Massachusetts Institute of Technology

Lower bounds on nonnegative rank
via nonnegative nuclear norms

Hamza Fawzi   Pablo A. Parrilo

January 28th, 2015

Abstract

The nonnegative rank of an entrywise nonnegative matrix $A \in \mathbb{R}_+^{m \times n}$ is the smallest integer r such that A can be written as $A = UV$ where $U \in \mathbb{R}_+^{m \times r}$ and $V \in \mathbb{R}_+^{r \times n}$ are both nonnegative. The nonnegative rank arises in different areas such as combinatorial optimization and communication complexity. Computing this quantity is NP-hard in general and it is thus important to find efficient bounding techniques especially in the context of the aforementioned applications.

In this paper we propose a new lower bound on the nonnegative rank which, unlike most existing lower bounds, does not solely rely on the matrix sparsity pattern and applies to nonnegative matrices with arbitrary support. The idea involves computing a certain nuclear norm with nonnegativity constraints which allows to lower bound the nonnegative rank, in the same way the standard nuclear norm gives lower bounds on the standard rank. Our lower bound is expressed as the solution of a copositive programming problem and can be relaxed to obtain polynomial-time computable lower bounds using semidefinite programming. We compare our lower bound with existing ones, and we show examples of matrices where our lower bound performs better than currently known ones.

1   Introduction

Given a nonnegative[1] matrix $A \in \mathbb{R}_+^{m \times n}$, the *nonnegative rank* of A is the smallest integer r such that A can be factorized as $A = UV$ where $U \in \mathbb{R}_+^{m \times r}$ and $V \in \mathbb{R}_+^{r \times n}$ are both nonnegative. The nonnegative rank of A is denoted by $\text{rank}_+(A)$ and it always satisfies:

$$\text{rank}(A) \leq \text{rank}_+(A) \leq \min(n, m).$$

The nonnegative rank appears in different areas such as in combinatorial optimization [Yan91] and communication complexity [KN06, Lov90, LS09]. Indeed, in combinatorial optimization a well-known result by Yannakakis [Yan91] shows that the nonnegative rank of

---

[1]Throughout the paper, a nonnegative matrix is a matrix whose entries are all nonnegative.

a suitable matrix characterizes the smallest number of linear inequalities needed to represent a given polytope. This quantity is very important in practice since the complexity of interior-point algorithms for linear programming explicitly depends on the number of linear inequalities. Another application of the nonnegative rank is in communication complexity where one is interested in the minimum number of bits that need to be exchanged between two parties in order to compute a binary function $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$, assuming that initially each party holds only one of the two arguments of the function. This quantity is known as the communication complexity of f and is tightly related to the nonnegative rank of the $|\mathcal{X}| \times |\mathcal{Y}|$ matrix M_f associated to f defined by $M_f(x, y) = f(x, y)$ [LS09, Lov90]. Finally it was also recently observed [Zha12, JSWZ13] that the logarithm of the nonnegative rank of a matrix A coincides with the minimum number of bits that need to be exchanged between two parties in order to sample from the bivariate probability distribution $p(x, y) = A_{x,y}$ represented by the matrix A (assuming A is normalized so that $\sum_{x,y} A_{x,y} = 1$).

Lower bounds on nonnegative rank Unfortunately, the nonnegative rank is hard to compute in general unlike the standard rank: For example it was shown in [Vav09] that the problem of deciding whether $\text{rank}_+(A) = \text{rank}(A)$ is NP-hard in general (see also [AGKM12] for further hardness results). Researchers have therefore developed techniques to find lower and upper bounds for this quantity. Existing lower bounds are not entirely satisfactory though, since in general they depend only on the sparsity pattern of the matrix A and not on the actual values of the entries, and as a consequence these bounds cannot be used when all the entries of A are strictly positive. In fact most of the currently known lower bounds are actually lower bounds on the *rectangle covering number* (also known as the *Boolean rank*) which is a purely combinatorial notion of rank that is itself a lower bound on the nonnegative rank. The *rectangle covering number* of a nonnegative matrix $A \in \mathbb{R}^{m \times n}_+$ is the smallest number of rectangles needed to cover the nonzero entries of A. More precisely it is the smallest integer r such that there exist r rectangles $R_i = I_i \times J_i \subseteq \text{support}(A)$ for $i = 1, \ldots, r$ such that

$$\text{support}(A) = \bigcup_{i=1}^{r} R_i,$$

where $\text{support}(A) = \{(i, j) \in \{1, \ldots, m\} \times \{1, \ldots, n\} : A_{i,j} \neq 0\}$ is the set of indices of the nonzero entries of A. It is easy to see that the rectangle covering number is always smaller than or equal to the nonnegative rank of A (each nonnegative rank 1 term in a decomposition of A corresponds to a rectangle). Some of the well-known lower bounds on the nonnegative rank such as the fooling set method or the rectangle size method (see [KN06, Section 1.3]) are in fact lower bounds on the rectangle covering number and they only depend on the sparsity pattern of A and not on the specific values of its entries. Recently, a new lower bound was proposed in [GG12] that does not rely solely on the sparsity pattern; however the lower bound depends on a new quantity called the *restricted nonnegative rank* which can be computed efficiently only when $\text{rank}(A) \leq 3$ but otherwise is NP-hard to compute in general. Also a non-combinatorial lower bound on the nonnegative rank known as the *hyperplane separation* bound was proposed recently and used in the remarkable result of Rothvoss [Rot14] on the matching polytope. The lower bound we propose in this paper has a similar flavor as the hyperplane separation bound except that it uses the Frobenius

norm instead of the entrywise infinity norm. Also one focus of the present paper is on computational approaches to compute the lower bound using sum-of-squares techniques and semidefinite programming. Finally, after the initial version of this paper was submitted, we extended some of the ideas presented here and we proposed in [FP14] new lower bounds that are invariant under scaling and that are related to hyperplane separation bounds and combinatorial bounds.

Contribution   In this paper we present an efficiently computable lower bound on the nonnegative rank that does not rely exclusively on the sparsity pattern and that is applicable to matrices that are strictly positive. Before we present our result, recall that a symmetric matrix $M \in \mathbb{R}^{n \times n}$ is said to be *copositive* if $x^\mathsf{T} M x \geq 0$ for all $x \in \mathbb{R}^n_+$. Our result can be summarized in the following:

Theorem 1. *Let $A \in \mathbb{R}^{m \times n}_+$ be a nonnegative matrix. Let $\nu_+(A)$ be the optimal value of the following convex optimization program:*

$$\nu_+(A) = \max_{W \in \mathbb{R}^{m \times n}} \left\{ \langle A, W \rangle \ : \ \begin{bmatrix} I & -W \\ -W^\mathsf{T} & I \end{bmatrix} \ copositive \right\}. \tag{1}$$

Then we have

$$\mathrm{rank}_+(A) \geq \left(\frac{\nu_+(A)}{\|A\|_F} \right)^2, \tag{2}$$

where $\|A\|_F := \sqrt{\sum_{i,j} A_{i,j}^2}$ is the Frobenius norm of A.

Note that $\nu_+(A)$ is defined as the solution of a conic program over the cone of copositive matrices. Copositive programming is known to be NP-hard in general (see e.g., [Dür10]), but fortunately one can obtain good approximations using semidefinite programming. For example one can obtain a lower bound to $\nu_+(A)$ by solving the following semidefinite program:

$$\nu_+^{[0]}(A) = \max_{W \in \mathbb{R}^{m \times n}} \left\{ \langle A, W \rangle \ : \ \begin{bmatrix} I & -W \\ -W^\mathsf{T} & I \end{bmatrix} \in \mathcal{N}^{n+m} + \mathcal{S}_+^{n+m} \right\} \tag{3}$$

where $\mathcal{S}_+^{n+m}$ denotes the cone of symmetric positive semidefinite matrices of size $n+m$, and $\mathcal{N}^{n+m}$ denotes the cone of symmetric nonnegative matrices of size $n+m$. Since in general the sum of a nonnegative matrix and a positive semidefinite matrix is copositive, we immediately see that $\nu_+(A) \geq \nu_+^{[0]}(A)$ for any A, and thus this yields a polynomial-time computable lower bound to $\mathrm{rank}_+(A)$:

$$\mathrm{rank}_+(A) \geq \left(\frac{\nu_+^{[0]}(A)}{\|A\|_F} \right)^2.$$

One can in fact obtain tighter estimates of $\nu_+(A)$ using semidefinite programming by considering hierarchies of approximations of the copositive cone, like e.g., the hierarchy developed in [Par00]. This is discussed in more detail later in the paper (Section 2.4).

Note from the definition of $\nu_+(A)$ (Equation (1)) that $\nu_+(A)$ is convex in A and is thus continuous on the interior of its domain. In Section 2.2 we show how the quantity $\nu_+(A)$

can in fact be used to obtain a lower bound on the nonnegative rank of any matrix $\widetilde{A}$ that is ϵ-close (in Frobenius norm) to A.

It is interesting to express the dual of the copositive program (1). The dual of the cone of copositive matrices is the cone of completely positive matrices. A symmetric matrix M is said to be *completely positive* if it admits a factorization $M = BB^\mathsf{T}$ where B is elementwise nonnegative. Note that a completely positive matrix is both nonnegative and positive semidefinite; however not every such matrix is necessarily completely positive (see e.g., [BSM03] for an example and for more information on completely positive matrices). The dual of the copositive program (1) defining $\nu_+(A)$ is the following completely positive program (both programs give the same optimal value by strong duality):

$$\nu_+(A) = \min_{\substack{X \in \mathbb{R}^{m \times m} \\ Y \in \mathbb{R}^{n \times n}}} \left\{ \frac{1}{2}(\text{trace}(X) + \text{trace}(Y)) \; : \; \begin{bmatrix} X & A \\ A^\mathsf{T} & Y \end{bmatrix} \text{ completely positive} \right\}.$$

We will revisit this completely positive program later in the paper in Section 2.3 when we discuss the relation between the quantity $\nu_+(A)$ and the nuclear norm.

Outline The paper is organized as follows. In Section 2 we give the proof of the lower bound of Theorem 1 and we also outline a connection between the quantity $\nu_+(A)$ and the nuclear norm of a matrix. We then discuss computational issues and we see how to obtain semidefinite programming approximations of the quantity $\nu_+(A)$. In Section 3 we look at specific examples of matrices and we show that our lower bound can be greater than the plain rank lower bound and the rectangle covering number. In general however our lower bound is uncomparable to the existing lower bounds, i.e., it can be either greater or smaller. Among the examples, we show that our lower bound is exact for the slack matrix of the hypercube, thus giving another proof that the extension complexity of the hypercube in n dimensions is equal to $2n$ (a combinatorial proof of this fact is given in [FKPT13]).

Notations Throughout the paper $\mathcal{S}^n$ denotes the vector space of real symmetric matrices of size n and $\mathcal{S}^n_+$ is the cone of positive semidefinite matrices of size n. The cone of symmetric elementwise nonnegative matrices of size n is denoted by $\mathcal{N}^n$. If $X \in \mathcal{S}^n$ we write $X \succeq 0$ to say that X is positive semidefinite. Given two matrices X and Y, their inner product is defined as $\langle X, Y \rangle = \text{trace}(X^\mathsf{T} Y) = \sum_{i,j} X_{i,j} Y_{i,j}$.

2 The lower bound

2.1 Proof of the lower bound

In this section we prove the lower bound of Theorem 1 in a slightly more general form:

Theorem 2. *Let $A \in \mathbb{R}^{m \times n}_+$ be a nonnegative matrix. Let $P \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{n \times n}$ be nonnegative symmetric matrices with strictly positive diagonal entries. Let $\nu_+(A; P, Q)$ be the optimal value of the following copositive program:*

$$\nu_+(A; P, Q) = \max_{W \in \mathbb{R}^{m \times n}} \left\{ \langle A, W \rangle \; : \; \begin{bmatrix} P & -W \\ -W^\mathsf{T} & Q \end{bmatrix} \text{ copositive} \right\} \tag{4}$$

Then we have

$$\text{rank}_+(A) \geq \left(\frac{\nu_+(A; P, Q)}{\sqrt{\text{trace}(A^\mathsf{T} PAQ)}} \right)^2. \tag{5}$$

Note that Theorem 1 in the Introduction corresponds to the special case where P and Q are the identity matrices. The more general lower bound of Theorem 2 has the advantage of being invariant under diagonal scaling: namely if $\widetilde{A} = D_1 A D_2$ is obtained from A by positive diagonal scaling (where D_1 and D_2 are positive diagonal matrices), then the lower bound produced by the previous theorem for both quantities $\text{rank}_+(A)$ and $\text{rank}_+(\widetilde{A})$ (which are equal) will coincide, for adequate choices of P and Q. In fact it is easy to verify that if we choose $P = D_1^{-2}$ and $Q = D_2^{-2}$ then

$$\left(\frac{\nu_+(\widetilde{A}; P, Q)}{\sqrt{\text{trace}(\widetilde{A}^\mathsf{T} P \widetilde{A} Q)}} \right)^2 = \left(\frac{\nu_+(A)}{\|A\|_F} \right)^2.$$

The dual of the copositive program defining $\nu_+(A; P, Q)$ is the following completely positive program where the matrices P and Q enter as weighting matrices in the objective function:

$$\nu_+(A; P, Q) = \min_{\substack{X \in \mathbb{R}^{m \times m} \\ Y \in \mathbb{R}^{n \times n}}} \left\{ \frac{1}{2}(\text{trace}(PX) + \text{trace}(QY)) \; : \; \begin{bmatrix} X & A \\ A^\mathsf{T} & Y \end{bmatrix} \text{ completely positive} \right\}.$$

Proof of Theorem 2. We introduce the shorthand notations $\|x\|_P = \sqrt{x^\mathsf{T} Px}$ when $x \in \mathbb{R}_+^m$ and $\|y\|_Q = \sqrt{y^\mathsf{T} Qy}$ when $y \in \mathbb{R}_+^n$ which are well defined by the assumptions on P and Q (note however that $\|\cdot\|_P$ and $\|\cdot\|_Q$ are not necessarily norms in the usual sense). Let A be an $m \times n$ nonnegative matrix with nonnegative rank $r \geq 1$ and consider a factorization $A = UV = \sum_{i=1}^r u_i v_i^\mathsf{T}$ where $u_i \in \mathbb{R}_+^m$ are the columns of U and $v_i^T \in \mathbb{R}_+^n$ the rows of V. Clearly the vectors u_i and v_i are nonzero for all $i \in \{1, \ldots, r\}$. By rescaling the u_i's and v_i's we can assume that $\|u_i\|_P = \|v_i\|_Q$ for all $i \in \{1, \ldots, r\}$ (simply replace u_i by $\widetilde{u}_i = \gamma_i u_i$ and v_i by $\widetilde{v}_i = \gamma_i^{-1} v_i$ where $\gamma_i = \sqrt{\|v_i\|_Q / \|u_i\|_P}$). Observe that by the Cauchy-Schwarz inequality, we have:

$$\frac{\sum_{i=1}^r \|u_i\|_P \|v_i\|_Q}{\sqrt{\sum_{i=1}^r \|u_i\|_P^2 \|v_i\|_Q^2}} \leq \sqrt{r} = \sqrt{\text{rank}_+(A)}$$

We will now show separately that the numerator of the left-hand side above is lower bounded by the quantity $\nu_+(A; P, Q)$, and that the denominator is upper bounded by $\sqrt{\text{trace}(A^\mathsf{T} PAQ)}$:

- Numerator: If W is such that $\begin{bmatrix} P & -W \\ -W^\mathsf{T} & Q \end{bmatrix}$ is copositive then, since $u_i, v_i \geq 0$, we have:

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix}^\mathsf{T} \begin{bmatrix} P & -W \\ -W^\mathsf{T} & Q \end{bmatrix} \begin{bmatrix} u_i \\ v_i \end{bmatrix} \geq 0$$

  and hence

$$u_i^\mathsf{T} W v_i \leq \frac{1}{2}(\|u_i\|_P^2 + \|v_i\|_Q^2) = \|u_i\|_P \|v_i\|_Q$$

where we used the fact that $\|u_i\|_P = \|v_i\|_Q$. Thus we get

$$\langle A, W \rangle = \left\langle \sum_{i=1}^{r} u_i v_i^\mathsf{T}, W \right\rangle = \sum_{i=1}^{r} u_i^\mathsf{T} W v_i \leq \sum_{i=1}^{r} \|u_i\|_P \|v_i\|_Q.$$

Note that this is true for any W such that $\begin{bmatrix} P & -W \\ -W^\mathsf{T} & Q \end{bmatrix}$ is copositive and thus we obtain

$$\nu_+(A; P, Q) \leq \sum_{i=1}^{r} \|u_i\|_P \|v_i\|_Q.$$

- Denominator: We now turn to finding an upper bound on $\sum_{i=1}^{r} \|u_i\|_P^2 \|v_i\|_Q^2$. Observe that we have

$$\begin{aligned}
\text{trace}(A^\mathsf{T} PAQ) = \langle PA, AQ \rangle &= \sum_{1 \leq i,j \leq r} \langle P u_i v_i^\mathsf{T}, u_j v_j^\mathsf{T} Q \rangle \\
&= \sum_{i=1}^{r} \|u_i\|_P^2 \|v_i\|_Q^2 + \sum_{i \neq j} (u_i^\mathsf{T} P u_j)(v_j^\mathsf{T} Q v_i) \\
&\geq \sum_{i=1}^{r} \|u_i\|_P^2 \|v_i\|_Q^2
\end{aligned}$$

where in the last inequality we used the fact that $u_i^\mathsf{T} P u_j \geq 0$ and $v_j^\mathsf{T} Q v_j \geq 0$ which is true since P and Q are nonnegative.

Now if we combine the two points above we finally get the desired inequality

$$\text{rank}_+(A) = r \geq \left(\frac{\sum_{i=1}^{r} \|u_i\|_P \|v_i\|_Q}{\sqrt{\sum_{i=1}^{r} \|u_i\|_P^2 \|v_i\|_Q^2}} \right)^2 \geq \left(\frac{\nu_+(A; P, Q)}{\sqrt{\text{trace}(A^\mathsf{T} PAQ)}} \right)^2.$$

$\square$

2.2 Lower bound on the approximate nonnegative rank

It is clear from the definition (1) that the function $A \in \mathbb{R}_+^{m \times n} \mapsto \nu_+(A)$ is convex and is thus continuous on the interior of its domain, unlike the nonnegative rank. A consequence of this is that the lower bound $(\nu_+(A)/\|A\|_F)^2$ will be small in general if A is close to a matrix with small nonnegative rank. This continuity property of $\nu_+(A)$ can be used to obtain a lower bound on the nonnegative rank of any matrix $\widetilde{A}$ that is close enough to A. Define the *approximate nonnegative rank* of A, denoted $\text{rank}_+^\epsilon(A)$, to be the smallest nonnegative rank among all nonnegative matrices that are ϵ-close to A in Frobenius norm:

$$\text{rank}_+^\epsilon(A) = \min \left\{ \text{rank}_+(\widetilde{A}) \ : \ \widetilde{A} \in \mathbb{R}_+^{m \times n} \text{ and } \|A - \widetilde{A}\|_F \leq \epsilon \right\}. \tag{6}$$

Approximate nonnegative factorizations and the approximate nonnegative rank have applications in lifts of polytopes [GPT13a] as well as in information theory [BJLP13].

The following theorem shows that one can obtain a lower bound on $\text{rank}_+^\epsilon(A)$ using the quantity $\nu_+(A)$:

Theorem 3. *Let $A \in \mathbb{R}_+^{m \times n}$ be a nonnegative matrix and let W be an optimal solution in the definition of $\nu_+(A)$ (cf. Equation (1)). Let ϵ be a positive constant with $\epsilon \leq \nu_+(A)/\|W\|_F$. Then we have:*

$$\text{rank}_+^\epsilon(A) \geq \left(\frac{\nu_+(A) - \epsilon \|W\|_F}{\|A\|_F + \epsilon} \right)^2.$$

Proof. Let $\widetilde{A}$ be any nonnegative matrix such that $\|A - \widetilde{A}\|_F \leq \epsilon$. Since W is a feasible point for the copositive program that defines $\nu_+(\widetilde{A})$ we clearly have

$$\nu_+(\widetilde{A}) \geq \langle \widetilde{A}, W \rangle = \langle A, W \rangle + \langle \widetilde{A} - A, W \rangle \geq \nu_+(A) - \epsilon \|W\|_F$$

Hence since $\nu_+(A) - \epsilon \|W\|_F \geq 0$ and $\|\widetilde{A}\|_F \leq \|A\|_F + \epsilon$ we get:

$$\text{rank}_+(\widetilde{A}) \geq \left(\frac{\nu_+(\widetilde{A})}{\|\widetilde{A}\|_F} \right)^2 \geq \left(\frac{\nu_+(A) - \epsilon \|W\|_F}{\|A\|_F + \epsilon} \right)^2.$$

Since this is valid for any $\widetilde{A}$ that is ϵ-close to A, we have

$$\text{rank}_+^\epsilon(A) \geq \left(\frac{\nu_+(A) - \epsilon \|W\|_F}{\|A\|_F + \epsilon} \right)^2.$$

$\square$

2.3   Connection with nuclear norm

In this section we discuss the connection between the quantity $\nu_+(A)$ and nuclear norms of linear operators. If A is an arbitrary (not necessarily nonnegative) matrix the *nuclear norm* of A is defined by [Jam87]:

$$\nu(A) = \min \left\{ \sum_i \|u_i\|_2 \|v_i\|_2 \ : \ A = \sum_i u_i v_i^\mathsf{T} \right\}. \tag{7}$$

It can be shown that the quantity $\nu(A)$ above is equal to the sum of the singular values $\sigma_1(A) + \cdots + \sigma_r(A)$ of A, and in fact this is the most commonly encountered definition of the nuclear norm. This latter characterization gives the following well-known lower bound on $\text{rank}(A)$ combining $\nu(A)$ and $\|A\|_F$:

$$\text{rank}(A) \geq \left(\frac{\sigma_1(A) + \cdots + \sigma_r(A)}{\sqrt{\sigma_1(A)^2 + \cdots + \sigma_r(A)^2}} \right)^2 = \left(\frac{\nu(A)}{\|A\|_F} \right)^2.$$

The characterization of nuclear norm given in Equation (7) can be very naturally adapted to nonnegative factorizations of A by restricting the vectors u_i and v_i in the decomposition of A to be nonnegative. The new quantity that we obtain with this restriction is in fact nothing but the quantity $\nu_+(A)$ introduced earlier, as we show in the Theorem below. Observe that this quantity $\nu_+(A)$ is always greater than or equal than the standard nuclear norm $\nu(A)$.

7

Theorem 4. *Let $A \in \mathbb{R}_+^{m \times n}$ be a nonnegative matrix. The following three quantities are equal to $\nu_+(A)$ as defined in Equation* (1):

(i) $\quad \min \left\{ \sum_i \|u_i\|_2 \|v_i\|_2 \; : \; A = \sum_i u_i v_i^\mathsf{T}, \; u_i, v_i \geq 0 \right\}$

(ii) $\min_{\substack{X \in \mathcal{S}^m \\ Y \in \mathcal{S}^n}} \left\{ \frac{1}{2}(\mathrm{trace}(X) + \mathrm{trace}(Y)) \; : \; \begin{bmatrix} X & A \\ A^\mathsf{T} & Y \end{bmatrix} \text{ completely positive} \right\}$

(iii) $\max_{W \in \mathbb{R}^{m \times n}} \left\{ \langle A, W \rangle \; : \; \begin{bmatrix} I & -W \\ -W^\mathsf{T} & I \end{bmatrix} \text{ copositive} \right\}$

Proof. Note that the conic programs in (ii) and (iii) are dual of each other; furthermore program (iii) is strictly feasible (simply take $W = 0$) thus it follows from strong duality that (ii) and (iii) have the same optimal value. We thus have to show only that (i) and (ii) are equal.

We start by proving that (ii) $\leq$ (i): If $A = \sum_{i=1}^k u_i v_i^\mathsf{T}$ where $u_i, v_i \geq 0$ and $u_i, v_i \neq 0$ for all $i \in \{1, \ldots, k\}$, then if we let

$$X = \sum_{i=1}^k \frac{\|v_i\|_2}{\|u_i\|_2} u_i u_i^\mathsf{T} \quad \text{and} \quad Y = \sum_{i=1}^k \frac{\|u_i\|_2}{\|v_i\|_2} v_i v_i^\mathsf{T}$$

then

$$\begin{bmatrix} X & A \\ A^\mathsf{T} & Y \end{bmatrix} = \sum_{i=1}^k \|u_i\|_2 \|v_i\|_2 \begin{bmatrix} u_i/\|u_i\|_2 \\ v_i/\|v_i\|_2 \end{bmatrix} \begin{bmatrix} u_i/\|u_i\|_2 \\ v_i/\|v_i\|_2 \end{bmatrix}^\mathsf{T}$$

and so $\begin{bmatrix} X & A \\ A^\mathsf{T} & Y \end{bmatrix}$ is completely positive since u_i and v_i are elementwise nonnegative. Furthermore we have

$$\mathrm{trace}(X) + \mathrm{trace}(Y) = \sum_{i=1}^k \|u_i\|_2 \|v_i\|_2 + \sum_{i=1}^k \|u_i\|_2 \|v_i\|_2 = 2 \sum_{i=1}^k \|u_i\|_2 \|v_i\|_2.$$

Hence this shows that (ii) $\leq$ (i).

We now show that (i) $\leq$ (ii). Assume that X and Y are such that $\begin{bmatrix} X & A \\ A^\mathsf{T} & Y \end{bmatrix}$ is completely positive and consider a decomposition

$$\begin{bmatrix} X & A \\ A^\mathsf{T} & Y \end{bmatrix} = \sum_{i=1}^k \begin{bmatrix} x_i \\ y_i \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix}^\mathsf{T}$$

where x_i and y_i are nonnegative. Then from this decomposition we have $A = \sum_{i=1}^k x_i y_i^\mathsf{T}$ and

$$\sum_{i=1}^k \|x_i\|_2 \|y_i\|_2 \leq \frac{1}{2} \left(\sum_{i=1}^k \|x_i\|_2^2 + \sum_{i=1}^k \|y_i\|_2^2 \right) = \frac{1}{2}(\mathrm{trace}(X) + \mathrm{trace}(Y)).$$

This shows that (i) $\leq$ (ii) and completes the proof. $\qquad \square$

For an arbitrary (not necessarily nonnegative) matrix A the nuclear norm $\nu(A)$ can be computed as the optimal value of the following primal-dual pair of semidefinite programs:

$$\begin{array}{ll}
\underset{\substack{X \in \mathcal{S}^m \\ Y \in \mathcal{S}^n}}{\text{minimize}} & \frac{1}{2}(\text{trace}(X) + \text{trace}(Y)) \\
\text{subject to} & \begin{bmatrix} X & A \\ A^\mathsf{T} & Y \end{bmatrix} \in \mathcal{S}_+^{m+n}
\end{array} \quad \left| \quad \begin{array}{ll}
\underset{W \in \mathbb{R}^{m \times n}}{\text{maximize}} & \langle A, W \rangle \\
\text{subject to} & \begin{bmatrix} I & -W \\ -W^\mathsf{T} & I \end{bmatrix} \in \mathcal{S}_+^{m+n}
\end{array} \right. \quad (8)$$

A solution of these semidefinite programs can be obtained from a singular value decomposition of A: namely if $A = U\Sigma V^\mathsf{T}$ is a singular value decomposition of A, then $X = U\Sigma U^\mathsf{T}$, $Y = V\Sigma V^\mathsf{T}$ and $W = UV^\mathsf{T}$ are optimal points of the semidefinite programs above, and the optimal value is $\text{trace}(\Sigma) = \nu(A)$.

Note that the primal-dual pairs that define $\nu(A)$ and $\nu_+(A)$ are very similar, and the only differences are in the cones used: for $\nu(A)$ it is the self-dual cone of positive semidefinite matrices, whereas for $\nu_+(A)$ it is the dual pair of completely positive / copositive cones. Note also that these optimization problems are related to the convex programs that arise in [DV13]. In the following proposition we give simple sufficient conditions on the singular value decomposition of A that allow to check if the two quantities $\nu_+(A)$ and $\nu(A)$ are actually equal. For the proposition recall that $\nu_+^{[0]}(A)$ is the first approximation of $\nu_+(A)$ defined in (3), where the copositive cone is replaced by the cone $\mathcal{N}^{n+m} + \mathcal{S}_+^{n+m}$. Using duality the quantity $\nu_+^{[0]}(A)$ is also equal to the solution of the following minimization problem:

$$\min_{X,Y} \left\{ \frac{1}{2}(\text{trace}(X) + \text{trace}(Y)) \; : \; \begin{bmatrix} X & A \\ A^\mathsf{T} & Y \end{bmatrix} \in \mathcal{S}_+^{2n} \cap \mathcal{N}^{2n} \right\} \quad (9)$$

where $\mathcal{S}_+^{2n} \cap \mathcal{N}^{2n}$ is the cone of doubly nonnegative matrices (i.e., the cone of matrices that are nonnegative and positive semidefinite).

Proposition 1. *Let $A \in \mathbb{R}_+^{m \times n}$ be a nonnegative matrix with a singular value decomposition $A = U\Sigma V^\mathsf{T}$ and let $\nu(A) = \text{trace}(\Sigma)$ be the nuclear norm of A.*
(i) If $U\Sigma U^\mathsf{T}$ and $V\Sigma V^\mathsf{T}$ are nonnegative then $\nu_+^{[0]}(A) = \nu(A)$.
(ii) Also, if the matrix

$$\begin{bmatrix} U\Sigma U^\mathsf{T} & A \\ A^\mathsf{T} & V\Sigma V^\mathsf{T} \end{bmatrix} \in \mathcal{S}^{n+m} \quad (10)$$

is completely positive, then $\nu_+(A) = \nu(A)$.

Proof. To prove point (i) of the proposition note first that we always have the inequality $\nu_+^{[0]}(A) \geq \nu(A)$. Now if $U\Sigma U^\mathsf{T}$ and $V\Sigma V^\mathsf{T}$ are nonnegative then $X = U\Sigma U^\mathsf{T}$ and $Y = V\Sigma V^\mathsf{T}$ are feasible for (9) and achieve the value $\text{trace}(\Sigma) = \nu(A)$. This shows that in this case $\nu_+^{[0]}(A) \leq \nu(A)$ and thus $\nu_+^{[0]}(A) = \nu_+(A)$. The proof of item (ii) is similar. $\square$

To finish this section we investigate properties of the solution of the completely positive/copositive pair that defines $\nu_+(A)$ (cf. (ii) and (iii) in Theorem 4). An interesting question is to know whether the solution of this primal-dual pair possesses some interesting properties like the singular value decomposition. The next proposition shows a property which relates the optimal W of the copositive program and a nonnegative decomposition $A = \sum_i u_i v_i^\mathsf{T}$ satisfying $\sum_i \|u_i\|_2 \|v_i\|_2 = \nu_+(A)$.

Proposition 2. *Let $A \in \mathbb{R}_+^{m \times n}$ be a nonnegative matrix. Assume $A = \sum_i \lambda_i u_i v_i^\mathsf{T}$ is a nonnegative decomposition of A that satisfies*

$$\sum_i \lambda_i = \nu_+(A)$$

where $\lambda_i \geq 0$ and $u_i \in \mathbb{R}_+^m$, $v_i \in \mathbb{R}_+^n$ have unit norm $\|u_i\|_2 = \|v_i\|_2 = 1$. Let W be an optimal point in the copositive program (1). Then we have for all i,

$$u_i = \Pi_+(Wv_i) \quad and \quad v_i = \Pi_+(W^\mathsf{T} u_i)$$

where $\Pi_+(x) = \max(x, 0)$ is the projection on the nonnegative orthant.

The proposition above can be seen as the analogue in the nonnegative world of the fact that, for arbitrary matrices A, the optimal W in the semidefinite program (8) satisfies $u_i = Wv_i$ and $v_i = W^\mathsf{T} u_i$ where u_i and v_i are respectively the left and right singular vectors of A.

Before proving the proposition we prove the following simple lemma:

Lemma 1. *If $W \in \mathbb{R}^{m \times n}$ is such that $\begin{bmatrix} I & -W \\ -W^\mathsf{T} & I \end{bmatrix}$ is copositive, then for any $v \in \mathbb{R}_+^n$ we have*

$$\|\Pi_+(Wv)\|_2 \leq \|v\|_2$$

where $\Pi_+(x) = \max(x, 0)$ is the projection on the nonnegative orthant.

Proof. Let $v \in \mathbb{R}_+^n$ and call $u = \Pi_+(Wv)$. By the copositivity assumption we have, since u and v are nonnegative:

$$\begin{bmatrix} u \\ v \end{bmatrix}^\mathsf{T} \begin{bmatrix} I & -W \\ -W^\mathsf{T} & I \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \geq 0$$

which can rewritten as:

$$(u - Wv)^\mathsf{T}(u - Wv) + v^\mathsf{T}(I - W^\mathsf{T} W)v \geq 0.$$

Since $u = \Pi_+(Wv)$ we get:

$$\|\Pi_+(Wv) - Wv\|_2^2 + \|v\|_2^2 - \|Wv\|_2^2 \geq 0.$$

Note that $\|\Pi_+(Wv) - Wv\|_2^2 - \|Wv\|_2^2 = -\|\Pi_+(Wv)\|_2^2$ and thus we get the desired inequality

$$\|\Pi_+(Wv)\|_2^2 \leq \|v\|_2^2.$$

$\square$

Using this lemma we now prove Proposition 2:

Proof of Proposition 2. Let $A = \sum_i \lambda_i u_i v_i^\mathsf{T}$ be a nonnegative decomposition of A that satisfies $\sum_i \lambda_i = \nu_+(A)$ and let W be an optimal point of the copositive program (1). Since

$\begin{bmatrix} I & -W \\ -W^{\mathsf{T}} & I \end{bmatrix}$ is copositive and $\|u_i\|_2 = \|v_i\|_2 = 1$, we have for each i, $u_i^{\mathsf{T}} W v_i \le 1$. Now since W is optimal we have

$$\sum_i \lambda_i u_i^{\mathsf{T}} W v_i = \langle A, W \rangle = \nu_+(A) = \sum_i \lambda_i,$$

and hence for each i we have necessarily $u_i^{\mathsf{T}} W v_i = 1$. Furthermore, we have the sequence of inequalities

$$1 = u_i^{\mathsf{T}} W v_i \le u_i^{\mathsf{T}} \Pi_+(W v_i) \le \|u_i\|_2 \|\Pi_+(W v_i)\|_2 \le \|u_i\|_2 \|v_i\|_2 = 1 \qquad (11)$$

where in the first inequality we used that $u \ge 0$, then we used Cauchy-Schwarz inequality and for the third inequality we used Lemma 1. Since the left-hand side and the right-hand side of (11) are equal this shows that all the intermediate inequalities are in fact equalities. In particular by the equality case in Cauchy-Schwarz we have that $\Pi_+(W v_i) = \rho_i u_i$ for some constant ρ_i. But since $u_i^{\mathsf{T}} \Pi_+(W v_i) = 1$ and $\|u_i\|_2 = 1$ we get that $\rho_i = 1$. This shows finally that $u_i = \Pi_+(W v_i)$. To prove that $v_i = \Pi_+(W^{\mathsf{T}} u_i)$ we use the same line of inequalities as in (11) starting from the fact that $1 = v_i^{\mathsf{T}} W^{\mathsf{T}} u_i$. $\qquad\square$

2.4 Approximations and semidefinite programming lower bounds

In this section we describe how one can use semidefinite programming to obtain polynomial-time computable lower bounds on $\nu_+(A)$, and thus on $\mathrm{rank}_+(A)$. We outline here the hierarchy of approximations of the copositive cone proposed by Parrilo in [Par00] using sums-of-squares techniques.

Recall that a symmetric matrix $M \in \mathcal{S}^n$ is copositive if $x^{\mathsf{T}} M x \ge 0$ for all $x \ge 0$. An equivalent way of formulating this condition is to say that the following polynomial of degree 4

$$\sum_{1 \le i,j \le n} M_{i,j} x_i^2 x_j^2$$

is globally nonnegative. The cone $\mathcal{C}$ of copositive matrices can thus be described as:

$$\mathcal{C} = \left\{ M \in \mathcal{S}^n \ : \ \text{the polynomial} \sum_{1 \le i,j \le n} M_{i,j} x_i^2 x_j^2 \text{ is nonnegative} \right\}.$$

The k'th order inner approximation of $\mathcal{C}$ proposed by Parrilo in [Par00] is defined as:

$$\mathcal{C}^{[k]} = \left\{ M \in \mathcal{S}^n \ : \ \left(\sum_{i=1}^n x_i^2 \right)^k \left(\sum_{1 \le i,j \le n} M_{i,j} x_i^2 x_j^2 \right) \text{ is a sum of squares} \right\}. \qquad (12)$$

It is clear that for any k we have $\mathcal{C}^{[k]} \subseteq \mathcal{C}$ and also $\mathcal{C}^{[k]} \subseteq \mathcal{C}^{[k+1]}$, thus the sequence $(\mathcal{C}^{[k]})_{k \in \mathbb{N}}$ forms a sequence of nested inner approximations to the cone of copositive matrices; furthermore it is known via Pólya's theorem that this sequence converges to the copositive cone, cf. [Par00]. A crucial property of this hierarchy is that each cone $\mathcal{C}^{[k]}$ can be represented

11

using linear matrix inequalities, which means that optimizing a linear function over any of these cones is equivalent to a semidefinite program. Of course, the size of the semidefinite program gets larger as k gets larger, and the size of the semidefinite program at the k'th level is $\binom{n+k+1}{k+2}$. Note however that the semidefinite programs arising from this hierarchy can usually be simplified so that they can be solved more efficiently, cf. [Par00] and [GP04, Section 8.1].

An interesting feature concerning this hierarchy is that the approximation of order $k = 0$ corresponds to

$$\mathcal{C}^{[0]} = \mathcal{N}^n + \mathcal{S}^n_+,$$

i.e., $\mathcal{C}^{[0]}$ is the set of matrices that can be written as the sum of a nonnegative matrix and a positive semidefinite matrix (this particular approximation was already mentioned in the introduction).

Using this hierarchy, we can now compute lower bounds to $\mathrm{rank}_+(A)$ using semidefinite programming:

Theorem 5. *Let $A \in \mathbb{R}^{m \times n}_+$ be a nonnegative matrix. For $k \in \mathbb{N}$, let $\nu^{[k]}_+(A)$ be the optimal value of the following semidefinite programming problem:*

$$\nu^{[k]}_+(A) = \max \left\{ \langle A, W \rangle \; : \; \begin{bmatrix} I & -W \\ -W^\top & I \end{bmatrix} \in \mathcal{C}^{[k]} \right\} \tag{13}$$

where the cone $\mathcal{C}^{[k]}$ is defined in Equation (12). Then we have:

$$\mathrm{rank}_+(A) \geq \left(\frac{\nu^{[k]}_+(A)}{\|A\|_F} \right)^2. \tag{14}$$

As k gets larger the quantity $(\nu^{[k]}_+(A)/\|A\|_F)^2$ will converge (from below) to $(\nu_+(A)/\|A\|_F)^2$ (where $\nu_+(A)$ is defined by the copositive program (1)) and this quantity can be strictly smaller than $\mathrm{rank}_+(A)$ as can be seen for instance in Example 6 given later in the paper. In general, the quantity $(\nu^{[k]}_+(A)/\|A\|_F)^2$ may not converge to $\mathrm{rank}_+(A)$, however it will always be a valid lower bound to $\mathrm{rank}_+(A)$. In summary, we can write for any $k \geq 0$:

$$\nu(A) \leq \nu^{[0]}_+(A) \leq \nu^{[k]}_+(A) \leq \nu_+(A) \leq \sqrt{\mathrm{rank}_+(A)}\|A\|_F$$

where $\nu(A)$ is the standard nuclear norm of A. The lower bounds (14) are thus always greater than the lower bound that uses the standard nuclear norm (in fact they can also be greater than the standard rank lower bound as we show in the examples later).

On the webpage `http://www.mit.edu/~hfawzi` we provide a MATLAB script to compute the quantity $\nu^{[k]}_+(A)$ and the associated lower bound on $\mathrm{rank}_+(A)$ for any nonnegative matrix A and approximation level $k \in \mathbb{N}$. The script uses the software YALMIP and its Sum-Of-Squares module in order to compute the quantity $\nu^{[k]}_+(A)$ [Löf04, Löf09]. On a standard computer, the script allows to compute the lower bound of level $k = 0$ for matrices up to size ≈ 50.

Note that another hierarchy of inner approximations of the copositive cone has been proposed by de Klerk and Pasechnik in [KP02]. This hierarchy is based on linear programming

(instead of semidefinite programming) and it leads in general to smaller programs that can be solved more efficiently. However the convergence of the linear programming hierarchy is in general much slower than the semidefinite programming hierarchy and one typically needs to take very large values of k (the hierarchy level) to obtain interesting bounds.

Program size reduction when A is symmetric If the nonnegative matrix A is square and symmetric, the semidefinite programs that define $\nu_+^{[k]}(A)$ can be simplified to obtain smaller optimization problems that are easier to solve. For simplicity, we describe below in detail the case of $k = 0$; the extensions to the general case can also be done. When A is symmetric, the dual program defining $\nu_+^{[0]}(A)$ is:

$$\text{minimize} \ \ \frac{1}{2}(\text{trace}(X) + \text{trace}(Y)) \ \ \text{subject to} \ \ \begin{bmatrix} X & A \\ A & Y \end{bmatrix} \in \mathcal{S}_+^{2n} \cap \mathcal{N}^{2n}$$

where $\mathcal{S}_+^{2n} \cap \mathcal{N}^{2n}$ is the cone of doubly nonnegative matrices. It is not difficult to see that one can always find an optimal solution of the SDP above where $X = Y$. Indeed if the matrix $\begin{bmatrix} X & A \\ A & Y \end{bmatrix}$ is feasible, then so is the matrix $\begin{bmatrix} Y & A \\ A & X \end{bmatrix}$ which also achieves the same objective function and thus by averaging we get a feasible point with the same matrices on the diagonal and with the same objective function. Therefore for symmetric A, the quantity $\nu_+^{[0]}(A)$ is given by the optimal value of:

$$\text{minimize} \ \ \text{trace}(X) \ \ \text{subject to} \ \ \begin{bmatrix} X & A \\ A & X \end{bmatrix} \in \mathcal{S}_+^{2n} \cap \mathcal{N}^{2n}.$$

Now using the known fact that (see e.g., [Ber09, Fact 8.11.8])

$$\begin{bmatrix} X & A \\ A & X \end{bmatrix} \succeq 0 \ \ \Leftrightarrow \ \ X - A \succeq 0 \text{ and } X + A \succeq 0$$

we can rewrite the program as:

$$\text{minimize} \ \ \text{trace}(X) \ \ \text{subject to} \ \ \begin{cases} X - A \succeq 0 \\ X + A \succeq 0 \\ X \text{ nonnegative} \end{cases}$$

This new program has two smaller positive semidefinite constraints, each of size n, and thus can be solved more efficiently than the previous definition which involved one large positive semidefinite constraint of size $2n$. Indeed most current interior-point solvers exploit block-diagonal structure in semidefinite programs.

One can also perform the corresponding symmetry reduction for the *primal* program that defines $\nu_+^{[0]}(A)$, and this gives:

$$\text{maximize} \ \ \langle A, W \rangle \ \ \text{subject to} \ \ \begin{cases} R - W \succeq 0 \\ R + W \succeq 0 \\ I - R \text{ nonnegative} \end{cases} \tag{15}$$

13

Observe that the constraint in the program above implies that $\left[\begin{smallmatrix} I & -W \\ -W^\top & I \end{smallmatrix}\right] \in \mathcal{S}_+^{2n} + \mathcal{N}^{2n}$ since:

$$\begin{bmatrix} I & -W \\ -W & I \end{bmatrix} = \begin{bmatrix} I - R & 0 \\ 0 & I - R \end{bmatrix} + \begin{bmatrix} R & -W \\ -W & R \end{bmatrix}$$

where in the right-hand side the first matrix is nonnegative and the second one is positive semidefinite.

3 Examples

In this section we apply our lower bound to some explicit matrices and we compare it to existing lower bounding techniques.

Example 1. *We start with an example where our lower bound exceeds the plain rank lower bound. Consider the following 4×4 nonnegative matrix from [CR93]:*

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

The rank of this matrix is 3 and its nonnegative rank is 4 as was noted in [CR93]. We can compute the exact value of $\nu_+^{[0]}(A)$ for this matrix A and we get $\nu_+^{[0]}(A) = 4\sqrt{2} \approx 5.65$. We thus see that the lower bound is sharp for this matrix:

$$4 = \operatorname{rank}_+(A) \geq \left(\frac{\nu_+^{[0]}(A)}{\|A\|_F} \right)^2 = \left(\frac{4\sqrt{2}}{\sqrt{8}} \right)^2 = 4.$$

The optimal matrix W in the semidefinite program[2] (3) for which $\nu_+^{[0]}(A) = \langle A, W \rangle$ is given by:

$$W = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix} = \frac{1}{\sqrt{2}}(2A - J) \tag{16}$$

Observe that the matrix W is obtained from A by replacing the ones with $\frac{1}{\sqrt{2}}$ and the zeros with $-\frac{1}{\sqrt{2}}$. The matrix W is feasible for the semidefinite program (3) and one can check that we have the following decomposition of $\begin{bmatrix} I & -W \\ -W^\top & I \end{bmatrix}$ into a nonnegative part and a positive semidefinite part:

---

[2] Note that the matrix considered in this example is symmetric, and so one could use the reduced semidefinite program (15) to simplify the computation of $\nu_+^{[0]}(A)$. However for simplicity and for illustration purposes, we used in this example the original formulation (3).

$$\left[\begin{array}{c|c} I & -W \\ \hline -W^\mathsf{T} & I \end{array}\right] = \left[\begin{array}{c|c} \begin{matrix} 0\ 0\ 0\ 1 \\ 0\ 0\ 1\ 0 \\ 0\ 1\ 0\ 0 \\ 1\ 0\ 0\ 0 \end{matrix} & 0 \\ \hline 0 & \begin{matrix} 0\ 0\ 0\ 1 \\ 0\ 0\ 1\ 0 \\ 0\ 1\ 0\ 0 \\ 1\ 0\ 0\ 0 \end{matrix} \end{array}\right]}_{nonnegative} + \underbrace{\left[\begin{array}{c|c} \begin{matrix} 1\ \ \ 0\ \ \ 0\ -1 \\ 0\ \ \ 1\ -1\ \ \ 0 \\ 0\ -1\ \ \ 1\ \ \ 0 \\ -1\ \ \ 0\ \ \ 0\ \ \ 1 \end{matrix} & -W \\ \hline -W^\mathsf{T} & \begin{matrix} 1\ \ \ 0\ \ \ 0\ -1 \\ 0\ \ \ 1\ -1\ \ \ 0 \\ 0\ -1\ \ \ 1\ \ \ 0 \\ -1\ \ \ 0\ \ \ 0\ \ \ 1 \end{matrix} \end{array}\right]}_{positive\ semidefinite} \qquad (17)$$

Example 2 (Slack matrix of the hypercube)**.** *The 4×4 matrix of the previous example is in fact the* slack matrix *of the square $[0,1]^2$ in the plane. Recall that the slack matrix [Yan91] of a polytope $P \subset \mathbb{R}^n$ with f facet inequalities $b_i - a_i^\mathsf{T} x \geq 0$, $i = 1, \ldots, f$ and v vertices $x_1, \ldots, x_v \in \mathbb{R}^n$ is the $f \times v$ matrix $S(P)$ given by:*

$$S(P)_{i,j} = b_i - a_i^\mathsf{T} x_j \quad \forall i = 1, \ldots, f, \ j = 1, \ldots, v.$$

The matrix $S(P)$ is clearly nonnegative since the vertices x_j belong to P and satisfy the facet inequalities. Yannakakis showed in [Yan91] that the nonnegative rank of $S(P)$ coincides with the smallest number of linear inequalities needed to represent the polytope[3] P. The minimal number of linear inequalities needed to represent P is also known as the extension complexity of P; the theorem of Yannakakis therefore states that $\mathrm{rank}_+(S(P))$ is equal to the extension complexity of P.

The hypercube $[0,1]^n$ in n dimensions has $2n$ facets and 2^n vertices. It is known that the extension complexity of the hypercube is equal to $2n$ and this was proved recently in [FKPT13, Proposition 5.9] using a combinatorial argument (in fact it was shown that any polytope in $\mathbb{R}^n$ that is combinatorially equivalent to the hypercube has extension complexity $2n$). Note that the trivial lower bound obtained from the rank of the slack matrix in this case is $n+1$: in fact for any full-dimensional polytope $P \subset \mathbb{R}^n$, the rank of the slack matrix of P is equal to $n+1$, see e.g., [GRT13, Lemma 3.1], and so the rank lower bound is in general not interesting in the context of slack matrices of polytopes. Also the lower bound of Goemans [Goe14] here is $\log_2(2^n) = n$ which is not tight.

Below we use the lower bound on the nonnegative rank introduced in this paper to give another proof that the extension complexity of the hypercube in n dimensions is $2n$.

Proposition 3. *Let $C_n = [0,1]^n$ be the hypercube in n dimensions and let $S(C_n) \in \mathbb{R}^{2n \times 2^n}$ be its slack matrix. Then*

$$\mathrm{rank}_+(S(C_n)) = \left(\frac{\nu_+^{[0]}(S(C_n))}{\|S(C_n)\|_F} \right)^2 = 2n.$$

---

[3]Note that the trivial representation of P uses f linear inequalities, where f is the number of facets of P. However by introducing new variables (i.e., allowing projections) one can sometimes reduce dramatically the number of inequalities needed to represent P. For example the cross-polytope $P = \{x \in \mathbb{R}^n : w^\mathsf{T} x \leq 1 \ \forall w \in \{-1,1\}^n\}$ has 2^n facets but can be represented using only $2n$ linear inequalities after introducing n additional variables: $P = \{x \in \mathbb{R}^n \ : \ \exists y \in \mathbb{R}^n \ y \geq x, \ y \geq -x, \ \mathbf{1}^\mathsf{T} y = 1\}$.

Proof. The proof is given in Appendix A. □

Example 3 (Matrix with strictly positive entries). *Consider the following matrix with strictly positive entries ($\epsilon > 0$):*

$$A_\epsilon = \begin{bmatrix} 1+\epsilon & 1+\epsilon & \epsilon & \epsilon \\ 1+\epsilon & \epsilon & 1+\epsilon & \epsilon \\ \epsilon & 1+\epsilon & \epsilon & 1+\epsilon \\ \epsilon & \epsilon & 1+\epsilon & 1+\epsilon \end{bmatrix}.$$

The matrix A_ϵ is obtained from the matrix of Example 1 by adding a constant $\epsilon > 0$ to each entry. The matrix A_ϵ has appeared before, e.g., in [Gil12, Equation 12] (under a slightly different form) and corresponds to the slack matrix for the pair of polytopes $[-1, 1]^2$ and $[-1 - 2\epsilon, 1 + 2\epsilon]^2$. It can be shown in this particular case, since A_ϵ is 4×4 and $\mathrm{rank}\, A_\epsilon = 3$, that the nonnegative rank of A_ϵ is the smallest r such that there is a polygon P with r vertices such that $[-1, 1]^2 \subset P \subset [-1 - 2\epsilon, -1 + 2\epsilon]^2$, cf. [Gil12] for more details.

Since A_ϵ is a small perturbation of the matrix A of Example 1 one can use the approach[4] of Theorem 3 to get a lower bound on $\nu_+^{[0]}(A_\epsilon)$. Let W be the matrix defined in (16). Since W is feasible for (3) we have $\nu_+^{[0]}(A_\epsilon) \geq \langle A_\epsilon, W \rangle$. It turns out in this example that the value of $\langle A_\epsilon, W \rangle$ does not depend on ϵ and is equal to $4\sqrt{2}$. This allows to obtain lower bounds on $\mathrm{rank}_+(A_\epsilon)$ without solving any additional optimization problem. For example for $\epsilon = 0.1$ we get:

$$\mathrm{rank}_+(A_{0.1}) \geq \frac{(4\sqrt{2})^2}{\|A_{0.1}\|_F^2} \approx 3.2$$

which shows that $\mathrm{rank}_+(A_{0.1}) = 4$.

Example 4 (Comparison with the Boolean rank lower bound). *We now show an example where the lower bound $(\nu_+^{[0]}(A)/\|A\|_F)^2$ is strictly greater than the rectangle covering lower bound (i.e., Boolean rank lower bound). Consider the 4×4 matrix*

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

Note that the rectangle covering number of A is 2 since $\mathrm{support}(A)$ can be covered with the two rectangles $\{1, 2\} \times \{2, 3, 4\}$ and $\{2, 3, 4\} \times \{1, 2\}$. If we compute the quantity $\nu_+^{[0]}(A)$ and the associated lower bound we get $\mathrm{rank}_+(A) \geq \lceil (\nu_+^{[0]}(A)/\|A\|_F)^2 \rceil = 3$ which is strictly greater than the rectangle covering number. In fact $\mathrm{rank}_+(A)$ is exactly equal to 3 since we have the factorization

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

---

[4]Since we are interested in obtaining a lower bound for the *specific* perturbation A_ϵ of A, we can obtain a better bound than the one of Theorem 3 by normalizing by $\|A_\epsilon\|_F^2$ directly.

The following simple proposition concerning diagonal matrices will be needed for the next example:

Proposition 4. *If $A \in \mathbb{R}^{n \times n}$ is a nonnegative diagonal matrix and P is a diagonal matrix with strictly positive elements on the diagonal, then*

$$\nu_+(A; P, P) = \langle A, P \rangle = \sum_{i=1}^{n} A_{i,i} P_{i,i}.$$

In particular, for $P = I$ we get $\nu_+(A) = \mathrm{trace}(A)$.

Proof. We first show that $\nu_+(A; P, P) \leq \langle A, P \rangle$: Observe that if W is such that $\begin{bmatrix} P & -W \\ -W & P \end{bmatrix}$ is copositive, then we must have $W_{i,i} \leq P_{i,i}$ for all $i \in \{1, \ldots, n\}$ (indeed, take e_i to be the i'th element of the canonical basis of $\mathbb{R}^n$, then by copositivity we must have $2e_i^\mathsf{T} P e_i - 2e_i^\mathsf{T} W e_i \geq 0$ which gives $W_{i,i} \leq P_{i,i}$). Hence, since A is diagonal we have $\langle A, W \rangle \leq \langle A, P \rangle$ for all feasible matrices W and thus $\nu_+(A; P, P) \leq \langle A, P \rangle$. Now if we take $W = P$, we easily see that $\begin{bmatrix} P & -W \\ -W & P \end{bmatrix}$ is copositive, and thus $\nu_+(A, P; P) \geq \langle A, P \rangle$. Thus we conclude that $\nu_+(A; P, P) = \langle A, P \rangle$. □

Example 5 (Matrix rescaling)**.** *The well-known lower bound on the rank*

$$\mathrm{rank}(A) \geq \left(\frac{\sigma_1(A) + \cdots + \sigma_r(A)}{\sqrt{\sigma_1(A)^2 + \cdots + \sigma_r(A)^2}} \right)^2$$

is sharp when all the singular values of A are equal, but it is loose when the matrix is not well-conditioned. The same phenomenon is expected to happen also for the lower bound of Theorem 1 on the nonnegative rank, even if it is not clear how to define the notion of condition number in the context of nonnegative matrices. If one considers the following diagonal matrix where $\beta > 0$

$$A = \begin{bmatrix} \beta & 0 & \ldots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \ldots & 0 & 1 \end{bmatrix} \in \mathbb{R}_+^{n \times n},$$

then we have from Proposition 4, $\nu_+(A) = \mathrm{trace}(A) = \beta + n - 1$, and thus for the lower bound we get:

$$\mathrm{rank}_+(A) \geq \left(\frac{\nu_+(A)}{\|A\|_F} \right)^2 = \left(\frac{\beta + n - 1}{\sqrt{\beta^2 + n - 1}} \right)^2.$$

If β is large and grows with n, say for example $\beta = n$, the lower bound on the right-hand side is $\leq O(1)$ whereas $\mathrm{rank}_+(A) = n$. Note also that the standard rank of A is equal to n.

To remedy this, one can use the more general lower bound of Theorem 2 with weight matrices P and Q to obtain a better lower bound, and in fact a sharp one. Indeed, if we let

$$P = Q = \begin{bmatrix} \epsilon & 0 & \dots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix}$$

then we have $\nu_+(A; P, Q)^2 = (n - 1 + \beta\epsilon)^2$ and $\mathrm{trace}(A^T PAQ) = n - 1 + (\beta\epsilon)^2$. Hence for $\epsilon = 1/\beta$, the ratio $\left(\frac{\nu_+(A;P,Q)}{\sqrt{\mathrm{trace}(A^T PAQ)}} \right)^2$ is equal to n which is equal to $\mathrm{rank}_+(A)$.

Example 6 (Derangement matrix)**.** *We now consider another example of a matrix that is not well-conditioned and where the lower bound of Theorem 1 is loose. Consider the derangement matrix*

$$D_n = \begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}$$

that has zeros on the diagonal and ones everywhere else. Then we have $\mathrm{rank}_+(D_n) = \mathrm{rank}(D_n) = n$, *but we will show that*

$$\left(\frac{\nu_+(D_n)}{\|D_n\|_F} \right)^2 \leq 4$$

for all n. Observe that the matrix D_n is not well-conditioned since it has one singular value equal to $n - 1$ while the other remaining singular values are all equal to 1.

To show that the lower bound of Theorem 1 is always bounded above by 4, note that the quantity $\nu_+(D_n)$ in this case is given by:

$$\nu_+(D_n) = \max \left\{ \sum_{i \neq j} W_{i,j} \; : \; \begin{bmatrix} I & -W \\ -W^T & I \end{bmatrix} \text{ copositive} \right\} \tag{18}$$

Observe that by symmetry, one can restrict the matrix W in the program above to have the form: [5]

$$W = \begin{bmatrix} a & b & \dots & b \\ b & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ b & \dots & b & a \end{bmatrix} = bJ_n + (a - b)I_n$$

---

[5]Indeed, observe first that if W is feasible for (18) then W^T is also feasible and has the same objective value. Thus by averaging one can assume that W is symmetric. Then note that for any feasible symmetric W and any permutation σ, the new matrix $W'_{i,j} = W_{\sigma(i),\sigma(j)}$ is also feasible and has the same objective value as W. Hence again by averaging we can assume W to be constant on the diagonal and constant on the off-diagonal.

where J_n is the $n \times n$ all ones matrix. For W of the form above we have for u and v arbitrary vectors in $\mathbb{R}^n$, $u^\mathsf{T} W v = (a-b)u^\mathsf{T} v + b(1^\mathsf{T} u)(1^\mathsf{T} v)$ and hence the condition that $\begin{bmatrix} I_n & -W \\ -W^\mathsf{T} & I_n \end{bmatrix}$ is copositive means that

$$\forall \, u, v \in \mathbb{R}^n_+, \;\; (a-b)u^\mathsf{T} v + b(1^\mathsf{T} u)(1^\mathsf{T} v) \;\le\; \frac{1}{2}(u^\mathsf{T} u + v^\mathsf{T} v)$$

Hence the quantity $\nu_+(D_n)$ can now be written as:

$$\nu_+(D_n) = \max_{a,b \in \mathbb{R}} \left\{ (n^2 - n)b \; : \; \forall u, v \in \mathbb{R}^n_+, \; (a-b)u^\mathsf{T} v + b(1^\mathsf{T} u)(1^\mathsf{T} v) \le \frac{1}{2}(u^\mathsf{T} u + v^\mathsf{T} v) \right\}$$

Let us call b_n the largest b in the problem above, so that $\nu_+(D_n) = (n^2 - n)b_n$. Assuming n is even let

$$u_n = (\underbrace{1, \ldots, 1}_{n/2}, 0, \ldots, 0) \in \mathbb{R}^n_+ \;\; and \;\; v_n = (0, \ldots, 0, \underbrace{1, \ldots, 1}_{n/2}) \in \mathbb{R}^n_+.$$

Then we have $u_n^\mathsf{T} v_n = 0$ and the optimal b_n must satisfy $b_n(1^\mathsf{T} u_n)(1^\mathsf{T} v_n) \le \frac{1}{2}(u_n^\mathsf{T} u_n + v_n^\mathsf{T} v_n)$, which gives $b_n \frac{n^2}{4} \le \frac{n}{2}$, i.e., $b_n \le 2/n$. Hence $\nu_+(D_n) \le (n^2 - n) \cdot 2/n = 2(n-1)$ and

$$\left(\frac{\nu_+(D_n)}{\|D_n\|_F} \right)^2 \le \frac{4(n-1)^2}{n^2 - n} = 4(1 - 1/n) \le 4.$$

The case when n is odd can be treated the same way and we also obtain the same upper bound of 4.

Summary of examples In the examples above we have shown that our lower bounds (both the exact and the first relaxation) are uncomparable to most existing bounds on the nonnegative rank:

- Examples 1 and 6 show that the bound can be either larger or smaller than the standard rank.

- Examples 4 and 5 show that the bound can be either larger or smaller than the rectangle covering number and the fooling set bound.

The following inequalities are however always satisfied for any nonnegative matrix A:

$$\mathrm{rank}_+(A) \ge \left(\frac{\nu_+(A)}{\|A\|_F} \right)^2 \ge \left(\frac{\nu_+^{[0]}(A)}{\|A\|_F} \right)^2 \ge \left(\frac{\nu(A)}{\|A\|_F} \right)^2.$$

4 Conclusion

In this paper we have presented new lower bounds on the nonnegative rank that can be computed using semidefinite programming. Unlike many of the existing bounds, our lower bounds do not solely depend on the sparsity pattern of the matrix and are applicable to matrices with strictly positive entries.

An interesting question is to know whether the techniques presented here can be strengthened to obtain sharper lower bounds. In particular the results given here rely on the ratio of the ℓ_1 and the ℓ_2 norms, however it is known that the ratio of the ℓ_1 and the ℓ_p norms for $1 < p < 2$ can yield better lower bounds.

Another question left open in this paper is the issue of scaling. As we saw in Section 2 and in the examples, one can choose scaling matrices P and Q to improve the bound. It is not clear however how to compute the optimal scaling P and Q that yields the best lower bound. In recent work [FP14] we extend some of the ideas presented in this paper to obtain a new lower bound on the nonnegative rank which is invariant under diagonal scaling, and which also satisfies other interesting properties (e.g., subadditivity, etc.). In fact the technique we propose in [FP14] applies to a large class of atomic cone ranks and can be used for example to obtain lower bounds on the cp-rank of completely positive matrices [BSM03].

Finally it is natural to ask whether the ideas presented here can be applied to obtain lower bounds on the *positive semidefinite (psd) rank*, a quantity which was introduced recently in [GPT13b] in the context of positive semidefinite lifts of polytopes. One main difficulty however is that the psd rank is not an "atomic" rank, unlike the nonnegative rank where the atoms correspond to nonnegative rank-one matrices. In fact it is this atomic property of the nonnegative rank which was crucial here to obtain the lower bounds in this paper.

A Proof of Proposition 3: Slack matrix of hypercube

In this appendix we prove Proposition 3 concerning the nonnegative rank of the slack matrix of the hypercube. We restate the proposition here for convenience:

Proposition. *Let $C_n = [0,1]^n$ be the hypercube in n dimensions and let $S(C_n) \in \mathbb{R}^{2n \times 2^n}$ be its slack matrix. Then*

$$\mathrm{rank}_+(S(C_n)) = \left(\frac{\nu_+^{[0]}(S(C_n))}{\|S(C_n)\|_F} \right)^2 = 2n.$$

Proof. The facets of the hypercube $C_n = [0,1]^n$ are given by the linear inequalities $\{x_k \geq 0\}$, $k = 1, \ldots, n$ and $\{x_k \leq 1\}$, $k = 1, \ldots, n$, and the vertices of C_n are given by the set $\{0,1\}^n$ of binary words of length n. It is easy to see that the slack matrix of the hypercube is a $0/1$ matrix: in fact, for a given facet F and vertex V, the (F,V)'th entry of $S(C_n)$ is given by:

$$S(C_n)_{F,V} = \begin{cases} 1 & \text{if } V \notin F \\ 0 & \text{if } V \in F \end{cases}.$$

Since the slack matrix of the hypercube $S(C_n)$ has $2n$ rows we clearly have

$$\left(\frac{\nu_+^{[0]}(S(C_n))}{\|S(C_n)\|_F} \right)^2 \le \operatorname{rank}_+(S(C_n)) \le 2n.$$

To show that we indeed have equality, we exhibit a particular feasible point W of the semidefinite program (3) and we show that this W satisfies $(\langle S(C_n), W \rangle / \|S(C_n)\|_F)^2 = 2n$.

Let

$$W = \frac{1}{\sqrt{2^{n-1}}}(2S(C_n) - J). \tag{19}$$

Observe that W is the matrix obtained from $S(C_n)$ by changing the ones into $\frac{1}{\sqrt{2^{n-1}}}$ and zeros into $-\frac{1}{\sqrt{2^{n-1}}}$. For this W we verify using a simple calculation that

$$\left(\frac{\langle S(C_n), W \rangle}{\|S(C_n)\|_F} \right)^2 = 2n.$$

It thus remains to prove that W is indeed a feasible point of the semidefinite program (3) and that the matrix

$$\begin{bmatrix} I & -W \\ -W^\mathsf{T} & I \end{bmatrix} \in \mathbb{R}^{(2n+2^n)\times(2n+2^n)}$$

can be written as the sum of a nonnegative matrix and a positive semidefinite one. This is the main part of the proof and for this we introduce some notations. Let $\mathcal{F}$ be the set of facets of the hypercube, and $\mathcal{V} = \{0,1\}^n$ be the set of vertices. If $F \in \mathcal{F}$ is a facet of the hypercube, we denote by $\overline{F}$ the opposite facet to F (namely, if F is given by $x_k \ge 0$, then $\overline{F}$ is the facet $x_k \le 1$ and vice-versa). Similarly for a vertex $V \in \mathcal{V}$ we denote by $\overline{V}$ the opposite vertex obtained by complementing the binary word V. Denote by $N_\mathcal{F} : \mathbb{R}^\mathcal{F} \to \mathbb{R}^\mathcal{F}$ and $N_\mathcal{V} : \mathbb{R}^\mathcal{V} \to \mathbb{R}^\mathcal{V}$ the "negation" maps, so that we have:

$$\forall g \in \mathbb{R}^\mathcal{F}, \ \forall F \in \mathcal{F}, \ (N_\mathcal{F} g)(F) = g(\overline{F}) \tag{20}$$

$$\forall h \in \mathbb{R}^\mathcal{V}, \ \forall V \in \mathcal{V}, \ (N_\mathcal{V} h)(V) = h(\overline{V}) \tag{21}$$

Note that in a suitable ordering of the facets and the vertices, the matrix representation of $N_\mathcal{F}$ and $N_\mathcal{V}$ take the following antidiagonal form ($N_\mathcal{F}$ is of size $2n \times 2n$ and $N_\mathcal{V}$ is of size $2^n \times 2^n$):

$$\begin{bmatrix} 0 & & & 1 \\ & & \cdot^{\cdot^\cdot} & \\ & \cdot^{\cdot^\cdot} & & \\ 1 & & & 0 \end{bmatrix}$$

Consider now the following decomposition of the matrix $\begin{bmatrix} I & -W \\ -W^\mathsf{T} & I \end{bmatrix}$:

$$\begin{bmatrix} I & -W \\ -W^\mathsf{T} & I \end{bmatrix} = \begin{bmatrix} N_\mathcal{F} & 0 \\ 0 & N_\mathcal{V} \end{bmatrix} + \begin{bmatrix} I - N_\mathcal{F} & -W \\ -W^\mathsf{T} & I - N_\mathcal{V} \end{bmatrix} \tag{22}$$

Clearly the first matrix in the decomposition is nonnegative. The next lemma states that the second matrix is actually positive semidefinite:

21

Lemma 2. *Let $\mathcal{F}$ and $\mathcal{V}$ be respectively the set of facets and vertices of the hypercube $C_n = [0,1]^n$ and let $\widehat{W} \in \mathbb{R}^{\mathcal{F} \times \mathcal{V}}$ be the matrix:*

$$\widehat{W}_{F,V} = \begin{cases} 1 & \text{if } V \notin F \\ -1 & \text{if } V \in F \end{cases} \quad \forall F \in \mathcal{F}, \ V \in \mathcal{V}$$

Then the matrix

$$\begin{bmatrix} I - N_{\mathcal{F}} & -\gamma \widehat{W} \\ -\gamma \widehat{W}^{\mathsf{T}} & I - N_{\mathcal{V}} \end{bmatrix} \tag{23}$$

is positive semidefinite for $\gamma = 1/\sqrt{2^{n-1}}$ (where $N_{\mathcal{F}}$ and $N_{\mathcal{V}}$ are defined in (20) and (21)).

Proof. We use the Schur complement to show that the matrix (23) is positive semidefinite. In fact we show that

1. $I - N_{\mathcal{V}} \succeq 0$

2. $\text{range}(\widehat{W}^{\mathsf{T}}) \subseteq \text{range}(I - N_{\mathcal{V}})$, and

3. $I - N_{\mathcal{F}} - \gamma^2 \widehat{W}(I - N_{\mathcal{V}})^{-1}\widehat{W}^{\mathsf{T}} \succeq 0$.

where $(I - N_{\mathcal{V}})^{-1}$ denotes the pseudo-inverse of $I - N_{\mathcal{V}}$.

Observe that for any $k \in \mathbb{N}$, the $2k \times 2k$ matrix given by:

$$I_{2k} - N_{2k} = \begin{bmatrix} 1 & & & -1 \\ & \ddots & \mathinner{\mkern1mu\raise1pt{.}\mkern2mu\raise4pt{.}\mkern2mu\raise7pt{.}} & \\ & \mathinner{\mkern1mu\raise1pt{.}\mkern2mu\raise4pt{.}\mkern2mu\raise7pt{.}} & \ddots & \\ -1 & & & 1 \end{bmatrix}$$

is positive semidefinite: in fact one can see that $\frac{1}{2}(I_{2k} - N_{2k})$ is the orthogonal projection onto the subspace spanned by its columns (i.e., the subspace of dimension k spanned by $\{e_i - e_{2k-i} \ : \ i = 1, \ldots, n\}$ where e_i is the i'th unit vector). Hence this shows that $I - N_{\mathcal{V}}$ is positive semidefinite, and it also shows that $(I - N_{\mathcal{V}})^{-1} = \frac{1}{4}(I - N_{\mathcal{V}})$.

Now we show that $\text{range}(\widehat{W}^{\mathsf{T}}) \subseteq \text{range}(I - N_{\mathcal{V}})$. For any $F \in \mathcal{F}$, the F'th column of $\widehat{W}^{\mathsf{T}}$ satisfies $(\widehat{W}^{\mathsf{T}})_{V,F} = -(\widehat{W}^{\mathsf{T}})_{\overline{V},F}$ for any $V \in \mathcal{V}$, and thus $\text{range}(\widehat{W}^{\mathsf{T}}) \subseteq \text{span}(e_V - e_{\overline{V}}, \ : \ V \in \mathcal{V}) = \text{range}(I - N_{\mathcal{V}})$.

It thus remains to show that

$$I - N_{\mathcal{F}} - \gamma^2 \widehat{W}(I - N_{\mathcal{V}})^{-1}\widehat{W}^{\mathsf{T}} \succeq 0$$

First note that since $\frac{1}{2}(I - N_{\mathcal{V}})$ is an orthogonal projection and that $\text{range}(\widehat{W}^{\mathsf{T}}) \subseteq \text{range}(I - N_{\mathcal{V}})$, we have $(I - N_{\mathcal{V}})^{-1}\widehat{W}^{\mathsf{T}} = \frac{1}{2}\widehat{W}^{\mathsf{T}}$. Thus we now have to show that

$$I - N_{\mathcal{F}} - \frac{\gamma^2}{2}\widehat{W}\widehat{W}^{\mathsf{T}} \succeq 0.$$

22

The main observation here is that the matrix $\widehat{W}\widehat{W}^{\mathsf{T}}$ is actually equal to $2^n(I - N_{\mathcal{F}})$. For any $F, G \in \mathcal{F}$, we have:

$$(\widehat{W}\widehat{W}^{\mathsf{T}})_{F,G} = \sum_{a \in \mathcal{V}} \widehat{W}_{F,a} \widehat{W}_{G,a} = \begin{cases} 2^n & \text{if } F = G \\ -2^n & \text{if } F = \overline{G} \\ 0 & \text{else} \end{cases}$$

First it is clear that if $F = G$, then $(\widehat{W}\widehat{W}^{\mathsf{T}})_{F,G} = 2^n$. Also if $F = \overline{G}$ then $(\widehat{W}\widehat{W}^{\mathsf{T}})_{F,G} = -2^n$ since if $F = \overline{G}$ then $a \in F \Leftrightarrow a \notin G$ hence $\widehat{W}_{F,a}\widehat{W}_{G,a} = -1$ for all $a \in \mathcal{V}$. In the case that $F \neq G$ and $F \neq \overline{G}$, it is easy to verify by simple counting that $\sum_{a \in \mathcal{V}} \widehat{W}_{F,a}\widehat{W}_{G,a} = 0$.

Hence we have $I - N_{\mathcal{F}} - \gamma^2 \widehat{W}(I - N_{\mathcal{V}})^{-1}\widehat{W}^{\mathsf{T}} = I - N_{\mathcal{F}} - \frac{\gamma^2}{2}\widehat{W}\widehat{W}^{\mathsf{T}} = (1 - \frac{\gamma^2}{2}2^n)(I - N_{\mathcal{F}})$ which is positive semidefinite for $\gamma = 1/\sqrt{2^{n-1}}$. $\qquad\square$

Using this lemma, Equation (22) shows that the matrix W is feasible for the semidefinite program (3), and thus that $\nu_+^{[0]}(S(C_n)) \geq \langle S(C_n), W \rangle = \sqrt{2^{n-1}} \cdot 2n$. Hence since $\|S(C_n)\|_F = \sqrt{2^{n-1} \cdot 2n}$ we get that

$$\text{rank}_+(S(C_n)) \geq \left(\frac{\nu_+^{[0]}(S(C_n))}{\|S(C_n)\|_F} \right)^2 \geq \left(\frac{\sqrt{2^{n-1}} \cdot 2n}{\sqrt{2^{n-1} \cdot 2n}} \right)^2 = 2n$$

which completes the proof. $\qquad\square$

References

[AGKM12] Sanjeev Arora, Rong Ge, Ravindran Kannan, and Ankur Moitra. Computing a nonnegative matrix factorization – provably. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 145–162. ACM, 2012. 2

[Ber09] Dennis S Bernstein. *Matrix mathematics: theory, facts, and formulas (Second Edition)*. Princeton University Press, 2009. 13

[BJLP13] Gábor Braun, Rahul Jain, Troy Lee, and Sebastian Pokutta. Information-theoretic approximations of the nonnegative rank. *ECCC preprint TR13-158*, 2013. 6

[BSM03] Abraham Berman and Naomi Shaked-Monderer. *Completely positive matrices.* World Scientific Pub Co Inc, 2003. 4, 20

[CR93] J.E. Cohen and U.G. Rothblum. Nonnegative ranks, decompositions, and factorizations of nonnegative matrices. *Linear Algebra and its Applications*, 190:149–168, 1993. 14

[Dür10] M. Dür. Copositive programming–a survey. *Recent Advances in Optimization and its Applications in Engineering*, pages 3–20, 2010. 3

[DV13]   X. Doan and S. Vavasis. Finding approximately rank-one submatrices with the nuclear norm and ℓ_1-norm. *SIAM Journal on Optimization*, 23(4):2502–2540, 2013. 9

[FKPT13]   Samuel Fiorini, Volker Kaibel, Kanstantsin Pashkovich, and Dirk Oliver Theis. Combinatorial bounds on nonnegative rank and extended formulations. *Discrete Mathematics*, 313(1):67 – 83, 2013. 4, 15

[FP14]   Hamza Fawzi and Pablo A. Parrilo. Self-scaled bounds for atomic cone ranks: applications to nonnegative rank and cp-rank. *arXiv preprint arXiv:1404.3240*, 2014. 3, 20

[GG12]   Nicolas Gillis and François Glineur. On the geometric interpretation of the nonnegative rank. *Linear Algebra and its Applications*, 437(11):2685–2712, 2012. 2

[Gil12]   Nicolas Gillis. Sparse and unique nonnegative matrix factorization through data preprocessing. *Journal of Machine Learning Research*, 13:3349–3386, 2012. 16

[Goe14]   Michel Goemans. Smallest compact formulation for the permutahedron. *Mathematical Programming*, 2014. 15

[GP04]   Karin Gatermann and Pablo A. Parrilo. Symmetry groups, semidefinite programs, and sums of squares. *Journal of Pure and Applied Algebra*, 192(1):95–128, 2004. 12

[GPT13a]   João Gouveia, Pablo A Parrilo, and Rekha R Thomas. Approximate cone factorizations and lifts of polytopes. *arXiv preprint arXiv:1308.2162*, 2013. 6

[GPT13b]   João Gouveia, Pablo A. Parrilo, and Rekha R. Thomas. Lifts of convex sets and cone factorizations. *Mathematics of Operations Research*, 38(2):248–264, 2013. 20

[GRT13]   João Gouveia, Richard Z. Robinson, and Rekha R. Thomas. Polytopes of minimum positive semidefinite rank. *Discrete & Computational Geometry*, 50(3):679–699, 2013. 15

[Jam87]   G.J.O. Jameson. *Summing and nuclear norms in Banach space theory*, volume 8. Cambridge University Press, 1987. 7

[JSWZ13]   R. Jain, Yaoyun Shi, Zhaohui Wei, and Shengyu Zhang. Efficient protocols for generating bipartite classical distributions and quantum states. *IEEE Transactions on Information Theory*, 59(8):5171–5178, 2013. 2

[KN06]   E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 2006. 1, 2

[KP02]   E. Klerk and D.V. Pasechnik. Approximation of the stability number of a graph via copositive programming. *SIAM Journal on Optimization*, 12(4):875–892, 2002. 12

[Löf04]   J. Löfberg. YALMIP: A toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004. 12

[Löf09]   J. Löfberg. Pre- and post-processing sum-of-squares programs in practice. *IEEE Transactions on Automatic Control*, 54(5):1007–1011, 2009. 12

[Lov90]   László Lovász. Communication complexity: A survey. *Paths, Flows, and VLSI-Layout*, page 235, 1990. 1, 2

[LS09]   Troy Lee and Adi Shraibman. *Lower bounds in communication complexity*, volume 3. NOW Publishers Inc, 2009. 1, 2

[Par00]   Pablo A. Parrilo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*. PhD thesis, California Institute of Technology, 2000. 3, 11, 12

[Rot14]   Thomas Rothvoss. The matching polytope has exponential extension complexity. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, pages 263–272. ACM, 2014. 2

[Vav09]   Stephen A. Vavasis. On the complexity of nonnegative matrix factorization. *SIAM Journal on Optimization*, 20(3):1364–1377, 2009. 2

[Yan91]   Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences*, 43(3):441–466, 1991. 1, 2, 15

[Zha12]   S. Zhang. Quantum strategic game theory. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pages 39–59. ACM, 2012. 2