Hardness of some optimization problems over correlation polyhedra

ALBERTO CAPRARA[1], FABIO FURINI[2], CLAUDIO GENTILE[3], LEO LIBERTI[4], ANDREA LODI[5,3]

[1] *Università di Bologna, Italy († 2012)*

[2] *Department of Computer, Control and Management Engineering "A. Ruberti" — Sapienza University of Rome. Via Ariosto 25, 00185 Roma, Italy*
Email:fabio.furini@uniroma1.it

[3] *Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti", Consiglio Nazionale delle Ricerche, Via dei Taurini 19, 00185 Roma, Italy*
Email:gentile@iasi.cnr.it

[4] *LIX CNRS, École Polytechnique, Institut Polytechnique de Paris, F-91128 Palaiseau, France*
Email:liberti@lix.polytechnique.fr

[5] *CornellTech, New York, USA*
Email:al748@cornell.edu

March 19, 2026

Abstract

We prove the **NP**-hardness, using Karp reductions, of some problems related to the correlation polytope and its corresponding cone, spanned by all of the $n \times n$ rank-one matrices over $\{0, 1\}$. The problems are: membership, rank of the decomposition, and a "relaxed rank" obtained from relaxing the zero-norm expression for the rank to an ℓ_1 norm. While membership and rank are natural problems for any matrix cone, the relaxed rank problem occurs in some signal processing and statistical applications.

Keywords: matrix decomposition, cone, cut polytope, boolean quadric polytope.

1 Introduction

For any finite set $\mathcal{M}$ of matrices one can define the cone $\mathsf{cone}(\mathcal{M})$ and the polytope $\mathsf{cvx}(\mathcal{M})$. The fundamental question in this setting is that of membership: given a matrix M', does it belong to the cone or the convex hull of $\mathcal{M}$? We note that $M' \in \mathsf{cone}(\mathcal{M})$ is equivalent to

$$\exists p \in \mathbb{R}_+^{|\mathcal{M}|} \quad M' = \sum_{M \in \mathcal{M}} p_M M, \tag{1}$$

while $M' \in \mathsf{cvx}(\mathcal{M})$ constrains p to be in the unit simplex, i.e. it adds the constraint $\sum_{\mu \in \mathcal{M}} p_\mu = 1$ to Eq. (1). A second question is that of rank: what is the minimum number of elements of $\mathcal{M}$ using which one can represent M? Or, equivalently, what is the minimum possible value of $\|p\|_0$ for M? A third relevant question is that of minimizing the relaxed rank $\|p\|_1$, given that the "mathematics of sparsity" [5] promises to turn a problem involving binary variables into one involving continuous variables only. Obviously, the relaxed rank question only makes sense for $\mathsf{cone}(\mathcal{M})$ since the convex hull fixes $\|p\|_1 = 1$ (recall that $p \geq 0$).

We consider the **NP**-hardness of these three questions for a special case where $\mathcal{M}$ is the set of rank-one matrices with entries in $\{0, 1\}$. The rank and relaxed rank minimization problems then become decision questions: is the rank, or relaxed rank, less than a given threshold or not? For rank and relaxed rank decision problems we prove **NP**-completeness "with promise", i.e. we assume that M satisfies Eq. (1)

or its convex hull counterpart, otherwise hardness follows trivially from membership. The fulfilment of the promise is given by a single bit of information: i.e. we know that M can be decomposed as per Eq. (1), but we do not have the weights p_μ in the decomposition. Therefore, the "rank of p" and "relaxed rank of p" $\leq$ threshold questions are non-trivial. We provide Karp reductions for every one of our hardness proofs: such reductions are stronger than Cook-Turing reductions (typically afforded by the optimization/separation/membership equivalence [14]) because they are based on a single query to the target problem.

We now consider $\mathcal{M}$ as a set of rank-1 symmetric matrices with entries in sets of two integers, either $\{0, 1\}$ or $\{-1, 1\}$. We focus mostly on the former, but obtain results on both. Given a symmetric matrix Γ of dimension $n \times n$, with entries $\Gamma_{ij} = \Gamma_{ji} \in \mathbb{R}$, we consider its decomposition in the following form:

$$\Gamma = \sum_{k \geq 0} p_k x_k x_k^\top, \tag{2}$$

where x_k is the vector in $\{0, 1\}^n$ providing the base-2 representation of the integer k, and the multiplier vector p is non-negative. From this, it follows that $0 \leq k \leq P_n = 2^n - 1$, so $p \in \mathbb{R}_+^{2^n}$, and that all $n \times n$ matrices $X^k = x_k x_k^\top$ are rank-1 symmetric and defined on $\{0, 1\}$. Therefore, no matrix Γ can be decomposed in the form Eq. (2) unless it is symmetric and all its entries are non-negative. For a given Γ, we are also interested in the smallest number of terms of the decomposition having $p_k > 0$, i.e. its rank. If we let $\mathcal{C}(n)$ be the cone spanned by all of the rank-one boolean symmetric matrices X^k, then the feasibility of Eq. (2) is the membership problem for the conic hull $\mathcal{C}(n) \triangleq \mathsf{cone}(X)$, where $X = \{X^k \mid 0 \leq k \leq P_n\}$.

Whenever Eq. (2) is subject to the constraint $\sum_{k \geq 0} p_k = 1$, we obtain the definition of the *correlation polytope* of order n, denoted by $\mathsf{COR}(n)$. This also establishes $\mathcal{C}(n)$ as the *correlation cone* of order n.

The rank problem for a given $\Gamma \in \mathcal{C}(n)$ can be formulated as the minimization of the zero norm of p subject to Eq. 2, i.e.

$$\min\{\|p\|_0 \mid \Gamma = \sum_{0 \leq k \leq P_n} p_k X^k \wedge p \geq 0\}, \tag{3}$$

which an obvious reformulation turns into an exponentially-sized Mixed-Integer Linear Program (MILP). We observe that the linear system Eq. (2) has $N = n(n + 1)/2$ equations, which implies that at most N of the exponentially many variables p_k need be nonzero: this provides an upper bound to the rank. Following the well-known relaxation of the zero-norm to ℓ_1 norm [6], we obtain an exponentially-sized Linear Programming (LP) formulation that computes the *relaxed rank* of a matrix $\Gamma \in \mathcal{C}(n)$, i.e.

$$\min\{\|p\|_1 \mid \Gamma = \sum_{0 \leq k \leq P_n} p_k X^k \wedge p \geq 0\}. \tag{4}$$

Note that, since $p \geq 0$, we have $\|p\|_1 = \sum_{0 \leq k \leq P_n} p_k = \langle \mathbf{1}, p \rangle$.

Membership and rank problems are fundamental problems for any matrix decomposition. The relaxed rank becomes meaningful whenever the results from the "mathematics of sparsity" [5, 19] apply. These results aim at finding sparse vectors as solutions of various problems. These results are relevant when the relaxed rank problem is used as a proxy to the rank problem: namely, this can be done when the number of constraints of a (large) LP is at least of the order of magnitude of the logarithm of the number of variables. Such an approximation is justified in our case because $N = O(n^2)$ and $\log(P_n) = O(n)$. This makes the relaxed rank problem likely to be easier to solve, in practice, than the rank problem, although both are equally hard in theory. We note that the relaxed rank problem appears in [7].

In summary, we consider the following decision problems:

1. conic hull membership: is $\Gamma \in \mathcal{C}(n)$?

2. conic hull rank: if $\Gamma \in \mathcal{C}(n)$, is $\|p\|_0 \leq \rho$?

3. conic hull relaxed rank: if $\Gamma \in \mathcal{C}(n)$, is $\|p\|_1 \leq \rho$?

4. convex hull membership: $\Gamma \in \mathsf{COR}(n)$?

5. convex hull rank: if $\Gamma \in \mathsf{COR}(n)$, is $\|p\|_0 \leq \rho$?

We recall that the relaxed rank of the convex hull is not considered because $\sum_k p_k = \|p\|_1 = 1$ by definition, so every instance in $\mathsf{COR}(n)$ has answer YES as long as $\rho \geq 1$, and every other instance has answer NO: the problem therefore reduces to checking whether $\rho \geq 1$. We state outright that Problem 4 in the above list is already known to be **NP**-hard by [24]. At first look, Problem 5 above would appear to be related to [1, Problem 5]; but, on closer inspection, the vertex set is part of the input in [1], which rules out our problems since Eq. (2) lists exponentially many terms in n.

In this paper we prove four main theorems, to the effect that membership, rank, and relaxed rank problems with respect to $\mathcal{C}(n)$, as well as rank with respect to $\mathsf{COR}(n)$, are all **NP**-hard. Together with the result in [24], this paper settles all the problems in the above list. As corollaries, we also prove **NP**-hardness of: membership and rank problems for the cut polytope [11]; membership and rank problems for cut and correlation polytopes without the zero element; and membership, rank, and relaxed rank problem for the cut cone. We note that the boolean quadric polytope (BQP) [23] is essentially the same as $\mathsf{COR}(n)$, so it inherits the same hardness results (see Sect. 2.2).

The rest of this paper is organized as follows. A literature review is given in Sect. 2, with the purpose of justifying our interest in the above problems, and clarifying the reasons why we need new hardness proofs for the problems in the list above. In Sect. 3, we give a few easy necessary conditions for matrices Γ to be in $\mathcal{C}(n)$. The four main results and their corollaries are given in Sect. 4. Some polynomial cases and other related results are discussed in Sect. 5. Sect. 6 concludes the paper.

2 Literature review

This section gives an overview of the literature about complexity in the correlation polytopes and some other related polytopes and polyhedra. Our study of the relevant literature shows that the **NP**-hardness of Problems 1-5 of the list in Sect. 1 is only clearly established, with Karp reductions, for Problem 4.

In the following, we shall cite [14] for two purposes: deriving **NP**-hardness proofs concerning two affinely isomorphic polytopes, and to support the polynomial-time equivalence, under Cook-Turing reductions, of linear optimization separation and membership for rational polyhedra.

2.1 The correlation polytope

The correlation polytope [24] is $\mathsf{COR}(n) \triangleq \mathsf{cvx}(X)$, i.e. the convex hull of X. Its formulation is

$$\left. \begin{aligned} \sum_{0 \leq k \leq P_n} p_k x_k x_k^\top &= \Gamma \\ \sum_{0 \leq k \leq P_n} p_k &= 1. \end{aligned} \right\} \tag{5}$$

In the space of the p variables, $\mathsf{COR}(n)$ is the intersection of the cone $\mathcal{C}(n)$ (right-hand-side of Eq. (2)) with the affine subspace $\langle \mathbf{1}, p \rangle = 1$. In Eq. (2), we note that the zero matrix bears zero contribution to the sum, independently of the value of p_0: instead of summing over $k \geq 0$, we could therefore sum over $k \geq 1$. This corresponds to removing the zero matrix from $\mathsf{COR}(n)$: we define the *normalized correlation polytope* as $\overline{\mathsf{COR}}(n) \triangleq \mathsf{cvx}(X \smallsetminus 0)$.

2.2 The boolean quadric polytope

The BQP of order n is the convex hull of pairs (x, y) for all $x \in \{0, 1\}^n$ and $y \in \{0, 1\}^{n(n-1)/2}$ that satisfy

$$\forall 1 \leq i \leq j \leq n \quad y_{ij} \quad \geq \quad 0 \tag{6}$$
$$\forall 1 \leq i \leq j \leq n \quad y_{ij} \quad \leq \quad x_i \tag{7}$$
$$\forall 1 \leq i \leq j \leq n \quad y_{ij} \quad \leq \quad x_j \tag{8}$$
$$\forall 1 \leq i \leq j \leq n \quad y_{ij} \quad \geq \quad x_i + x_j - 1. \tag{9}$$

For any $x \in \{0, 1\}$, it is well-known [12] that $y_{ij} = x_i x_j$ for all $i < j \leq n$. Note that, should we extend y to also include y_{ii}, we would obtain $y_{ii} = x_i x_i = x_i^2 = x_i$ since $x_i \in \{0, 1\}$. Thus, the BQP can also be written as the convex hull of all matrices Y of the form

$$Y = \begin{pmatrix} x_1 & y_{12} & \cdots & y_{1n} \\ y_{21} & x_2 & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & x_n \end{pmatrix} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nn} \end{pmatrix},$$

the first being simply the trivial symmetrization of the original (x, y) vector, and the second being due to $y_{ii} = x_i$ for all $i \leq n$. This form shows that $Y = xx^\top$. Then, the BQP can be written as the convex hull of all Y matrices derived as $xx^\top$ where x ranges over $\{0, 1\}^n$, which is exactly the definition of $\mathsf{COR}(n)$. We note that the above mapping is a polynomial-time computable rational linear isomorphism. By [14], this gives a valid polynomial-time reduction between corresponding decision problems. Therefore, the **NP**-hardness on $\mathsf{COR}(n)$ problems (with or without zero element) transfers to the boolean quadratic polytope (with or without zero element).

2.3 The cut polytope and its cone

The correlation polytope is also related to the cut polytope $\mathsf{CUT}(n)$ [11], defined similarly to Eq. (5), but with all matrices $Y^k = y_k y_k^\top$ where $y_k \in \{-1, 1\}^n$ for all $0 \leq k \leq P_n$. Note that each vector $y_k \in \{-1, 1\}^n$ can be written as $2x_k - \mathbf{1}_n$: through this correspondence we can define y_k as the vector whose corresponding x_k encodes the integer k in binary. There is, however, a twist: while swapping zeros and ones in $x_k \in \{0, 1\}$ one obtains another vector $x_h \in \{0, 1\}$ such that $X^k \neq X^h$, if one swaps ± 1 in y_k one obtains another vector $-y_k$ that yields the same matrix Y^k: for example, the all-zero and all-one matrices in $\mathsf{COR}(n)$ both map to the all-one matrix in $\mathsf{CUT}(n)$ under the transformation $2x_k - \mathbf{1}$. Hence, while $\mathcal{C}(n)$ and $\mathsf{COR}(n)$ have 2^n generators, $\mathsf{CUT}(n)$ has 2^{n-1} generators: half with respect to $\mathsf{COR}(n)$. Cardinality-wise, $\mathsf{COR}(n)$ therefore corresponds to $\mathsf{CUT}(n + 1)$ [10].

Analogously to the correlation cone, the cut cone is defined as $\mathsf{cone}(Y)$ where $Y = \{Y^k = y_k y_k^\top \mid y_k \in \{-1, 1\} \wedge 0 \leq k \leq P_n\}$ is the set of all symmetric $n \times n$ rank-1 matrices with entries over $\{-1, 1\}$. It is shown in [16] that the separation problem for the cut cone is **NP**-hard.

2.4 Linear optimization

It is known that linear optimization over $\mathsf{COR}(n)$ and $\mathsf{CUT}(n)$ is **NP**-hard [23, 11].

We first consider Quadratic Boolean Optimization (QuBO)

$$\max\{x^\top Q x + c^\top x \mid x \in \{0, 1\}^n\},$$

where we linearize each quadratic product, introducing a variable matrix $X = xx^\top$. This yields a linear optimization problem $\max\langle Q, X \rangle$ over the set of vertices $X = xx^\top$, where $x \in \{0, 1\}^n$ of the $\mathsf{COR}(n)$

polytope. Since linear optimization on a set of points has the same optima as linear optimization on the convex hull of the points, linear optimization on $\mathsf{COR}(n)$ is as hard as QuBO, which is known to be **NP**-hard. This reduction is detailed but implicit in [23, Eq. (1)-(8)], and succinct but explicit in [11, Eq. (5.1.4), p. 54].

Next, we consider the MAX CUT formulation [11]

$$\max\Big\{\sum_{i<j} w_{ij}\frac{1-y_iy_j}{2} \mid y \in \{-1,1\}^n\Big\},$$

where w are a set of given weights on unordered pairs $\{i,j\}$. We re-write this formulation as a linear matrix optimization problem $\max\{\langle W,Y\rangle \mid Y = yy^\top \wedge y \in \{-1,1\}^n\}$, where $W = (w_{ij})$ is an $n \times n$ symmetric matrix. Since the feasible set $\{yy^\top \mid y \in \{-1,1\}^n\}$ is the vertex set of $\mathsf{CUT}(n)$, we can take the convex hull of this vertex set and obtain the exponentially-sized LP $\max\{\langle W,Y\rangle \mid Y \in \mathsf{CUT}(n)\}$. Again, since linear optimization on a set of points has the same optima as linear optimization on the convex hull of the points, linear optimization on $\mathsf{CUT}(n)$ is as hard as MAX CUT, which is known to be **NP**-hard. This reduction is mentioned in [11, Eq. (4.1.4), p. 38].

Even though these two linear optimization problems are **NP**-hard, we cannot conclude that all their possible subsets of instances form **NP**-hard problems. We therefore need to explicitly prove **NP**-hardness for rank problems with respect to $\mathsf{COR}(n)$ and $\mathsf{CUT}(n)$, as well as relaxed rank problems for the corresponding cones (see Sect. 4).

2.5 Membership

We already mentioned that testing membership in $\mathsf{COR}(n)$ is **NP**-hard by [24] (Problem 4 from the list in Sect. 1). We note that [24] provides a Karp reduction. We mention in passing that a Cook-Turing reduction is readily established by Sect. 2.4 and the polynomial-time equivalence of linear optimization, separation, and membership over $\mathsf{COR}(n)$ [14].

In [11], the **NP**-hardness of the corresponding membership problem in $\mathsf{CUT}(n)$ is derived as follows: MAX CUT proves **NP**-hardness of linear optimization over $\mathsf{CUT}(n)$ (Sect. 2.4). By [14], linear optimization separation and membership in rational polytopes are polynomial-time equivalent problems under Cook-Turing reductions: whence membership in $\mathsf{CUT}(n)$ is **NP**-hard. Similarly, the fact that separation in the cut cone is **NP**-hard [16] implies that membership in that cone is also **NP**-hard. We note that these are Cook-Turing reductions: we will provide Karp reductions in Sect. 4.

2.6 Rank

It is known that solving sparse linear systems of equations (SPARSE LINEAR SYSTEM) is **NP**-hard [22]. More precisely, there is a reduction from EXACT COVER BY 3-SETS (X3C) to the (decision version of the) problem of computing a vector p satisfying $\|Ap - b\|_2 \leq \epsilon$ for given $A, b, \epsilon > 0$ such that $\|p\|_0$ is minimum. Given the real numbers setting due to the presence of the ℓ_2 norm, the reduction assumes the Real RAM computational model [25, §1.4]. The requirement of a real computational model and the absence of the constraint $p \geq 0$ makes it doubtful that SPARSE LINEAR SYSTEM is a valid candidate for a source problem in view to reduce to rank minimization for $\mathsf{COR}(n)$ or $\mathcal{C}(n)$. Indeed, we use reductions from other problems in Sect. 4.

3   Some necessary conditions for membership in $\mathcal{C}(n)$

We start with some easy preliminary results in order to establish some necessary conditions for Γ to belong to the cone $\mathcal{C}(n)$. A *positive semidefinite* (PSD) matrix is a square symmetric matrix where all the eigenvalues are non-negative.

3.1 Lemma
If $\Gamma \in \mathcal{C}(n)$, then it is PSD.

Proof. The k-th term of the sum decomposition Eq. (2) consists of a non-negative scalar p_k which multiplies a rank-one matrix $X^k = x_k x_k^\top$, where $x_k \in \{0,1\}^n$, for all $k \leq P_n$. For any $y \in \mathbb{R}^n$, we have $y^\top X^k y = y^\top x_k x_k^\top y = (x_k^\top y)^\top (x_k^\top y) = \|x_k^\top y\|_2^2 \geq 0$. Hence, X^k is a PSD matrix for all $k \leq P_n$. The non-negative sum of PSD matrices is PSD, whence the result. $\qquad\square$

A *doubly non-negative* (DNN) matrix is a PSD matrix with non-negative entries.

3.2 Corollary
If $\Gamma \in \mathcal{C}(n)$, then it is DNN.

Proof. By Lemma 3.1 and the fact that every term of the RHS of Eq. (2) has non-negative entries. $\square$

A *completely positive* (CP) matrix Γ is such that $\Gamma = AA^\top$ where A is $n \times r$ and A is non-negative, i.e.

$$\Gamma = \sum_{k=1}^{r} A_k A_k^\top. \tag{10}$$

The minimum r for which Γ can be represented as in Eq. (10) is the *CP rank* of Γ. Any CP matrix is trivially DNN, but the converse does not hold in general [2].

3.3 Proposition
If $\Gamma \in \mathcal{C}(n)$, then it is CP.

Proof. By definition, we have $\Gamma = \sum_{k \leq r} p_K x_k x_k^\top$ with $p \geq 0$ for some $r \leq P_n$. Since $p \geq 0$, $\sqrt{p_k}$ is real for all k. Hence,

$$\Gamma = \sum_{k \leq r} (\sqrt{p_K} x_k)(\sqrt{p_k} x_k)^\top = \sum_{k=1}^{r} A_k A_k^\top,$$

where $A_k = \sqrt{p_k} x_k$, as claimed. $\qquad\square$

Note that the r in the proof of Prop. 3.3 does not correspond to the CP rank: while we show that membership of Γ in $\mathcal{C}(n)$ also proves that Γ is CP, there may exist other non-negative factorizations of Γ in terms of an $n \times r'$ non-negative matrix B with $r' < r$. The proof of Prop. 3.3 only shows that the CP rank cannot exceed r, but this is not a tight bound [2, Thm. 3.5].

Proving that a matrix is DNN can be done in polynomial time up to a given precision tolerance e.g. by computing eigenvalues and inspecting the matrix entries. By contrast, proving that a matrix is CP is **NP**-hard by reduction from MAX CLIQUE via the Motzkin-Straus formulation [20], which can be transformed into a linear optimization problem over the CP matrix cone [21].

4 New NP-hardness results

In this section, we prove the **NP**-hardness of membership, rank, and relaxed rank in $\mathcal{C}(n)$, as well as of rank in $\mathsf{COR}(n)$, using Karp reductions. We also consider many corollaries of these four main theorems, and derive **NP**-hardness results for problems related to normalized polytopes and cut cones.

4.1 Membership in $\mathcal{C}(n)$

We prove **NP**-hardness of the $\mathcal{C}(n)$ membership problem by reduction from the problem of establishing membership in the correlation polytope $\mathsf{COR}(n)$, known to be **NP**-hard by [24] by reduction from 1-IN-3SAT: given a SAT formula ψ in conjunctive normal form, each of whose clause consists of three literals, does there exist a satisfying solution for ψ with the additional constraint that exactly one literal for each clause evaluates to TRUE?

Note that membership in $\mathsf{COR}(n)$ simply imposes a convex hull constraint $\sum_k p_k = 1$ to the cone $\mathcal{C}(n)$. So, in our reduction, we model the satisfaction of this constraint as part of the cone. For this, we need one more dimension.

4.1 Theorem
*Membership in $\mathcal{C}(n)$ is **NP**-hard, by inclusion of the case where Γ has its last row and column identical to the diagonal, and its last element equal to 1.*

Proof. All along the proof, we will employ symbols associated to "z" for the origin problem $\mathsf{COR}(n)$, and symbols associated to "x" for the target problem $\mathcal{C}(n)$.

For an instance Z of $\mathsf{COR}(n)$ we will define a corresponding instance of $\mathcal{C}(n+1)$. More precisely, we let Γ have its upper-left $n \times n$ block equal to Z, then $\Gamma_{i,n+1} = \Gamma_{n+1,i} = Z_{ii}$ for all $i \leq n$, and finally $\Gamma_{n+1,n+1} = 1$. This yields an affine map $\mathscr{L} : Z \mapsto \Gamma$ (also see Sect. 4.4). Note that this definition makes Γ have identical last row, last column, and diagonal.

We assume that $Z \in \mathsf{COR}(n)$: then, there exists a vector $\lambda \geq 0$ in $\mathbb{R}^{P_n}$ such that $Z = \sum_{k \leq P_n} \lambda_k z_k z_k^\top$ holds. For each vector $z_k \in \{0,1\}^n$, we define $x_k \in \{0,1\}^{n+1}$ as $x_k = (z_k, 1)$ and $p_k = \lambda_k$ for all $k \leq P_n$. Now we have

- for $1 \leq i \leq j \leq n$:
$$\Gamma_{ij} = Z_{ij} = \sum_{k \leq P_n} \lambda_k Z_{ij}^k = \sum_{k \leq P_n} p_k Z_{ij}^k = \sum_{k \leq P_n} p_k X_{ij}^k,$$
  where $X^k = x_k x_k^\top$;

- for $1 \leq i \leq n$ (last column):
$$\Gamma_{i,n+1} = Z_{ii} = \sum_{k \leq P_n} \lambda_k Z_{ii}^k = \sum_{k \leq P_n} \lambda_k z_{ki}^2,$$
  and note that $z_{ki}^2 = z_{ki}$ (because $z \in \{0,1\}$), that $z_{ki} \times 1 = x_{ki} x_{k,n+1}$ (by definition of x_k), and that $x_{ki} x_{k,n+1} = X_{i,n+1}^k$ (by definition of X^k), so $\Gamma_{i,n+1} = \sum_{k \leq P_n} p_k X_{i,n+1}^k$ (by definition of p_k);

- the last row of Γ is identical to its last column by symmetry;

- lastly, we check that this construction is consistent for $i = j = n+1$, i.e.
$$1 = \Gamma_{n+1,n+1} = \sum_{k \leq P_{n+1}} p_k X_{n+1,n+1}^k = \sum_{k \leq P_n} (p_k \times 1) = \sum_{k \leq P_n} p_k = \sum_{k \leq P_n} \lambda_k = 1$$

  by definition of $\Gamma_{n+1,n+1}$, because $X_{n+1,n+1}^k = x_{k,n+1}^2 = 1$ whenever $x_k = (z_k, 1)$ and 0 otherwise, by definition of $p_k = \lambda_k$ for all $k \leq P_n$, and because Z is a YES instance of $\mathsf{COR}(n)$.

Therefore, $p \geq 0$ is such that $\Gamma = \sum_{k \leq P_{n+1}} p_k X^k$, as claimed.

Next, we assume that Γ (constructed from Z as above) is an $(n+1) \times (n+1)$ YES instance of $\mathcal{C}(n)$. Thus, $\Gamma = \sum_{k \leq P_{n+1}} p_k X^k$, with $X^k = x_k x_k^\top$. For each k, let $x_k = (z_k, b)$ be a decomposition of x_k into its first n elements vector z_k, followed by the last element b. We know, by definition of Γ, that it has identical last row, last column, and diagonal. Hence, subtracting the last column from the diagonal, we have

$$0 = \sum_{k \leq P_{n+1}} p_k (X_{ii}^k - X_{i,n+1}^k) = \sum_{k \leq P_{n+1}} p_k x_{ki}(1 - x_{k,n+1}) \tag{11}$$

because $x_{ki}^2 = x_{ki} \in \{0,1\}$. Since $p \geq 0$, by Eq. (11) we have $x_{ki}(1 - x_{k,n+1}) = 0$ for all k, i, so if $x_{k,n+1} = b = 0$, then $x_{ki} = 0$ for all $i \leq n$, whence $b = 1$ for all non-zero boolean vectors x_k. This implies

$$1 = \Gamma_{n+1,n+1} = \sum_{k \leq P_{n+1}} p_k X_{n+1,n+1}^k = \sum_{k \leq P_{n+1}} p_k x_{k,n+1}^2 = \sum_{k \leq P_n} p_k, \tag{12}$$

where the last sum consists of only P_n terms because the vectors x_k having zero $(n+1)$-st components are identically the zero vector, as argued above. Finally, the $n \times n$ top left block of Γ being equal to Z by construction, we have $Z = \sum_{k \leq P_n} p_k X^k$ with $\sum_{k \leq P_n} p_k = 1$ by Eq. (12). Therefore, Y is a YES instance of $\mathsf{COR}(n)$, as claimed. $\qquad\square$

The proof of Thm. 4.1 is such that, potentially, both source and target problems may have exponentially-sized YES certificates: this does not invalidate the reduction presented in the theorem, but it might be puzzling to some readers. By Carathéodory's theorem, for any long representation there must also exist a compact one. Moreover, both source and target problems are essentially LP feasibility problems: membership in $\mathsf{COR}(n)$ can be formulated as an LP feasibility with $N + 1$ equality constraints, and membership in $\mathcal{C}(n)$ has N equality constraints. Therefore, both problems admit basic feasible solutions with at most $N + 1$ nonzeroes. These facts also prove that membership of $\mathcal{C}(n)$ is in **NP**.

4.2   Rank in $\mathcal{C}(n)$

In this section, we prove the **NP**-hardness of the rank problem associated to $\mathcal{C}(n)$.

Given a positive integer q and a symmetric $n \times n$ matrix Γ, find a vector $p \in \mathbb{R}_+^{P_n}$ such that

$$\begin{align}
\|p\|_0 &\leq q \tag{13} \\
\Gamma &= \sum_{k \leq P_n} p_k x_k x_k^\top. \tag{14}
\end{align}$$

We reduce from LINEAR EXACT COVER BY 3-SETS (LINEAR-X3C): given a set U with $|U| = 3q$ and a set $\mathcal{S} = \{S_1, \ldots, S_m\}$ of m subsets of U each having exactly three elements, and such that every unordered pair $\{e, f\} \subset U$ occurs in at most one of the subsets in $\mathcal{S}$, does $\mathcal{S}$ contain an exact cover (i.e. a partition) of U using exactly q out of the m subsets? If this problem seems contrived it is because its purpose is exactly to serve as reduction source for other, more interesting problems with challenging constraints. A non-peer-reviewed (but reasonably clear) proof of the **NP**-hardness of LINEAR-X3C can be found in the first answer to question 20386 on `cstheory.stackexchange.com`. A peer-reviewed proof can be found in [9, §2], under the problem name PSTS-1-RES: given a partial Steiner triple system on $3q$ points (where $q = 2\ell + 1$ for some integer ℓ), does it contain a parallel class, i.e. a set of q triplets such that each unordered pair of elements from U occurs in at most one triplet of $\mathcal{S}$? The only limitation that q should be odd can be removed by adding three new elements to U and adding a triplet to $\mathcal{S}$ that contains exactly those new elements.

4.2 Theorem
*The rank problem for $\mathcal{C}(n)$ is **NP**-hard.*

Proof. Given an instance of LINEAR-X3C with $|U| = 3q$ for some integer $q \geq 1$, we assume without loss of generality that $U = \{1, \ldots, n-1\}$, so $n = 3q + 1$. We construct Γ as follows:

- for each $i < n$: $\Gamma_{in} = \Gamma_{ni} = 1$;

- for each $i < j < n$: $\Gamma_{ij} = \Gamma_{ji} = \begin{cases} 1 & \text{if } \exists S \in \mathcal{S} \text{ with } \{i, j\} \in S, \\ 0 & \text{otherwise}; \end{cases}$

- for $i \in U$: $\Gamma_{ii} = 1$;

- $\Gamma_{nn} = q$.

First, we assume that the LINEAR-X3C instance is YES. So, there exists an exact cover $\mathcal{T} \subset \mathcal{S}$ that partitions U: for every $i < n$ there is a unique $S \in \mathcal{T}$ such that $i \in S$. Since $|U| = 3q$ and it has a partition into triplets, the number of triplets in the partition is q. So we can write $\mathcal{T} = \{T_1, \ldots, T_q\}$, where each $T_k \in \mathcal{T}$ is equal to some $S_j \in \mathcal{S}$. For each $k \leq q$ we define $x_k \in \{0, 1\}^n$ as follows:

$$x_{kn} = 1 \qquad \text{and} \qquad \forall i < n \quad x_{ki} = \begin{cases} 1 & \text{if } i \in T_k, \\ 0 & \text{otherwise}. \end{cases}$$

Moreover, we let $p_k = 1$ for all $k \leq q$. Let $\bar{\Gamma} = \sum_{k \leq q} p_k x_k x_k^\top$: we claim that $\bar{\Gamma} = \Gamma$. For each $i < n$ exactly one triplet T_k in $\mathcal{T}$ contains i, hence exactly one boolean vector x_k has $x_{ki} = x_{kn} = 1$, which implies $\bar{\Gamma}_{in} = 1 = \Gamma_{in}$. The same consideration also implies that $\bar{\Gamma}_{ii} = \Gamma_{ii}$ for all $i < n$. For $i < j < n$ we have $\bar{\Gamma}_{ij} = 1$ iff i, j both belong to a given triplet T_k (for some $k \leq q$). By definition of LINEAR-X3C, T_k must be the only triplet to contain both i, j: hence $\bar{\Gamma}_{ij} = \Gamma_{ij}$. Therefore, Γ satisfies Eq. (14) with exactly q nonzero coefficients, whence the rank of Γ is $\leq q$.

Next, we assume that (q, Γ) is a YES instance of the rank problem for $\mathcal{C}(n)$, i.e. Γ has rank $r \leq q$, whence $\Gamma = \sum_{k \leq r} p_k x_k x_k^\top$ with $p > 0$. For each $k \leq r$ we define the support sets $R_k = \{i < n \mid x_{ki} = 1\}$ of each x_k. For each $i < n$ we have

$$1 = \Gamma_{ii} = \sum_{k \leq r} p_k x_{ki}^2 = \sum_{\substack{k \leq r \\ i \in R_k}} p_k. \tag{15}$$

Moreover, since $\Gamma_{ni} = \Gamma_{in} = 1$ for all $i < n$, we have

$$1 = \Gamma_{in} = \sum_{k \leq r} p_k x_{kn} x_{ki} = \sum_{\substack{k \leq r \\ i, n \in R_k}} p_k.$$

Subtracting one equation from the other, we infer that any x_k such that $x_{ki} = 1$ must also have $x_{kn} = 1$, otherwise the two sums could not both be equal to 1. Hence, for any $k \leq r$, if R_k contains $i < n$ it also contains n. Now, for any $k \leq r$ consider $i < j < n \in R_k$: the k-th term contributes $p_k > 0$ to Γ_{ij}. Since $\Gamma \in \{0, 1\}^{n \times n}$ and $p \geq 0$, $\Gamma_{ij} = 1$. By construction of Γ, there must be $S_h \in \mathcal{S}$ such that $i, j \in S_h$. By definition of LINEAR-X3C, all pairs in R_k must be in a single triplet, so $R_k \subseteq S_h$, implying $|R_k| \leq 3$.

At this point, we claim that $r = q$, all $p_k = 1$, and $|R_k| = 3$. We sum over the last column of Γ and obtain $\sum_{i < n} \Gamma_{in} = 3q$, since $\Gamma_{in} = 1$ for all $i < n$, and $|U| = |\{i \mid i < n\}| = 3q$. We also have

$$\sum_{i < n} \Gamma_{in} = \sum_{i < n} \sum_{k \leq r} p_k x_{ki} x_{kn} = \sum_{k \leq r} p_k |R_k| x_{kn} = \sum_{k \leq r} p_k |R_k|$$

by definition of Γ, by definition of R_k, and because $x_{kn} = 1$ whenever $R_k \neq \varnothing$. Note that $|R_k| \leq 3$ implies

$$3q = \sum_{k \leq r} p_k |R_k| \leq 3 \sum_{k \leq r} p_k. \tag{16}$$

Now, by $\Gamma_{nn} = q$ we have

$$q = \Gamma_{nn} = \sum_{k \le r} p_k x_{kn}^2 = \sum_{k \le r} p_k x_{kn} = \sum_{k \le r} p_k, \qquad (17)$$

again by definition of Γ, by $x_{kn} \in \{0,1\}$, and because $x_{kn} = 1$ whenever $p_k > 0$. Combining Eq. (16)-(17) we obtain

$$3q \le 3 \sum_{k \le r} p_k = 3q,$$

whence equality must hold in Eq. (16). Therefore $|R_k| = 3$ for all k with $p_k > 0$. To settle the claim, we remark the following facts: (i) $|R_k| = 3$ for all $k \le r$, (ii) $\sum_{k \le r} p_k = q$, (iii) $r \le q$, and (iv) $p_k > 0$ for all $k \le r$. These imply that $r = q$ and $p_k = 1$ for all $k \le r$, as claimed.

Finally, by Eq. (15) and the fact that $p_k = 1$ for all $k \le r$, for each $i < n$ we have

$$1 = \sum_{\substack{k \le r \\ i \in R_k}} 1,$$

so each $i < n$ belongs to exactly one R_k. Therefore, $R_1, \dots R_r$ form a partition of U into q disjoint triplets: then $\mathcal{T} = \{R_1, \dots R_r\}$ is an exact cover of U, which implies that the LINEAR-X3C is a YES instance, as claimed. $\qquad \square$

4.3   Relaxed rank in $\mathcal{C}(n)$

The relaxed rank problem with respect to $\mathcal{C}(n)$ is as follows. Given a symmetric $n \times n$ matrix $\Gamma \in \mathcal{C}(n)$ (1-bit promise), find $p \ge 0$ such that

$$\left. \begin{array}{rcl} \sum\limits_{0 \le k \le P_n} p_k X^k & = & \Gamma \\ \sum\limits_{0 \le k \le P_n} p_k & \le & \rho. \end{array} \right\} \qquad (18)$$

4.3 Theorem
*The relaxed rank problem for $\mathcal{C}(n)$ is **NP**-hard.*

Proof. We reduce from the $\mathsf{COR}(n)$ membership problem by exploiting its simplex constraint $\sum_k p_k = 1$. For any $\rho > 0$, we define the scaled polytope $\rho\mathsf{COR}(n)$ as

$$\left. \begin{array}{rcl} \sum\limits_{k \ge 0} p_k X^k & = & \Gamma \\ \sum\limits_{k \ge 0} p_k & = & \rho. \end{array} \right\} \qquad (19)$$

Since scaling by ρ is linear and invertible, it is a linear isomorphism, and therefore, by [14], it provides a Karp reduction from $\mathsf{COR}(n)$ membership to $\rho\mathsf{COR}(n)$ membership, making the latter **NP**-hard.

Next, we consider the membership problem:

$$\left. \begin{array}{rcl} \sum\limits_{k \ge 0} p_k X^k & = & \Gamma \\ \sum\limits_{k \ge 0} p_k & \le & \rho. \end{array} \right\} \qquad (20)$$

We prove that satisfying Eq. (20) is **NP**-hard, by reduction from $\sigma\mathsf{COR}(n)$ for some $0 < \sigma \le \rho$. Let $\Gamma \in \sigma\mathsf{COR}(n)$. Then, Γ trivially satisfies Eq. (20) because $\sum_{k \ge 0} p_k = \sigma \le \rho$. Conversely, take a Γ that satisfies Eq. (20), and assume (without loss of generality) that $p_0 = 0$. Therefore, there is some $\tau > 0$ such that $\sum_{k \ge 1} p_k = \tau \le \rho$. We set $p_0 = \rho - \tau$. Now $\sum_{k \ge 0} p_k = p_0 + \sum_{k \ge 1} p_k = \rho - \tau + \tau = \rho$. This shows that

$\Gamma \in \rho\mathsf{COR}(n)$. Finally, we scale by $\sigma/\rho \leq 1$ to show that $\Gamma \in \sigma\mathsf{COR}(n)$, which concludes the reduction. This shows that Eq. (20) is **NP**-hard. Lastly, we note that Eq. (20) is also the definition of the relaxed rank problem for $\mathcal{C}(n)$, which is therefore **NP**-hard. □

A non-geometric, longer, and more informative proof by reduction from the FRACTIONAL CLIQUE COVER problem can be found in Appendix A.

4.4 Rank in $\mathsf{COR}(n)$

In order to prove that the rank decision problem for $\mathsf{COR}(n)$ is **NP**-hard, we use reduction techniques from Sect. 4.1-4.2. More precisely, we reduce from the rank problem in $\mathcal{C}(n)$ and use the same lifting as in the reduction from $\mathsf{COR}(n)$ membership to $\mathcal{C}(n+1)$ membership. For an $n \times n$ matrix Z, we employ the affine map $\mathscr{L}$ defined in the proof of Thm. 4.1:

$$\mathscr{L}(Z) = \begin{pmatrix} Z & \mathsf{diag}(Z) \\ \mathsf{diag}(Z)^\top & 1 \end{pmatrix}.$$

4.4 Lemma
For any $n \times n$ symmetric matrix Z and any integer $\rho \geq 1$, we have

  (i) *if $Z \in \mathsf{COR}(n)$, then $\mathscr{L}(Z) \in \mathcal{C}(n+1)$;*

  (ii) *$Z \in \mathsf{COR}(n)$ and the $\mathsf{COR}(n)$-rank of Z is $\leq \rho \iff \mathscr{L}(Z) \in \mathcal{C}(n+1)$ and the $\mathcal{C}(n+1)$-rank of $\mathscr{L}(Z)$ is $\leq \rho$.*

Proof. ($\Rightarrow$) Assume $Z = \sum_{k=0}^{r} \lambda_k z_k z_k^\top$, $\lambda \geq 0$, $\sum_{k=0}^{r} \lambda_k = 1$, $r \leq \rho$. Let $x_k = (z_k, 1) \in \{0,1\}^{n+1}$. Then,

$$x_k x_k^\top = \begin{pmatrix} z_k z_k^\top & z_k \\ z_k^\top & 1 \end{pmatrix}$$

$$\Rightarrow \sum_{k=0}^{r} \lambda_k x_k x_k^\top = \begin{pmatrix} \sum_k \lambda_k z_k z_k^\top & \sum_k \lambda_k z_k \\ \sum_k \lambda_k z_k^\top & \sum_k \lambda_k \end{pmatrix} = \begin{pmatrix} Z & \mathsf{diag}(Z) \\ \mathsf{diag}(Z)^\top & 1 \end{pmatrix} = \mathscr{L}(Z) = \Gamma$$

So $\Gamma \in \mathcal{C}(n+1)$ and uses $r \leq \rho$ generators.

($\Leftarrow$) Assume $\Gamma = \sum_{k=0}^{r} p_k x_k x_k^\top$, $p \geq 0$, $x_k \in \{0,1\}^{n+1}$ for all $k \leq r$, and $r \leq \rho$. The $(n+1, n+1)$-th entry of Γ is 1, and in particular $1 = \Gamma_{n+1,n+1} = \sum_{k=0}^{r} p_k x_{k,n+1}^2 = \sum_{k=0}^{r} p_k x_{k,n+1}$ because booleans are square-invariant. Since $x_{k,n+1} \in \{0,1\}$ and $p \geq 0$, we have $x_{k,n+1} = 1$ for every k with $p_k > 0$, and $\sum_{k=0}^{r} p_k = 1$. Again, since $x_k = (z_k, 1)$, the upper-left block of $\mathscr{L}(Z)$ is $Z = \sum_{k=0}^{r} p_k x_k x_k^\top$ with $\sum_{k=0}^{r} p_k = 1$, whence $Z \in \mathsf{COR}(n)$ and the rank of Z is $r \leq \rho$, as claimed. □

4.5 Theorem
*The rank problem for $\mathsf{COR}(n)$ is **NP**-hard.*

Proof. We reduce from the rank problem in $\mathcal{C}(n+1)$, which is **NP**-hard by Thm. 4.2. By Lemma 4.4, the map $\mathscr{L} : Z \mapsto \mathscr{L}(Z)$ is affine, can be carried out in polynomial time, and maps rational instances to rational instances. Therefore, by [14], it provides a valid polynomial reduction from $\mathcal{C}(n+1)$ to $\mathsf{COR}(n)$, as claimed. We note that, by Lemma 4.4, Z has $\mathsf{COR}(n)$-rank at most ρ iff $\mathscr{L}(Z)$ has $\mathcal{C}(n+1)$-rank at most ρ. □

4.5   Hardness corollaries

In this section we list several corollaries of the theorems in Sect. 4.

4.5.1   Membership problems in normalized polytopes

We first introduce the membership problems for the normalized polytopes for correlation and cut cones, which will give us a useful affine isomorphism to be used later. Note that $\mathsf{COR}(n)$ does not satisfy the definition of "convex hull of rank-one boolean matrices" syntactically, since the zero matrix has rank zero, not one. By contrast, $\overline{\mathsf{COR}}(n)$ does. We can reduce membership in $\mathsf{COR}(n)$ to membership in $\overline{\mathsf{COR}}(n+1)$ by the mapping $\mathscr{L}'$ (reflection of $\mathscr{L}$ along the skew diagonal, see the proof of Thm. 4.1) that sends $\Gamma \in \mathsf{COR}(n)$ to the bordered matrix

$$\Gamma' = \begin{pmatrix} 1 & \mathsf{diag}(\Gamma)^\top \\ \mathsf{diag}(\Gamma) & \Gamma \end{pmatrix} \in \overline{\mathsf{COR}}(n+1), \tag{21}$$

where the presence of the 1 in the upper-left corner ensures that the affine map never sends any $\Gamma \in \mathsf{COR}(n)$ to the zero matrix. Thus, $\mathscr{L}'$ is a polytime-computable and rational affine isomorphism between a $\mathsf{COR}(n)$ and a face of $\overline{\mathsf{COR}}(n+1)$. Therefore, by [24] and [14], we obtain a Karp reduction from $\mathsf{COR}(n)$ to $\overline{\mathsf{COR}}(n+1)$ (by inclusion of the relevant face), showing that membership in $\overline{\mathsf{COR}}(n)$ is also **NP**-hard.

Starting from $\mathsf{CUT}(n)$, we define a *normalized cut polytope*, denoted $\overline{\mathsf{CUT}}(n)$, which removes the extreme point $Y^0 = y_0 y_0^\top = (-\mathbf{1})(-\mathbf{1})^\top = \mathbf{1}\mathbf{1}^\top = Y^{P_n}$. Similarly to the case of $\mathsf{COR}(n)$, we reduce $\mathsf{CUT}(n)$ to $\overline{\mathsf{CUT}}(n+1)$ by means the affine isomorphism in Eq. (21), which, again by [14], provides a reduction from $\mathsf{CUT}(n)$ to $\overline{\mathsf{CUT}}(n+1)$, which is therefore **NP**-hard. We remark that, at this point of our paper, the **NP**-hardness of $\mathsf{CUT}(n)$ is based on a Cook-Turing reduction (see Sect. 2.5). We correct the situation in Sect. 4.5.2 below.

4.5.2   Karp reduction for membership in $\mathsf{CUT}(n)$

Since we promised Karp reductions, but some of the results we cited from the literature are based on Cook-Turing reductions (Sect. 2.5), we provide a Karp reduction for proving **NP**-hardness of membership in $\mathsf{CUT}(n)$. We reduce from $\mathsf{COR}(n)$, extending the cardinality-wise correspondence between $\mathsf{COR}(n)$ and $\mathsf{CUT}(n+1)$ given in Sect. 2.3 to an affine isomorphism. Consider the affine transformation from $x_k \in \{0,1\}$ to $y_k \in \{-1,1\}$ vectors again: for any $0 \leq k \leq P_n$ we have $y_k = 2x_k - \mathbf{1}$. From this, we obtain

$$y_k y_k^\top = 4x_k x_k^\top - 2(x_k \mathbf{1}^\top + \mathbf{1}x_k^\top) + \mathbf{1}\mathbf{1}^\top. \tag{22}$$

And from the inverse mapping $x_k = (y_k + \mathbf{1})/2$ we have

$$x_k x_k^\top = (y_k y_k^\top + y_k \mathbf{1}^\top + \mathbf{1}y_k^\top + \mathbf{1}\mathbf{1}^\top)/4. \tag{23}$$

Using Eq. (22)-(23) and $X^k = x_k x_k^\top$, $Y^k = y_k y_k^\top$, we can define an affine isomorphism $\mathscr{C}$ from $\mathsf{COR}(n)$ and $\mathsf{CUT}(n+1)$ as follows:

- for $X^k \in X$ define $Y_{00}^k = 1$, $Y_{0i}^k = 2X_{ii}^k - 1$ for $1 \leq i \leq n$, and $Y_{ij}^k = 4X_{ij}^k - 2X_{ii}^k - 2X_{jj}^k + 1$ for $1 \leq i < j \leq n$;

- the inverse map is given by $X_{ij}^k = (1 + Y_{0i}^k + Y_{0j}^k + Y_{ij}^k)/4$ for all $1 \leq i \leq j \leq n$,

where we note that Y is a symmetric matrix (which defines the remaining entries).

It is easy to show that this map is affine [10]: it suffices to verify that it holds for a generic matrix Γ within $\mathsf{COR}(n)$ by simply applying the transformations to each term of the decomposition of Γ in

terms of the correlation polytope. Note, moreover, that this affine map can be applied in polynomial time, and maps rational instances of $\mathsf{COR}(n)$ to rational instances of $\mathsf{COR}(n+1)$. By [14], a polynomial-time computable affine isomorphism with polynomially bounded encoding size gives a valid polynomial-time reduction between the corresponding decision problems. We therefore obtain a Karp reduction establishing the **NP**-hardness of the membership problem for $\mathsf{CUT}(n)$.

4.5.3   Rank problem for the normalized correlation polytope

The **NP**-hardness of membership in the normalized correlation polytope $\overline{\mathsf{COR}}(n)$ was argued in Sect. 4.5.1. The rank problem in $\mathsf{COR}(n)$ is **NP**-hard by Thm. 4.5, so we reduce from $\mathsf{COR}(n)$-rank to $\overline{\mathsf{COR}}(n)$-rank in order to prove **NP**-hardness of the latter. To this purpose, we consider the same mapping $\mathscr{L}'$ defined in Sect. 4.5.1, which maps every $Z \in \mathsf{COR}(n)$ to $\mathscr{L}'(Z) \in \overline{\mathsf{COR}}(n+1)$, a bordered matrix with Z in the upper left corner, $\mathsf{diag}(Z)$ in the borders, and the scalar 1 in the lower-right entry. Since the 1 entry is fixed, $\mathscr{L}'$ never maps to the zero matrix: hence $\mathscr{L}'$ maps into $\overline{\mathsf{COR}}(n+1)$. Moreover, by Lemma 4.4 (with all sum quantifiers starting from $k = 1$, which has no impact on the value of the sums since $X^0 = 0$), $\mathscr{L}'$ preserves the rank exactly. This provides a valid polynomial reduction.

4.5.4   Rank problems for the cut and normalized cut polytopes

The **NP**-hardness of the rank problems for $\mathsf{CUT}(n)$ and $\overline{\mathsf{CUT}}(n)$ follows from those of $\mathsf{COR}(n)$ and $\overline{\mathsf{COR}}(n)$ and the affine mapping $\mathscr{C}$ given in Eq. (22)-(23), since affine isomorphisms such as $\mathscr{C}$ preserve rank exactly. This provides valid polynomial reductions.

4.5.5   Membership, rank, relaxed rank for the cut cone

The **NP**-hardness of membership, rank, and relaxed rank in the cut cone follows from those in the correlation cone $\mathcal{C}(n)$ (Sections 4.1, 4.2, 4.3) and the affine mapping $\mathscr{C}$ given in Eq. (22)-(23), since affine isomorphisms preserve rank. Moreover, because an isomorphism between cones must be linear (no translation terms), the mapping $\mathscr{C}$ applied to cones also preserves the relaxed rank $\|p\|_1$ that only depends on the conic hull and its generators, which are both preserved under the mapping. This provides valid polynomial reductions.

5   Related results

In this section, we look at easy polynomial cases of the problems that we proved to be **NP**-hard. Finally, we also show that the exponential extension complexity of the $\mathsf{COR}(n)$ membership problem implies the same for the $\mathcal{C}(n)$ membership problem. The fact that $\mathsf{COR}(n)$ has exponential extension complexity is not overly surprising given that the problem is **NP**-hard, but [15] discriminates between super-polynomial and exponential, and unconditionally with respect to the conjecture $\mathbf{P} \neq \mathbf{NP}$.

5.1   Simple polynomial cases

Let $G = (V, E)$ be the support graph of the $n \times n$ symmetric matrix Γ.

5.1.1   Forests

If G is a forest (e.g. tree, matching, star), then Γ can be covered by boolean vector support sets having cardinality limited to $1, 2$, as long as

$$\forall i \leq n \quad s_i = \Gamma_{ii} - \sum_{\substack{j \leq n \\ \{i,j\} \in E}} \Gamma_{ij} \geq 0. \tag{24}$$

If Eq. (24) holds, then we have

$$\Gamma = \sum_{\{i,j\} \in E} \Gamma_{ij} X^{\{i,j\}} + s_i X^i,$$

where $X^{\{i,j\}}$ are the rank-one boolean matrices obtained by support vectors with two nonzero entries corresponding to indices i, j, and X^i are those with one nonzero entry corresponding to index i.

5.1.2   Bounded treewidth

If G has bounded treewidth, we can solve the three problems (membership, rank, relaxed rank) in polytime by Dynamic Programming (DP) on the tree decomposition. Suppose that the treewidth of G is equal to t: then every clique in G has size bounded by $t + 1$, and every clique is contained in some node (called a "bag", and representing a set of vertices of the original graph) of any tree decomposition of G having width t. These are the premises to a general DP algorithm based on bounded treewidth graphs [4].

Moreover, it can be shown that the exponentially long sums in the equality constraints in Eq. (2) whenever $r = P_n$ become polynomial when G has bounded treewidth. Enumerate all cliques of G by listing all non-empty subsets C in each bag of a nice tree decomposition of G (i.e. a rooted tree decomposition where every node has at most two children and every node belongs to one of four types [17]), then verify if they are cliques. Since the treewidth t is a constant, the set $\mathcal{K}$ of these cliques has cardinality $O(2^t n) = O(n)$. Now we introduce the $O(n)$ decision variables $p_C \geq 0$ for each clique C, and the $O(n^2)$ equality constraints:

$$\Gamma_{ij} = \sum_{\substack{C \in \mathcal{K} \\ i,j \in C}} p_C \tag{25}$$

$$\Gamma_{ii} = \sum_{\substack{C \in \mathcal{K} \\ i \in C}} P_C. \tag{26}$$

Eq. (25)-(26) can replace Eq. (2) in the three exponential LP formulations for membership and relaxed rank, yielding polynomially-sized LP formulations for membership and relaxed rank. Applied to the exact rank problem, we obtain a polynomially-sized MILP formulation with $O(n)$ variables, which can be solved by the methods given in [8]. The exact rank problem can also be solved by DP [4] because $\|p\|_0$ is additive over the covering cliques from the bags of the tree decomposition.

5.1.3   Chordality

If G is chordal, then every clique is contained in a unique maximal clique, the maximal cliques form a clique tree, and there are at most n maximal cliques [3], which can therefore be listed in polynomial time. Thus, both of the exponential LPs for membership and relaxed rank become polynomially-sized: both can then be solved in weakly polynomial time using e.g. the interior point method for LP.

5.1.4   Perfect graphs

If G is a perfect graph, the maximum weight clique is polynomial-time solvable. The primal LP formulation of the relaxed rank problem is:

$$\min\{\sum_{C \in \mathcal{K}} p_C \mid \forall i \le j \le n \ \Gamma_{ij} = \sum_{\substack{C \in \mathcal{K} \\ i,j \in C}} p_C \wedge p \ge 0\},$$

where $\mathcal{K}$ is the set of all cliques in G. Its dual is

$$\max\{\langle Y, \Gamma \rangle \mid \forall C \in \mathcal{K} \ \sum_{i,j \in C} Y_{ij} \le 1\},$$

which has an exponential number of constraints. These can be handled in polynomial time by separation, iteratively solving the separation subproblem "given Y, find a clique $C \in \mathcal{K}$ maximizing $\sum_{ij \in C} Y_{ij}$" [13].

5.2   Extension complexity

Let $\mathsf{xc}(P)$ be the extension complexity of a linear programming formulation of some given problem P. Consider the affine slice

$$\Sigma = \{\Gamma \in \mathcal{C}(n+1) \mid \Gamma_{n+1,n+1} = 1\}$$

of $\mathcal{C}(n+1)$, which is affinely isomorphic to $\mathsf{COR}(n)$ by the mapping $\mathscr{L}$ introduced in the proof of Thm. 4.1. We have:

$$\mathsf{xc}(\mathcal{C}(n+1)) \ge \mathsf{xc}(\Sigma) = \mathsf{xc}(\mathsf{COR}(n)) = 2^{\Omega(n)},$$

where the last equation follows by [15]. Since Eq. (2) has extension complexity 2^n, we conclude that

$$\exists c > 0 \quad 2^{cn} \le \mathsf{xc}(\mathcal{C}(n)) \le 2^n.$$

6   Conclusion

We proved **NP**-hardness of three fundamental problems in polyhedral theory (membership, rank, and a relaxed rank derived from relaxing the zero-norm to the rank-norm) when applied to the correlation cone and polytope, the boolean quadric cone and polytope, and the cut cone and polytope, as well as to the normalized versions of the mentioned polytopes. All of these results stem from five reductions on the correlation cone and polytope, one of which (membership in the correlation polytope) was already known [24], while we believe that the other four are new.

Acknowledgments

References

[1] S. Bereg and M. Haghpanah. Computing the Carathéodory number of a point. In M. Keil and D. Mondal, editors, *Proceedings of the 32nd Canadian Conference on Computational Geometry*, CCCG, pages 182–188, Saskatoon, 2020. University of Saskatchewan.

[2] A. Berman and N. Shaked-Monderer. *Completely Positive Matrices*. World Scientific, Singapore, 2003.

[3] J. Blair and B. Peyton. An introduction to chordal graphs and clique trees. In A. George *et al.*, editor, *Graph Theory and Sparse Matrix Computation*, pages 1–29. Springer, New York, 1993.

[4] H. Bodlaender. Dynamic programming on graphs with bounded treewidth. In T. Lepistö and A. Salomaa, editors, *International Conference on Automata, Languages and Programming (ICALP)*, volume 317 of *LNCS*, pages 105–118, Berlin, 1988. Springer.

[5] E. Candès. The mathematics of sparsity. In S.Y. Jang, Y.R. Kim, D.-W. Lee, and I. Yie, editors, *Proceedings of the International Congress of Mathematicians*, volume I. Kyung Moon SA, Seoul, 2014.

[6] E. Candès and T. Tao. Decoding by Linear Programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

[7] A. Caprara, F. Furini, A. Lodi, M. Mangia, R. Rovatti, and G. Setti. Generation of antipodal random vectors with prescribed non-stationary 2-nd order statistics. *IEEE Transactions on Signal Processing*, 62(6):1603–1612, 2014.

[8] T. Chan, J. Cooper, M. Koutecký, D. Král, and K. Pekárková. Matrices of optimal tree-depth and a row-invariant parameterized algorithm for integer programming. *SIAM Journal on Computing*, 51(3):664–700, 2022.

[9] B. P. Ching Li and M. Toulouse. Some NP-completenesss results on partial Steiner triple systems and parallel classes. *Ars Combinatoria*, 80:45–51, 2006.

[10] C. De Simone. The cut polytope and the boolean quadric polytope. *Discrete Mathematics*, 79:71–75, 1990.

[11] M. Deza and M. Laurent. *Geometry of Cuts and Metrics*. Number 15 in Algorithms and Combinatorics. Springer, Heidelberg, 1997.

[12] R. Fortet. Applications de l'algèbre de Boole en recherche opérationelle. *Revue Française de Recherche Opérationelle*, 4:17–26, 1960.

[13] M. Grötschel, L. Lovász, and A. Schrijver. Polynomial algorithms for perfect graphs. *Annals of Discrete Mathematics*, 21:325–356, 1984.

[14] M. Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*. Number 2 in Algorithm and Combinatorics. Springer, Berlin, 2nd edition, 1993.

[15] V. Kaibel and S. Weltge. A short proof that the extension complexity of the correlation polytope grows exponentially. *Discrete and Computational Geometry*, 53:397–401, 2015.

[16] A. Karzanov. Metrics and undirected cuts. *Mathematical Programming*, 32:183–198, 1985.

[17] T. Kloks. *Treewidth Computations and Approximations*. Number 842 in LNCS. Springer, Berlin, 1994.

[18] C. Lund and M. Yannakakis. On the hardness of approximating minimization problems. *Journal of the ACM*, 41(5):960–981, 1994.

[19] A. Moitra. *Algorithmic aspects of Machine Learning*. CUP, Cambridge, 2018.

[20] T. Motzkin and E. Straus. Maxima for graphs and a new proof of a theorem of Turán. *Canadian Journal of Mathematics*, 17:533–540, 1965.

[21] K. Murty and S. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39:117–129, 1987.

[22] B. Natarajan. Sparse approximate solutions to linear systems. *SIAM Journal of Computing*, 24(2):227–234, 1995.

[23] M. Padberg. The boolean quadric polytope: some characteristics, facets and relatives. *Mathematical Programming*, 45:139–172, 1989.

[24] I. Pitowsky. Correlation polytopes: Their geometry and complexity. *Mathematical Programming*, 50:395–414, 1991.

[25] F. Preparata and M. Shamos. *Computational Geometry.* Texts and Monographs in Computer Science. Springer, New York, 1985.

Appendix A   Alternative reduction for the $\mathcal{C}(n)$ relaxed rank

Let K be the complete undirected graph on the vertex set $V(\mathsf{K}) = \{1, \ldots, n\}$, with loops. We denote non-loop edges by $\{i, j\} \in E(\mathsf{K})$ and loops by the singletons $\{i\}$. For $y \in \{0, 1\}^n$ we let $Y = yy^\top$ and $K(y)$ be the complete graph on the support set of the vector y, so that Y is the adjacency matrix of $K(y)$.

We re-cast the relaxed rank problem with respect to $\mathcal{C}(n)$ as follows. Given a scalar $\rho > 0$ and a matrix Γ, find a collection $\mathcal{V}$ of vectors in $\{0, 1\}^n$ and a vector $p \geq 0$ such that

$$\sum_{0 \leq k \leq P_n} p_k \;\leq\; \rho \tag{27}$$

$$\forall \{i, j\} \in E(\mathsf{K}) \quad \sum_{\substack{x_k \in \mathcal{V} \\ \{i,j\} \in K(x_k)}} p_k \;=\; \Gamma_{ij}. \tag{28}$$

We prove the **NP**-hardness of the relaxed rank problem related to $\mathcal{C}(n)$ by reduction from the FRACTIONAL CLIQUE COVER problem, which is also known as FRACTIONAL VERTEX COLORING on the complemented graph [18].

> FRACTIONAL CLIQUE COVER (FCC). Given an undirected graph $G = (V, E)$ and a value $t > 0$, determine whether there exists a set $\mathcal{K}$ of cliques of G, and a vector $w \in \mathbb{R}_+^{|\mathcal{K}|}$, such that
>
> $$\sum_{C \in \mathcal{K}} w_C \;\leq\; t \tag{29}$$
>
> $$\forall v \in V \quad \sum_{\substack{C \in \mathcal{K} \\ v \in C}} w_C \;=\; 1. \tag{30}$$

The reduction maps Eq. (29) to Eq. (27) and Eq. (30) to Eq. (28).

A.1 Theorem
*The relaxed rank problem for $\mathcal{C}(n)$ is **NP**-hard by inclusion of the case where the off-diagonal entries of Γ are in $\{0, \alpha\}$ for some $\alpha > 0$.*

Proof. For a given scalar $t > 0$ and a graph $G = (V, E)$, let (t, G) be an instance of the FCC. We construct an instance of the relaxed rank problem with $n = |V| + 1$, so we assume that $V = \{1, \ldots, n-1\}$: for $i, j \in V, i \neq j$, let $\Gamma_{ij} = 1/n^2$ for $\{i, j\} \in E$ and $\Gamma_{ij} = 0$ for $\{i, j\} \notin E$. Moreover, for $i \in V$, let $\Gamma_{ii} = 1/n$. Then let $\Gamma_{in} = 1/n^2$ for $i \in V$ and $\Gamma_{nn} = t/n^2$. Finally, we let $\rho = \frac{2t}{n^2} + \frac{3n-1}{2n} = \frac{3n^2 - n + 4t}{2n^2}$. We prove

that the instance (ρ, Γ) is YES iff (t, G) is a YES instance of the FCC. Note that the size of the relaxed rank instance is polynomially bounded in the size of the FCC instance.

We assume that the solution of the YES instance (t, G) of the FCC is a collection $\mathcal{C}$ of (say) m cliques of G with weights $w_C > 0$ for $C \in \mathcal{C}$, satisfying Eq. (29)-(30). We note that m is polynomially bounded in the size of the FCC instance, since NP certificates are polytime verifiable; therefore, m is also polynomially bounded in the size of the relaxed rank instance.

We construct a solution of the corresponding relaxed rank instance (ρ, Γ) consisting of a set $\mathcal{K}$ of cliques of the complete graph K, and a vector p indexed by cliques in $\mathcal{K}$. Initially, we set $\mathcal{K} = \varnothing$. We first deal with the last column of Γ: let $\mathcal{K} = \{C \cup \{n\} \mid C \in \mathcal{C}\}$. For a clique $K = C \cup \{n\} \in \mathcal{K}$, we define $p_K = w_C / n^2$. Condition (30) guarantees that

$$\forall i \le n \quad \sum_{\substack{K \in \mathcal{K} \\ \{i,n\} \subseteq K}} p_K = \sum_{\substack{C \in \mathcal{C} \\ i \in C}} \frac{w_C}{n^2} = \frac{1}{n^2} = \Gamma_{in}.$$

Moreover, the contribution of these cliques to Γ_{ij} for $\{i, j\} \notin E$ is zero, as desired, i.e.

$$\forall \{i, j\} \notin E \quad \sum_{\substack{K \in \mathcal{K} \\ \{i,j\} \subseteq K}} p_K = \sum_{\substack{C \in \mathcal{C} \\ \{i,j\} \subseteq C}} \frac{w_C}{n^2} = 0 = \Gamma_{ij},$$

since all sets in $\mathcal{C}$ are cliques of G, and $\{i, j\} \notin E$. Next, the condition $\sum_{C \in \mathcal{C}} w_C \le t$ guarantees that

$$\sum_{\substack{K \in \mathcal{K} \\ n \in K}} p_K = \sum_{C \in \mathcal{C}} \frac{w_C}{n^2} \le \frac{t}{n^2} = \Gamma_{nn}.$$

If the inequality is strict, we extend $\mathcal{K}$ with the 1-clique $\{n\}$ having weight $p_{\{n\}} = \frac{t}{n^2} - \sum_{C \in \mathcal{C}} \frac{w_C}{n^2}$, yielding

$$\sum_{\substack{K \in \mathcal{K} \\ n \in K}} p_K = \Gamma_{nn}.$$

We note that the clique $\{n\}$ is not yet part of $\mathcal{K}$, so we add it to $\mathcal{K}$.

We now consider Eq. (2) for $\{i, j\} \in E$ (for both $i = j$ and $i \ne j$). We want to achieve

$$\forall \{i, j\} \in E \quad \sum_{\substack{K \in \mathcal{K} \\ \{i,j\} \subseteq K}} p_K = \Gamma_{ij}.$$

With the current definition of $\mathcal{K}$, we have

$$\sum_{\substack{K \in \mathcal{K} \\ \{i,j\} \subseteq K}} p_K = \sum_{\substack{C \in \mathcal{C} \\ \{i,j\} \subseteq C}} \frac{w_C}{n^2} \le \sum_{\substack{C \in \mathcal{C} \\ i \in C}} \frac{w_C}{n^2} = \frac{1}{n^2}.$$

In order to reach the required $\Gamma_{ij} = 1/n^2$, it is sufficient to consider the 2-clique $\{i, j\}$ with weight

$$p_{\{i,j\}} = \frac{1}{n^2} - \sum_{\substack{K \in \mathcal{K} \\ \{i,j\} \subseteq K}} p_K.$$

We note that the 2-clique $\{i, j\}$ is not yet in $\mathcal{K}$, so we add it to $\mathcal{K}$. This settles all Γ_{ij} for $\{i, j\} \in E$.

For $i = j$, note that

$$\forall i \in V \quad \sum_{\substack{K \in \mathcal{K} \\ i \in K}} p_K = \frac{1}{n^2}$$

after the initialization of the weights p_K for $K \in \mathcal{K}$. In the (possible) successive 2-clique extensions, this sum is increased by at most $1/n^2$ for each edge incident to i (there are at most $n-1$ such edges). This means that, after the extension, we have

$$\sum_{\substack{K \in \mathcal{K} \\ i \in K}} p_K \leq \frac{1}{n^2} + \frac{n-1}{n^2} = \frac{1}{n}.$$

Hence, the required weight Γ_{ii} is reached by further extending $\mathcal{K}$ (if needed) with the 1-clique $\{i\}$ having weight

$$p_{\{i\}} = \frac{1}{n} - \sum_{\substack{K \in \mathcal{K} \\ i \in K}} p_K.$$

None of these 1- and 2-clique extensions affect the correct entries already achieved for Γ.

The construction of $\mathcal{K}$ ensures that its size, say r, is polynomially bounded in m, which is polynomially bounded in the size of the relaxed rank instance.

Finally, we look at Eq. (27). We note that $\mathcal{W} = \sum_{K \in \mathcal{K}} p_K$ is initially at most t/n^2. The possible extensions of $\mathcal{K}$ increase $\mathcal{W}$ by at most: $\frac{t}{n^2}$ for $p_{\{n\}}$, $1/n^2$ for each incident edge to v, and $1/n$ for each $v \in V$. So the total contribution $\mathcal{W}'$ from the extra cliques is bounded above by

$$\mathcal{W}' = \frac{t}{n^2} + \frac{n(n-1)}{2n^2} + n\frac{1}{n} = \frac{t}{n^2} + \frac{3n-1}{2n},$$

and therefore $\mathcal{W} \leq \frac{t}{n^2} + \mathcal{W}' = \frac{2t}{n^2} + \frac{3n-1}{2n} = \rho$, as claimed. Thus, the reduction maps a feasible instance of the FCC to a feasible relaxed rank instance.

Conversely, consider a feasible solution of the relaxed rank instance, defined by the collection of cliques $\mathcal{K}$ of the complete graph K on n vertices, with each clique having weight p_K for $K \in \mathcal{K}$. Note that, for each $K \in \mathcal{K}$, $K \setminus \{n\}$ is a clique of G: this occurs because, for all $\{i,j\} \notin E$, $\Gamma_{ij} = 0$ guarantees that $\{i,j\} \not\subseteq K$. Let $K_1, \ldots, K_m$ be the cliques in $\mathcal{K}$ such that $n \in K$ and $K \cap V \neq \varnothing$. We construct a feasible solution of the FCC instance (t, G): we let $C = \{C_1, \ldots, C_m\}$, where $C_i = K_i \setminus \{n\}$ with weight $w_{C_i} = n^2 p_{K_i}$ for $i \in \{1, \ldots, m\}$. The condition

$$\Gamma_{nn} = \frac{t}{n^2} = \sum_{\substack{K \in \mathcal{K} \\ n \in K}} p_K \geq \sum_{K \in \{K_1, \ldots, K_m\}} p_K$$

implies

$$\sum_{C \in \mathcal{C}} w_C \leq t,$$

so the solution has value at most t. Finally, the condition

$$\forall i \in V \quad \Gamma_{in} = 1/n^2 = \sum_{\substack{K \in \mathcal{K} \\ \{i,n\} \in K}} p_K = \sum_{\substack{K \in \{K_1, \ldots, K_m\} \\ i \in K}} p_K$$

implies (30), so the solution is feasible. $\square$

We observe that the relaxed rank problem for $\mathcal{C}(n)$ is in **NP**, given that YES instances can be certified by a number of strictly positive p_k that is at most $N = n(n+1)/2$.