Exploring the bounds on the positive semidefinite rank

Andrii Riazanov*
andrii.riazanov@gmail.com

Mikhail Vyalyi†
vyalyi@gmail.com

Abstract

The nonnegative and positive semidefinite (PSD-) ranks are closely connected to the nonnegative and positive semidefinite extension complexities of a polytope, which are the minimal dimensions of linear and SDP programs which represent this polytope. Though some exponential lower bounds on the nonnegative [FMP+12] and PSD- [LRS15] ranks has recently been proved for the slack matrices of some particular polytopes, there are still no tight bounds for these quantities. We explore some existing bounds on the PSD-rank and prove that they cannot give exponential lower bounds on the extension complexity. Our approach consists in proving that the existing bounds are upper bounded by the polynomials of the regular rank of the matrix, which is equal to the dimension of the polytope (up to an additive constant). As one of the implications, we also retrieve an upper bound on the mutual information of an arbitrary matrix of a joint distribution, based on its regular rank.

1 Introduction

Linear optimization plays an important role in computer science and mathematics. Though there exist efficient algorithms of linear optimization over convex sets, for the polytopes with exponential number of facets they still work too long in general case. That is why one may want to represent such "hard" convex set as a projection (linear map) of some "easier" convex set, for example of some affine slice of the cone of nonnegative orthant or the cone of positive semidefinite matrices, since on slices of both these cones linear optimization has efficient algorithms. Such representations are called *the nonnegative* and *the positive semidefinite (PSD-) extensions*, respectively.

Since many problems of combinatorial optimization can be represented as linear programs over a polytope, studying the extensions of convex polytopes is an important and challenging problem. The natural question is to find the minimal dimension for which there exists an extension of the given polytope. It can be also formulated as determining the smallest dimensions of LP or SDP programs which represent optimization over the given polytope, and such sizes are called *the nonnegative* and *the semidefinite extension complexities*, respectively.

In the context of P $\neq$ NP we do not expect to find small nonnegative or PSD- extension complexities for NP-hard problems, since that would mean that there exist polynomial algorithms for solving these problems. However, there is still no general approach for proving the lower bounds on these quantities, and only a few exponential lower bounds for some particular problems has recently been proved. All such results use the connection between extension complexity and matrix factorizations, which was first discovered in [Yan91] for the nonnegative extension complexity and nonnegative matrix factorizations. Further, this approach was extended in [GPT13] for the general case of cone factorizations, and the same result for PSD-factorizations was also obtained in [FMP+12]. This instrument gave an opportunity to explore the nonnegative and PSD- extension complexities of polytopes via studying some characteristics of their slack matrices called *the nonnegative* and *the PSD-* ranks. For example, in the 1980s there were attempts to prove P = NP by providing the polynomial-sized linear program to solve the NP-hard travelling salesman problem (TSP). However, using the described approach, Yannakakis proved in [Yan91] that any *symmetric* LP which solves TSP has exponential size, which meant invalidity of all such attempts, since all the presented LPs were symmetric. The extension of this result for *any* (not only symmetric) TSP was first presented in [FMP+12], where the authors used the connection between the nonnegative rank of the matrix and the nondeterministic communication complexity of its support. In this work, the exponential lower bounds on the nonnegative

*Skolkovo Institute of Science and Technology; Moscow Institute of Physics and Technology (State University).

†Dorodnicyn Computing Centre, FRC CSC RAS; Moscow Institute of Physics and Technology (State University); National Research University Higher School of Economics. The study has been funded by the Russian Academic Excellence Project '5-100'.

rank were also proved for CUT and Stable Set polytopes. The first analogical bounds for the PSD-extension complexity were presented in [LRS15] using the sum-of-squares SDP hierarchy.

Since exponential lower bounds were obtained for some particular cases only, it is still a challenging problem to obtain reasonable estimations and bounds for the nonnegative and PSD- ranks. This problem is widely discussed during the last decade. For instance, exponential bounds on the nonnegative rank, and thus on the nonnegative extension complexity, were proved in [Rot14] for the matching polytope , where the author used the extension of Razborov's result [Raz90]. We address the reader to the review [FGP$^+$15] for more details about recent research on the PSD-rank.

There is also a problem of determining the computational complexity of computing the nonnegative and PSD- ranks. Both problems are known to be NP-hard, and recent research [Shi16] shows that the problem of computing the PSD-rank is complete in $\exists\mathbb{R}$ – the existential theory of the reals.

Contribution

In this paper we explore the lower bounds on the PSD-rank introduced in [LWdW16], which we will further address as *bounding functionals (of a matrix)*. We show that these functionals cannot give exponential bounds on the PSD-rank, and thus on the positive semidefinite extension complexity. Our approach consists in proving that the bounding functionals of the slack matrix are bounded above by the polynomial of the regular rank of this matrix and the logarithm of the matrix size. Since for any polytope P we have $\operatorname{rank} S_P = \dim(P) + 1$, it would mean that the bounds are polynomial in the dimension of the polytope.

As one of the implications of our approach, we achieve the upper bound on the mutual information for an arbitrary matrix of a joint distribution. More precisely, we show that the mutual information is bounded above by the logarithm of the rank of the matrix.

Outline of the paper

This paper is organized as follows. In Sect. 2 we introduce all the necessary notations and explain some connections between the PSD-rank and the quantum communication complexity. In Sect. 3 we present the bounding functionals from [LWdW16] and explain how the lower bound on the PSD-rank can be obtained via the mutual information. Finally, in Sect. 4 the upper bounds on the bounding functionals are proved. In particular, Theorem 4.1 shows that the mutual information of two discrete random variables is bounded above by the logarithm of the regular rank of the matrix of their joint distribution.

2 Preliminaries

2.1 Nonnegative and PSD- matrix factorizations

The nonnegative matrix factorization of the nonnegative matrix $A \in \mathbb{R}^{m\times n}$ is the decomposition $A = BC$, where $B \in \mathbb{R}^{m\times k}$, $C \in \mathbb{R}^{k\times n}$, and B, C are nonnegative matrices. Alternatively, such factorization can be thought of as two sets of vectors $\{b_i\}_{i=1}^m,\ \{c_j\}_{j=1}^n,\ b_i, c_j \in \mathbb{R}_+^k$, such that $A(i,j) = \langle b_i, c_j\rangle$. Then *the nonnegative rank* of A, denoted $\operatorname{rank}_+ A$, is the smallest $k \in \mathbb{N}$ for which such nonnegative factorization of A exists.

Similarly, *the positive semidefinite rank* $\operatorname{rank}_{\text{psd}} A$ is the minimal integer r for which there exist two sets of complex Hermitian positive semidefinite matrices $\{B_i\}_{i=1}^m,\ \{C_j\}_{j=1}^n,\ B_i, C_j \in \mathbf{S}_+^r$, such that $A(i,j) = \langle B_i, C_j\rangle = \operatorname{Tr}(B_iC_j)$. Such factorization is called *the positive semidefinite factorization*, and it has many applications in combinatorial optimization and communication complexity. If to restrict the matrices in the factorization to be real symmetric positive semidefinite, one will obtain the definition of *the real PSD-rank* $\operatorname{rank}_{psd}^{\mathbb{R}}$. It can be shown ([LWdW16]), that the restriction for matrices to be real can increase $\operatorname{rank}_{\text{psd}}$ at most by the factor of 2, e.g. $\operatorname{rank}_{\text{psd}} \le \operatorname{rank}_{psd}^{\mathbb{R}} \le 2\operatorname{rank}_{\text{psd}}$. Since in our context we only study asymptotic bounds on the ranks, there is no difference between considering $\operatorname{rank}_{\text{psd}}$ or $\operatorname{rank}_{psd}^{\mathbb{R}}$.

We would like to emphasize that rescaling the nonnegative matrix by multiplying its rows or columns by any positive factors does not change its nonnegative and PSD- ranks. Indeed, multiplication of the i^{th} row of A by α corresponds to the multiplication of b_i by the same factor α in the nonnegative factorization.

Similarly, it corresponds to the multiplication of B_i by α in the PSD-factorization. Obviously, the situation with the columns of A is the same.

2.2 Extension complexity

The nonnegative extension complexity of the polytope P is the smallest number d such that P can be expressed as a projection of an affine slice of the nonnegative d-dimensional orthant $\mathbb{R}^d_+$. Similarly, *the semidefinite (PSD-) extension complexity* of P is the minimum number r for which there exists an affine slice of the cone of complex Hermitian $r \times r$ positive semidefinite matrices $\mathbf{S}^r_+$ that projects onto P.

In other words, for optimizing over some polytope $P \in \mathbb{R}^d$ one may want to represent is as $P = \pi(K \cap L)$, where $K \subseteq \mathbb{R}^n$ is some close convex cone, L is some affine subspace of $\mathbb{R}^n$, and π is a linear map (projection). Such representations are called K*-lifts*, ([GPT13]), or K*-extensions*. If to choose K from the families of the cones of nonnegative orthants $\mathbb{R}^k_+$ or positive semidefinite matrices $\mathbf{S}^r_+$, the nonnegative and PSD- extension complexities for the given polytope correspond to minimal k and r for which such representations exist.

2.3 Factorization theorem

As it was discussed in Introduction, [Yan91], [GPT13], and [FMP+12] proved that the extension complexities and matrix factorizations are interconnected. Here we present the Factorization theorem, which explains the relations between these two notions.

Let P be a polytope in $\mathbb{R}^d$ with n vertices and m facets, thus $P = \{x \in \mathbb{R}^d \mid \langle x, a_j \rangle \leq b_j,\ j \in \overline{1,m}\}$. Then *the slack matrix of the polytope* P is defined as the nonnegative matrix $S_P \in \mathbb{R}^{n\times m}$ with $S_P(i,j) = b_j - \langle v_i, a_j \rangle$, where v_i is the i^{th} vertex of P. Then the Factorization theorem can be formulated as follows:

Factorization Theorem. *The nonnegative extension complexity of P is equal to* $\operatorname{rank}_+ S_P$. *Similarly, the PSD-extension complexity of P is equal to* $\operatorname{rank}_{\mathrm{psd}} S_P$.

This approach allows applying techniques for estimating or bounding such algebraic notions as sizes of matrix factorizations to answer geometrical questions about the complexities of the polytopes.

2.4 Quantum communication complexity

In this section, we describe the connection between the quantum communication complexity and $\operatorname{rank}_{\mathrm{psd}}$. First, we will consider *one-way quantum communication protocol*.

A quantum state ρ is a positive semidefinite matrix with $\operatorname{Tr}\rho = 1$. A measurement $\mathcal{E}$ is the set of positive semidefinite matrices $\{E_i\}_{i\in\Omega}$, indexed by the finite set of nonnegative real numbers Ω, with the condition $\Sigma_{i\in\Omega} E_i = I$. The measurements are also called POVM ("Positive Operator Value Measure") in the literature. POVMs work in the following way: when we apply the measurement $\mathcal{E}$ to the state ρ, the outcome is i with probability $\operatorname{Tr}(E_i \rho)$.

Then the process of communication is set as follows: initially, Alice has the integer x, and Bob has y. Then Alice sends an $r\times r$-dimensional quantum state ρ_x to Bob, who measures it with POVM $\mathcal{E}_y$ and outputs the result. We say that such a protocol computes the nonnegative matrix M in expectation, if the expected value of Bob's output on the input (x, y) is equal to $M(x, y)$ (the entry of the matrix M in x^{th} row and y^{th} column). Then *the quantum communication complexity* of the matrix M is the logarithm of such a minimal size of dimension r, for which there exists a one-way quantum protocol which computes M in expectation.

Fiorini et. al. [FMP+12] and Jain et. al. [JSWZ13] proved that the minimal amount of quantum information needed for Alice and Bob to generate the nonnegative matrix M is completely determined by the PSD-rank of this matrix. More precisely, they showed that the quantum communication complexity of M is equal to $\lceil \log \operatorname{rank}_{\mathrm{psd}} M \rceil$.

3 Bounding functionals on the PSD-rank

In this section, we present some existing general lower bounds on $\operatorname{rank}_{\mathrm{psd}}$ from [LWdW16], which we address as *bounding functionals*. All these functionals are natural lower bounds of the quantum communication complexity, so understanding their power is also interesting for quantum information theory.

Except for the bound via mutual information, the bounding functionals are introduced here without justification. We address the reader to the original article for more details on the bounds. For convenience, we preserve the notations for the bounding functionals from the original article.

3.1 Bound via Mutual Information

If X and Y are two random variables, then *the mutual information* is defined as follows:

$$I(A:B) = H(A) + H(B) - H(A,B) = H(A) - H(A|B) = H(B) - H(B|A),$$

where H is Shannon entropy. The mutual information can be interpreted as the number of bits of information about A that are revealed by the value of B. We will now use Holevo's theorem [Wat11] to bound the mutual information. It claims that the number of classical bits of information that Alice can communicate to Bob by sending n qubits does not exceed n. From the previous passage we know that we need exactly $\lceil \log \operatorname{rank}_{\mathrm{psd}} M \rceil$ qubits of information to compute the matrix M. Normalizing M and considering it as a matrix of joint distribution $\mathbb{P}(A,B)$, we then have:

Fact 3.1. *Let M be a matrix of a joint distribution of two discrete random variables A, B with finite support, $M(a,b) = \mathbb{P}[B = b, A = a]$. Then*

$$\operatorname{rank}_{\mathrm{psd}} M \geq B_2(P) = 2^{I(A:B)}.$$

3.2 Bounding functionals from [LWdW16]

For two probability distributions $p = \{p_i\}_{i=1}^n$ and $q = \{q_i\}_{i=1}^n$ fidelity is defined as $F(p,q) = \Sigma_{i=1}^n \sqrt{p_i q_i}$.

Recall that *the left stochastic matrix* is the matrix with nonnegative entries, with each column summing to 1. Further in the text we will omit "left" and just use the term "stochastic matrix" instead.

Then we have the following lower bounds:

Fact 3.2. *Let $M \in \mathbb{R}^{n \times m}$ be a stochastic matrix. Then*

$$\operatorname{rank}_{\mathrm{psd}} M \geq B_3(M) = \max_{\{q_i\}_{i=1}^m} \frac{1}{\sum_{i,j=1}^m q_i q_j F(M_i, M_j)^2}$$

where the max *is taken over all probability distributions $q = \{q_i\}_{i=1}^m$, and M_i is the i^{th} column of M.*

Fact 3.3. *Let $M \in \mathbb{R}^{n \times m}$ be a stochastic matrix. Then*

$$\operatorname{rank}_{\mathrm{psd}} M \geq B_4(M) = \sum_{i=1}^n \max_j M(i,j).$$

Fact 3.4. *Let $M \in \mathbb{R}^{n \times m}$ be a stochastic matrix. Then*

$$\operatorname{rank}_{\mathrm{psd}} M \geq B_5(M) = \sum_{i=1}^n \max_{\{q_j^{(i)}\}_{j=1}^m} \frac{\sum_{k=1}^m q_k^{(i)} M(i,k)}{\sqrt{\sum_{s,t=1}^m q_s^{(i)} q_t^{(i)} F(M_s, M_t)^2}}$$

where the max *is taken over all probability distributions $q^{(i)} = \{q_j^{(i)}\}_{j=1}^m$, and M_i is the i^{th} column of M.*

4 Upper bounds on the bounding functionals

All the bounds from section 3 were explored and compared in [LWdW16]. It turned out that in different cases B_2, B_3, B_4, or B_5 can give better bounds on $\operatorname{rank}_{\mathrm{psd}}$ than others, and some of them can be tight in some particular cases. However, the key question of whether these functions can give exponential lower bounds on the PSD-rank with respect to the regular rank was not addressed. In this section we answer this question negatively.

In the context of combinatorial optimization, we would like to show that for the polytope of some NP-hard problem the semidefinite extension complexity is exponential in the dimension. Following the arguments from Section 2.2, it suffices to show that the PSD-rank of the corresponding slack matrix is exponential. It is easy to show ([GGK+13]) that the regular rank of the slack matrix equals to the dimension of the polytope plus one: $\operatorname{rank} S_P = \dim P + 1$. For all the presented bounding functionals we provide the upper bounds polynomial in the regular rank of the matrix and the logarithm of the matrix size, which means that they cannot be exponential in the dimension.

4.1 Row elimination transformation

We will now describe the row elimination transformation, which will be used for proving the required bounds.

Let $M \in \mathbb{R}^{n \times m}$ be a nonnegative matrix with $\operatorname{rank} M = r < n$. Without loss of generality, assume that first $r+1$ rows $\overline{m_1}, \overline{m_2} \ldots, \overline{m_{r+1}}$ are non-zero. They are linearly dependent, so there exists a nontrivial set of real numbers $\{\alpha_i\}_{i=1}^{r+1}$, such that $\sum_{i=1}^{r+1} \alpha_i \overline{m_i} = \overline{0}$. Since all entries of M are nonnegative, there are both negative and positive numbers among $\{\alpha_i\}_{i=1}^{r+1}$. For such a set of real numbers $\{\alpha_i\}$ we denote by Δ_α the closed interval $\Delta_\alpha = \left[-\frac{1}{\max_i \alpha_i}, -\frac{1}{\min_i \alpha_i}\right]$, which is properly defined due to the last remark.

Then we define the matrix M^ε as follows: for $1 \le i \le (r+1)$ the i-th row of M^ε equals $\overline{m_i}(1+\varepsilon\alpha_i)$, for $i > (r+1)$ the i-th row of M^ε coincides with the i-th row of M. We call the matrix M^ε *ε-transformation* of M.

First of all, note that $\quad (1+\varepsilon\alpha_i) \ge 0 \quad \forall i \in \overline{1,(r+1)} \quad \Leftrightarrow \quad \varepsilon \in \Delta_\alpha$. Moreover, it holds that when ε is equal to one of the ends of Δ_α, at least one of the coefficients $(1+\varepsilon\alpha_i)$ is equal to zero. It means that for $\varepsilon \in \Delta_\alpha$ the matrix M^ε is nonnegative matrix, and when ε is either the left or the right end of Δ_α, M^ε has more zero rows than M.

Next, we prove that sums of columns do not change after row elimination transformation. Indeed,

$$\sum_{i=1}^{n} m^\varepsilon_{ij} = \sum_{i=1}^{r+1} m^\varepsilon_{ij} + \sum_{i=r+2}^{n} m_{ij} = \sum_{i=1}^{r+1} m_{ij}(1+\varepsilon\alpha_i) + \sum_{i=r+2}^{n} m_{ij} = \sum_{i=1}^{n} m_{ij} + \varepsilon \underbrace{\sum_{i=1}^{r+1} \alpha_i m_{ij}}_{0} = \sum_{i=1}^{n} m_{ij}.$$

In particular, it means that if M is stochastic, then for $\varepsilon \in \Delta_\alpha$ M^ε is also stochastic. Similarly, if M is a matrix of a joint distribution, then M^ε is also a matrix of some joint distribution for ε from Δ_α.

4.2 Upper bound on B_2 (Mutual Information)

Let $M \in \mathbb{R}^{n \times m}$ be the matrix of a joint distribution of two discrete random variables X, Y :

$$m_{ij} = \mathbb{P}\left[X = x_i, Y = y_j\right] \ge 0, \qquad \sum_{i=1,j=1}^{n,m} m_{ij} = 1.$$

Let $p_i,\ i \in \overline{1,n}$, and $q_j,\ j \in \overline{1,m}$, be the marginal probabilities of X and Y respectively:

$$p_i = \mathbb{P}\left[X = x_i\right] = \sum_{j=1}^{m} m_{ij},\ i \in \overline{1,n}; \qquad q_j = \mathbb{P}\left[Y = y_j\right] = \sum_{i=1}^{n} m_{ij},\ j \in \overline{1,m}.$$

Then the mutual information between X and Y can also be defined as:

$$I(X:Y) = D_{KL}\left(p(X,Y) \,||\, p(X)p(Y)\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_j) \log_2\left(\frac{p(x_i,y_j)}{p(x_i)p(y_j)}\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} m_{ij} \log_2\left(\frac{m_{ij}}{p_i q_j}\right),$$

where we set $0 \log \frac{0}{q} = 0$ (the logarithm here and further is to the base 2). We also denote $I(M) = I(X:Y)$.

Theorem 4.1. *Let $M \in \mathbb{R}^{n \times m}$ be the matrix of a joint distribution of X and Y. Then*

$$B_2(M) = 2^{I(X:Y)} \le \operatorname{rank} M.$$

Proof. Denote $r = \operatorname{rank} M$. We will now transform the original matrix M in such a way, that the mutual information will not decrease, but the new matrix $\widetilde{M}$ will have at most r non-zero rows.

Suppose M has more than r non-zero rows. Then we apply the row elimination transformtaion and consider the ε-transformation M^ε of the original matrix. Since we have already shown that it is also a matrix of some joint distribution, we explore how the mutual information changes after such transformations.

First, since the ε-transformation does not change the sums in the columns of M, we have $q_j^\varepsilon = q_j$. Then, since p_i^ε is the sum of entries in the i-th row, we obtain $p_i^\varepsilon = p_i(1 + \varepsilon\alpha_i)$.

Note that since M and M^ε coincide on rows with indexes larger than $r+1$, we may omit the summation over these rows:

$$I(M^\varepsilon) - I(M) = \sum_{i=1}^{r+1}\sum_{j=1}^{s}\left[m_{ij}^\varepsilon \log\left(\frac{m_{ij}^\varepsilon}{p_i^\varepsilon q_j^\varepsilon}\right) - m_{ij}\log\left(\frac{m_{ij}}{p_i q_j}\right)\right] =$$

$$= \sum_{i=1}^{r+1}\sum_{j=1}^{s}\left[m_{ij}(1+\varepsilon\alpha_i)\log\left(\frac{m_{ij}\cancel{(1+\varepsilon\alpha_i)}}{p_i\cancel{(1+\varepsilon\alpha_i)}q_j}\right) - m_{ij}\log\left(\frac{m_{ij}}{p_i q_j}\right)\right] =$$

$$= \sum_{i=1}^{r+1}\sum_{j=1}^{s}\left[\varepsilon\alpha_i m_{ij}\log\left(\frac{m_{ij}}{p_i q_j}\right)\right] = \varepsilon\cdot\Lambda.$$

Now recall that the ε-transformation is valid for $\varepsilon \in \Delta_\alpha$, where the left end of Δ_α is negative, and the right end is positive. It means that we can choose an end of the interval of Δ_α such that $I(M^\varepsilon) \geq I(M)$. It only remains to note that with the chosen value of ε at least one of the first $(r+1)$ rows in M^ε becomes zero.

To get an upper bound on the mutual information, we apply ε-transformations with such suitable ε's that the number of non-zero rows strictly decreases and the mutual information does not decrease. At the end of such procedure we obtain the matrix $\widetilde{M}$ with at most r non-zero rows for which $I(M) \leq I(\widetilde{M})$. Since $\widetilde{M}$ is the matrix of joint distribution, we have $I(\widetilde{M}) = I(\widetilde{X} : \widetilde{Y})$, where the support of $\widetilde{X}$ has cardinality at most r. Using the equality $I(\widetilde{X} : \widetilde{Y}) = H(\widetilde{X}) - H(\widetilde{X}|\widetilde{Y})$ and the non-negativity of the conditional entropy, we finally have:

$$I(M) \leq I(\widetilde{M}) = I(\widetilde{X} : \widetilde{Y}) \leq H(\widetilde{X}) \leq \log|\operatorname{supp}(\widetilde{X})| \leq \log r.$$

□

4.3 Upper bound on B_3

We will show that $B_3(M)$ is upper bounded by *poly*$(\operatorname{rank}(M), \ln m)$:

Theorem 4.2. *Let $M \in \mathbb{R}^{n\times m}$ be a stochastic matrix,* $\operatorname{rank} M = r$*. Then*

$$B_3(M) \leq (\ln m + 1)^2 r^2.$$

We start with proving the following well-known fact:

Lemma 4.1. *For distributions p, q it holds $F(p,q) \geq 1 - \frac{|p-q|}{2}$, where $|p-q|$ is l_1−norm of the vector $(p-q)$, and thus $\frac{|p-q|}{2}$ is the statistical distance between the distributions.*

Proof.

$$1 - \sum_{k=1}^{m}\sqrt{p_k q_k} = \frac{1}{2}\left(\sum p_k + \sum q_k - 2\sum\sqrt{p_k q_k}\right) = \frac{1}{2}\sum|\sqrt{p_k} - \sqrt{q_k}|^2 \leq \frac{1}{2}\sum|p_k - q_k|$$

$$\Rightarrow F(p,q) = \sum_{k=1}^{m}\sqrt{p_k q_k} \geq 1 - \frac{1}{2}\sum|p_k - q_k| = 1 - \frac{|p-q|}{2}.$$

□

Now, we have

$$B_3(M) = \max_{\{q_i\}_{i=1}^m} \frac{1}{\sum_{i,j} q_i q_j F(M_i, M_j)^2} = \frac{1}{\min\limits_{\{q_i\}_{i=1}^m} \sum_{i,j} q_i q_j F(M_i, M_j)^2}. \tag{1}$$

Then we need to prove the lower bound on $\min\limits_{q \in \Delta_m} \sum_{i,j} q_i q_j F(M_i, M_j)^2$.

We will find the lower bound on this quadratic form for an arbitrary distribution q. Without loss of generality, assume $q_1 \geq q_2 \geq \cdots \geq q_n$.

Lemma 4.2. *There exists $s \in \overline{1, m}$ such that $sq_s \geq \frac{1}{\ln m + 1}$.*

Proof. Suppose the opposite: $sq_s \leq \frac{1}{\ln m+1}$ $\forall s \in \overline{1, m}$. Then

$$1 = q_1 + q_2 + \cdots + q_m \leq \frac{1}{\ln m + 1} + \frac{1}{2\,(\ln m + 1)} + \cdots + \frac{1}{m\,(\ln m + 1)} =$$

$$= \frac{1}{\ln m + 1}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{m}\right) < \frac{1}{\ln m + 1}\left(1 + \int_1^m \frac{1}{x} dx\right) = 1.$$

□

Then we have

$$\sum_{i,j=1}^{m} q_i q_j F(M_i, M_j)^2 \geq \sum_{i,j=1}^{s} q_i q_j F(M_i, M_j)^2 \geq \sum_{i,j=1}^{s} q_s^2 F(M_i, M_j)^2 =$$

$$= s^2 q_s^2 \cdot \frac{\sum\limits_{i,j=1}^{s} F(M_i, M_j)^2}{s^2} \geq \frac{1}{(\ln m + 1)^2} \cdot \frac{\sum\limits_{i,j=1}^{s} F(M_i, M_j)^2}{s^2} \tag{2}$$

Now, using the RMS-AM inequality and Lemma 4.1, we get:

$$\frac{\sum\limits_{i,j=1}^{s} F(M_i, M_j)^2}{s^2} \geq \left(\frac{\sum\limits_{i,j=1}^{s} F(M_i, M_j)}{s^2}\right)^2 \geq \left(\frac{\sum\limits_{i,j=1}^{s}\left(1 - \frac{|M_i - M_j|}{2}\right)}{s^2}\right)^2 = \left(1 - \frac{\frac{1}{2}\sum\limits_{i,j=1}^{s} |M_i - M_j|}{s^2}\right)^2 \tag{3}$$

For any stochastic matrix $M \in \mathbb{R}^{n \times m}$ denote $S(M) = \frac{\frac{1}{2}\sum\limits_{i,j=1}^{m} |M_i - M_j|}{m^2}$ – the arithmetic mean of statistical distances between m columns of M. It now suffices to show the upper bound on $S(M)$.

Lemma 4.3. *Let $M \in \mathbb{R}^{n \times m}$ be a stochastic matrix with* $\operatorname{rank}(M) = r$*. Then there exists a stochastic matrix $\widetilde{M} \in \mathbb{R}^{r \times m}$ such that $S(M) \leq S(\widetilde{M})$.*

Proof. We apply the row elimination algorithm. Suppose M has more then r non-zero rows. Consider then the ε-transformation M^ε of the original matrix. Since the ε-transformation does not change the sums of entries in every column of the matrix, M^ε is also stochastic. We now explore how $S(M)$ changes after the ε-transformation:

$$
\begin{aligned}
S(M^\varepsilon) - S(M) &= \frac{1}{2m^2}\left(\sum_{i,j=1}^{m}\left(|M_i^\varepsilon - M_j^\varepsilon| - |M_i - M_j|\right)\right) = \\
&= \frac{1}{2m^2}\left(\sum_{k=1}^{n}\left[\sum_{i,j=1}^{m}\left(|m_{ki}^\varepsilon - m_{kj}^\varepsilon| - |m_{ki} - m_{kj}|\right)\right]\right) = \\
&= \frac{1}{2m^2}\left(\sum_{k=1}^{r+1}\left[\sum_{i,j=1}^{m}\left(|m_{ki} - m_{kj}|(1+\varepsilon\alpha_k) - |m_{ki} - m_{kj}|\right)\right]\right) = \\
&= \frac{1}{2m^2}\left(\sum_{k=1}^{r+1}\left[\sum_{i,j=1}^{m}|m_{ki} - m_{kj}|\varepsilon\alpha_k\right]\right) = \varepsilon\cdot\Lambda.
\end{aligned}
$$

So, the difference $S(M^\varepsilon) - S(M)$ is linear in terms of ε. Remind again that the ε-transformation is valid for $\varepsilon \in \Delta_\alpha$, where the left end of Δ_α is negative, and the right end is positive. It means that we can choose an end of the interval of Δ_α such that $S(M^\varepsilon) \geq S(M)$ and with the chosen value of ε at least one of the first $(r+1)$ rows in M^ε becomes zero. When we apply such ε-transformations with suitable ε's, the number of non-zero rows strictly decreases, and $S(M)$ does not decrease. At the end of such procedure we will obtain the matrix $\widetilde{M}$ with at most r non-zero rows for which $S(M) \leq S(\widetilde{M})$.

□

Lemma 4.4. *Let $M \in \mathbb{R}^{r\times m}$ be a stochastic matrix. Then*

$$S(M) \leq 1 - \frac{1}{r}.$$

Proof. If $m \leq r$, then $\dfrac{\frac{1}{2}\sum_{i,j=1}^{m}|M_i - M_j|}{m^2} \leq \dfrac{m^2 - m}{m^2} = 1 - \dfrac{1}{m} \leq 1 - \dfrac{1}{r}$, where we just used $|M_i - M_j| \leq 2$.

Now suppose $m > r$. Denote $Z(M) = \frac{1}{2}\sum_{i,j=1}^{m}|M_i - M_j| = \frac{1}{2}\sum_{k=1}^{r}\sum_{i,j=1}^{m}|m_{ki} - m_{kj}|$.

We now construct the matrix B by sorting every row of M. Obviously, $Z(M) = Z(B)$, since it is just a permutation of terms. Then

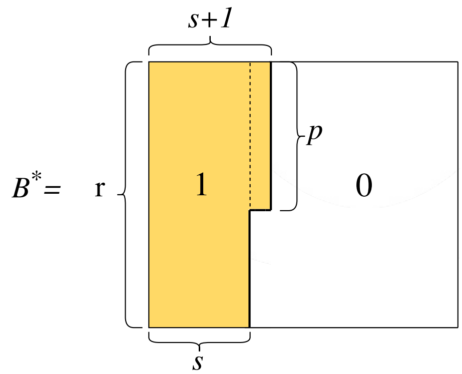
$$Z(M) = Z(B) = \frac{1}{2}\sum_{k=1}^{r}\sum_{i,j=1}^{m}|b_{ki} - b_{kj}| = \sum_{k=1}^{r}\sum_{i=1}^{m}\sum_{j=i}^{m}(b_{ki} - b_{kj}).$$

For each b_{ki} in this sum it occurs $(m-i)$ times with the sign $(+1)$ and $(i-1)$ times with the sign (-1). Hence,

$$
\begin{aligned}
Z(M) &= \sum_{k=1}^{r}\left((m-1)b_{k1} + (m-3)b_{k2} + \cdots - (m-3)b_{k(m-1)} - (m-1)b_{km}\right) = \\
&= (m-1)\sum_{k=1}^{r}b_{k1} + (m-3)\sum_{k=1}^{r}b_{k2} + \cdots - (m-3)\sum_{k=1}^{r}b_{k(m-1)} - (m-1)\sum_{k=1}^{r}b_{km} \qquad (4)
\end{aligned}
$$

Clearly, $Z(M)$ takes its maximal value when the sum in the first columns of B is maximal. Since $b_{ki} \leq 1$ and the sums of all the entries in B and M coincide and are equal to m, to maximize $Z(M)$ we need to have m ones in total in the first columns of B. Denote $m = sr + p,\ p < r$. If $r = 1$, then the matrix M consists of ones only (since it is stochastic), then $S(M) = 0$ and the inequality in the lemma is obvious. If $r > 1$, then it is easy to show that $(s+1) \leq \lceil \frac{m}{2} \rceil$. Note that exactly first $\lceil \frac{m}{2} \rceil$ summands are nonnegative in (4), so to maximize $Z(M)$ first $(s+1)$ columns of B should be filled with ones:

$$B^* = \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & 0 \\ \vdots & \vdots & \cdots & 1 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & 0 \\ 1 & 1 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$



Such matrix B^* would correspond to the following matrix M^*:

$$M^* = \begin{pmatrix} 1 & & & & & 1 & & & & & & 1 & & \\ & 1 & & & & & 1 & & & & \cdots & & \ddots & \\ & & \ddots & & & & & \ddots & & & \cdots & & & 1 \\ & & & \ddots & & & & & \ddots & & \cdots & & & 0 \\ & & & & 1 & & & & & 1 & & & & 0 \end{pmatrix}.$$

$$Z(M^*) = r\left((m-1)+(m-3)+\cdots+(m+1-2s)\right)+p(m-1-2s) = r(m-s)s+pm-p-2sp =$$
$$= m(sr+p)-rs^2-p-2sp = m^2 - \frac{(rs)^2+2rsp+pr}{r} \leq m^2 - \frac{(rs)^2+2rsp+p^2}{r} = m^2\left(1-\frac{1}{r}\right).$$

Then

$$S(M) \leq S(M^*) = \frac{Z(M^*)}{m^2} \leq \left(1-\frac{1}{r}\right).$$

□

Proof of Theorem 4.2. The first s columns of M form the matrix $M' \in \mathbb{R}^{n\times s}$ with $\operatorname{rank}(M') = r' \leq r$. Using Lemma 4.3, we conclude that there exists $\widetilde{M'} \in \mathbb{R}^{r'\times s}$ such that $S(M') \leq S(\widetilde{M'})$. Applying Lemma 4.4 we get $S(M') \leq S(\widetilde{M'}) \leq \left(1-\frac{1}{r'}\right) \leq \left(1-\frac{1}{r}\right)$. Then from (3):

$$\frac{\sum\limits_{i,j=1}^{s,s} F(M_i, M_j)^2}{s^2} \geq \frac{1}{r^2}.$$

Then from (2) for every distribution q we obtain:

$$\sum_{i,j}^{m} q_i q_j F(M_i, M_j)^2 \geq \frac{1}{(\ln m+1)^2 r^2}. \tag{5}$$

And finally, using (1),

$$B_3(P) \leq (\ln m+1)^2 r^2.$$

□

4.4 Upper bound on B_4

Theorem 4.3. *Let $M \in \mathbb{R}^{n\times m}$ be a stochastic matrix,* $\operatorname{rank} M = r$. *Then*

$$B_4(M) \le r. \tag{6}$$

Proof. Again, we apply the row elimination transformation. Note that since every row in the matrix M after this transformation is either multiplied by some nonnegative factor α or remains unchanged, the maximal element in this row is, obviously, multiplied by the same factor α or remains constant as well.

Suppose M has at least $r+1$ non-zero rows, and without loss of generality, suppose that these are the first $r+1$ rows of M. Now consider the ε-transformation M^ε of M, and explore how the functional B_4 changes after such transformation, taking the last remark into consideration:

$$B_4(M^\varepsilon) - B_4(M) = \sum_{i=1}^{n}\left(\max_j M^\varepsilon(i,j) - \max_j M(i,j)\right) = \sum_{i=1}^{n}\left((1+\alpha_i\varepsilon)\max_j M(i,j) - \max_j M(i,j)\right) =$$

$$= \sum_{i=1}^{n}\left(\alpha_i\varepsilon \max_j M(i,j)\right) = \varepsilon\cdot\Lambda.$$

Similarly to previous proofs, B_4 is linear in terms of ε, and therefore when ε equals one of the ends of Δ_α, the difference between $B_4(M^\varepsilon)$ and $B_4(M)$ is nonnegative, while M^ε has strictly less non-zero rows, then M. Again, applying such transformations with suitable ε's, at the end we obtain the matrix $\widetilde{M}$ with at most r non-zero rows, for which $B_4(M) \le B_4(\widetilde{M})$. It only remains to note that in the formula for $B_4(\widetilde{M})$ there are at most r non-zero summands, each less or equal than 1 (since $\widetilde{M}$ is also stochastic). Therefore, we have $B_4(M) \le B_4(\widetilde{M}) \le r$.

□

4.5 Upper bound on B_5

Theorem 4.4. *Let $M \in \mathbb{R}^{n\times m}$ be a stochastic matrix,* $\operatorname{rank} M = r$. *Then*

$$B_5(M) \le r^2(\ln m + 1).$$

Proof. Simply applying (5) and (6) , we get:

$$B_5(M) \le \sum_{i=1}^{n}\max_{\{q_j^{(i)}\}_{j=1}^m}\left((\ln m+1)r\sum_{k=1}^{m} q_k^{(i)} M(i,k)\right) = (\ln m+1)r\sum_{i=1}^{n}\max_k M(i,k) \le (\ln m+1)r^2.$$

The last inequality is due to Theorem 4.3. □

5 Discussion

The results presented here are rather negative: we proved that several specific functionals cannot be used for retrieving superpolynomial lower bounds on the PSD-extension complexity. However, our analysis shows several directions of future research:

- The bounding functionals B_3, B_4, and B_5 seem to be too topic-specific, so it is unlikely they can be directly used in another context. But the mutual information (B_2) is the well-known and relatively widely used notion in the Information Theory and Communication Complexity. Thus, the bound on the mutual information (4.1) might have applications in these fields.
- In the paper we showed that the functionals $B_2(M)$, $S(M)$, and $B_4(M)$ are linear during row elimination transformations, though they have completely nothing in common. It indicates an importance of these transformations. There are several interesting questions about them:

– An interesting question is to explore which other functionals are also linear during such transformations. Another assumption is that, in fact, one may use row elimination transformation to obtain new lower bounds for the PSD-rank. For example, it might be achieved if one will find functional, which is concave relatively to all such transformations.

– The PSD-rank itself is constant on the segment defined by a row elimination transformation except the extreme points. So, the equivalence relation defined by moves inside the segments is worth to study. It gives a possibility to apply implicit lower bounds of the PSD-rank obtained via quantifier elimination theory (see [FGP+15] for a brief introduction) to explicit matrices.

– In the extreme points of the segments, the PSD-rank might diminish drastically. How large is the difference of the PSD-ranks inside a segment and at the extreme points? An answer on this question might give a useful tool to evaluate the PSD-rank.

References

[FGP+15] Hamza Fawzi, João Gouveia, Pablo A. Parrilo, Richard Z. Robinson, and Rekha R. Thomas. Positive semidefinite rank. *Mathematical Programming*, 153(1):133–177, Jul 2015.

[FMP+12] Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans Raj Tiwary, and Ronald de Wolf. Linear vs. semidefinite extended formulations. In *Proceedings of the 44th symposium on Theory of Computing - STOC'12*. Association for Computing Machinery (ACM), 2012.

[GGK+13] João Gouveia, Roland Grappe, Volker Kaibel, Kanstantsin Pashkovich, Richard Z. Robinson, and Rekha R. Thomas. Which nonnegative matrices are slack matrices? *Linear Algebra and its Applications*, 439(10):2921–2933, nov 2013.

[GPT13] João Gouveia, Pablo A. Parrilo, and Rekha R. Thomas. Lifts of convex sets and cone factorizations. *Mathematics of Operations Research*, 38(2):248–264, May 2013.

[JSWZ13] Rahul Jain, Yaoyun Shi, Zhaohui Wei, and Shengyu Zhang. Efficient protocols for generating bipartite classical distributions and quantum states. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1503–1512. Society for Industrial & Applied Mathematics (SIAM), Jan 2013.

[LRS15] James R. Lee, Prasad Raghavendra, and David Steurer. Lower bounds on the size of semidefinite programming relaxations. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing - STOC'15*. Association for Computing Machinery (ACM), 2015.

[LWdW16] Troy Lee, Zhaohui Wei, and Ronald de Wolf. Some upper and lower bounds on PSD-rank. *Mathematical Programming*, 162(1-2):495–521, Jul 2016.

[Raz90] A. A. Razborov. On the distributional complexity of disjointness. In *Automata, Languages and Programming*, pages 249–253. Springer Nature, 1990.

[Rot14] Thomas Rothvoss. The matching polytope has exponential extension complexity. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing - STOC'14*. Association for Computing Machinery (ACM), 2014.

[Shi16] Yaroslav Shitov. The complexity of positive semidefinite matrix factorization. 11(1), Jun 2016.

[Wat11] John Watrous. Lecture notes in theory of quantum information, 2011.

[Yan91] Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences*, 43(3):441–466, Dec 1991.