

Lower bounds on the nonnegative rank using a nested polytopes formulation

Julien Dewez and François Glineur

UCLouvain - ICTEAM/INMA and CORE
Avenue Georges Lemaître, 4, Louvain-la-Neuve - Belgium

Abstract. Computing the nonnegative rank of a nonnegative matrix has been proven to be, in general, NP-hard [1]. However, this quantity has many interesting applications, e.g., it can be used to compute the extension complexity of polytopes [2]. Therefore researchers have been trying to approximate this quantity as closely as possible with strong lower and upper bounds. In this work, we introduce a new lower bound on the nonnegative rank based on a representation of the matrix as a pair of nested polytopes. The nonnegative rank then corresponds to the minimum number of vertices of any polytope nested between these two polytopes. Using the geometric concept of supporting corner, we introduce a parametrized family of computable lower bounds and present preliminary numerical results on slack matrices of regular polygons.

1 Nonnegative rank and nested polytope problem

Introduction The nonnegative rank of an entry-wise nonnegative matrix, denoted rank_+ , is the minimum number of nonnegative rank-one matrices whose sum is equal to that matrix. Equivalently, for a matrix $M \in \mathbb{R}_+^{m \times n}$ it is the smallest positive integer r such that there exists $U \in \mathbb{R}_+^{m \times r}$ and $V \in \mathbb{R}_+^{r \times n}$ such that $M = UV$. The nonnegative rank features several interesting applications, such as providing a way to compute the extension complexity of a polytope (which is equal to the nonnegative rank of its slack matrix¹ [2]). Computing the nonnegative rank and a corresponding factorization has been proven to be, in general, NP-hard [1]. Hence several tractable lower bounds have been developed to approximate it as closely as possible, see for example [3, 4, 5, 6]. In this work, we introduce a new lower bound on the nonnegative rank based on a representation of the matrix as a pair of nested polytopes.

Nested polytope problem We first describe how a nonnegative matrix can be represented by a pair of nested polytopes, as introduced in [7], and show that the nonnegative rank of that matrix can be found by solving a geometric problem involving that representation, called the nested polytope problem (NPP).

Definition 1 (NPP). *Let O be a polytope given by its facets and let I be a polytope given by its vertices such that $I \subseteq O$. Find the minimum number of points belonging to O whose convex hull contains I , i.e., find the minimum k such that there exists k points $\{p_i\}$ satisfying $I \subseteq T = \text{conv}\{p_1, \dots, p_k\} \subseteq O$.*

¹A polytope can be defined either as the convex hull of its vertices, $\text{conv}\{v_1, \dots, v_n\}$ or an intersection of half-spaces, $\{x : a_i^\top x \leq b_i, i = 1, \dots, m\}$. The slack matrix of such a polytope is defined as the $m \times n$ nonnegative matrix S such that $S_{i,j} = b_i - a_i^\top v_j$.

While there exists a polynomial time algorithm to solve this problem if I and O are polygons, i.e., 2-dimensional polytopes [8], it has been recently proven that this problem is in general $\exists\mathbb{R}$ -complete [9]. This complexity class consists of problems that can be polynomially reduced to the problem of deciding whether a system of polynomial equalities and inequalities admits a solution.

Representation as nested polytopes Let $M \in \mathbb{R}_+^{m \times n}$ be a nonnegative matrix. Without loss of generality, let us assume that M has no column containing only zeros (as those do not modify the nonnegative rank) and is column-stochastic (as applying a positive scaling to the columns of a matrix does not modify its nonnegative rank [7]). Note that we could alternatively assume that M is row-stochastic and use its transpose to obtain another representation as a pair of nested polytopes.

We build a pair of nested polytopes corresponding to such a matrix M in the following way. First, we pose the problem in \mathbb{R}^m and define the outer polytope O to be the simplex given by $O = \{x \in \mathbb{R}^m \mid x_i \geq 0 \forall i \text{ and } \sum_{i=1}^m x_i = 1\}$. We then use the columns of M as the n points whose convex hull defines the inner polytope I (and indeed we can check that $I \subseteq O$).

Theorem 1. *Let M be a column-stochastic matrix and let $I \subseteq O$ be the pair of nested polytopes corresponding to M . Let k be the solution of this NPP instance, then $\text{rank}_+(M) = k$. (all proofs are omitted due to space restrictions)*

Another way of representing a nonnegative matrix as a NPP instance was introduced in [6]. In that representation, the solution of the NPP instance corresponds to the so-called restricted nonnegative rank, a variation of the nonnegative rank which requires that the column-space of the first factor U is the same as the column-space of the original matrix M . In that representation, the inner and outer polytopes have the same affine hull dimension, which is not always the case in the representation described in this work.

As it has been recently proven that NPP is $\exists\mathbb{R}$ -complete, solving the NPP instance corresponding to a nonnegative matrix is not a priori easier than computing the nonnegative rank directly. However, we can derive a lower bound on the nonnegative rank from geometric properties of the NPP instance.

2 Lower bounds

2.1 Supporting corners and disjoint corners bound

First recall the concept of support function.

Definition 2. *Let d be a direction in \mathbb{R}^m , i.e., a nonzero unit vector, and let P be a polytope. Denote the usual inner product by $\langle \cdot, \cdot \rangle$. The support function of P with respect to d is given by $\sigma(d, P) = \max_{x \in P} \langle d, x \rangle$.*

Let us introduce the following notion illustrated on Figure 1.

Definition 3. *Let $I \subseteq O \subseteq \mathbb{R}^m$ be a pair of nested polytopes defining a NPP instance and let d be a direction in \mathbb{R}^m . Define a supporting corner as*

$$C(d) = \{x \in O \mid \langle d, x \rangle \geq \sigma(d, I)\}.$$

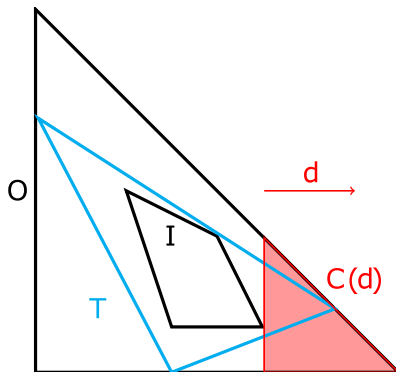


Fig. 1: NPP instance (black) with solution T (blue) and supporting corner $C(d)$ (red). Instance shown is for $m = 3$, but depicted after projection on (x_1, x_2) , i.e., $O = \{(x_1, x_2) : x_1, x_2, x_3 \geq 0, x_1 + x_2 + x_3 = 1 \Leftrightarrow x_1, x_2 \geq 0, x_1 + x_2 \leq 1\}$.

Note that set $\{x \in \mathbb{R}^m \mid \langle d, x \rangle = \sigma(d, I)\}$ is a supporting hyperplane of I , which means that a supporting corner can also be defined using any supporting hyperplane of I (in which case the direction is the normal to the hyperplane). We now introduce the main property on which our new lower bound is based.

Property 1. *Let $I \subseteq O$ be a pair of nested polytopes and T any polytope nested in between, i.e., such that $I \subseteq T \subseteq O$. Any supporting corner must contain at least one vertex of T .*

Using this property, we can introduce several ways to compute lower bounds on the solution of the NPP, which translates into lower bounds on the nonnegative rank. A first idea consists in finding the maximum number of disjoint supporting corners.

Theorem 2. *Let $I \subseteq O$ be a pair of nested polytopes. Let d_1, \dots, d_k be k directions such that the corresponding supporting corners $C(d_i)$ are pairwise disjoint. Then, any polytope T such that $I \subseteq T \subseteq O$ must have at least k vertices.*

However, if the inner polytope is small relative to the outer polytope, it can be difficult to find a large number of pairwise disjoint supporting corners, leading to a weak bound. The next section introduces a stronger bound.

Infinite dimensional reformulation We first propose an equivalent formulation of NPP using the notion of supporting corner: the minimum number of vertices of a nested polytope is equal to the minimum number of points that simultaneously satisfy Property 1 for all directions/supporting corners.

Theorem 3. *Let $I \subseteq O$ be a pair of polytopes defining a NPP instance. Let \mathcal{D} be the set of all directions. The minimum number of vertices of a polytope T such that $I \subseteq T \subseteq O$ is equal to the minimum number of points $p_1, \dots, p_k \in O$ such that $\forall d \in \mathcal{D}, \exists i \in \{1, \dots, k\}$ such that $p_i \in C(d)$.*

2.2 Relaxation and additional bounds

This formulation is infinite-dimensional (both because of its variables, which are all points in O , and of its constraints, indexed by all directions in \mathcal{D}) and cannot be solved exactly. Hence we propose to relax it, which will provide lower bounds on its optimal value. First, we select $\hat{\mathcal{D}} \subset \mathcal{D}$, a finite subset of directions, implying we will only consider a finite number of supporting corners (and a finite number of constraints). Now, each points in O will belong to a certain list of supporting corners and, for the purpose of the reformulation of Theorem 3, two points belonging to the same list of supporting corners are completely equivalent. Therefore we can partition all points in O into a finite number of equivalence classes, denoted by \hat{O} . Finally, we introduce binary variables x_j (to select the points/equivalence classes) and end up with the following (where $\mathcal{M}_{d,j}$ is a binary indicator of inclusion $x_j \in C(d)$)

$$L_{0,1}(\hat{O}, \hat{\mathcal{D}}) = \min_{x_j \in \{0,1\}, j \in \hat{O}} \sum_{j \in \hat{O}} x_j \text{ s.t. } \sum_{j \in \hat{O}} \mathcal{M}_{d,j} x_j \geq 1, \forall d \in \hat{\mathcal{D}}$$

This finite integer linear optimization problem (or its continuous relaxation with $x_j \geq 0$ instead of $x_j \in \{0,1\}$, denoted $L_{\mathbb{R}}(\hat{O}, \hat{\mathcal{D}})$) can now be solved numerically, and will give a lower bound $L_{\bullet}(\hat{O}, \hat{\mathcal{D}}) \leq \text{rank}_+(M)$ for any set of directions $\hat{\mathcal{D}}$.

The dual of this linear program is also interesting, because its interpretation links back to the initial lower bound from Theorem 2, i.e., finding the largest number of pairwise disjoint corners. The dual is given by

$$\Gamma_{\mathbb{R}}(\hat{O}, \hat{\mathcal{D}}) = \max_{y_d \geq 0, d \in \hat{\mathcal{D}}} \sum_{d \in \hat{\mathcal{D}}} y_d \text{ s.t. } \sum_{d \in \hat{\mathcal{D}}} \mathcal{M}_{d,j} y_d \leq 1, \forall j \in \hat{O}$$

In the integer version of this problem, we want to find the largest number of directions/supporting corners such that each point/equivalence class is contained in at most one, i.e., this consists exactly in finding the largest number of pairwise disjoint corners. Finally, the optimal solutions of these problems satisfy the following relation

$$\Gamma_{0,1}(\hat{O}, \hat{\mathcal{D}}) \leq \Gamma_{\mathbb{R}}(\hat{O}, \hat{\mathcal{D}}) = L_{\mathbb{R}}(\hat{O}, \hat{\mathcal{D}}) \leq L_{0,1}(\hat{O}, \hat{\mathcal{D}}) \leq L_{0,1}(O, \mathcal{D}) = \text{rank}_+(M)$$

This shows that solving any of these problems (with any set of directions $\hat{\mathcal{D}}$) gives a lower bound on the nonnegative rank of the input matrix.

3 Numerical experiments

3.1 Computing lower bounds numerically

In this section, we perform some preliminary numerical experiments. To compute a lower bound, we must generate a set of directions. We use the following set $\hat{\mathcal{D}}_N$ which attempts to spread directions over the whole space (we also tried random generation of directions, and observed similar but slightly worse results)

$$\hat{\mathcal{D}}_N = \left\{ d \in \mathbb{R}^m : \sum_{i=1}^m d_i^2 = 1, d_i^2 = \frac{k}{N} \text{ for some integer } k \in [0, N] \right\}.$$

To define the set \hat{O} of equivalent classes of points, we must in principle compute all intersections between sets of supporting corners. However, the number of regions and the computation time then grows exponentially with the number of directions. To avoid this, we use instead a fixed discretization of the outer polytope, independent of the set of directions. We chose to use small hypercubes (which may be truncated) as regions by dividing each interval $[0,1]$ on the axes of the space in K subintervals of same length.

$$\hat{O}_K = \left\{ C_1 \times \dots \times C_m : C_i = \left[\frac{k}{K}, \frac{k+1}{K} \right], \text{ for some } k = 0, \dots, K-1 \right\} \cap O.$$

We then adapt the definition of the $L_{0,1}(\hat{O}, \hat{D})$ linear program to preserve the lower bound property: binary indicators of inclusion are now equal to one as soon as a region in \hat{O}_K features a nonempty intersection with a corner.

3.2 Numerical results on slack matrices

We tested the lower bound on the slack matrices of regular polygons, from 4-gon to 8-gon. The slack matrix of a regular n -gon is a circulant matrix and its first row is symmetric ($c_k = c_{n-k-1}$) with $c_k = \cos\left(\frac{\pi}{n}\right) - \cos\left(\frac{(2k+1)\pi}{n}\right)$. As $L_{0,1}(\hat{O}, \hat{D}) = [L_{\mathbb{R}}(\hat{O}, \hat{D})]$ always holds in our tests, we only focus on $L_{\mathbb{R}}(\hat{O}, \hat{D})$.

ID	4-gon	5-gon	6-gon	7-gon	8-gon
rank ₊	4	5	5	6	6
$ \hat{O}_7 / \hat{D}_6 $	88/608	236/1970	562/5336	1219/12642	2452/27008
$L_{\mathbb{R}}(\hat{O}_7, \hat{D}_6)$	4	4.25	4.05	3.5	3.25
Time[s]	0.81	1.71	6.87	38.4	247
$L_{\mathbb{R}}(\hat{O}_7, \hat{D})$	4	4.33	4.5	3.5	-
Time[s]	7	357	19072	*	-
$ \hat{D} $	8	36	46	164	-
From [4]/[5]	4/4	5/5	4.69/5	5.03/6	5.15/6

Table 1: NPP bound for regular polygons. In bold, values that close the gap with the true value of rank₊ (possibly after rounding up)

We fix the direction parameter $N = 6$ and outer discretization parameter $K = 7$, and display the corresponding numbers of corners/regions in the table. Computed values of $L_{\mathbb{R}}(\hat{O}_7, \hat{D}_6)$ for each slack matrix are reported, with CPU time, in the third row of Table 1; we observe that some lower bounds are tight (in bold). Bounds for the 7- and 8-gon are weaker, and we suspect that a larger value of parameter K is needed to close the gap (however the bound is not monotone in K : for example, for the 5-gon, we obtain $L_{\mathbb{R}}(\hat{O}_8, \hat{D}_6) = 4 \not\geq 4.25$).

Dynamic generation of directions When moving to larger slack matrices, the number of directions grows, increasing our computational effort. To counter this effect, we tested a procedure that successively adds directions to the problem (instead of considering a fixed set \hat{D}_N). New directions are chosen so that they

reject the previous optimal solution, hoping to increase the value of the bound; they are computed using a mixed-integer linear program (using one binary variable per region in \hat{O}_K). We stop when the (rounded) bound is equal the true nonnegative rank, or when no new direction can be found.

Results are displayed on the fourth row of Table 1. This dynamic approach appears to require more computational effort. Nevertheless, we observe for 4-, 5- and 6-gons that it finds lower bounds that are as good as the fixed \hat{D}_6 bound, but with much fewer directions (for instance 46 directions instead of more than five thousands for the 6-gon). In essence this provides us with more compact certificates for those nonnegative ranks. The case of the dynamic approach for the 7-gon is slightly different: it was started with the full set of directions \hat{D}_6 , which gives a bound equal to 3.5, and the goal was to improve that bound. Instead, the procedure found that no direction can improve the bound, implying that improving the bound requires to increase the outer discretization parameter K . This seems to support our claim that the higher the value of n , the finer the discretization is needed to be to obtain good lower bounds.

4 Conclusion

In this paper, we introduced new lower bounds on the nonnegative rank based on a nested polytopes formulation. Numerical experiments on small slack matrices demonstrate that this approach is promising, as the bound can be close to the true nonnegative rank or even tight. In the future, we plan to focus on the computational efficiency of our procedure, which will allow tests on larger matrices with higher values of the discretization parameters N and K .

References

- [1] Stephen A. Vavasis. On the complexity of nonnegative matrix factorization. *SIAM Journal on Optimization*, 20(3):1364–1377, 2009.
- [2] Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. *Journal of computer and system sciences*, 43:441–466, 1991.
- [3] Hamza Fawzi and Pablo A Parrilo. Lower bounds on nonnegative rank via nonnegative nuclear norms. *Mathematical Programming*, 153(1):41–66, 2015.
- [4] Hamza Fawzi and Pablo A Parrilo. Self-scaled bounds for atomic cone ranks: applications to nonnegative rank and cp-rank. *Mathematical Programming*, 158(1-2):417–465, 2016.
- [5] 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.
- [6] Nicolas Gillis and François Glineur. On the geometric interpretation of the nonnegative rank. *Linear Algebra and its Applications*, 437(11):2685–2712, 2012.
- [7] Joel E. Cohen and Uriel G. Rothblum. Nonnegative ranks, decompositions, and factorizations of nonnegative matrices. *Linear Algebra and its Applications*, 190:149–168, 1993.
- [8] Alok Aggarwal, Heather Booth, Joseph O’Rourke, Suri Surhash, and Chee K. Yap. Finding minimal convex nested polygons. *Information and Computation*, 83(1):98–110, 1989.
- [9] Michael G Dobbins, Andreas Holmsen, and Tillmann Miltzow. A universality theorem for nested polytopes. *arXiv preprint arXiv:1908.02213*, 2019.