathrm{Tr}(G_{ij}G_{kl})\right)\\ &= \frac{1}{n}\left(n - \delta_{ik}\delta_{jl}n\right) = 1 - \delta_{ik}\delta_{jl}.\end{aligned}$$

□

We now turn to the case that n is even. The result is slightly worse here.

Lemma 7 *Let n be even. Then there are $n^2 - 1$ Hermitian matrices G_{ij} such that*

- $\mathrm{Tr}(G_{ij}) = \mathrm{Tr}(G_{kl})$ *for all* i,j,k,l.
- $\mathrm{Tr}(G_{ij}G_{kl}) = \delta_{ik}\delta_{jl}n$.
- $G_{ij}^2 = I_n$.

[3] A Latin square is an n-by-n matrix in which each row and each column is a permutation of $\{0,\dots,n-1\}$.

Proof The construction is similar. Again let V be the Vandermonde matrix of roots of unity. This time we take a symmetric Latin square L' which is different from the L used above. The entries of L' are from $\{0, \dots, n-1\}$ and the diagonal has all entries equal to 0. Note that constructing this kind of L' is always possible, and can be obtained as follows. We first find a symmetric Latin square A of size $n-1$ with entries from $\{1, 2, \dots, n-1\}$, whose diagonal entries are distinct (this kind of matrices exists according to the proof for Lemma 6). Then we add an n-th row and n-th column to A by setting $A(n,k) = A(k,n) = A(k,k)$ for $1 \le k < n$. Finally we change all the diagonal entries to 0 and let L' be the resulting matrix, which is a symmetric Latin square.

For $i > 0$, the matrix G_{ij} is defined as before, i.e., G_{ij} is nonzero only in those (k,l)-entries where $L'(k,l) = i$, and the nonzero entries are filled in from the j-th row of V. Since $i > 0$, according to the construction of L' the nonzero entries of G_{ij} will not be on the diagonal. Once again, we choose conjugate pairs from these entries and put each pair in symmetrical positions to make the matrix obtained Hermitian. An additional subtlety is that if j is odd then $V(j,0) = 1$ and $V(j, n/2) = -1$. They are not conjugate, but we choose them as a pair and change them to be $(\mathrm{i}, -\mathrm{i})$ in the matrix G_{ij}, where i denotes the imaginary unit (not to be confused with the index i). Then the new pair is conjugate, and it can be verified that this change does not affect the inner product between this matrix and the others.

For $i = 0$, all the nonzero entries of G_{ij} will be on the diagonal. For each j, we fill these entries with a real unit vector that sums to 0. Furthermore, the vectors chosen for different j are orthogonal to each other. By induction, it can be proved that the size of the largest set of such n-dimensional vectors is $n-1$. In this way, we construct $n^2 - 1$ matrices that satisfy the requirements. □

As with the case where n is odd, we have the following theorem based on Lemma 7.

Theorem 13 *Suppose n is even, and let A_{n^2-1} be the nonequality matrix of size n^2-1. Then it holds that* $\mathtt{rank}_{\mathtt{psd}}(A_{n^2-1}) \le n$.

The nonequality function gives a family of matrices where PSD-rank is smaller than the real PSD-rank.

Example 9 We have seen that for odd n, the PSD-rank of the nonequality matrix of size n^2 is at most n. This is tight by Fact 3, since the rank of the nonequality matrix of this size is n^2. On the other hand, also by Fact 3, the real PSD-rank is at least $\lceil \sqrt{2}n - 1/2 \rceil$. This shows a multiplicative gap of approximately $\sqrt{2}$ between the real and complex PSD-rank. The rank lower bound on the real PSD-rank is in fact tight, as shown by [16, Example 5.1].

Fawzi-Gouveia-Parrilo-Robinson-Thomas [16, Section 2.2] independently observed that the real and complex PSD-rank are not the same, showing that the 4-by-4 derangement matrix has complex PSD-rank 2, while by Fact 3 the real PSD-rank is at least 3.

It should be pointed out that the results in the current subsection reveal a fundamental difference between PSD-rank and the normal rank. Recall that for the normal rank we have that $\mathtt{rank}(A-B) \ge \mathtt{rank}(B) - \mathtt{rank}(A)$. Thus if A is a rank-one matrix, the ranks of $A - B$ and B cannot be very different. The results above, on the other hand, indicate that the situation is very different for PSD-rank, where $A - B$ and B can have vastly different PSD-ranks even for a rank-one matrix A. This fact shows that the PSD-rank is not as robust to perturbations as the normal rank, a contributing reason to why the PSD-rank is difficult to bound.

Proposition 1 *There exist nonnegative matrices A and B, such that $A - B$ is also nonnegative, and*

$$\mathtt{rank}_{\mathtt{psd}}(A-B) < \mathtt{rank}_{\mathtt{psd}}(B) - \mathtt{rank}_{\mathtt{psd}}(A).$$

Proof Choose $A = J$ and $B = I$, where their common size is n, and J is the all-one matrix. Then we have that $\mathtt{rank}_{\mathtt{psd}}(A-B) \approx \sqrt{n}$, $\mathtt{rank}_{\mathtt{psd}}(B) = n$, while $\mathtt{rank}_{\mathtt{psd}}(A) = 1$. Choosing n large enough gives the desired separation. □

6.2 Approximations of the identity

Here we first consider the PSD-rank of *approximations* of the identity. We say that an n-by-n matrix A is an ϵ-approximation of the identity if $A(i,i) = 1$ for all $i \in [n]$ and $0 \le A(i,j) \le \epsilon$ for all $i \ne j$. The usual rank of approximations of the identity has been well studied by Alon [17].

In particular, it is easy to show that if A is an ϵ-approximation of the identity then

$$\mathtt{rank}(A) \ge \frac{n}{1 + \epsilon^2(n-1)}.$$

Using the bound B_4 we can show a very analogous result for PSD-rank.

Theorem 14 *If an n-by-n matrix A is an ϵ-approximation of the identity, then*

$$\texttt{rank}_{\texttt{psd}}(A) \geq \frac{n}{1+\epsilon(n-1)}.$$

In particular, if $\epsilon \leq 1/n$ then $\texttt{rank}_{\texttt{psd}}(A) > n/2$.

Proof We first normalize each column of A to a probability distribution, obtaining a stochastic matrix P. Each column will be divided by a number at most $1+\epsilon(n-1)$. Thus the largest entry of each column is at least $1/(1+\epsilon(n-1))$. Hence the method B_4 gives the claimed bound. □

We now show that this bound is tight in the case of small ϵ. If $\epsilon \geq 1/(n-1)^2$, then by Theorem 6 the PSD-rank of the n-by-n matrix with ones on the diagonal and ϵ off the diagonal is not full. On the other hand, if $\epsilon < 1/(n-1)^2$ then any ϵ-approximation of the identity has full PSD-rank, by Theorem 14. This gives the following proposition.

Proposition 2 *Suppose $A(i,i) = 1$ for all $i \in [n]$ and $A(i,j) = \epsilon$ for $i \neq j$, then $\texttt{rank}_{\texttt{psd}}(A) = n$ if and only if $\epsilon < 1/(n-1)^2$.*

Combining this proposition and Lemma 3, we immediately have the following proposition.

Proposition 3 *Let m divide n and consider the m-by-m matrix B where $B(i,i) = 1$ and $B(i,j) = 1/(m-1)^2$. Then $A = I_{n/m} \otimes B$ is an ϵ-approximation of the identity, and $\texttt{rank}_{\texttt{psd}}(A) \leq n - \frac{n}{m}$, where $\epsilon = 1/(m-1)^2$.*

Proof Note that $\texttt{rank}_{\texttt{psd}}(B) \leq m-1$. □

As a generalization of approximations of the identity with the same off-diagonal entries, we now turn to consider the PSD-rank of the following class of matrices

$$M_c = \begin{bmatrix} c & 1 & 1 & \cdots & 1 & 1 \\ 1 & c & 1 & \cdots & 1 & 1 \\ 1 & 1 & c & \cdots & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & c & 1 \\ 1 & 1 & 1 & \cdots & 1 & c \end{bmatrix},$$

where c could be any nonnegative real number, and suppose the size of M_c is n-by-n. For $c = 0$, M_c is exactly the matrix corresponding to the Nonequality function. Besides, if $c > (n-1)^2$, Proposition 2 implies that the PSD-rank of M_c will be full. In both of these two cases, our results are very tight. Then a natural question is, how about the case when $0 < c < (n-1)^2$ (excluding $c = 1$)? For this case, it turns out that we have the following theorem. Combined with $B_1(M_c) = \sqrt{n}$, this result indicates that when c is not very large, $\texttt{rank}_{\texttt{psd}}(M_c)$ is very small, which is much stronger than Proposition 3.

Theorem 15 *If $c > 2$, $\texttt{rank}^{\mathbb{R}}_{\texttt{psd}}(M_c) \leq 2\lceil c \rceil \cdot \lceil \sqrt{n} \rceil$. If $c \in [0,2]$, $\texttt{rank}^{\mathbb{R}}_{\texttt{psd}}(M_c) \leq \lceil \sqrt{2n} \rceil + 1$.*

Proof Note that $\texttt{rank}^{\mathbb{R}}_{\texttt{psd}}(M_c) \leq n$, which means that when $c \geq \sqrt{n}/2$, the above theorem is trivially true. Therefore, we only consider the case that $c < \sqrt{n}/2$. We first suppose c is an integer in the interval $(2, \sqrt{n}/2)$. For a fixed $r \geq c$, we consider the largest set $\mathcal{S}$ of subsets of $[r]$ such that every subset has exactly c elements and the intersection of any two subsets contains at most one element in $[r]$. Suppose the cardinality of $\mathcal{S}$ is $p(r,c)$, and the elements of $\mathcal{S}$ are $\{S_1, S_2, \ldots, S_{p(r,c)}\}$, i.e., for any $i \in [p(r,c)]$, S_i is a subset of $[r]$ with size c.

For any $i \in [p(r,c)]$, we now construct two r-by-r matrices E_i and F_i based on S_i as follows. In E_i, we first choose the submatrix whose row index set and column index set are S_i, and let this submatrix be a c-by-c all-one matrix. All the other entries of E_i are set to 0. F_i is similar to E_i except that all its diagonal entries are 1. Thus, for every i, both E_i and F_i are positive semidefinite.

It is not difficult to verify that for any $x, y \in [p(r,c)]$, if $x = y$ then $\mathrm{Tr}(E_x F_y) = c^2$, and if $x \neq y$ then $\mathrm{Tr}(E_x F_y) = c$. That is, if we choose r properly such that $p(r,c) \geq n$, then $\{\frac{1}{c}E_1, \ldots, \frac{1}{c}E_n\}$ and $\{F_1, \ldots, F_n\}$ form a size-r PSD-factorization of M_c, which shows that $\texttt{rank}^{\mathbb{R}}_{\texttt{psd}}(M_c) \leq r$. We have the following lemma to provide bounds on $p(r,c)$.

Lemma 8 *Let c be a positive integer and $r \geq c$ be a prime number. There exists a family of r^2 many c-element sets over a universe of size cr, such that any two distinct sets from this family intersect in at most one point.*

Proof Since r is a prime number, $\mathbb{F}_r$ is a finite field. With each $(a,b) \in \mathbb{F}_r \times \mathbb{F}_r$ we associate the following set in the universe $[c] \times \mathbb{F}_r$. It is a c-element subset of the graph of the line $y = ax + b$.

$$S_{ab} = \{(x, ax+b) : x \in [c]\}.$$

We have r^2 such sets, one for each choice of a, b. Since two distinct lines can intersect in at most one (x,y)-pair, we have $|S_{ab} \cap S_{a'b'}| \leq 1$ if $(a,b) \neq (a',b')$. □

Let us go back to the proof for Theorem 15. Let r be the smallest prime number greater than or equal to $\lceil\sqrt{n}\rceil$, then we know $r \leq 2\lceil\sqrt{n}\rceil$. Now $r > c$, and by the above lemma there exist $r^2 \geq n$ c-element sets over a universe of size cr. This results in a PSD-factorization for M_c of size cr, hence $\mathtt{rank}^{\mathbb{R}}_{\mathtt{psd}}(M_c) \leq cr \leq 2c \cdot \lceil\sqrt{n}\rceil$.

We now turn to the case that $c \in (2, \sqrt{n}/2)$ and c is not an integer. Firstly, we construct the PSD-factorization for $M_{\lceil c\rceil}$ as above. Then we replace all the nonzero off-diagonal entries of the E_i's (which are 1's) by $a = \frac{c-1}{\lceil c\rceil - 1}$, and obtain E_i''s. Now $\{E_1', \ldots, E_n'\}$ and $\{F_1, \ldots, F_n\}$ form a PSD-factorization for M_c.

Finally, in order to settle the case that $c \in [0,2]$, we first focus on the special case that $c = 2$. It is easy to see that in this case, if r is a positive integer, $p(r,c) = \frac{1}{2}r(r-1)$. Thus if we choose $r = \lceil\sqrt{2n}\rceil + 1$, it holds that $p(r,c) \geq n$, and we have $\mathtt{rank}^{\mathbb{R}}_{\mathtt{psd}}(M_2) \leq \lceil\sqrt{2n}\rceil + 1$. When $c \in [0,2)$, we replace all the nonzero off-diagonal entries of the E_i's (which are 1's) by $c-1$, and obtain E_i''s. It can be verified that $\{E_1', \ldots, E_n'\}$ and $\{F_1, \ldots, F_n\}$ form a valid PSD-factorization for M_c. □

We now consider a more general approximation of the identity than M_c, where the diagonal entries do not have to be 1, and the off-diagonal entries do not have to be equal. Alon [17] proved:

Theorem 16 *([17]) There exists an absolute positive constant c so that the following holds. Let $A = [a(i,j)]$ be an n-by-n real matrix with $|a(i,i)| \geq 1/2$ for all i and $|a(i,j)| \leq \epsilon$ for any $i \neq j$, where $\frac{1}{2\sqrt{n}} \leq \epsilon \leq 1/4$. Then the rank of A satisfies*

$$\mathtt{rank}(A) \geq \frac{c \log n}{\epsilon^2 \log(1/\epsilon)}.$$

Combining the above theorem and Fact 3, we immediately obtain that

Theorem 17 *There exists an absolute positive constant c so that the following holds. Let $A = [a(i,j)]$ be an n-by-n real matrix with $|a(i,i)| \geq 1/2$ for all i and $|a(i,j)| \leq \epsilon$ for any $i \neq j$, where $\frac{1}{2\sqrt{n}} \leq \epsilon \leq 1/4$. Then the PSD-rank of A satisfies*

$$\mathtt{rank}_{\mathtt{psd}}(A) \geq \frac{c\sqrt{\log n}}{\epsilon\sqrt{\log(1/\epsilon)}}.$$

We do not know if this lower bound on PSD-rank is tight. It is not hard to show that there are ε-approximations of the n-by-n identity matrix with $\varepsilon \approx 1/2$ for which the *nonnegative* rank is $O(\log n)$. For example, we can take a set of n random ℓ-bit words $C_1, \ldots, C_n \in \{0,1\}^\ell$. For $\ell = c \log n$ and c a sufficiently large constant, $\langle C_i | C_j\rangle$ will be close to $\ell/2$ for all $i = j$ and close to $\ell/4$ for all $i \neq j$. Hence if we associate both the ith row and the ith column with the ℓ-dimension vector $\sqrt{\frac{2}{\ell}} C_i$, we get an $\ell = O(\log n)$-dimensional nonnegative factorization of an approximation of the identity.

6.3 The inner product function

Let $x, y \in \{0,1\}^n$ be two n-bit strings. The inner product function is defined as $\mathrm{IP}(x,y) = \sum_{i=1}^n x_i y_i \bmod 2$. We denote the corresponding N-by-N matrix by IP_n, where $N = 2^n$. We have the following theorem.

Theorem 18 $\mathtt{rank}^{\mathbb{R}}_{\mathtt{psd}}(\mathrm{IP}_n) \leq c\sqrt{N}$*, where $c = 2$ if n is even, and $c = 2\sqrt{2}$ if n is odd.*

Proof We will design a one-way quantum protocol to compute IP_n in expectation and then invoke the equivalence between $\mathtt{rank}_{\mathtt{psd}}$ and communication complexity mentioned in Section 2.2. We will actually prove the bound for more general 0/1-matrices, of which IP_n is a special case. Let W be an N-by-N 0/1-matrix, with rows and columns indexed by n-bit strings x and y respectively. We first consider the case that n is even. View $x = x_0 x_1$ as concatenation of two $n/2$-bit strings x_0 and x_1. Suppose there exist two Boolean functions $f, g : \{0,1\}^{n/2+n} \to \{0,1\}$ such that $W(x,y) = f(x_0, y) + g(x_1, y) \bmod 2$. Then IP_n is a special case of such a W, where $f(x_0, y) = \text{IP}(x_0, y_0)$ and $g(x_1, y) = \text{IP}(x_1, y_1)$. We now show there exists a one-way quantum protocol that computes W in expectation and whose quantum communication complexity is at most $n/2 + 1$ qubits. This implies $\texttt{rank}_{\texttt{psd}}(W) \leq 2^{n/2+1} = 2\sqrt{N}$.

For any input x, Alice sends the following state of $1 + n/2$ qubits to Bob:

$$|\psi_x\rangle := \frac{1}{\sqrt{2}}(|0, x_0\rangle + |1, x_1\rangle).$$

Then by a unitary operation, Bob turns the state into

$$|\psi_{xy}\rangle := \frac{1}{\sqrt{2}}((-1)^{f(x_0,y)}|0, x_0\rangle + (-1)^{g(x_1,y)}|1, x_1\rangle).$$

Bob then applies the Hadamard gate to the last $n/2$ qubits and measures those in the computational basis. If he gets any outcome other than $0^{n/2}$, he outputs 0. With probability $1/\sqrt{2^n}$ he gets outcome $0^{n/2}$, and then the first qubit will have become $\frac{1}{\sqrt{2}}((-1)^{f(x_0,y)}|0\rangle + (-1)^{g(x_1,y)}|1\rangle)$. By another Hadamard gate and a measurement in the computational basis, Bob learns the bit $f(x_0, y) + g(x_1, y) \bmod 2 = W(x, y)$. Then he outputs that bit times $\sqrt{2^n}$. The expected value of the output is $\frac{1}{\sqrt{2^n}} \cdot (W(x, y) \cdot \sqrt{2^n}) = W(x, y)$.

For the case that n is odd, Alice can make the length of x even by appending the bit 0 to the end of x, and Bob can do the same change to y. Then we go back to the case where the inputs have even length, and the inner product remains unchanged. Now the quantum communication complexity is at most $(n + 1)/2 + 1$ qubits, implying for odd n that $\texttt{rank}_{\texttt{psd}}(W) \leq 2^{(n+1)/2+1} = 2\sqrt{2} \cdot \sqrt{N}$. □

We give another proof of this theorem by explicitly providing a PSD-factorization for IP_n. Note that the factors in the following PSD-factorization are rank-1 real matrices.

Theorem 19 $\texttt{rank}^{\mathbb{R}}_{\texttt{psd}}(\text{IP}_n) \leq c\sqrt{N}$. *If n is even,* $c = 2$, *and if n is odd,* $c = \frac{3}{2}\sqrt{2}$.

Proof For any k we have $\text{IP}_{k+1} = \begin{bmatrix} \text{IP}_k & \text{IP}_k \\ \text{IP}_k & J_k - \text{IP}_k \end{bmatrix}$, where J_k is the k-by-k all-one matrix. Using this relation twice, we have that

$$\text{IP}_{k+2} = \begin{bmatrix} \text{IP}_k & \text{IP}_k & \text{IP}_k & \text{IP}_k \\ \text{IP}_k & J_k - \text{IP}_k & \text{IP}_k & J_k - \text{IP}_k \\ \text{IP}_k & \text{IP}_k & J_k - \text{IP}_k & J_k - \text{IP}_k \\ \text{IP}_k & J_k - \text{IP}_k & J_k - \text{IP}_k & \text{IP}_k \end{bmatrix}.$$

Repeating this procedure, it can be seen that IP_n can be expressed as a block matrix with each block being IP_k or $J_k - \text{IP}_k$ for some $k < n$ to be chosen later. We now consider a new block matrix M_n with the same block configuration as IP_n generated as follows. The blocks in the first block row of M_n are the same as IP_n, that is they are IP_k's. In the rest of the block rows, if a block of IP_n is IP_k, then we choose the corresponding block of M_n to be $-\text{IP}_k$, and if a block of IP_n is $J_k - \text{IP}_k$, the corresponding block of M_n is also $J_k - \text{IP}_k$. It is not difficult to check that $M_n \circ \overline{M}_n = \text{IP}_n$, and since M_n is real, we have that $\texttt{rank}^{\mathbb{R}}_{\texttt{psd}}(\text{IP}_n) \leq \texttt{rank}(M_n)$.

In order to upper bound the rank of M_n, we add its first block row to the other block rows, and obtain another matrix M'_n, with the same rank as M_n, in which all the blocks are 0 or J_k except those in the first row are still IP_k's. Since the rank of M'_n can be upper bounded by the sum of the rank of the first block row and that of the remaining block rows, we have that

$$\texttt{rank}^{\mathbb{R}}_{\texttt{psd}}(\text{IP}_n) \leq \texttt{rank}(M_n) = \texttt{rank}(M'_n) \leq 2^k - 1 + \frac{N}{2^k},$$

where $2^k - 1$ comes from the rank of IP_k, and $\frac{N}{2^k}$ comes from the number of blocks in every row of M'_n. If n is even, we choose $k = n/2$, and the inequality above is $\texttt{rank}^{\mathbb{R}}_{\texttt{psd}}(\text{IP}_n) \leq 2\sqrt{N} - 1$. If n is odd, we choose $k = (n + 1)/2$, and the inequality becomes $\texttt{rank}^{\mathbb{R}}_{\texttt{psd}}(\text{IP}_n) \leq (\frac{3}{2}\sqrt{2})\sqrt{N} - 1$. □

Acknowledgments. We would like to thank Rahul Jain for helpful discussions, and Hamza Fawzi, Richard Robinson, and Rekha Thomas for sharing their results on the derangement matrix. Troy Lee and Zhaohui Wei are supported in part by the Singapore National Research Foundation under NRF RF Award No. NRF-NRFF2013-13. Ronald de Wolf is partially supported by ERC Consolidator Grant QPROGRESS and by the EU STREP project QALGO (Grant agreement no. 600700).

References

1. J. Gouveia, P. Parrilo, and R. Thomas. Lifts of convex sets and cone factorizations. *Mathematics of Operations Research*, 38(2):248–264, 2013. arXiv:1111.3164.
2. S. Fiorini, S. Massar, S. Pokutta, H. R. Tiwary, and R. de Wolf. Exponential lower bounds for polytopes in combinatorial optimization. *Journal of the ACM*, 16(2), 2015. Earlier version in STOC'12. arXiv:1111.0837.
3. M. Yannakakis. Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences*, 43(3):441–466, 1991.
4. T. Rothvoß. The matching polytope has exponential extension complexity. In *Proceedings of 46th ACM STOC*, pages 263–272, 2014.
5. J. Lee, D. Steurer, and P. Raghavendra. Lower bounds on the size of semidefinite programming relaxations. In *Proceedings of 47th ACM STOC*, pages 567–576, 2015. arXiv:1411.6317.
6. T. Lee and Z. Wei. The square root rank of the correlation polytope is exponential, 2014. arXiv:1411.6712.
7. M. Conforti, Y. Faenza, S. Fiorini, and H. R. Tiwary. Extended formulations, non-negative factorizations and randomized communication protocols. In *2nd International Symposium on Combinatorial Optimization*, pages 129–140, 2012. arXiv:1105.4127.
8. S. Zhang. Quantum strategic game theory. In *Proceedings of the 3rd Innovations in Theoretical Computer Science*, pages 39–59, 2012. arXiv:1012.5141.
9. R. Jain, Y. Shi, Z. Wei, and S. Zhang. Efficient protocols for generating bipartite classical distributions and quantum states. *IEEE Transactions on Information Theory*, 59:5171–5178, 2013.
10. E. Kushilevitz and N. Nisan. *Communication complexity*. Cambridge University Press, Cambridge, 1997.
11. T. Lee and D. O. Theis. Support based bounds for positive semidefinite rank, 2012. arXiv:1203.3961.
12. J. Sikora, A. Varvitsiotis, and Z. Wei. On the minimum dimension of a Hilbert space needed to generate a quantum correlation, 2015. arXiv:1507.00213.
13. M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.
14. A. Ben-Tal and A. Nemirovski. On polyhedral approximations of the second-order cone. *Mathematics of Operations Research*, 26(2):193–205, 2001.
15. H. Fawzi and P. Parrilo. Lower bounds on nonnegative rank via nonnegative nuclear norms. *Mathematical Programming, Series B*, 153(1):41–66, 2015. arXiv:1210.6970.
16. H. Fawzi, J. Gouveia, P. Parrilo, R. Robinson, and R. Thomas. Positive semidefinite rank. *Mathematical Programming, Series B*, 153(1):133–177, 2015. arXiv:1407.4095.
17. N. Alon. Perturbed identity matrices have high rank: proof and applications. *Combinatorics, Probability, and Computing*, 18:3–15, 2009.