Linear vs. Semidefinite Extended Formulations: Exponential Separation and Strong Lower Bounds

Samuel Fiorini*
Université libre de Bruxelles, Brussels, Belgium
sfiorini@ulb.ac.be

Serge Massar†
Université libre de Bruxelles, Brussels, Belgium
smassar@ulb.ac.be

Sebastian Pokutta
Friedrich-Alexander-Universität, Erlangen-Nürnberg, Germany
sebastian.pokutta@math.uni-erlangen.de

Hans Raj Tiwary‡
Université libre de Bruxelles, Brussels, Belgium
htiwary@ulb.ac.be

Ronald de Wolf§
CWI and University of Amsterdam
Amsterdam, The Netherlands.
rdewolf@cwi.nl

ABSTRACT

We solve a 20-year old problem posed by Yannakakis and prove that there exists no polynomial-size linear program (LP) whose associated polytope projects to the traveling salesman polytope, even if the LP is not required to be symmetric. Moreover, we prove that this holds also for the cut polytope and the stable set polytope. These results were discovered through a new connection that we make between one-way quantum communication protocols and semidefinite programming reformulations of LPs.

Categories and Subject Descriptors

F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems; G.2.0 [**Discrete Mathematics**]: General

General Terms

Theory

*Partially supported by the *Actions de Recherche Concertées* (ARC) fund of the French community of Belgium.

†Partially supported by the European Commission under the project QCS (Grant No. 255961).

‡Postdoctoral Researcher of the *Fonds National de la Recherche Scientifique* (F.R.S.–FNRS).

§Partially supported by a Vidi grant from the Netherlands Organization for Scientific Research (NWO), and by the European Commission under the project QCS (Grant No. 255961)

STOC'12, May 19–22, 2012, New York, New York, USA.

Keywords

Combinatorial optimization, linear programming, communication complexity, semidefinite programming, quantum communication complexity

1. INTRODUCTION

In 1986–1987 there were attempts [42] to prove P = NP by giving a polynomial-size LP that would solve the traveling salesman problem (TSP). Due to the large size and complicated structure of the proposed LP for the TSP, it was difficult to show directly that the LP was erroneous. In a groundbreaking effort to refute all such attempts, Yannakakis [47] proved that every symmetric LP for the TSP has exponential size (see [48] for the journal version). Here, an LP is called *symmetric* if every permutation of the cities can be extended to a permutation of all the variables of the LP that preserves the constraints of the LP. Because the proposed LP for the TSP was symmetric, it could not possibly be correct.

In his paper, Yannakakis left as a main open problem the question of proving that the TSP admits no polynomial-size LP, *symmetric or not.* We solve this question by proving a super-polynomial lower bound on the number of inequalities in *every* LP for the TSP. We also prove such unconditional super-polynomial lower bounds for the maximum cut and maximum stable set problems. Therefore, it is impossible to prove P = NP by means of a polynomial-size LP that expresses any of these problems. Our approach is inspired by a close connection between semidefinite programming reformulations of LPs and one-way quantum communication protocols that we introduce here.

1.1 State of the Art

Solving a Problem Through an LP.

A combinatorial optimization problem such as the TSP comes with a natural set of binary variables. When we say that an LP solves the problem, we mean that there exists an LP over this set of variables plus extra variables that returns the correct objective function value for all instances over the same set of natural variables, that is, for *all* choices of weights for the natural variables.

From Problems to Polytopes.

When encoded as 0/1-points in $\mathbb{R}^d$, the feasible solutions of a combinatorial optimization problem yield a polytope that is the convex hull of the resulting points (see Appendix A for background on polytopes). Solving an instance of the problem then amounts to optimizing a linear objective function over this polytope.

For example, the *TSP polytope* $\mathrm{TSP}(n)$ is the convex hull of all points $x \in \{0,1\}^{\binom{n}{2}}$ that correspond to a Hamiltonian cycle in the complete n-vertex graph K_n. If we want to solve a TSP instance with edge-weights w_{ij}, the goal would be to minimize $\sum_{i<j} w_{ij} x_{ij}$ for $x \in \mathrm{TSP}(n)$. This minimum is attained at a vertex of the polytope, i.e., at an $x \in \{0,1\}^{\binom{n}{2}}$ that corresponds to a Hamiltonian cycle.

The idea of representing the set of feasible solutions of a problem by a polytope forms the basis of a standard and powerful methodology in combinatorial optimization, see, e.g., [39].

Extended Formulations and Extensions.

Even for polynomially solvable problems, the associated polytope may have an exponential number of facets. By working in an extended space, it is often possible to decrease the number of constraints. In some cases, a polynomial increase in dimension can be traded for an exponential decrease in the number of constraints. This is the idea underlying extended formulations.

Formally, an *extended formulation* (EF) of a polytope $P \subseteq \mathbb{R}^d$ is a linear system

$$E^{\leqslant} x + F^{\leqslant} y \leqslant g^{\leqslant},\ E^{=} x + F^{=} y = g^{=} \tag{1}$$

in variables $(x,y) \in \mathbb{R}^{d+k}$ such that $x \in P$ if and only if there exists y such that (1) holds. The *size* of an EF is defined as its number of *inequalities* in the system.[1] Optimizing any objective function $f(x)$ over all $x \in P$ amounts to optimizing $f(x)$ over all $(x,y) \in \mathbb{R}^{d+k}$ satisfying (1), provided (1) defines an EF of P.

An *extension* of the polytope P is another polytope[2] $Q \subseteq \mathbb{R}^e$ such that P is the image of Q under a linear map. We define the *size* of an extension Q as the number of facets of Q. If P has an extension of size r, then it has an EF of size r. Conversely, it is known that if P has an EF of size r, then it has an extension of size at most r (see Theorem 3 below). In this sense, the concepts of EF and extension are essentially equivalent.

The Impact of Extended Formulations.

EFs have pervaded the areas of discrete optimization and approximation algorithms for a long time. For instance, Balas' disjunctive programming [5], the Sherali-Adams hierarchy [41], the Lovász-Schrijver closures [32], lift-and-project [6], and configuration LPs are all based on the idea of working in an extended space. Recent surveys on EFs in the context of combinatorial optimization and integer programming are [11, 43, 23, 46].

Symmetry Matters.

Yannakakis [48] proved a $2^{\Omega(n)}$ lower bound on the size of any *symmetric* EF of the TSP polytope $\mathrm{TSP}(n)$ (defined above and in Section 3.4). Although he remarked that he did "not think that asymmetry helps much", it was recently shown by Kaibel *et al.* [24] (see also [35]) that symmetry is a restriction in the sense that there exist polytopes that have polynomial-size EFs but no polynomial-size symmetric EF. This revived Yannakakis's tantalizing question about unconditional lower bounds. That is, bounds which apply to the *extension complexity* of a polytope P, defined as the minimum size of an EF of P.

0/1-Polytopes with Large Extension Complexity.

The strongest unconditional lower bounds so far were obtained by Rothvoß [37]. By an elegant counting argument inspired by Shannon's theorem [40], it was proved that there exist 0/1-polytopes in $\mathbb{R}^d$ whose extension complexity is at least $2^{d/2-o(d)}$. However, Rothvoß's technique does not provide *explicit* 0/1-polytopes with an exponential extension complexity.

The Factorization Theorem.

Yannakakis [48] discovered that the extension complexity of a polytope P is determined by certain factorizations of an associated matrix, called the *slack matrix* of P, that records for each pair (F,v) of a facet F and vertex v, the algebraic distance of v to a hyperplane supporting F. Defining the *nonnegative rank* of a matrix M as the smallest natural number r such that M can be expressed as $M = TU$ where T and U are nonnegative matrices (i.e., matrices whose elements are all nonnegative) with r columns (in case of T) and r rows (in case of U), respectively, it turns out that the extension complexity of every polytope P is exactly the nonnegative rank of its slack matrix.

This *factorization theorem* led Yannakakis to explore connections between EFs and communication complexity. Let S denote the slack matrix of the polytope P. He proved that: (i) every deterministic communication protocol of complexity k computing S gives rise to an EF of P of size at most 2^k, provided S is a 0/1-matrix; (ii) the nondeterministic communication complexity of the *support matrix* of S (i.e., the binary matrix that has 0-entries exactly where S is 0) yields a lower bound on the extension complexity of P, or more generally, the nondeterministic communication complexity of the support matrix of every nonnegative matrix M yields a lower bound on the nonnegative rank of M.[3]

[1] Another possible definition of size is the sum of the number of variables and total number of constraints (equalities plus inequalities) defining the EF. This makes little difference because if $P \subseteq \mathbb{R}^d$ has an EF with r inequalities, then it has an EF with $d+r$ variables, r inequalities and at most $d+r$ equalities (see the proof of Theorem 3 below).

[2] We could allow unbounded polyhedra here, but it again makes little difference because it can be shown that every extension of a polytope with a minimum number of facets is also a polytope.

[3] The classical nondeterministic communication complexity of a binary communication matrix is defined as $\lceil \log B \rceil$, where B is the minimum number of monochromatic 1-rectangles that cover the matrix, see [26]. This last quantity is also known as the *rectangle covering bound*. It is easy to see that the rectangle covering bound of the support matrix of any matrix M lower bounds the nonnegative rank of M (see Theorem 4 below).

Tighter Communication Complexity Connection.
Faenza *et al.* [15] proved that the base-2 logarithm of the nonnegative rank of a matrix equals, up to a small additive constant, the minimum complexity of a randomized communication protocol *with nonnegative outputs* that computes the matrix *in expectation*. In particular, every EF of size r can be regarded as such a protocol of complexity $\log r + O(1)$ bits that computes a slack matrix in expectation.

The Clique vs. Stable Set Problem.
When P is the stable set polytope STAB(G) of a graph G (see Section 3.3), the slack matrix of P contains an interesting row-induced 0/1-submatrix that is the communication matrix of the *clique vs. stable set problem* (also known as the *clique vs. independent set problem*): its rows correspond to the cliques and its columns to the stable sets (or independent sets) of graph G. The entry for a clique K and stable set S equals $1 - |K \cap S|$. Yannakakis [48] gave an $O(\log^2 n)$ deterministic protocol for the clique vs. stable set problem, where n denotes the number of vertices of G. This gives a $2^{O(\log^2 n)} = n^{O(\log n)}$ size EF for STAB(G) whenever the whole slack matrix is 0/1, that is, whenever G is a perfect graph.

A notoriously hard open question is to determine the communication complexity (in the deterministic or nondeterministic sense) of the clique vs. stable set problem. (For recent results that explain why this question is hard, see [27, 28].) The best lower bound to this day is due to Huang and Sudakov [22]: they obtained a $\frac{6}{5}\log n - O(1)$ lower bound.[4] Furthermore, they state a graph-theoretical conjecture that, if true, would imply a $\Omega(\log^2 n)$ lower bound, and hence settle the communication complexity of the clique vs. stable set problem. Moreover it would give a worst-case $n^{\Omega(\log n)}$ lower bound on the extension complexity of stable set polytopes. However, a solution to the Huang-Sudakov conjecture seems only a distant possibility.

1.2 Contribution

Our contribution in this paper is three-fold.

- First, inspired by earlier work [45], we define a $2^n \times 2^n$ matrix $M = M(n)$ and show that the nonnegative rank of M is $2^{\Omega(n)}$ because the nondeterministic communication complexity of its support matrix is $\Omega(n)$. The latter was proved in [45] using the well-known disjointness lower bound of Razborov [36]. We use the matrix M to prove a $2^{\Omega(n)}$ lower bound on the extension complexity of the cut polytope CUT(n) (Section 3.2). That is, we prove that *every* EF of the cut polytope has an exponential number of inequalities. Via reductions, we infer from this: (i) an infinite family of graphs G such that the extension complexity of the corresponding stable set polytope STAB(G) is $2^{\Omega(n^{1/2})}$, where n denotes the number of vertices of G (Section 3.3); (ii) that the extension complexity of the TSP polytope TSP(n) is $2^{\Omega(n^{1/4})}$ (Section 3.4).

  In addition to settling simultaneously the above-mentioned open problems of Yannakakis [48] and Rothvoß [37], our results provide a lower bound on the extension complexity of stable set polytopes that goes beyond what is implied by the Huang-Sudakov conjecture (thanks to the fact that we consider a different part of the slack matrix). Although our lower bounds are strong, unconditional and apply to explicit polytopes that are well-known in combinatorial optimization, they have very accessible proofs.

- Second, we generalize the factorization theorem to *conic* EFs, that is, reformulations of an LP through a conic program. In particular, this implies a factorization theorem for *semidefinite* EFs: the *semidefinite extension complexity* of a polytope equals the *positive semidefinite rank* (*PSD rank*) of its slack matrix.

- Third, we generalize the tight connection between linear[5] EFs and classical communication complexity found by Faenza *et al.* [15] to a tight connection between semidefinite EFs and quantum communication complexity. We show that any *rank-r PSD factorization* of a (nonnegative) matrix M gives rise to a one-way quantum protocol computing M in expectation that uses $\log r + O(1)$ qubits and, *vice versa*, that any one-way quantum protocol computing M in expectation that uses q qubits results in a PSD factorization of M of rank 2^q. Via the semidefinite factorization theorem, this yields a characterization of the semidefinite extension complexity of a polytope in terms of the minimum complexity of quantum protocols that compute the corresponding slack matrix in expectation.

  Then, we give a complexity $\log r + O(1)$ quantum protocol for computing a nonnegative matrix M in expectation, whenever there exists a rank-r matrix N such that M is the entry-wise square of N. This implies in particular that every d-dimensional polytope with 0/1 slacks has a semidefinite EF of size $O(d)$.

  Finally, we obtain an exponential separation between classical and quantum protocols that compute our specific matrix $M = M(n)$ in expectation. On the one hand, our quantum protocol gives a rank-$O(n)$ PSD factorization of M. On the other hand, the nonnegative rank of M is $2^{\Omega(n)}$ because the nondeterministic communication complexity of the support matrix of M is $\Omega(n)$. Thus we obtain an exponential separation between nonnegative rank and PSD rank.

We would like to point out that some of our results in the two last sections were also obtained by Gouveia, Parillo and Thomas. This applies to Theorem 13, Corollary 15, Theorem 18 and Corollary 19. We were aware of the fact that they had obtained Theorem 13 and Corollary 15 prior to writing this paper. However, their proofs were not yet publicly available at that time. Theorem 18 and Corollary 19 were obtained independently, and in a different context. All their results are now publicly available, see [21].

1.3 Related Work

Yannakakis's paper has deeply influenced the TCS community. In addition to the works cited above, it has inspired a whole series of papers on the quality of restricted *approximate* EFs, such as those defined by the Sherali-Adams hi-

[4] All logarithms in this paper are computed in base 2.

[5] In this paragraph, and later in Sections 4 and 5, an EF (in the sense of the previous section) is called a *linear* EF. The use of adjectives such as "linear", "semidefinite" or "conic" will help distinguishing the different types of EFs.

erarchies and Lovász-Schrijver closures starting with [3] ([4] for the journal version), see, e.g., [9, 38, 16, 10, 19, 18, 7].

We would also like to point out that the lower bounds on the extension complexity of polytopes established in Section 3 were obtained by first finding an efficient PSD factorization or, equivalently, an efficient one-way quantum communication protocol for the matrix $M = M(n)$. In this sense our classical lower bounds stem from quantum considerations somewhat similar in style to [25, 1, 2]. See [14] for a survey of this line of work.

1.4 Organization

The discovery of our lower bounds on extension complexity crucially relied on finding the right matrix M and the right polytope whose slack matrix contains M. In our case, we found these through a connection with quantum communication. However, these quantum aspects are not strictly necessary for the resulting lower bound proof itself. Hence, in order to make the main results more accessible to those without background or interest in quantum computing, we start by giving a purely classical presentation of those lower bounds.

In Section 2 we define our matrix M and lower bound the nondeterministic communication complexity of its support matrix. In Section 3 we embed M in the slack matrix of the cut polytope in order to lower bound its extension complexity; further reductions then give lower bounds on the extension complexities of the stable set, and TSP polytopes. In Section 4 we state and prove the factorization theorem for arbitrary closed convex cones. In Section 5 we establish the equivalence of PSD factorizations of a (nonnegative) matrix M and one-way quantum protocols that compute M in expectation, and give an efficient quantum protocol in the case where some entry-wise square root of M has small rank. This is then used to provide an exponential separation between quantum and classical protocols for computing a matrix in expectation (equivalently, an exponential separation between nonnegative rank and PSD rank). Concluding remarks are given in Section 6.

2. A SIMPLE MATRIX WITH LARGE RECTANGLE COVERING BOUND

In this section we consider the following $2^n \times 2^n$ matrix $M = M(n)$ with rows and columns indexed by n-bit strings a and b, and real nonnegative entries:

$$M_{ab} := (1 - a^\intercal b)^2.$$

Note for later reference that M_{ab} can also be written as

$$M_{ab} = 1 - \langle 2\,\mathrm{diag}(a) - aa^\intercal, bb^\intercal\rangle, \tag{2}$$

where $\langle\cdot,\cdot\rangle$ denotes Frobenius inner product[6] and $\mathrm{diag}(a)$ is the $n \times n$ diagonal matrix with the entries of a on its diagonal. Let us verify this identity, using $a, b \in \{0,1\}^n$:

$$\begin{aligned} 1-&\langle 2\,\mathrm{diag}(a) - aa^\intercal, bb^\intercal\rangle \\ &= 1 - 2\langle \mathrm{diag}(a), bb^\intercal\rangle + \langle aa^\intercal, bb^\intercal\rangle \\ &= 1 - 2a^\intercal b + (a^\intercal b)^2 = (1 - a^\intercal b)^2. \end{aligned}$$

[6]The *Frobenius inner product* is the component-wise inner product of two matrices. For matrices X and Y of the same dimensions, this equals $Tr\,[X^\intercal Y]$. When X is symmetric this can also be written $Tr\,[XY]$.

Let $\mathrm{suppmat}(M)$ be the binary support matrix of M, so

$$\mathrm{suppmat}(M)_{ab} = \begin{cases} 1 & \text{if } M_{ab} \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

De Wolf [45] proved that an exponential number of (monochromatic) rectangles are needed to cover all the 1-entries of the support matrix of M. Equivalently, the corresponding function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ has nondeterministic communication complexity of $\Omega(n)$ bits. For the sake of completeness we repeat the proof here:

THEOREM 1 ([45]). *Every 1-monochromatic rectangle cover of* $\mathrm{suppmat}(M)$ *has size* $2^{\Omega(n)}$.

PROOF. Let $R_1, \dots, R_k$ be a 1-cover for f, i.e., a set of (possibly overlapping) 1-monochromatic rectangles in the matrix $\mathrm{suppmat}(M)$ that together cover all 1-entries in $\mathrm{suppmat}(M)$.

We use the following result from [26, Example 3.22 and Section 4.6], which is essentially due to Razborov [36]:

> There exist sets $A, B \subseteq \{0,1\}^n \times \{0,1\}^n$ and probability distribution μ on $\{0,1\}^n \times \{0,1\}^n$ such that all $(a,b) \in A$ have $a^\intercal b = 0$, all $(a,b) \in B$ have $a^\intercal b = 1$, $\mu(A) = 3/4$, and there are constants $\alpha, \delta > 0$ (independent of n) such that for all rectangles R,
>
> $$\mu(R \cap B) \geqslant \alpha \cdot \mu(R \cap A) - 2^{-\delta n}.$$
>
> (For sufficiently large n, $\alpha = 1/135$ and $\delta = 0.017$ will do.)

Since the R_i are 1-rectangles, they cannot contain elements from B. Hence $\mu(R_i \cap B) = 0$ and $\mu(R_i \cap A) \leqslant 2^{-\delta n}/\alpha$. However, since all elements of A are covered by the R_i, we have

$$\frac{3}{4} = \mu(A) = \mu\left(\bigcup_{i=1}^{k}(R_i \cap A)\right) \leqslant \sum_{i=1}^{k} \mu(R_i \cap A) \leqslant k \cdot \frac{2^{-\delta n}}{\alpha}.$$

Hence $k \geqslant 2^{\Omega(n)}$. □

3. STRONG LOWER BOUNDS ON EXTENSION COMPLEXITY

Here we use the matrix $M = M(n)$ defined in the previous section to prove that the (linear) extension complexity of the cut polytope of the n-vertex complete graph is $2^{\Omega(n)}$, i.e., every (linear) EF of this polytope has an exponential number of inequalities. Then, via reductions, we prove super-polynomial lower bounds for the stable set polytopes and the TSP polytopes. To start, let us define more precisely the slack matrix of a polytope. For a matrix A, let A_i denote the ith row of A and A^j to denote the jth column of A.

DEFINITION 2. *Let* $P = \{x \in \mathbb{R}^d \mid Ax \leqslant b\} = \mathrm{conv}(V)$ *be a polytope, with* $A \in \mathbb{R}^{m\times d}$, $b \in \mathbb{R}^m$ *and* $V \subseteq \mathbb{R}^d$, $V = \{v_1, \dots, v_n\}$. *Then* $S \in \mathbb{R}_+^{m\times n}$ *defined as* $S_{ij} := b_i - A_i v_j$ *with* $i \in [m] := \{1, \dots, m\}$ *and* $j \in [n] := \{1, \dots, n\}$ *is the* slack matrix *of P w.r.t. $Ax \leqslant b$ and V. We sometimes refer to the submatrix of the slack matrix induced by rows corresponding to facets and columns corresponding to vertices simply as* the *slack matrix of P, denoted by $S(P)$.*

Let $P \subseteq \mathbb{R}^d$ be a polytope. Recall that:

1. an EF of P is a linear system in variables (x, y) such that $x \in P$ if and only if there exists y satisfying the system;

2. an extension of P is a polytope $Q \subseteq \mathbb{R}^e$ such that there is a linear map $\pi : \mathbb{R}^e \to \mathbb{R}^d$ with $\pi(Q) = P$;

3. the extension complexity of P is the minimum size (i.e., number of inequalities) of an EF of P.

We denote the extension complexity of P by $\mathrm{xc}(P)$.

3.1 The Factorization Theorem

A *rank-r nonnegative factorization* of a (nonnegative) matrix M is a factorization $M = TU$ where T and U are nonnegative matrices with r columns (in case of T) and r rows (in case of U), respectively. The nonnegative rank of M (denoted by: $\mathrm{rank}_+(M)$) is thus simply the minimum rank of a nonnegative factorization of M. Note that $\mathrm{rank}_+(M)$ is also the minimum r such that M is the sum of r nonnegative rank-1 matrices. In particular, the nonnegative rank of a matrix M is at least the nonnegative rank of any submatrix of M.

The following factorization theorem was proved by Yannakakis (see also [17]). It can be stated succinctly as: $\mathrm{xc}(P) = \mathrm{rank}_+(S)$ whenever P is a polytope and S a slack matrix of P. We include a sketch of the proof for completeness.

THEOREM 3 ([48]). *Let $P = \{x \in \mathbb{R}^d \mid Ax \leqslant b\} = \mathrm{conv}(V)$ be a polytope with $\dim(P) \geqslant 1$, and let S denote the slack matrix of P w.r.t. $Ax \leqslant b$ and V. Then the following are equivalent for all positive integers r:*

(i) S has nonnegative rank at most r;

(ii) P has an extension of size at most r;

(iii) P has an EF of size at most r.

PROOF (SKETCH). It should be clear that (ii) implies (iii). We prove that (i) implies (ii), and then that (iii) implies (i).

First, consider a rank-r nonnegative factorization $S = TU$ of the slack matrix of P. Notice that we may assume that no column of T is zero, because otherwise r can be decreased. We claim that P is the image of

$$Q := \{(x, y) \in \mathbb{R}^{d+r} \mid Ax + Ty = b,\ y \geqslant \mathbf{0}\}$$

under the projection π_x onto the x-space. We see immediately that $\pi_x(Q) \subseteq P$ since $Ty \geqslant \mathbf{0}$. To prove the inclusion $P \subseteq \pi_x(Q)$, it suffices to remark that for each point $v_j \in V$ the point (v_j, U^j) is in Q since

$$Av_j + TU^j = Av_j + b - Av_j = b \text{ and } U^j \geqslant \mathbf{0}.$$

Since no column of T is zero, Q is a polytope. Moreover, Q has at most r facets, and is thus an extension of P of size at most r. This proves that (i) implies (ii).

Second, suppose that the system

$$E^{\leqslant}x + F^{\leqslant}y \leqslant g^{\leqslant},\ E^{=}x + F^{=}y = g^{=}$$

with $(x, y) \in \mathbb{R}^{d+k}$ defines an EF of P with r inequalities. Let $Q \subseteq \mathbb{R}^{d+k}$ denote the set of solutions to this system. Thus Q is a (nonnecessarily bounded) *polyhedron*. For each point $v_j \in V$, pick $w_j \in \mathbb{R}^k$ such that $(v_j, w_j) \in Q$. Because

$$Ax \leqslant b \iff \exists y : E^{\leqslant}x + F^{\leqslant}y \leqslant g^{\leqslant},\ E^{=}x + F^{=}y = g^{=},$$

each inequality in $Ax \leqslant b$ is valid for all points of Q. Let S_Q be the nonnegative matrix that records the slacks of the points (v_j, w_j) with respect to the inequalities of $E^{\leqslant}x + F^{\leqslant}y \leqslant g^{\leqslant}$, and then of $Ax \leqslant b$. By construction, the submatrix obtained from S_Q by deleting the r first rows is exactly S, thus $\mathrm{rank}_+(S) \leqslant \mathrm{rank}_+(S_Q)$. Furthermore, it follows from Farkas's lemma (here we use $\dim(P) \geqslant 1$) that every row of S_Q is a nonnegative combination of the first r rows of S_Q. Thus, $\mathrm{rank}_+(S_Q) \leqslant r$. Therefore, $\mathrm{rank}_+(S) \leqslant r$. Hence (iii) implies (i). ☐

We will prove a generalization of Theorem 3 for arbitrary closed convex cones in Section 4, but for now this special case is all we need.

We would like to emphasize that we will not restrict the slack matrix to have rows corresponding only to the facet-defining inequalities. This is not an issue since appending rows corresponding to redundant[7] inequalities does not change the nonnegative rank of the slack matrix. This fact was already used in the second part of the previous proof (see also Lemma 14 below).

Theorem 3 shows in particular that we can lower bound the extension complexity of P by lower bounding the nonnegative rank of its slack matrix S; in fact it suffices to lower bound the nonnegative rank of any submatrix of the slack matrix S corresponding to an implied system of inequalities. To that end, Yannakakis made the following connection with nondeterministic communication complexity. Again, we include the (easy) proof for completeness.

THEOREM 4 ([48]). *Let M be any matrix with nonnegative real entries and* $\mathrm{suppmat}(M)$ *its support matrix. Then* $\mathrm{rank}_+(M)$ *is lower bounded by the rectangle covering bound for* $\mathrm{suppmat}(M)$.

PROOF. If $M = TU$ is a rank-r nonnegative factorization of M, then S can be written as the sum of r nonnegative rank-1 matrices:

$$S = \sum_{k=1}^{r} T^k U_k.$$

Taking the support on each side, we find

$$\begin{aligned} \mathrm{supp}(S) &= \bigcup_{k=1}^{r} \mathrm{supp}(T^k U_k) \\ &= \bigcup_{k=1}^{r} \mathrm{supp}(T^k) \times \mathrm{supp}(U_k). \end{aligned}$$

Therefore, $\mathrm{suppmat}(M)$ has a 1-monochromatic rectangle cover with r rectangles. ☐

3.2 Cut Polytopes

Let $K_n = (V_n, E_n)$ denote the n-vertex complete graph. For a set X of vertices of K_n, we let $\delta(X)$ denote the set of edges of K_n with one endpoint in X and the other in its

[7] An inequality of a linear system is called *redundant* if removing the inequality from the system does not change the set of solutions.

complement $\bar{X}$. This set $\delta(X)$ is known as the *cut* defined by X. For a subset F of edges of K_n, we let $\chi^F \in \mathbb{R}^{E_n}$ denote the *characteristic vector* of F, with $\chi^F_e = 1$ if $e \in F$ and $\chi^F_e = 0$ otherwise. The *cut polytope* $\mathrm{CUT}(n)$ is defined as the convex hull of the characteristic vectors of all cuts in the complete graph $K_n = (V_n, E_n)$. That is,

$$\mathrm{CUT}(n) := \mathrm{conv}\{\chi^{\delta(X)} \in \mathbb{R}^{E_n} \mid X \subseteq V_n\}.$$

We will not deal with the cut polytopes directly, but rather with 0/1-polytopes that are linearly isomorphic to them. The *correlation polytope* (or *boolean quadric polytope*) $\mathrm{COR}(n)$ is defined as the convex hull of all the rank-1 binary symmetric matrices of size $n \times n$. In other words,

$$\mathrm{COR}(n) := \mathrm{conv}\{bb^\intercal \in \mathbb{R}^{n\times n} \mid b \in \{0,1\}^n\}.$$

We use the following known result:

Theorem 5 ([12]). *For all n, $\mathrm{COR}(n)$ is linearly isomorphic to $\mathrm{CUT}(n+1)$.*

Because M is nonnegative, Eq. (2) gives us a linear inequality that is satisfied by all vertices $bb^\intercal$ of $\mathrm{COR}(n)$, and hence (by convexity) is satisfied by all points of $\mathrm{COR}(n)$:

Lemma 6. *For all $a \in \{0,1\}^n$, the inequality*

$$\langle 2\,\mathrm{diag}(a) - aa^\intercal, x\rangle \leqslant 1 \tag{3}$$

is valid for $\mathrm{COR}(n)$. Moreover, the slack of vertex $x = bb^\intercal$ with respect to (3) is precisely M_{ab}.

We remark that (3) is weaker than the *hypermetric inequality* [13] $\langle \mathrm{diag}(a) - aa^\intercal, x\rangle \leqslant 0$, in the sense that the face defined by the former is strictly contained in the face defined by the latter. Nevertheless, we persist in using (3). Now, we prove the main result of this section.

Theorem 7. *There exists some constant $C > 0$ such that, for all n,*

$$\mathrm{xc}(\mathrm{CUT}(n+1)) = \mathrm{xc}(\mathrm{COR}(n)) \geqslant 2^{Cn}\ .$$

In particular, the extension complexity of $\mathrm{CUT}(n)$ is $2^{\Omega(n)}$.

Proof. The equality is implied by Theorem 5. Now, consider any system of linear inequalities describing $\mathrm{COR}(n)$ starting with the 2^n inequalities (3), and a slack matrix S w.r.t. this system and $\{bb^\intercal \mid b \in \{0,1\}^n\}$. Next delete from this slack matrix all rows except the 2^n first rows. By Lemma 6, the resulting $2^n \times 2^n$ matrix is M. Using Theorems 3, 4, and 1, and the fact that the nonnegative rank of a matrix is at least the nonnegative rank of any of its submatrices, we have

$$\begin{aligned}\mathrm{xc}(\mathrm{COR}(n)) &= \mathrm{rank}_+(S)\\ &\geqslant \mathrm{rank}_+(M)\\ &\geqslant 2^{Cn}\end{aligned}$$

for some positive constant C. □

3.3 Stable Set Polytopes

A *stable set* S (also called an *independent set*) of a graph $G = (V, E)$ is a subset $S \subseteq V$ of the vertices such that no two of them are adjacent. For a subset $S \subseteq V$, we let $\chi^S \in \mathbb{R}^V$ denote the *characteristic vector* of S, with $\chi^S_v = 1$ if $v \in S$ and $\chi^S_v = 0$ otherwise. The *stable set polytope*, denoted $\mathrm{STAB}(G)$, is the convex hull of the characteristic vectors of all stable sets in G, i.e.,

$$\mathrm{STAB}(G) := \mathrm{conv}\{\chi^S \in \mathbb{R}^V \mid S \text{ stable set of } G\}.$$

Recall that a polytope Q is an extension of a polytope P if P is the image of Q under a linear projection.

Lemma 8. *For each n, there exists a graph H_n with $O(n^2)$ vertices such that $\mathrm{STAB}(H_n)$ contains a face that is an extension of $\mathrm{COR}(n) \cong \mathrm{CUT}(n+1)$.*

Proof. Consider the complete graph K_n with vertex set $V_n := [n]$. For each vertex i of K_n we create two vertices labeled $ii, \overline{ii}$ in H_n and an edge between them. For each edge ij of K_n, we add to H_n four vertices labeled $ij, \overline{ij}, \underline{ij}, \underline{\overline{ij}}$ and all possible six edges between them. We further add the following eight edges to H_n:

$$\{ij,\overline{ii}\}, \{ij,\overline{jj}\}, \{\overline{ij}, ii\}, \{\overline{ij},\overline{jj}\},$$
$$\{\underline{ij},\overline{ii}\}, \{\underline{ij}, jj\}, \{\underline{\overline{ij}}, ii\}, \{\underline{\overline{ij}}, jj\}.$$

See Fig. 1 for an illustration. The number of vertices in H_n is $2n + 4\binom{n}{2}$.

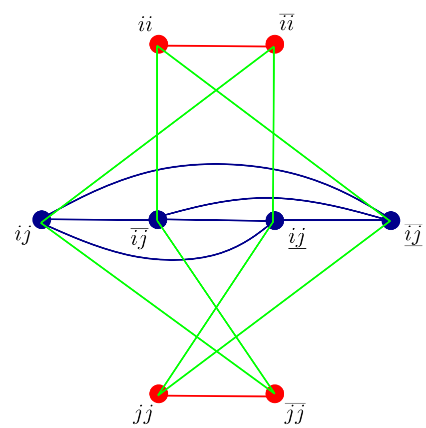

Figure 1: The edges and vertices of H_n corresponding to vertices i, j and edge ij of K_n.

Thus the vertices and edges of K_n are represented by cliques of size 2 and 4 respectively in H_n. We will refer to these as *vertex-cliques* and *edge-cliques* respectively. Consider the face $F = F(n)$ of $\mathrm{STAB}(H_n)$ whose vertices correspond to the stable sets containing exactly one vertex in each vertex-clique and each edge-clique. (The vertices in this face correspond exactly to stable sets of H_n with maximum cardinality.)

Consider the linear map $\pi : \mathbb{R}^{V(H_n)} \to \mathbb{R}^{n\times n}$ mapping a point $x \in \mathbb{R}^{V(H_n)}$ to the point $y \in \mathbb{R}^{n\times n}$ such that $y_{ij} = y_{ji} = x_{ij}$ for $i \leqslant j$. In this equation, the subscripts in y_{ij} and y_{ji} refer to an ordered pair of elements in $[n]$, while the subscript in x_{ij} refers to a vertex of H_n that corresponds either to a vertex of K_n (if $i = j$) or to an edge of K_n (if $i \neq j$).

We claim that the image of F under π is $\mathrm{COR}(n)$, hence F is an extension of $\mathrm{COR}(n)$. Indeed, pick an arbitrary stable set S of H_n such that $x := \chi^S$ is on face F. Then define $b \in \{0,1\}^n$ by letting $b_i := 1$ if $ii \in S$ and $b_i := 0$ otherwise (i.e., $\overline{ii} \in S$). Notice that for the edge ij of K_n we have $ij \in S$ if and only if both vertices ii and jj belong to S. Hence, $\pi(x) = y = bb^\intercal$ is a vertex of $\mathrm{COR}(n)$. This proves $\pi(F) \subseteq \mathrm{COR}(n)$. Now pick a vertex $y := bb^\intercal$ of $\mathrm{COR}(n)$ and consider the unique maximum stable set S that contains vertex ii if $b_i = 1$ and vertex $\overline{ii}$ if $b_i = 0$. Then $x := \chi^S$ is a vertex of F with $\pi(x) = y$. Hence, $\pi(F) \supseteq \mathrm{COR}(n)$. Thus $\pi(F) = \mathrm{COR}(n)$. This concludes the proof. □

Our next lemma establishes simple monotonicity properties of the extension complexity used in our reduction.

LEMMA 9. *Let P, Q and F be polytopes. Then the following hold:*

(i) if F is an extension of P, then $\mathrm{xc}(F) \geqslant \mathrm{xc}(P)$;

(ii) if F is a face of Q, then $\mathrm{xc}(Q) \geqslant \mathrm{xc}(F)$.

PROOF. The first part is obvious because every extension of F is in particular an extension of P. For the second part, notice that a slack matrix of F can be obtained from the (facet-vs-vertex) slack matrix of Q by deleting columns corresponding to vertices not in F. □

Using previous results, we can prove the following result about the worst-case extension complexity of the stable set polytope.

THEOREM 10. *For all n, one can construct a graph G_n with n vertices such that the extension complexity of the stable set polytope $\mathrm{STAB}(G_n)$ is $2^{\Omega(n^{1/2})}$.*

PROOF. W.l.o.g., we may assume $n \geqslant 18$. For an integer $p \geqslant 3$, let $f(p) := |V(H_p)| = 2p + 4\binom{p}{2}$. Given $n \geqslant 18$, we define p as the largest integer with $f(p) \leqslant n$. Now let G_n be obtained from H_p by adding $n - f(p)$ isolated vertices. Then $\mathrm{STAB}(H_p)$ is linearly isomorphic to a face of $\mathrm{STAB}(G_n)$. Using Theorem 7 in combination with Lemmas 8 and 9, we find that

$$\begin{aligned} \mathrm{xc}(\mathrm{STAB}(G_n)) &\geqslant \mathrm{xc}(\mathrm{STAB}(H_p)) \\ &\geqslant \mathrm{xc}(\mathrm{COR}(p)) \\ &= 2^{\Omega(p)} \\ &= 2^{\Omega(n^{1/2})}. \end{aligned}$$

□

3.4 TSP Polytopes

Recall that $\mathrm{TSP}(n)$, the *traveling salesman polytope* or *TSP polytope* of $K_n = (V_n, E_n)$, is defined as the convex hull of the characteristic vectors of all subsets $F \subseteq E_n$ that define a tour of K_n. That is,

$$\mathrm{TSP}(n) := \mathrm{conv}\{\chi^F \in \mathbb{R}^{E_n} \mid F \subseteq E_n \text{ is a tour of } K_n\}.$$

It is known that for every p-vertex graph G, $\mathrm{STAB}(G)$ is the linear projection of a face of $\mathrm{TSP}(n)$ with $n = O(p^2)$, see [48, Theorem 6]. Combining with Theorem 10 gives:

THEOREM 11. *The extension complexity of the TSP polytope $\mathrm{TSP}(n)$ is $2^{\Omega(n^{1/4})}$.*

4. CONIC AND SEMIDEFINITE EFs

In this section we extend Yannakakis's factorization theorem (Theorem 3) to arbitrary closed convex cones. The proof of that theorem shows that, in the linear case, any EF can be brought in the form $Ex + Fy = g$, $y \geqslant \mathbf{0}$ without increasing its size. This form is the basis of our generalization: we replace the nonnegativity constraint $y \geqslant \mathbf{0}$ by a general conic constraint $y \in C$.

Let $Q = \{(x, y) \in \mathbb{R}^{d+k} \mid Ex + Fy = g,\ y \in C\}$ for an arbitrary closed convex cone $C \subseteq \mathbb{R}^k$, where $E \in \mathbb{R}^{p\times d}, F \in \mathbb{R}^{p\times k}$, and $g \in \mathbb{R}^p$. Let $C^* := \{z \in \mathbb{R}^k \mid z^\intercal y \geqslant 0, \forall y \in C\}$ denote the *dual cone* of C. We define the *projection cone of* Q as

$$C_Q := \{\mu \in \mathbb{R}^p \mid F^\intercal \mu \in C^*\}$$

and

$$\mathrm{proj}_x(Q) := \{x \in \mathbb{R}^d \mid \mu^\intercal Ex \leqslant \mu^\intercal g, \forall \mu \in C_Q\}.$$

In a first step we show that $\mathrm{proj}_x(Q)$ equals

$$\pi_x(Q) := \{x \in \mathbb{R}^d \mid \exists y \in \mathbb{R}^k : (x, y) \in Q\},$$

the projection of Q to the x-space.

LEMMA 12. *With the above notation, we have* $\pi_x(Q) = \mathrm{proj}_x(Q)$.

PROOF. Let $\alpha \in \pi_x(Q)$. Then there exists $y \in C$ with $E\alpha + Fy = g$. Pick any $\mu \in C_Q$. Then, $\mu^\intercal E\alpha + \mu^\intercal Fy = \mu^\intercal g$ holds. Since $F^\intercal\mu \in C^*$ and $y \in C$ we have that $(F^\intercal\mu)^\intercal y = \mu^\intercal Fy \geqslant 0$. Therefore $\mu^\intercal E\alpha \leqslant \mu^\intercal g$ holds for all $\mu \in C_Q$. We conclude $\alpha \in \mathrm{proj}_x(Q)$ and as α was arbitrary $\pi_x(Q) \subseteq \mathrm{proj}_x(Q)$ follows.

Now suppose $\pi_x(Q) \neq \mathrm{proj}_x(Q)$. Then there exists α such that $\alpha \in \mathrm{proj}_x(Q)$ but $\alpha \notin \pi_x(Q)$. In other words there is no $y \in C$ such that $Fy = g - E\alpha$ or, equivalently, the convex cone $F(C) := \{Fy \mid y \in C\}$ does not contain the point $g - E\alpha$. Since C is a closed cone, so is $F(C)$. Therefore, by the Strong Separation Theorem there exists $\mu \in \mathbb{R}^p$ such that $\mu^\intercal z \geqslant 0$ is valid for $F(C)$ but $\mu^\intercal(g - E\alpha) < 0$. Then $\mu^\intercal z = \mu^\intercal(Fy) = (\mu^\intercal F)y \geqslant 0$ is valid for C, i.e., $(\mu^\intercal F)y \geqslant 0$ holds for all $y \in C$, implying $F^\intercal\mu \in C^*$. Because $\mu^\intercal(g - E\alpha) < 0$ we have $\mu^\intercal E\alpha > \mu^\intercal g$. On the other hand we have $F^\intercal\mu \in C^*$ so that $\mu \in C_Q$ implying $\mu^\intercal E\alpha \leqslant \mu^\intercal g$; a contradiction. Hence, $\pi_x(Q) = \mathrm{proj}_x(Q)$ follows. □

Let $P = \{x \in \mathbb{R}^d \mid Ax \leqslant b\} = \mathrm{conv}(V)$ be a polytope, with $A \in \mathbb{R}^{m\times d}$, $b \in \mathbb{R}^m$ and $V = \{v_1, \ldots, v_n\} \subseteq \mathbb{R}^d$. Then $Ex + Fy = g$, $y \in C$ is a *conic EF* of P whenever $x \in P$ if and only if there exists $y \in C$ such that $Ex + Fy = g$. The set $Q = \{(x, y) \in \mathbb{R}^{d+k} \mid Ex + Fy = g,\ y \in C\}$ is then called a *conic extension* of P w.r.t. C.

We now prove a factorization theorem for the slack matrix of polytopes that generalizes Yannakakis's factorization theorem in the linear case. Yannakakis's result can be obtained as a corollary of our result by taking $C = \mathbb{R}^k_+$, and using Theorem 13 together with the fact that $(\mathbb{R}^k_+)^* = \mathbb{R}^k_+$.

THEOREM 13. *Let $P = \{x \in \mathbb{R}^d \mid Ax \leqslant b\} = \mathrm{conv}(V)$ be a polytope with $\dim(P) \geqslant 1$ defined by m inequalities and n points respectively, and let S be the slack matrix of P w.r.t. $Ax \leqslant b$ and V. Also, let $C \subseteq \mathbb{R}^k$ be a closed convex cone. Then, the following are equivalent:*

(i) There exist T, U such that (the transpose of) each row of T is in C^*, each column of U is in C, and $S = TU$.

(ii) There exists a conic extension $Q = \{(x, y) \in \mathbb{R}^{d+k} \mid Ex + Fy = g,\ y \in C\}$ with $P = \pi_x(Q)$.

Before proving the theorem, we prove a lemma which will allow us to get rid of rows of a slack matrix that correspond to redundant inequalities. Below, we call a factorization as in (i) a factorization of S *w.r.t.* C.

LEMMA 14. *Let $P \subseteq \mathbb{R}^d$ be a polytope with $\dim(P) \geqslant 1$, let S and S' be two slack matrices of P, and let $C \subseteq \mathbb{R}^k$ be a closed convex cone. Then S has a factorization w.r.t. C iff S' has a factorization w.r.t. C.*

PROOF. It suffices to prove the theorem when S' is the submatrix of S induced by the rows corresponding to facet-defining inequalities and the columns corresponding to vertices, that is, when $S' = S(P)$. One implication is clear: if S has a factorization w.r.t. C, then S' also because S' is a submatrix of S.

For the other implication, consider a system $Ax \leqslant b$ of m inequalities and a set $V = \{v_1, \ldots, v_n\}$ of n points such that $P = \{x \in \mathbb{R}^d \mid Ax \leqslant b\} = \text{conv}(V)$. Assume that the f first inequalities of $Ax \leqslant b$ are facet-defining, while the remaining $m - f$ are not, and that the v first points of V are vertices, while the remaining $n - v$ are not.

Consider an inequality $A_i x \leqslant b_i$ with $i > f$. Suppose first that the inequality is redundant. By Farkas's lemma (using $\dim(P) \geqslant 1$), there exist nonnegative coefficients $\mu_{i,k}$ ($k \in [f]$) such that $A_i = \sum_{k \in [f]} \mu_{i,k} A_k$ and $b_i = \sum_{k \in [f]} \mu_{i,k} b_k$ as P is a polytope. If the inequality is not redundant, since it is not facet-defining, it is satisfied with equality by all points of P. In this case, we let $\mu_{i,k} := 0$ for all $k \in [f]$. Finally, for $i \leqslant f$ we let $\mu_{i,k} := 1$ if $i = k$ and $\mu_{i,k} := 0$ otherwise.

Next, consider a point v_j with $j > v$. Because v_j is in P, it can be expressed as a convex combination of the vertices of P: $v_j = \sum_{\ell \in [v]} \lambda_{j,\ell} v_\ell$, where $\lambda_{j,\ell}$ ($\ell \in [v]$) are nonnegative coefficients that sum up to 1. Similarly as above, for $j \leqslant v$ we let $\lambda_{j,\ell} := 1$ if $j = \ell$ and $\mu_{j,\ell} := 0$ otherwise.

Now, let $S' = TU$ be a factorization of S' w.r.t. C. That is, we have row vectors $T_1, \ldots, T_f$ with $(T_k)^\intercal \in C^*$ (for $k \in [f]$) and column vectors $U^1, \ldots, U^v$ with $U^\ell \in C$ (for $\ell \in [v]$) such that $b_k - A_k v_\ell = S'_{k\ell} = T_k U^\ell$ for $k \in [f], \ell \in [v]$.

We extend the factorization of S' into a factorization of S by letting $T_i := \sum_{k \in [f]} \mu_{i,k} T_k$ and $U^j := \sum_{\ell \in [v]} \lambda_{j,\ell} U^\ell$ for $i > f$ and $j > v$. Given our choice of coefficients, these equations also hold for $i \leqslant f$ and $j \leqslant v$. Clearly, each T_i (transposed) is in C^* and each U^j is in C. A straightforward computation then shows $T_i U^j = S_{ij}$ for all $i \in [m], j \in [n]$. Therefore, T_i ($i \in [m]$) and U^j ($j \in [n]$) define a factorization of S w.r.t. C. ☐

PROOF OF THEOREM 13. We first show that a factorization induces a conic extension. Suppose there exist matrices T, U as above. We claim that Q with $E := A$, $F := T$ and $g := b$ has the desired properties. Let $v_j \in V$, then $S^j = TU^j = b - Av_j$ and so it follows that $(v_j, U^j) \in Q$ and $v_j \in \pi_x(Q)$. Now let $x \in \pi_x(Q)$. Then, there exists $y \in C$ such that $Ax + Ty = b$. Since $T_i y \geqslant 0$ for all $i \in [m]$, we have that $x \in P$. This proves the first implication.

For the converse, suppose $P = \pi_x(Q)$ with Q being a conic extension of P. By Lemma 12, $\pi_x(Q) = \{x \in \mathbb{R}^d \mid \mu^\intercal Ex \leqslant \mu^\intercal g, \forall \mu \in C_Q\}$, where $C_Q = \{\mu \in \mathbb{R}^p \mid F^\intercal \mu \in C^*\}$. By Lemma 14, it suffices to prove that the submatrix of S induced by the rows corresponding to the inequalities of $Ax \leqslant b$ that define facets of P admits a factorization w.r.t. C. Thus, we assume for the rest of the proof that all rows of S correspond to facets of P. Then, for any facet-defining inequality $A_i x \leqslant b_i$ of P there exists $\mu_i \in C_Q$ such that $\mu_i^\intercal Ex \leqslant \mu_i^\intercal g$ defines the same facet as $A_i x \leqslant b_i$. (This follows from the fact that C_Q is closed; see also [29, Theorem 4.3.4].) Scaling μ_i if necessary, this means that $\mu_i^\intercal E = A_i + c^\intercal$ and $\mu_i^\intercal g = b_i + \delta$, where $c^\intercal x = \delta$ is satisfied for *all* points of P. We define $T_i := \mu_i^\intercal F$ for all i; in particular $(T_i)^\intercal \in C^*$ as $\mu_i \in C_Q$. Now let $v_j \in V$. Since $P = \pi_x(Q)$, there exists a $y_j \in C$ such that $Ev_j + Fy_j = g$ and so $\mu_i^\intercal Ev_j + \mu_i^\intercal Fy_j = \mu_i^\intercal g$. With the above we have $A_i v_j + c^\intercal v_j + T_i y_j = b_i + \delta$, hence $A_i v_j + T_i y_j = b_i$ and as $v_j \in \pi_x(Q)$ we deduce $T_i y_j \geqslant 0$. The slack of v_j w.r.t. $A_i x \leqslant b_i$ is $b_i - A_i v_j = \mu_i^\intercal g - \mu_i^\intercal Ev_j = \mu_i^\intercal Fy_j = T_i y_j$. This implies the factorization $S = TU$ with $T_i = \mu_i^\intercal F$ and $U^j = y_j$. ☐

For a positive integer r, we let $\mathbb{S}_+^r$ denote the cone of $r \times r$ symmetric positive semidefinite matrices embedded in $\mathbb{R}^{r(r+1)/2}$ in such a way that, for all $y, z \in \mathbb{S}_+^r$, the scalar product $z^\intercal y$ is the Frobenius product of the corresponding matrices. A *semidefinite EF (resp. extension) of size* r is simply a conic EF (resp. extension) w.r.t. $C = \mathbb{S}_+^r$. The *semidefinite extension complexity* of polytope P, denoted by $\text{xc}_{SDP}(P)$, is the minimum r such that P has a semidefinite EF of size r. Observe that $(\mathbb{S}_+^r)^* = \mathbb{S}_+^r$. Hence, taking $C := \mathbb{S}_+^k$ and $k := r(r+1)/2$ in Theorem 13, we obtain the following factorization theorem for semidefinite EFs.

COROLLARY 15. *Let $P = \{x \in \mathbb{R}^d \mid Ax \leqslant b\} = \text{conv}(V)$ be a polytope. Then the slack matrix S of P w.r.t. $Ax \leqslant b$ and V has a factorization $S = TU$ so that $(T_i)^\intercal, U^j \in \mathbb{S}_+^r$ if and only if there exists a semidefinite extension $Q = \{(x, y) \in \mathbb{R}^{d+r(r+1)/2} \mid Ex + Fy = g,\ y \in \mathbb{S}_+^r\}$ with $P = \pi_x(Q)$.*

Analogous to nonnegative factorizations and nonnegative rank, we can define PSD factorizations and PSD rank. A *rank-r PSD factorization* of an $m \times n$ matrix M is a collection of $r \times r$ symmetric positive semidefinite matrices $T_1, \ldots, T_m$ and $U^1, \ldots, U^n$ such that the Frobenius product $\langle T_i, U^j \rangle = Tr\left[(T_i)^\intercal U^j\right] = Tr\left[T_i U^j\right]$ equals M_{ij} for all $i \in [m], j \in [n]$. The *PSD rank* of M is the minimum r such that M has a rank-r PSD factorization. We denote this $\text{rank}_{PSD}(M)$. By Corollary 15 (and also Lemma 14), the semidefinite extension complexity of a polytope P is equal to the PSD rank of any slack matrix of P: $\text{xc}_{SDP}(P) = \text{rank}_{PSD}(S)$ whenever S is a slack matrix of P. In the next section we will show that $\text{rank}_{PSD}(M)$ can be expressed in terms of the amount of communication needed by a one-way quantum communication protocol for computing M in expectation (Corollary 17).

5. QUANTUM COMMUNICATION AND PSD FACTORIZATIONS

In this section we explain the connection with quantum communication. This yields results that are interesting in their own right, and also clarifies where the matrix M of Section 2 came from.

For a general introduction to quantum computation we refer to [34, 33], and for quantum communication complexity we refer to [44, 8]. For our purposes, an r-dimensional *quantum state* ρ is an $r \times r$ PSD matrix of trace 1.[8] A *k-qubit state*

[8]For simplicity we restrict to real rather than complex entries, which doesn't significantly affect the results.

is a state in dimension $r = 2^k$. If ρ has rank 1, it can be written as an outer product $|\phi\rangle\langle\phi|$ for some unit vector $|\phi\rangle$, which is sometimes called a *pure state*. We use $|i\rangle$ to denote the pure state vector that has 1 at position i and 0s elsewhere. A quantum measurement (POVM) is described by a set of PSD matrices $\{E_\theta\}_{\theta\in\Theta}$, each labeled by a real number θ, and summing to the r-dimensional identity: $\sum_{\theta\in\Theta} E_\theta = I$. When measuring state ρ with this measurement, the probability of outcome θ equals $Tr\,[E_\theta\rho]$. Note that if we define the PSD matrix $E := \sum_{\theta\in\Theta}\theta E_\theta$, then the *expected value* of the measurement outcome is $\sum_{\theta\in\Theta}\theta Tr\,[E_\theta\rho] = Tr\,[E\rho]$.

5.1 Quantum Protocols

A *one-way quantum protocol with r-dimensional messages* can be described as follows. On input i, Alice sends Bob an r-dimensional state ρ_i. On input j, Bob measures the state he receives with a POVM $\{E^j_\theta\}$ for some *nonnegative* values θ, and outputs the result. We say that such a protocol *computes a matrix M in expectation*, if the expected value of the output on respective inputs i and j, equals the matrix entry M_{ij}. Analogous to the equivalence between classical protocols and nonnegative factorizations of M established by Faenza *et al.* [15], such quantum protocols are essentially equivalent to PSD factorizations of S:

Theorem 16. *Let $M \in \mathbb{R}^{m\times n}_+$ be a matrix. Then the following holds:*

(i) A one-way quantum protocol with r-dimensional messages that computes M in expectation, gives a rank-r PSD factorization of M.

(ii) A rank-r PSD factorization of M gives a one-way quantum protocol with $(r+1)$-dimensional messages that computes M in expectation.

Proof. The first part is straightforward. Given a quantum protocol as above, define $E^j := \sum_{\theta\in\Theta}\theta E^j_\theta$. Clearly, on inputs i and j the expected value of the output is $Tr\left[\rho_i E^j\right] = M_{ij}$.

For the second part, suppose we are given a PSD factorization of a matrix M, so we are given PSD matrices $T_1,\ldots,T_m$ and $U^1,\ldots,U^n$ satisfying $Tr\left[T_iU^j\right] = M_{ij}$ for all i,j. In order to turn this into a quantum protocol, define $\tau = \max_i Tr\,[T_i]$. Let ρ_i be the $(r+1)$-dimensional quantum state obtained by adding a $(r+1)$st row and column to T_i/τ, with $1 - Tr\,[T_i]/\tau$ as $(r+1)$st diagonal entry, and 0s elsewhere. Note that ρ_i is indeed a PSD matrix of trace 1, so it is a well-defined quantum state. For input j, derive Bob's $(r+1)$-dimensional POVM from the PSD matrix U^j as follows. Let λ be the largest eigenvalue of U^j, and define $E^j_{\tau\lambda}$ to be U^j/λ, extended with a $(d+1)$st row and column of 0s. Let $E^j_0 = I - E^j_{\tau\lambda}$. These two operators together form a well-defined POVM. The expected outcome (on inputs i,j) of the protocol induced by the states and POVMs that we just defined, is

$$\tau\lambda Tr\left[E^j_{\tau\lambda}\rho_i\right] = Tr\left[T_iU^j\right] = M_{ij},$$

so the protocol indeed computes M in expectation. □

We obtain the following corollary which summarizes the characterization of semidefinite EFs:

Corollary 17. *For a polytope P with slack matrix S, the following are equivalent:*

(i) P has a semidefinite extension $Q = \{(x,y) \in \mathbb{R}^{d+r(r+1)/2} \mid Ex + Fy = g, y \in \mathbb{S}^r_+\}$;

(ii) the slack matrix S has a rank-r PSD factorization;

(iii) there exists a one-way quantum communication protocol with $(r+1)$-dimensional messages (i.e., using $\lceil\log(r+1)\rceil$ qubits) that computes S in expectation (for the converse we consider r-dimensional messages).

5.2 A General Upper Bound on Quantum Communication

Now, we provide a quantum protocol that efficiently computes a nonnegative matrix M in expectation, whenever there exists a low rank matrix N whose entry-wise square is M. The quantum protocol is inspired by [45, Section 3.3].

Theorem 18. *Let M be a matrix with nonnegative real entries, N be a rank-r matrix of the same dimensions such that $M_{ij} = N^2_{ij}$. Then there exists a one-way quantum protocol using $(r+1)$-dimensional pure-state messages that computes M in expectation.*

Proof. Let $N^\intercal = U\Sigma V$ be the singular value decomposition of the transpose of N; so U and V are unitary, Σ is a matrix whose first r diagonal entries are nonzero while its other entries are 0, and $\langle j|U\Sigma V|i\rangle = N_{ij}$. Define $|\phi_i\rangle = \Sigma V|i\rangle$. Since only its first r entries can be nonzero, we will view $|\phi_i\rangle$ as an r-dimensional vector. Let $\Delta_i = \||\phi_i\rangle\|$ and $\Delta = \max_i \Delta_i$. Add one additional dimension and define the normalized $(r+1)$-dimensional pure quantum states $|\psi_i\rangle = (|\phi_i\rangle/\Delta, \sqrt{1-\Delta_i^2/\Delta^2})$. Given input i, Alice sends $|\psi_i\rangle$ to Bob. Given input j, Bob applies a 2-outcome POVM $\{E^j_{\Delta^2}, E^j_0 = I - E^j_{\Delta^2}\}$ where $E^j_{\Delta^2}$ is the projector on the pure state $U^*|j\rangle$ (which has no support in the last dimension of $|\psi_i\rangle$). If the outcome of the measurement is $E^j_{\Delta^2}$ then Bob outputs Δ^2, otherwise he outputs 0. Accordingly, the expected output of this protocol on input (i,j) is

$$\begin{aligned}\Delta^2\,\Pr[\text{outcome } E^j_{\Delta^2}] &= \Delta^2\langle\psi_i|E^j_{\Delta^2}|\psi_i\rangle = \langle\phi_i|E^j_{\Delta^2}|\phi_i\rangle\\ &= |\langle j|U|\phi_i\rangle|^2 = |\langle j|U\Sigma V|i\rangle|^2 = N^2_{ij} = M_{ij}.\end{aligned}$$

The protocol only has two possible outputs: 0 and Δ^2, both nonnegative. Hence it computes M in expectation with an $(r+1)$-dimensional quantum message. □

Note that if M is a 0/1-matrix then we may take $N = M$, hence any low-rank 0/1-matrix can be computed in expectation by an efficient quantum protocol. We obtain the following corollary (implicit in Theorem 4.2 of [20]) which also implies a compact (i.e., polynomial size) semidefinite EF for the stable set polytope of perfect graphs, reproving the previously known result by [30, 31]. We point out that the result still holds when $\dim(P)+2$ is replaced by $\dim(P)+1$, see [21]. (This difference is due to normalization.)

Corollary 19. *Let P be a polytope such that $S(P)$ is a 0/1 matrix. Then $\mathrm{xc}_{SDP}(P) \leqslant \dim(P) + 2$.*

5.3 Quantum vs Classical Communication, and PSD vs Nonnegative Factorizations

We now give an example of an exponential separation between quantum and classical communication in expectation, based on the matrix M of Section 2. This result actually preceded and inspired the results in Section 3.

Theorem 20. *For each n, there exists a nonnegative matrix $M \in \mathbb{R}^{2^n \times 2^n}$ that can be computed in expectation by a quantum protocol using $\log n + O(1)$ qubits, while any classical randomized protocol needs $\Omega(n)$ bits to compute M in expectation.*

Proof. Consider the matrix $N \in \mathbb{R}^{2^n \times 2^n}$ whose rows and columns are indexed by n-bit strings a and b, respectively, and whose entries are defined as $N_{ab} = 1 - a^\intercal b$. Define $M \in \mathbb{R}_+^{2^n \times 2^n}$ by $M_{ab} = N_{ab}^2$. This M is the matrix from Section 2. Note that N has rank $r \leqslant n+1$ because it can be written as the sum of $n+1$ rank-1 matrices. Hence Theorem 18 immediately implies a quantum protocol with $(n+2)$-dimensional messages that computes M in expectation.

For the classical lower bound, note that a protocol that computes M in expectation has positive probability of giving a nonzero output on input a, b iff $M_{ab} > 0$. With a message m in this protocol we can associate a rectangle $R_m = A \times B$ where A consists of all inputs a for which Alice has positive probability of sending m, and B consists of all inputs b for which Bob, when he receives message m, has positive probability of giving a nonzero output. Together these rectangles will cover exactly the nonzero entries of M. Accordingly, a c-bit protocol that computes M in expectation induces a rectangle cover for the support matrix of M of size 2^c. Theorem 1 lower bounds the size of such a cover by $2^{\Omega(n)}$, hence $c = \Omega(n)$. □

Together with Theorem 16 and the equivalence of randomized communication complexity (in expectation) and nonnegative rank established in [15], we immediately obtain an exponential separation between the nonnegative rank and the PSD rank.

Corollary 21. *For each n, there exists $M \in \mathbb{R}_+^{2^n \times 2^n}$, with $\operatorname{rank}_+(M) = 2^{\Omega(n)}$ and $\operatorname{rank}_{PSD}(M) = O(n)$.*

In fact a simple rank-$(n+1)$ PSD factorization of M is the following: let $T_a := \binom{1}{-a}\binom{1}{-a}^\intercal$ and $U^b := \binom{1}{b}\binom{1}{b}^\intercal$, then $\text{Tr}[(T_a)^\intercal U^b] = (1 - a^\intercal b)^2 = M_{ab}$.

6. CONCLUDING REMARKS

In addition to proving the first unconditional super-polynomial lower bounds on the size of linear EFs for the cut polytope, stable set polytope and TSP polytope, we demonstrate that the rectangle covering bound can prove strong results in the context of EFs. In particular, it can be super-polynomial in the dimension and the logarithm of the number of vertices of the polytope, settling an open problem of Fiorini *et al.* [17].

The exponential separation between nonnegative rank and PSD rank that we prove here (Theorem 20) actually implies more than a super-polynomial lower bound on the extension complexity of the cut polytope. As noted in Theorem 5, the polytopes $\text{CUT}(n)$ and $\text{COR}(n-1)$ are affinely isomorphic. Let $Q(n)$ denote the polyhedron isomorphic (under the same affine map) to the polyhedron defined by (3) for $a \in \{0,1\}^n$. Then (i) *every* polytope (or polyhedron) that contains $\text{CUT}(n)$ and is contained in $Q(n)$ has exponential extension complexity; (ii) there exists a low complexity spectrahedron that contains $\text{CUT}(n)$ and is contained in $Q(n)$. (A *spectrahedron* is an intersection of the positive semidefinite cone with an affine subspace, or any projection of such convex set.)

An important problem also left open in [48] is whether the perfect matching polytope has a polynomial-size linear EF. Yannakakis proved that every *symmetric* EF of this polytope has exponential size, a striking result given the fact that the perfect matching problem is solvable in polynomial time. He conjectured that asymmetry also does not help in the case of the perfect matching polytope. Because it is based on the rectangle covering bound, our argument would not yield a super-polynomial lower bound on the extension complexity of the perfect matching polytope. Even though a polynomial-size linear EF of the matching polytope would not prove anything as surprising as P=NP, the existence of a polynomial-size EF or an unconditional super-polynomial lower bound for it remains open.

We hope that the new connections developed here will inspire more research, in particular about approximate EFs. Here are two concrete questions left open for future work: (i) find a *slack matrix* that has an exponential gap between nonnegative rank and PSD rank; (ii) prove that the cut polytope has no polynomial-size *semidefinite* EF (that would rule out SDP-based algorithms for optimizing over the cut polytope, in the same way that this paper ruled out LP-based algorithms).

Acknowledgements

We thank Giannicola Scarpa for useful discussions. We also thank Monique Laurent for information about hypermetric inequalities. Finally, we thank the three anonymous referees for suggesting improvements to the text.

7. REFERENCES

[1] S. Aaronson. Lower bounds for local search by quantum arguments. In *Proc. STOC 2004*, pages 465–474, 2004.

[2] D. Aharonov and O. Regev. Lattice problems in NP ∩ coNP. In *Proc. FOCS 2004*, pages 362–371, 2004.

[3] S. Arora, B. Bollobás, and L. Lovász. Proving integrality gaps without knowing the linear program. In *Proc. FOCS 2002*, pages 313–322, 2002.

[4] S. Arora, B. Bollobás, L. Lovász, and I. Tourlakis. Proving integrality gaps without knowing the linear program. *Theory Comput.*, 2:19–51, 2006.

[5] E. Balas. Disjunctive programming and a hierarchy of relaxations for discrete optimization problems. *SIAM J. Algebraic Discrete Methods*, 6:466–486, 1985.

[6] E. Balas, S. Ceria, and G. Cornuéjols. A lift-and-project algorithm for mixed 0-1 programs. *Math. Programming*, 58:295–324, 1993.

[7] S. Benabbas and A. Magen. Extending SDP integrality gaps to Sherali-Adams with applications to quadratic programming and MaxCutGain. In *Proc. IPCO 2010*, pages 299–312, 2010.

[8] H. Buhrman, R. Cleve, S. Massar, and R. de Wolf. Nonlocality and communication complexity. *Rev. Modern Phys.*, 82:665, 2010.

[9] J. Buresh-Oppenheim, N. Galesi, S. Hoory, A. Magen, and T. Pitassi. Rank bounds and integrality gaps for cutting planes procedures. *Theory Comput.*, 2:65–90, 2006.

[10] M. Charikar, K. Makarychev, and Y. Makarychev. Integrality gaps for Sherali-Adams relaxations. In *Proc. STOC 2009*, pages 283–292, 2009.

[11] M. Conforti, G. Cornuéjols, and G. Zambelli. Extended formulations in combinatorial optimization. *4OR*, 8:1–48, 2010.

[12] C. De Simone. The cut polytope and the Boolean quadric polytope. *Discrete Math.*, 79:71–75, 1989/90.

[13] M.M. Deza and M. Laurent. *Geometry of cuts and metrics*, volume 15 of *Algorithms and Combinatorics*. Springer-Verlag, 1997.

[14] A. Drucker and R. de Wolf. Quantum proofs for classical theorems. *Theory Comput., Graduate Surveys*, 2, 2011.

[15] Y. Faenza, S. Fiorini, R. Grappe, and H. R. Tiwary. Extended formulations, non-negative factorizations and randomized communication protocols. arXiv:1105.4127, 2011.

[16] W. Fernandez de la Vega and C. Mathieu. Linear programming relaxation of Maxcut. In *Proc. SODA 2007*, pages 53–61, 2007.

[17] S. Fiorini, V. Kaibel, K. Pashkovich, and D. O. Theis. Combinatorial bounds on nonnegative rank and extended formulations. arXiv:1111.0444, 2011.

[18] K. Georgiou, A. Magen, T. Pitassi, and I. Tourlakis. Integrality gaps of $2 - o(1)$ for vertex cover SDPs in the Lovász-Schrijver hierarchy. *SIAM J. Comput.*, 39:3553–3570, 2010.

[19] K. Georgiou, A. Magen, and M. Tulsiani. Optimal Sherali-Adams gaps from pairwise independence. In *Proc. APPROX-RANDOM 2009*, pages 125–139, 2009.

[20] J. Gouveia, P.A. Parrilo, and R. Thomas. Theta bodies for polynomial ideals. *SIAM J. Optim.*, 20:2097–2118, 2010.

[21] J. Gouveia, P.A. Parrilo, and R. Thomas. Lifts of convex sets and cone factorizations. arXiv:1111.3164, 2011.

[22] H. Huang and B. Sudakov. A counterexample to the alon-saks-seymour conjecture and related problems. arXiv:1002.4687, 2010.

[23] V. Kaibel. Extended formulations in combinatorial optimization. *Optima*, 85:2–7, 2011.

[24] V. Kaibel, K. Pashkovich, and D.O. Theis. Symmetry matters for the sizes of extended formulations. In *Proc. IPCO 2010*, pages 135–148, 2010.

[25] I. Kerenidis and R. de Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. In *Proc. STOC 2003*, pages 106–115, 2003.

[26] E. Kushilevitz and N. Nisan. *Communication complexity*. Cambridge University Press, 1997.

[27] E. Kushilevitz and E. Weinreb. The communication complexity of set-disjointness with small sets and 0-1 intersection. In *Proc. FOCS 2009*, pages 63–72, 2009.

[28] E. Kushilevitz and E. Weinreb. On the complexity of communication complexity. In *Proc. STOC 2009*, pages 465–474, 2009.

[29] C. Lemaréchal and J.B. Hiriart-Urruty. *Convex analysis and minimization algorithms I*. Springer, 1996.

[30] L. Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25:1–7, 1979.

[31] L. Lovász. Semidefinite programs and combinatorial optimization. In *Recent advances in algorithms and combinatorics*, volume 11 of *CMS Books Math./Ouvrages Math. SMC*, pages 137–194. Springer, 2003.

[32] L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM J. Optim.*, 1:166–190, 1991.

[33] N.D. Mermin. *Quantum computer science: an introduction*. Cambridge University Press, 2007.

[34] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2000.

[35] K. Pashkovich. Symmetry in extended formulations of the permutahedron. arXiv:0912.3446, 2009.

[36] A. A. Razborov. On the distributional complexity of disjointness. *Theoret. Comput. Sci.*, 106(2):385–390, 1992.

[37] T. Rothvoß. Some 0/1 polytopes need exponential size extended formulations. arXiv:1105.0036, 2011.

[38] G. Schoenebeck, L. Trevisan, and M. Tulsiani. Tight integrality gaps for Lovasz-Schrijver LP relaxations of vertex cover and max cut. In *Proc. STOC 2007*, pages 302–310, 2007.

[39] A. Schrijver. *Combinatorial optimization. Polyhedra and efficiency*. Springer-Verlag, 2003.

[40] C. E. Shannon. The synthesis of two-terminal switching circuits. *Bell System Tech. J.*, 25:59–98, 1949.

[41] H. D. Sherali and W. P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM J. Discrete Math.*, 3:411–430, 1990.

[42] E. R. Swart. P = NP. Technical report, University of Guelph, 1986; revision 1987.

[43] F. Vanderbeck and L. A. Wolsey. Reformulation and decomposition of integer programs. In M. Jünger et al., editor, *50 Years of Integer Programming 1958-2008*, pages 431–502. Springer, 2010.

[44] R. de Wolf. Quantum communication and complexity. *Theoret. Comput. Sci.*, 287:337–353, 2002.

[45] R. de Wolf. Nondeterministic quantum query and communication complexities. *SIAM J. Comput.*, 32:681–699, 2003.

[46] L. A. Wolsey. Using extended formulations in practice. *Optima*, 85:7–9, 2011.

[47] M. Yannakakis. Expressing combinatorial optimization problems by linear programs (extended abstract). In *Proc. STOC 1988*, pages 223–228, 1988.

[48] M. Yannakakis. Expressing combinatorial optimization problems by linear programs. *J. Comput. System Sci.*, 43(3):441–466, 1991.

[49] G. M. Ziegler. *Lectures on Polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer-Verlag, 1995.

APPENDIX

A. BACKGROUND ON POLYTOPES

A *(convex) polytope* is a set $P \subseteq \mathbb{R}^d$ that is the convex hull $\operatorname{conv}(V)$ of a finite set of points V. Equivalently, P is a polytope if and only if P is bounded and the intersection of a finite collection of closed halfspaces. This is equivalent to saying that P is bounded and the set of solutions of a finite system of linear inequalities and possibly equalities (each of which can be represented by a pair of inequalities).

Let $P \subseteq \mathbb{R}^d$ be a polytope. A closed halfspace H^+ that contains P is said to be *valid* for P. In this case the hyperplane H that bounds H^+ is also said to be *valid* for P. A *face* of P is either P itself or the intersection of P with a valid hyperplane. Every face of a polytope is again a polytope. A face is called *proper* if it is not the polytope itself. A *vertex* is a minimal nonempty face. A *facet* is a maximal proper face. An inequality $c^\intercal x \leqslant \delta$ is said to be *valid* for P if it is satisfied by all points of P. The face it defines is $F := \{x \in P \mid c^\intercal x = \delta\}$. The inequality is called *facet-defining* if F is a facet. The *dimension* of a polytope P is the dimension of its affine hull $\mathrm{aff}(P)$.

Every (finite or infinite) set V such that $P = \mathrm{conv}(V)$ contains all the vertices of P. Conversely, letting $\mathrm{vert}(P)$ denote the vertex set of P, we have $P = \mathrm{conv}(\mathrm{vert}(P))$. Suppose now that P is *full-dimensional*, i.e., $\dim(P) = d$. Then, every (finite) system $Ax \leqslant b$ such that $P = \{x \in \mathbb{R}^d \mid Ax \leqslant b\}$ contains all the facet-defining inequalities of P, up to scaling by positive numbers. Conversely, P is described by its facet-defining inequalities.

If P is not full-dimensional, these statements have to be adapted as follows. Every (finite) system describing P contains all the facet-defining inequalities of P, up to scaling by positive numbers and adding an inequality that is satisfied with equality by *all* points of P. Conversely, a linear description of P can be obtained by picking one inequality per facet and adding a system of equalities describing $\mathrm{aff}(P)$.

A 0/1-*polytope* in $\mathbb{R}^d$ is simply the convex hull of a subset of $\{0,1\}^d$.

A (convex) *polyhedron* is a set $P \subseteq \mathbb{R}^d$ that is the intersection of a finite collection of closed halfspaces. A polyhedron P is a polytope if and only if it is bounded.

For more background on polytopes and polyhedra, see the standard reference [49].