AMS Short Course on Sums of Squares: Theory and Practice

Notes for lectures on *Lifts of convex sets*

Hamza Fawzi*

December 14, 2018

These notes are for the lectures on *Lifts of convex sets* part of the AMS Short Course on Sum of Squares: Theory and Practice, Baltimore, January 14-15, 2019. The present notes are mostly based on Chapters 1 and 2 of [Faw16].

1 Lifts of convex sets

1.1 Introduction

In this short introductory section we give a brief overview of the lifting method and discuss its importance in optimization when combined with the idea of *convex reformulations*. We illustrate the power of the lifting method on a simple example and discuss in an informal way linear programming and semidefinite programming lifts.

Convex reformulations Consider the following problem where we want to minimize a linear function $\ell(x)$ subject to the constraint $x \in X$:

$$\text{minimize} \quad \ell(x) \quad \text{subject to} \quad x \in X. \tag{1}$$

Here X is an arbitrary subset of $\mathbb{R}^n$ and need not be convex. It is a well-known fact that, since the objective function is linear, the optimal value of (1) remains unchanged if we change the constraint "$x \in X$" by "$x \in \text{conv}(X)$", where $\text{conv}(X)$ denotes the convex hull of X (see Figure 1 for an illustration):

$$\text{minimize} \quad \ell(x) \quad \text{subject to} \quad x \in \text{conv}(X). \tag{2}$$

Recall that the convex hull of X is the set of all possible convex combinations of elements of X:

$$\text{conv}(X) = \left\{ \sum_{i=1}^m \lambda_i x_i : m \in \mathbb{N}, \lambda_1, \dots, \lambda_m \geq 0, \sum_{i=1}^m \lambda_i = 1, x_1, \dots, x_m \in X \right\}.$$

Using this definition the equality of the optimal values of (1) and (2) is straightforward to verify. Note that problem (2) is now formally *convex* since the cost function is linear and the feasible set is convex.

The transformation from (1) to (2) seems to rely heavily on the fact that ℓ is a linear function. It turns out however that a similar transformation can be applied more generally, even if the objective is not linear, by introducing additional variables and constraints. To illustrate this, assume that

*Department of Applied Mathematics and Theoretical Physics, University of Cambridge. Email: h.fawzi@damtp.cam.ac.uk.

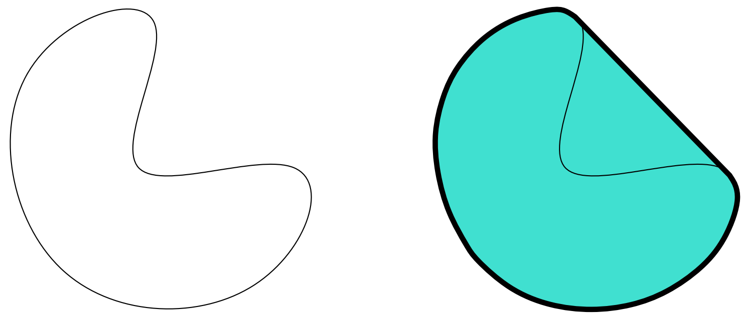

Figure 1: A nonconvex set X and its convex hull

our objective in (1) was quadratic instead of being linear, i.e., we are interested in minimizing $q(x) = \sum_{i \le j} q_{ij} x_i x_j$ subject to $x \in X$:

$$\text{minimize} \quad q(x) = \sum_{i \le j} q_{ij} x_i x_j \quad \text{subject to} \quad x \in X. \tag{3}$$

If we introduce additional variables y_{ij} playing the role of $x_i x_j$ we can reformulate the problem above as follows:

$$\text{minimize} \quad \sum_{i \le j} q_{ij} y_{ij} \quad \text{subject to} \quad y \in Y \tag{4}$$

where Y is defined as

$$Y = \left\{ (y_{ij})_{i \le j} : \exists x \in X \text{ s.t. } y_{ij} = x_i x_j \ \forall 1 \le i \le j \le n \right\}.$$

The objective function of (4) is now linear and thus by the same reasoning as above the constraint "$y \in Y$" can be changed to "$y \in \mathrm{conv}(Y)$". In both cases we have transformed the original problem to a new problem that is convex, at least formally. In order to solve the problem however we need to find a tractable representation of the set $\mathrm{conv}(X)$ (or $\mathrm{conv}(Y)$).

Lifts of convex sets When the set X is finite (e.g., in discrete and combinatorial optimization), the corresponding convex set $P = \mathrm{conv}(X)$ is called a *polytope* and can be described using a finite number of linear inequalities. The problem of optimizing a linear function over a polytope P is known as linear programming. Interior-point methods are a popular class of algorithms for linear programming, and the complexity of such algorithms typically depend on the size of the inequality description of the polytope. The size of the trivial such description is equal to the number of *facets* of P, a geometric quantity associated to P. Unfortunately, in many cases of interest, the number of facets of P is prohibitively large to enumerate directly in a linear programming formulation.

The idea of *lifting* consists in expressing the polytope P as the projection of a higher-dimensional polytope Q that has much fewer facets than P. We say in this case that Q is a (linear programming) *lift* of P. For the purpose of optimization one can then work over Q rather than working over P. Indeed if we are interested in minimizing ℓ over P and if $P = \pi(Q)$ where π is a linear (projection) map then we have:

$$\min_{x \in P} \ell(x) = \min_{y \in Q} \ell \circ \pi(y). \tag{5}$$

To give a simple example of a lift let P denote the ℓ_1 ball in $\mathbb{R}^n$:

$$P = \{x \in \mathbb{R}^n : \|x\|_1 \le 1\}.$$

The trivial description of P obtained by enumerating the facets has size 2^n since the facets of the ℓ_1 ball are given by all possible inequalities of the form $\pm x_1 \pm x_2 \pm \cdots \pm x_n \le 1$, see Figure 2. It is not difficult however to see that the ℓ_1 ball admits a description using only $2n$ linear inequalities, namely:

$$P = \left\{ x \in \mathbb{R}^n : \exists y \in \mathbb{R}^n \text{ s.t. } -y_i \le x_i \le y_i, \sum_{i=1}^n y_i = 1 \right\}. \tag{6}$$

In this case the higher-dimensional polytope Q lives in $\mathbb{R}^{2n}$ and consists of vectors (x, y) that satisfy the constraints $-y_i \le x_i \le y_i$ and $\sum_{i=1}^n y_i = 1$. The projection map is $\pi : (x, y) \mapsto x$. The key point to note here is that the number of inequalities in the description (6) is now $2n$, instead of the 2^n we had before. Observe that the description (6) is the one that we often use in practice when solving ℓ_1 optimization problems. Later we will give other examples of polytopes where lifting allows us to get a description that is much smaller than the trivial one.

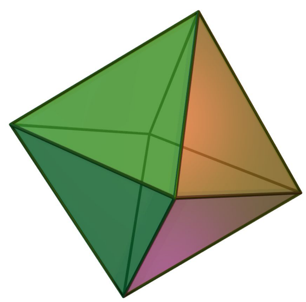

Figure 2: The ℓ_1 ball in $\mathbb{R}^3$ has 8 facets – it is an octahedron. The ℓ_1 ball in $\mathbb{R}^n$ has 2^n facets. Figure from Wikipedia. Credits: "*Original uploader was Cyp at English Wikipedia. Later versions were uploaded by Fropuff at en.wikipedia. - Transferred from en.wikipedia to Commons., CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=38706*"

We have discussed lifts whereby the polytope P is expressed as the projection of another polytope Q. What about if the convex set to start with is not a polytope? There is a more general class of lifts known as *semidefinite programming* lifts. Here the goal is to express a convex set C as the projection of the feasible set of a *semidefinite program* (SDP), i.e., a convex set Q that can be described using linear matrix inequalities:

$$Q = \{ y \in \mathbb{R}^m : F_0 + y_1 F_1 + \cdots + y_m F_m \succeq 0 \}. \tag{7}$$

$F_0, \ldots, F_m$ are $d \times d$ real symmetric matrices and the constraint in (7) indicates that the matrix $F_0 + y_1 F_1 + \cdots + y_m F_m$ is positive semidefinite. Such a convex set is called a *spectrahedron*. When the matrices $F_0, \ldots, F_m$ are diagonal the set Q is a polytope; in general however Q is not necessarily a polytope. It is clear from this observation that SDP lifts form a broader class than LP lifts, even when C is a polytope (see e.g., Figure 7 for a SDP lift of a polytope). One of the main results we will prove is a connection between semidefinite lifts and sums of squares. More precisely we will prove in Theorem 5 that a convex set has a small semidefinite lift if and only if all its valid linear inequalities admit sum-of-squares certificates from a *low-dimensional* subspace.

History The idea of lifting (also known as *extended formulations*) is well known in optimization. The first paper that studies lifts in a systematic way, proving nonexistence of small lifts for certain polytopes, is due to Yannakakis in 1991 [Yan91]. In his paper Yannakakis showed that the traveling salesman polytope and the matching polytope do not have polynomial-size symmetric linear programming lifts.

The recent years have witnessed a resurgence of interest in this topic. From the lower bounds point of view several results have been proved concerning the nonexistence of small lifts for polytopes arising in combinatorial optimization. Fiorini et al. [FMP+12] resolved a conjecture left open by Yannakakis and proved that the traveling salesman polytope does not admit any polynomial-size linear programming lifts (without any symmetry requirement). Later, Lee, Raghavendra, Steurer [LRS15] showed that the traveling salesman polytope has no polynomial-size semidefinite programming lift. Another major result was also obtained recently by Rothvoß [Rot14] where he showed that the matching polytope has no polynomial-size linear programming lift (again, with no symmetry requirement). This result of Rothvoß is particularly striking since the matching polytope is known to have a polynomial-time separation oracle. Several results have also been obtained concerning approximate lifts, see for example [BFPS12, CLRS13, BM13].

From the upper bounds point of view new methods have been proposed to construct improved lifts for certain classes of polytopes and convex sets, see for example [KP11, Goe15, KPT12, FSP16, SPW14]. The problem of constructing semidefinite programming lifts for algebraic sets has been of specific importance in the area of *convex algebraic geometry* [BPT13, GT12, Las09]. In [HN09] Helton and Nie conjectured that any convex semialgebraic set admits a semidefinite programming lift, however this was recently disproved in [Sch18].

1.2 Linear programming lifts and Yannakakis theorem

The main goal of this section is to give a concrete way to think about lifts of a polytope in terms of nonnegative factorizations of a certain matrix associated to P. This matrix, known as the *slack matrix* encodes information about the extreme points and the facet inequalities of P. We first need to recall some basic and important facts about polytopes, and linear programming.

Polytopes A polytope is the convex hull of a finite number of points. Any polytope can also be described using a finite number of linear inequalities

$$P = \{x \in \mathbb{R}^N : b + Ax \geq 0\}. \tag{8}$$

Here $A \in \mathbb{R}^{d\times N}$, $b \in \mathbb{R}^d$ and the inequality $b + Ax \geq 0$ is understood componentwise. The number of inequalities d is the number of *facets* of P, see Figure 3. Linear programming is the problem of

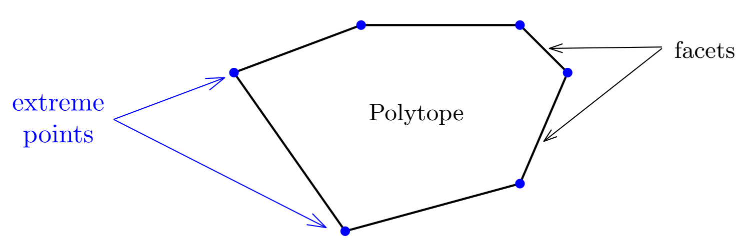

Figure 3: A polytope is the convex hull of a finite number of points, called *extreme points* or *vertices*. Any polytope has also a dual description in terms of a finite number of linear inequalities.

minimizing (or maximizing) a linear function subject to linear inequalities. The standard form of linear programming is

$$\min_{z\in\mathbb{R}^N} \quad c^T z \quad \text{subject to} \quad Az = b,\ z \geq 0. \tag{9}$$

Again, the inequality $z \geq 0$ is understood componentwise. It is not difficult to verify (exercise!) that minimizing a linear function over a polytope can be modeled as a linear program (9) where N (the dimension of the variable z in (9)) is equal to the number of facets of a polytope. A natural question is to ask whether one can do better. This motivates the following definition:

Definition 1. *Let P be a polytope in $\mathbb{R}^n$. We say that P has a* LP lift of size d *if P can be written as $P = \pi(\mathbb{R}^d_+ \cap L)$ where $\pi : \mathbb{R}^d \to \mathbb{R}^n$ is a linear map, $\mathbb{R}^d_+ = \{x \in \mathbb{R}^d : x \geq 0\}$ and L is an affine subspace of $\mathbb{R}^d$. The size of the smallest LP lift of P is called the* LP extension complexity *of P and denoted $\mathrm{xc}_{\mathrm{LP}}(P)$.*

If a polytope $P \subset \mathbb{R}^n$ has an LP lift of size d then any linear optimization problem over P can be formulated as a linear program in standard form (9) of size d. Indeed, assume $P = \pi(\mathbb{R}^d_+ \cap L)$. For any $c \in \mathbb{R}^n$ we have

$$\min \left\{ c^T x : x \in P \right\} \; = \; \min_{z \in \mathbb{R}^d} \left\{ \tilde{c}^T z : z \in L, z \geq 0 \right\} \tag{10}$$

where $\tilde{c} = \pi^*(c)$ and π^* is the adjoint of π. The problem on the right-hand side of (10) is precisely of the form (9) when A, b are chosen to describe the affine subspace L.

One can equivalently define an LP lift as a representation $P = \pi(Q)$ where Q is a polytope with d facets. We leave it as an exercise to the reader to verify this. Figure 4 shows an LP lift of size 5 for the regular hexagon in the plane.

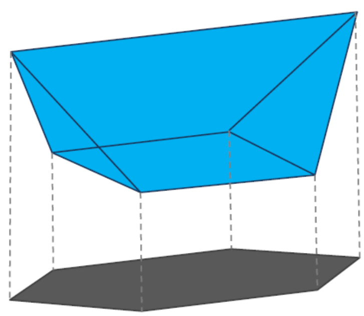

Figure 4: LP Lift of a hexagon of size 5. Note that hexagon has 6 facets whereas the higher-dimensional polytope has 5 facets in $\mathbb{R}^3$.

Examples of LP lifts We now give some examples of polytopes P that admit nontrivial lifts.

- We saw earlier the example of the ℓ_1 ball in $\mathbb{R}^n$ which has 2^n facets and which admits a simple lift of size $2n$.
- Another example of nontrivial lift is for the *permutahedron*. The permutahedron $P \subset \mathbb{R}^n$ is defined as the convex hull of all possible permutations of the vector $(1, 2, \ldots, n)$, i.e.:

  $$P = \mathrm{conv}\left\{(\sigma(1), \ldots, \sigma(n)) : \sigma \in \mathfrak{S}_n\right\}$$

  where $\mathfrak{S}_n$ is the set of permutations on $\{1, \ldots, n\}$. This polytope arises naturally in ordering problems such as in gene sequencing, see e.g. [LW14]. It is known that the permutahedron has an exponential number of facets, precisely $2^n - 2$. However it is not very difficult to construct a lift of the permutahedron of size n^2. Indeed let Q denote the convex hull of *permutation* *matrices* in $\mathbb{R}^{n\times n}$. The Birkhoff-von Neumann theorem asserts that Q is precisely the set of doubly stochastic matrices, i.e.:

$$Q = \Big\{ M \in \mathbb{R}^{n\times n} \ : M_{ij} \geq 0 \ \ \forall i,j = 1,\dots,n, \\ \sum_{i=1}^{n} M_{ij} = 1, \ \forall j = 1,\dots,n, \\ \sum_{j=1}^{n} M_{ij} = 1, \ \forall i = 1,\dots,n \Big\}.$$

It is easy to see that P is a projection of Q: indeed if we let $\pi : \mathbb{R}^{n\times n} \to \mathbb{R}^n$ defined by $\pi(M) = Mu$ where $u = (1,2,\dots,n)^T$ then we get that $\pi(Q) = P$. This lift has size n^2 because Q requires exactly n^2 inequalities for its description. It turns out however that this lift is not optimal. Goemans showed in [Goe15] that the permutahedron admits a lift of size $O(n \log n)$ which is optimal (i.e., there is no smaller possible lift). His construction however is more complicated and makes use of sorting networks.

- Let $P \subset \mathbb{R}^2$ be the regular N-gon in the plane, i.e., the convex hull of the N complex roots of unity. Even though P has N facets (and N vertices), a result of Ben-Tal and Nemirovski [BTN01] shows that the regular N-gon admits a lifted description with only $O(\log N)$ inequalities. This construction was used in [BTN01] to obtain polyhedral approximations of the second-order cone. Figure 4 shows a lift of the regular hexagon of size 5.

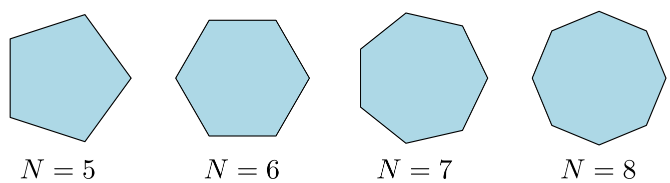

Figure 5: Ben-Tal and Nemirovski showed in [BTN01] that the regular N-gon admits a LP lift of size $O(\log N)$.

- For other examples of lifts from the combinatorial optimization literature we refer the reader to the surveys [CCZ10, Kai11].

Yannakakis' theorem We mentioned earlier that any polytope $P \subset \mathbb{R}^n$ is described by a finite set of linear inequalities $\ell(x) \leq \ell_{\max}$ where ℓ is a linear function and $\ell_{\max} := \max_{x\in P} \ell(x)$. The set of *facet inequalities* is the minimal set of inequalities needed to completely describe P. For example Figure 6 shows a facet inequality of the regular hexagon: in this case $\ell(x) = x + \frac{1}{\sqrt{3}} y$ and $\ell_{\max} = 1$. We now introduce the crucial concept of *slack matrix* needed for the statement of Yannakakis theorem:

Definition 2 (Slack matrix)**.** *Let P be a polytope in $\mathbb{R}^n$. The* slack matrix *of P is a matrix where rows are indexed by extreme points of P and columns are indexed by facets of P, and is defined as follows: The value at the (x,ℓ) entry of P (where x is an extreme point and $\ell \leq \ell_{\max}$ is a facet inequality of P) is given by:*

$$S_{x,\ell} = \ell_{\max} - \ell(x).$$

The slack matrix of P is not, strictly speaking, uniquely defined since it depends on the ordering of the facets/extreme points and the scaling of the facet inequalities. The results stated here however will be independent of the choice of slack matrix and that is why we will often talk about "the" slack matrix of P.

Example 1 (Slack matrix of the regular hexagon)**.** *To illustrate the definition of the slack matrix consider the case of the regular hexagon in* $\mathbb{R}^2$*. This polytope has* $|X| = 6$ *vertices, as well as 6 facets so its slack matrix is* 6×6*. Figure 6 (left) gives the slack matrix.*

$$\left[\begin{array}{c|c|c|c|c|c} 0 & 1 & 2 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 2 & 1 \\ 1 & 0 & 0 & 1 & 2 & 2 \\ 2 & 1 & 0 & 0 & 1 & 2 \\ 2 & 2 & 1 & 0 & 0 & 1 \\ 1 & 2 & 2 & 1 & 0 & 0 \end{array}\right] \tag{11}$$

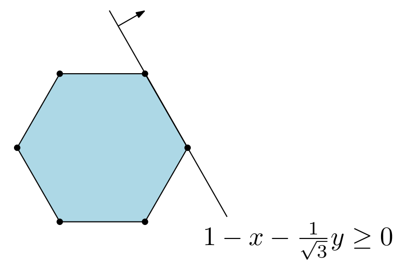

Figure 6: Slack matrix of the regular hexagon

For example to see how the first column is formed consider the facet inequality of the hexagon given by $1-x-y/\sqrt{3} \geq 0$*. The k'th vertex of the hexagon has coordinates* $(\cos(2(k-1)\pi/6), \sin(2(k-1)\pi/6))$*. The k'th component of the first column is thus given by* $1 - \cos(2(k-1)\pi/6) - \sin(2(k-1)\pi/6)/\sqrt{3}$*. Note that each column has exactly two zeros: this is because each facet of the regular hexagon passes through exactly two vertices.*

The entries of a slack matrix are all nonnegative, by definition. We now introduce a second important definition needed for the statement of Yannakakis theorem.

Definition 3 (Nonnegative matrix factorization / nonnegative rank)**.** *Let* $S \in \mathbb{R}_+^{p\times q}$ *be a matrix with nonnegative entries. We say that* S *has a* nonnegative factorization *of size* r *if we can write* $S = AB$ *where* $A \in \mathbb{R}_+^{p\times r}$ *and* $B \in \mathbb{R}_+^{r\times q}$ *have nonnegative entries. The smallest* r *for which such a factorization exists is called the* nonnegative rank *of* S *and denoted* $\operatorname{rank}_+(S)$*.*

One can show that if S is a nonnegative matrix then $\operatorname{rank}(S) \leq \operatorname{rank}_+(S)$ and that the inequality can be strict. Exercise: check that $\operatorname{rank}(S) < \operatorname{rank}_+(S)$ for the 4×4 matrix

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

Yannakakis' theorem We are now ready to state Yannakakis' theorem which establishes a connection between LP lifts of a polytope, and nonnegative factorizations of its slack matrix.

Theorem 1 (Yannakakis [Yan91])**.** *Let* P *be a full-dimensional polytope and let* S *be its slack matrix. Then* P *has a LP lift of size* d *if, and only if,* S *has a nonnegative factorization of size* d*. As a consequence, the smallest size of a LP lift of* P *is equal to* $\operatorname{rank}_+(S)$*.*

Before presenting the proof of the theorem we discuss an example and an interpretation of Yannakakis' theorem that will be useful later.

Example 2 (Regular hexagon continued)**.** *One can show that the slack matrix of the regular hexagon has the following nonnegative factorization of size 5:*

$$\underbrace{\left[\begin{array}{c|c|c|c|c|c} 0 & 1 & 2 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 2 & 1 \\ 1 & 0 & 0 & 1 & 2 & 2 \\ 2 & 1 & 0 & 0 & 1 & 2 \\ 2 & 2 & 1 & 0 & 0 & 1 \\ 1 & 2 & 2 & 1 & 0 & 0 \end{array}\right]}_{\textit{slack matrix of hexagon}} = \left[\begin{array}{c|c|c|c|c} 0 & 1 & 0 & 0 & 1 \\ 0 & 2 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{array}\right] \left[\begin{array}{c|c|c|c|c|c} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 1 & 0 & 0 \end{array}\right]. \tag{12}$$

Note that the matrices on the right-hand side are elementwise nonnegative. By Yannakakis' theorem this factorization says that the regular hexagon has a LP lift of size 5, see Figure 4. One can actually show that the regular hexagon has no smaller LP lift.

Interpretation of Yannakakis' theorem in terms of certificates of nonnegativity Let P be a polytope and let X be the set of extreme points of the polytope P. If $\ell \leq \ell_{\max}$ is a facet inequality for P then this implies that $\ell_{\max} - \ell|_X$ (i.e., the restriction of $\ell_{\max} - \ell$ to X) is a nonnegative function on X. Yannakakis' theorem can be restated as follows:

Theorem 2 (Yannakakis [Yan91]; restatement of Theorem 1)**.** *Let $P = \operatorname{conv}(X) \subset \mathbb{R}^n$ be a full-dimensional polytope. Then P has an LP lift of size d if and only if, there exist d nonnegative functions on X, $a_1, \ldots, a_d : X \to \mathbb{R}_+$ such that the following holds: for any facet inequality $\ell(x) \leq \ell_{\max}$ of P there exist nonnegative coefficients $b_1, \ldots, b_d \geq 0$ such that*

$$\ell_{\max} - \ell|_X = \sum_{i=1}^{d} b_i a_i. \tag{13}$$

To make the connection with Theorem 1, the functions $a_1, \ldots, a_D : X \to \mathbb{R}_+$ are nothing but the columns of A in a nonnegative factorization of the slack matrix $S = AB$, and the coefficients $b_1, \ldots, b_d$ are the entries in the ℓ'th column of B.

Equation (13) is an equality of functions on X. It should be interpreted as a certificate of nonnegativity of $\ell_{\max} - \ell$ on X: indeed the right hand side is a linear combination with nonnegative weights (the b_i's) of nonnegative functions $a_1, \ldots, a_d$, and so is "obviously" nonnegative on X. The size of the certificate, here the number of functions $a_1, \ldots, a_d$, gives us the size of the LP lift. The main question in constructing a lift is of course to come up with the nonnegative functions $a_1, \ldots, a_d$. This point of view will be useful later when we consider semidefinite programming lifts since it will allow us to establish the connection with sums of squares.

Proof of Yannakakis' theorem We now present a proof of Yannakakis' theorem.

Proof of Theorem 1. We first prove necessity. Assume $P = \pi(\mathbb{R}^d_+ \cap L)$ is an LP lift of P of size d, where L is an affine subspace of $\mathbb{R}^d$ and π a linear map. Let $\ell_{\max} \leq \ell$ be any facet inequality of P. Since $\pi(\mathbb{R}^d_+ \cap L) \subseteq P$ it is clear that the following implication holds for any y:

$$\begin{cases} y \geq 0 \\ y \in L \end{cases} \Rightarrow \ell \circ \pi(y) \leq \ell_{\max}.$$

One can show using Farkas' lemma/strong duality for LP, that there exist coefficients $b_1, \dots, b_d \geq 0$ and an affine form γ that vanishes on L such that

$$\ell_{\max} - \ell \circ \pi(y) = \sum_{i=1}^{d} b_i y_i + \gamma(y). \tag{14}$$

Note that Equation (14) automatically implies that $\ell_{\max} - \ell \circ \pi$ is nonnegative on $\mathbb{R}^d_+ \cap L$. To see why (14) is true let L_0 be the linear subspace of $\mathbb{R}^d$ parallel to L and let $y_0 \in \mathbb{R}^d$ such that $L = y_0 + L_0$. Consider the following primal/dual pair of LPs whose value is $\ell_{\max}$:

$$\begin{array}{lll} \max\limits_{y \in \mathbb{R}^d} & (\ell \circ \pi)(y) \\ \text{s.t.} & y \in \mathbb{R}^d_+ \\ & y - y_0 \in L_0 \end{array} \qquad \begin{array}{lll} \min\limits_{b,h \in \mathbb{R}^d} & -\langle h, y_0 \rangle \\ \text{s.t.} & -\ell \circ \pi = b + h \\ & b \in \mathbb{R}^d_+, \ h \in L_0^\perp \end{array} \tag{15}$$

By strong duality, there exists $b \in \mathbb{R}^d_+$ and $h \in L_0^\perp$ such that $-\ell \circ \pi = b + h$ and $-\langle h, y_0 \rangle = \ell_{\max}$. Thus this means that

$$\ell_{\max} - \ell \circ \pi = b + h - \langle h, y_0 \rangle$$

which is exactly (14) with $\gamma(y) = \langle h, y - y_0 \rangle$. Now for any $x \in X$ we know that there exists $A(x) = (a_1(x), \dots, a_d(x)) \in \mathbb{R}^d_+ \cap L$ such that $\pi(A(x)) = x$. By evaluating Equation (14) at $y = A(x)$ we get (using the fact that $A(x) \in L$ and so $\gamma(A(x)) = 0$):

$$\ell_{\max} - \ell(x) = \sum_{i=1}^{d} b_i a_i(x) \quad \forall x \in X.$$

Thus this proves the claim.

We now show sufficiency. Assume $P = \{x \in \mathbb{R}^n : Fx \leq g\}$ is a facet description of our polytope P where $F \in \mathbb{R}^{N \times n}$ and $g \in \mathbb{R}^N$. Assume $S = AB$ is a nonnegative factorization of the slack matrix of size d, where $A \in \mathbb{R}^{|X| \times d}_+$ and $B \in \mathbb{R}^{d \times N}_+$. Then it is easy to verify that P can be written as:

$$P = \{x \in \mathbb{R}^n : \exists a \in \mathbb{R}^d \text{ s.t. } a \geq 0, \ g - Fx = B^T a\}. \tag{16}$$

To see why the inclusion "$\subseteq$" holds, note that if $x \in X$ is a vertex of P then by letting a be the row of A indexed by x the constraints on the right-hand side are satisfied. The inclusion "$\supseteq$" is trivial since $B^T a \geq 0$ for $a \geq 0$. The proof is almost complete since the right-hand side of (16) is defined using only d linear inequalities. To be sure we just need to show that (16) can be put in the form $P = \pi(\mathbb{R}^d_+ \cap L)$ for some linear map π and affine subspace $L \subset \mathbb{R}^d$. Since P is bounded and $\dim(P) > 0$, we know that $\operatorname{rank}(F) = n$ and $g \notin \operatorname{Im}(F)$. Since $\operatorname{rank}(F) = n$ the equation $Fx = g - B^T a$ (in x) has a unique solution x_a if $g - B^T a \in \operatorname{Im}(F)$ and no solution otherwise. It is easy to see that the map that sends a to x_a, defined on the affine subspace $L = \{a \in \mathbb{R}^d : g - B^T a \in \operatorname{Im}(F)\}$, is affine. Since 0 does not belong to L (this is because $g \notin \operatorname{Im}(F)$) this affine map can be extended to a linear map π on the whole space. We thus finally get that $P = \pi(\mathbb{R}^d_+ \cap L)$. This shows that P admits a LP lift of size d. □

1.3 Semidefinite programming lifts, positive semidefinite factorizations, and sums of squares

In this section we extend the concepts and results to *semidefinite programming* lifts.

Semidefinite programming A *semidefinite program* (SDP) in standard form is an optimization problem of the form:

$$\min_{Y \in \mathbf{S}^d} \operatorname{Tr}(CY) \quad \text{subject to} \quad \mathcal{A}(Y) = b, \;\; Y \in \mathbf{S}_+^d \tag{17}$$

where $\mathbf{S}^d$ is the space of $d \times d$ real symmetric matrices, and $\mathbf{S}_+^d$ is the cone of positive semidefinite matrices. A semidefinite program is completely specified by the triple $(C, \mathcal{A}, b)$ where the matrix $C \in \mathbf{S}^d$ specifies the objective function, $\mathcal{A} : \mathbf{S}^d \to \mathbb{R}^m$ is a linear map and $b \in \mathbb{R}^m$. The feasible set of a semidefinite program is called a *spectrahedron* and is an intersection of the positive semidefinite cone $\mathbf{S}_+^d$ with an affine subspace $\{Y : \mathcal{A}(Y) = b\}$.

A natural question is to understand which optimization problems can be put in semidefinite programming form. This motivates the following definition of SDP lift:

Definition 4. *Let C be a convex set. We say that C has a SDP lift of size d if it can be written as $C = \pi(\mathbf{S}_+^d \cap L)$ where $\mathbf{S}_+^d$ is the cone of $d \times d$ real symmetric positive semidefinite matrices, and L is an affine subspace of $\mathbf{S}^d$. The size of the smallest SDP lift of C is called the* SDP extension complexity *of C and denoted $\operatorname{xc}_{\mathrm{SDP}}(C)$.*

If C has an SDP lift of size d, then any linear optimization problem over C can be modeled as an SDP in standard form of size d. Indeed if $C = \pi(\mathbf{S}_+^d \cap L)$ for some affine subspace L of $\mathbf{S}^d$ and linear map π, then we have:

$$\min \left\{ c^T x : x \in C \right\} \;=\; \min_{Y \in \mathbf{S}^d} \left\{ \operatorname{Tr}(\tilde{C} Y) : Y \in L, Y \in \mathbf{S}_+^d \right\}$$

where $\tilde{C} = \pi^*(c)$ with π^* being the adjoint of π. The right-hand side is an SDP in standard form (17) with an appropriate choice of $\mathcal{A}, b$.

Example 3 (Disk)**.** *Let $D = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \le 1\}$. One can check that D has an SDP lift of size 2, namely $D = \pi(\mathbf{S}_+^2 \cap L)$ where $L = \left\{Y \in \mathbf{S}^2 : \operatorname{Tr}(Y) = 2\right\}$ and $\pi\left(\left[\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right]\right) = ((c-a)/2, b)$. We leave it as an exercise to the reader to verify this.*

Example 4. *Figure 7 illustrates a SDP lift of the square $[-1,1]^2$ of size 3, given by:*

$$[-1,1]^2 = \left\{ (x_1, x_2) \in \mathbb{R}^2 : \exists u \in \mathbb{R} \begin{bmatrix} 1 & x_1 & x_2 \\ x_1 & 1 & u \\ x_2 & u & 1 \end{bmatrix} \succeq 0 \right\}. \tag{18}$$

Using the notations of Definition 4, $L = \{Y \in \mathbf{S}^3 : Y_{11} = Y_{22} = Y_{33} = 1\}$ and $\pi(Y) = (Y_{11}, Y_{12})$. To see why (18) is true note that if $(x_1, x_2) \in \{-1, 1\}^2$ then by letting $u = x_1 x_2$ we have:

$$\begin{bmatrix} 1 & x_1 & x_2 \\ x_1 & 1 & u \\ x_2 & u & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}^T \succeq 0.$$

Since the right-hand side of (18) is convex this shows that $\operatorname{conv}(\{-1,1\}^2) = [-1,1]^2$ is contained in it. Conversely if (x_1, x_2) belongs to the right-hand side of (18) then by looking at the 2×2 minors of the 3×3 positive semidefinite matrix we easily get that $1 - x_1^2 \ge 0$ and $1 - x_2^2 \ge 0$ i.e., $(x_1, x_2) \in [-1,1]^2$.

Our goal now is to state a result similar to Yannakakis theorem to characterize *SDP lifts*. Before we do so, we need to recall some basic facts about general convex sets.

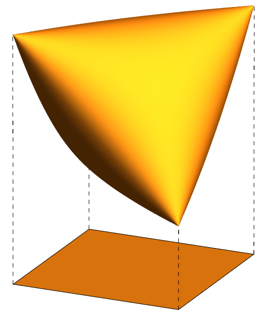

Figure 7: SDP lift of the square $[-1,1]^2$ of size 3 (cf. Equation (18)). The three-dimensional convex set shown in the figure is the set of (x_1, x_2, u) such that the 3×3 symmetric matrix on the right-hand side of (18) is positive semidefinite. Projecting this set onto (x_1, x_2) yields the square $[-1,1]^2$.

Polars of convex sets and biduality Let $C \subset \mathbb{R}^n$ be a closed convex set and assume (without loss of generality) that C has non-empty interior and that the origin is in its interior. Any such convex set will be called a *convex body*. The following are two important definitions related to convex sets:

Definition 5 (Extreme point)**.** *Let C be a convex set. A point $x \in C$ is called an* extreme point *if whenever $x = (a+b)/2$ with $a, b \in C$ then $a = b = x$. The set of extreme points of C will be denoted* extreme(C)*.*

The *Minkowski theorem* (also called the Krein-Milman theorem) says that if C is a closed and bounded convex set in $\mathbb{R}^n$ then C is the convex hull of its extreme points.

Definition 6 (Polar)**.** *The* polar *of a convex body C is defined as:*

$$C^\circ = \{y \in \mathbb{R}^n : \langle y, x\rangle \le 1 \; \forall x \in C\}.$$

In words, C° is the set of linear forms $\langle y, \cdot\rangle$ whose maximum on C is at most 1. The *biduality theorem* tells us that these linear forms give a complete inequality description of C, namely:

$$C = \{x \in \mathbb{R}^n : \langle y, x\rangle \le 1 \; \forall y \in C^\circ\} = (C^\circ)^\circ. \tag{19}$$

One should compare the equation above with the inequality description of a polytope $P = \{x : Ax \le b\}$. By convexity it is enough in Equation (19) to consider $y \in \text{extreme}(C^\circ)$, i.e., we can write $C = \{x \in \mathbb{R}^n : \langle y, x\rangle \le 1 \; \forall y \in \text{extreme}(C^\circ)\}$. If C is a polytope then C° has finitely many extreme points which are in one-to-one correspondence with the facets of C.

Slack matrix of convex sets In the previous section we defined the slack matrix of a polytope. We are now going to generalize this definition to convex bodies.

Definition 7 (Slack matrix of a convex body)**.** *Let $C \subset \mathbb{R}^n$ be a convex body. The slack matrix of C is a (potentially infinite) matrix S whose rows are indexed by the extreme points of C, columns are indexed by the extreme points of C° and is defined as follows:*

$$S_{x,y} = 1 - \langle y, x\rangle \quad \forall x \in \text{extreme}(C), y \in \text{extreme}(C^\circ).$$

Note that if C is a polytope then we recover the same definition from the previous section. This is because when C is a polytope the facet inequalities are precisely of the form $\langle y, x\rangle \leq 1$ where $y \in \operatorname{extreme}(C^\circ)$.

Positive semidefinite factorization We now introduce the notion of positive semidefinite factorization of a nonnegative matrix introduced in [GPT13].

Definition 8 (Positive semidefinite factorization [GPT13, FGP+15])**.** *Let $S \in \mathbb{R}_+^{p\times q}$ be a matrix with nonnegative entries. We say that S has a* positive semidefinite factorization (psd factorization) *of size d if there exist positive semidefinite matrices $A_1, \dots, A_p \in \mathbf{S}_+^d$ and $B_1, \dots, B_q \in \mathbf{S}_+^d$ such that $S_{ij} = \langle A_i, B_j\rangle$ for all $i = 1, \dots, p$ and $j = 1, \dots, q$. The size of the smallest psd factorization of S is called the* psd rank *of S and denoted* $\operatorname{rank}_{\mathrm{psd}}(S)$.

We are now ready to state a result similar to that of Yannakakis to characterize the existence of SDP lifts. It is due to Gouveia, Parrilo, Thomas [GPT13].

Theorem 3 (Gouveia, Parrilo, Thomas [GPT13])**.** *Let C be a convex body and let S be its slack matrix. Then C has a SDP lift of size d if, and only if, S has a positive semidefinite factorization of size d. As a consequence, the smallest size of a SDP lift of P is equal to* $\operatorname{rank}_{\mathrm{psd}}(S)$.

Example 5 (Disk)**.** *Consider the example of the disk $D = \{(x_1, x_2) : x_1^2 + x_2^2 \leq 1\}$. The polar of D is D itself, $D^\circ = D$. One can verify that the slack matrix of D is the infinite matrix $S_{x,y} = 1 - \langle x, y\rangle$ where rows and columns are indexed by $x, y \in \mathbb{R}^2$ with $\|x\|_2 = \|y\|_2 = 1$. We can give a positive semidefinite factorization of size 2 of S as follows: $S_{x,y} = \langle A(x), B(y)\rangle$ where*

$$A(x) = \begin{bmatrix} 1 - x_1 & x_2 \\ x_2 & 1 + x_1 \end{bmatrix} \quad B(y) = \frac{1}{2}\begin{bmatrix} 1 + y_1 & -y_2 \\ -y_2 & 1 - y_1 \end{bmatrix}.$$

Since $\|x\|_2 = \|y\|_2 = 1$ we have $A(x), B(y) \succeq 0$ and

$$\langle A(x), B(y)\rangle = \frac{1}{2}\left[(1 - x_1)(1 + y_1) - 2x_2y_2 + (1 + x_1)(1 - y_1)\right] = 1 - \langle x, y\rangle = S_{x,y}$$

as desired. Theorem 3 tells us that we must have a SDP lift of size 2, which we have already seen in Example 3.

Interpretation of Theorem 3 in terms of certificates of nonnegativity As for the case of LP lifts, we can give an interpretation of Theorem 3 in terms of certificates of nonnegativity for valid linear inequalities.

Theorem 4 (Gouveia, Parrilo, Thomas, [GPT13]; restatement of Theorem 3)**.** *Let C be a convex body and let X be its set of extreme points. Then C has a SDP lift of size d if, and only if, there exists a map $A : X \to \mathbf{S}_+^d$ such that the following holds: for any valid linear inequality $\ell \leq \ell_{\max}$ of C there exists $B \in \mathbf{S}_+^d$ such that*

$$\ell_{\max} - \ell(x) = \langle A(x), B\rangle \quad \forall x \in X. \tag{20}$$

Note that Equation (20) is an equality of functions on X. Just like in the LP case, it should be understood as a certificate of nonnegativity for $\ell_{\max} - \ell|_X$. Indeed the function $x \mapsto \langle A(x), B\rangle$ is "obviously" nonnegative on X since $A(x) \in \mathbf{S}_+^d$ and $B \in \mathbf{S}_+^d$.

Example 6. *In Example 4 (see also Figure 7) we saw that the square* $[-1,1]^2$ *has an SDP lift of size 3. Let us exhibit the function* $A : X \to \mathbf{S}^3_+$ *in this case corresponding to this lift. In this example the vertex set is* $X = \{-1,1\}^2$*. Consider the map* A *given by:*

$$A(x) = \begin{bmatrix} 1 & x_1 & x_2 \\ x_1 & 1 & x_1x_2 \\ x_2 & x_1x_2 & 1 \end{bmatrix}.$$

Note that $A(x) \in \mathbf{S}^3_+$ *for* $x \in X$ *since we have, for any* $x \in X$ *(using the fact that* $x_1^2 = x_2^2 = 1$*):*

$$A(x) = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}^T \succeq 0.$$

To show that the condition of Theorem 4 is satisfied consider the facet inequality $1 - x_1 \geq 0$*. Define*

$$B = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}^T$$

and note that $B \succeq 0$*. Then we have for any* $x \in \{-1,1\}^2$

$$\langle A(x), B \rangle = \left\langle \begin{bmatrix} 1 & x_1 & x_2 \\ x_1 & 1 & x_1x_2 \\ x_2 & x_1x_2 & 1 \end{bmatrix}, \frac{1}{2} \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right\rangle = \frac{1}{2}(1 - 2x_1 + 1) = 1 - x_1.$$

Similarly one can show that the three other facet inequalities $1 + x_1 \geq 0, 1 - x_2 \geq 0$ *and* $1 + x_2 \geq 0$ *can be written as* $\langle A(x), B \rangle$ *for a suitable choice of* B*. This shows that the map* A *satisfies the condition of Theorem 4.*

Sums of squares In this section we show that SDP lifts can also be interpreted in terms of *sum of squares* certificates of the facet inequalities $\ell_{\max} - \ell|_X$. Such a certificate consists in expressing $\ell_{\max} - \ell|_X$ as a sum of squares of functions on X. More formally we have:

Theorem 5. *Let* C *be a convex body and let* X *be its extreme points. Assume there is a subspace* V *of* $\mathbb{R}^X$ *(the space of functions* $X \to \mathbb{R}$*) such that the following holds:*

(*) for any valid linear inequality $\ell \leq \ell_{\max}$ *of* C *there are elements* $h_1, \ldots, h_J \in V$ *such that*

$$\ell_{\max} - \ell|_X = \sum_{j=1}^{J} h_j^2. \tag{21}$$

Then C *has a SDP lift of size* $\dim V$*.*

Conversely if C *has a SDP lift of size* d*, then there is a subspace* V *of* $\mathbb{R}^X$ *of dimension at most* d^2 *such that condition (*) holds.*

Proof. We show how to prove Theorem 5 from Theorem 4. We start by proving the first part. Assume we have a subspace V of dimension d such that condition (*) holds. Let $f_1, \ldots, f_d$ be a basis of this subspace and define the map $A : X \to \mathbf{S}^d_+$ as follows:

$$A(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_d(x) \end{bmatrix} \begin{bmatrix} f_1(x) \\ \vdots \\ f_d(x) \end{bmatrix}^T = [f_i(x) f_j(x)]_{1 \leq i,j \leq d}.$$

Clearly $A(x) \succeq 0$ (note also that $A(x)$ is rank-one for any $x \in X$). Now let $\ell \leq \ell_{\max}$ be a valid inequality for C. By (*) we know that there exists elements $h_1, \ldots, h_J \in V$ such that (21) holds. We will show that there exists $B \in \mathbf{S}^d_+$ such that $\ell_{\max} - \ell|_X = \langle A(\cdot), B\rangle$. Since each h_j is in V we can write $h_j(x) = b_j^T f(x)$ where $b_j \in \mathbb{R}^d$ and $f(x)$ is the column vector $[f_1(x); \ldots; f_d(x)]$. Now note that $(h_j(x))^2 = (b_j^T f(x))^2 = \langle f(x)f(x)^T, b_j b_j^T\rangle = \langle A(x), b_j b_j^T\rangle$. Thus if we let $B = \sum_{j=1}^J b_j b_j^T$ we get

$$\ell_{\max} - \ell(x) = \langle A(x), B\rangle \quad \forall x \in X$$

which is what we wanted. Thus using Theorem 4, this shows that P has an SDP lift of size d.

We now prove the other direction. Assume that C has a SDP lift of size d. According to Theorem 4 this means that there exists $A : X \to \mathbf{S}^d_+$ such that the following holds: for any valid inequality $\ell \leq \ell_{\max}$ there is $B \in \mathbf{S}^d_+$ such that $\ell_{\max} - \ell(x) = \langle A(x), B\rangle$, for all $x \in X$. Since $A(x) \in \mathbf{S}^d_+$ we can factorize it as $A(x) = R(x)R(x)^T$. Let V be the subspace of $\mathbb{R}^X$ spanned by the entries of R, i.e., $V = \operatorname{span}(x \mapsto R_{ij}(x), i, j = 1, \ldots, d) \subset \mathbb{R}^X$ and note that $\dim V \leq d^2$. Now given $\ell \leq \ell_{\max}$ a valid inequality of P we know that there exists $B \in \mathbf{S}^d_+$ such that $\ell_{\max} - \ell(x) = \langle A(x), B\rangle$ for all $x \in X$. We can write B as $B = CC^T$ to get

$$\ell_{\max} - \ell(x) = \langle R(x)R(x)^T, CC^T\rangle = \|C^T R(x)\|_F^2.$$

Since each entry of $x \mapsto C^T R(x)$ is an element of V the previous equation gives a sum-of-squares certificate of $\ell_{\max} - \ell|_X$ using functions from V. This completes the proof. $\square$

Example 7. *The lift of the square* $[-1, 1]^2$ *discussed earlier can be explained in terms of sum-of-squares certificates. Note that the valid inequality* $1 - x_1 \geq 0$ *can be certified using sum-of-squares as follows:*

$$1 - x_1 = \frac{1}{2}(1 - x_1)^2 \quad \forall x \in \{-1, 1\}^2.$$

It is crucial to note that the equality above is understood on $\{-1, 1\}^2$ *(the equality is of course not true globally because the left hand side is a polynomial of degree 1 whereas the right-hand side is a polynomial of degree 2). Since we are working on* $\{-1, 1\}^2$ *the right-hand side expands to* $\frac{1}{2}(1 - 2x_1 + x_1^2) = \frac{1}{2}(1 - 2x_1 + 1) = 1 - x_1$ *where we used the fact that* $x_1^2 = 1$*. Similarly one can show that the other facet inequalities have the following sum-of-squares certificates:*

$$\begin{aligned}
1 - x_2 &= \frac{1}{2}(1 - x_2)^2 \quad \forall x \in \{-1, 1\}^2,\\
1 + x_1 &= \frac{1}{2}(1 + x_1)^2 \quad \forall x \in \{-1, 1\}^2,\\
1 + x_2 &= \frac{1}{2}(1 + x_2)^2 \quad \forall x \in \{-1, 1\}^2.
\end{aligned}$$

Thus if we let V *be the space of polynomials of degree at most 1 on* $\{-1, 1\}^2$ *(i.e., by an abuse of notation* $V = \operatorname{span}(1, x_1, x_2)$*) the condition (*) of Theorem 5 holds true. Thus this shows that* $[-1, 1]^2$ *has a SDP lift of size* $\dim V = 3$*.*

Proof of the factorization theorem for SDP lifts We now give a proof of Theorem 4, which is essentially the same as the proof in [GPT13, Theorem 1].

Proof of Theorem 4. The proof follows the same steps as the proof of Theorem 2 for LP lifts. The only difference is that we use the generalized Farkas' lemma/strong duality for SDPs.

Assume that $C = \pi(\mathbf{S}_+^d \cap L)$ is an SDP lift of C of size d where L is an affine subspace of $\mathbf{S}^d$ and π a linear map. Let $\ell \leq \ell_{\max}$ be a valid linear inequality for C. Since $\pi(\mathbf{S}_+^d \cap L) \subseteq C$ the following implication holds for all $Y \in \mathbf{S}^d$:

$$\begin{cases} Y \succeq 0 \\ Y \in L \end{cases} \Rightarrow \ell_{\max} - \ell \circ \pi(Y) \geq 0.$$

We will now show using Farkas' lemma/strong duality for SDP that there is a positive semidefinite matrix $B \in \mathbf{S}_+^d$, an affine form γ that vanishes on L such that

$$\ell_{\max} - \ell \circ \pi(Y) = \langle B, Y \rangle + \gamma(Y) \quad \forall Y \in \mathbf{S}^d. \tag{22}$$

To see why (22) is true let L_0 be the linear space in $\mathbf{S}^d$ parallel to L and let $Y_0 \in \mathbf{S}^d$ such that $L = Y_0 + L_0$. The following problems are dual to each other and the value of the primal (maximization) problem is equal to $\ell_{\max}$:

$$\begin{array}{lll} \displaystyle\max_{Y \in \mathbf{S}^d} & (\ell \circ \pi)(Y) \\ \text{s.t.} & Y \in \mathbf{S}_+^d \\ & Y - Y_0 \in L_0 \end{array} \qquad\qquad \begin{array}{ll} \displaystyle\min_{B, H \in \mathbf{S}^d} & -\langle H, Y_0 \rangle \\ \text{s.t.} & -\ell \circ \pi = B + H \\ & B \in \mathbf{S}_+^d, \; H \in L_0^\perp \end{array} \tag{23}$$

We can assume that the intersection of L with the interior of $\mathbf{S}_+^d$ is nonempty (otherwise the intersection lies on a strict face of $\mathbf{S}_+^d$ which means that one can reduce the size of the SDP lift). In this case strong duality holds, the optimal values of the two SDPs (23) are equal to $\ell_{\max}$ and the dual (minimization) problem is attained. Let B, H be the optimal points of the dual problem in (23). From dual feasibility we have $-\ell \circ \pi = B + H$ and so since $\ell_{\max} = -\langle H, Y_0 \rangle$ we get that:

$$\ell_{\max} - \ell \circ \pi = B + H - \langle H, Y_0 \rangle.$$

Note that this shows (22) where the affine map γ is $\gamma(Y) = \langle H, Y - Y_0 \rangle$. For $x \in X$ let $A(x)$ be any element in $\mathbf{S}_+^d \cap L$ such that $\pi(A(x)) = x$. Evaluating (22) at $A(x)$, for any $x \in X$ we get:

$$\ell_{\max} - \ell(x) = \langle B, A(x) \rangle$$

where we used the fact that $\pi(A(x)) = x$ and that $\gamma(A(x)) = 0$ since $A(x) \in L$. This proves our claim.

We now prove the converse. We show how to construct an SDP lift of P from a psd factorization of its slack matrix. By bipolarity we know that $C = \{x \in \mathbb{R}^n : \ell(x) \leq 1 \; \forall \ell \in C^\circ\}$ is an inequality description of C. Let S be the slack matrix of C and let $S_{x,\ell} = \langle A(x), B(\ell) \rangle$ be a psd factorization of S of size d, where $A(x), B(\ell) \in \mathbf{S}_+^d$ (here $x \in X$ is an extreme point of C and ℓ is an extreme point of C°). It is easy to verify that we have the following description of C:

$$C = \left\{ x \in \mathbb{R}^n : \exists A \in \mathbf{S}^d, A \succeq 0 \text{ and } 1 - \ell(x) = \langle A, B_\ell \rangle \; \forall \ell \in \operatorname{extreme}(C^\circ) \right\} \tag{24}$$

To see why the inclusion "$\subseteq$" holds let $x \in X$ be an extreme point of C and take $A = A(x)$ in the right-hand side. The reverse inclusion "$\supseteq$" follows immediately by observing that $\langle A, B_\ell \rangle \geq 0$ since A and B_ℓ are positive semidefinite. It remains to show that one can put (24) into the form $C = \pi(\mathbf{S}_+^d \cap L)$ for some linear map π and affine subspace L. This can be done in a way very similar to the proof of Yannakakis theorem for LP lifts (see paragraph following Equation (16)). $\square$

1.4 Summary

We summarize briefly the main results in this section.

- Let $P = \operatorname{conv}(X)$ be a polytope. The slack matrix of P is defined as $S_{x,\ell} = \ell_{\max} - \ell(x)$ where rows are indexed by extreme points $x \in X$ and columns are indexed by facets $\ell \leq \ell_{\max}$ of P. The entries of S are nonnegative.

- Finding an LP lift of P is equivalent to finding a nonnegative factorization of S (Yannakakis' theorem, Theorem 1) of the form $S = AB$ where A and B are elementwise nonnegative matrices. Thus the size of the smallest LP lift of P is equal to the nonnegative rank of S.

- Finding an SDP lift of P is equivalent to finding a *positive semidefinite* factorization of S (Theorem 3) of the form $S_{x,\ell} = \langle A(x), B(\ell) \rangle$ where $A(x), B(\ell)$ are real symmetric $d \times d$ matrices that are positive semidefinite. Thus, the size of the smallest semidefinite programming lift is equal to the positive semidefinite rank of S.

- Sum-of-squares lifts (which are a special case of SDP lifts when the map $A(x)$ is rank-one) consist in finding a subspace V of $\mathbb{R}^X$ such that any facet inequality $\ell_{\max} - \ell|_X \geq 0$ has a certificate:
$$\ell_{\max} - \ell|_X = \sum_{j=1}^{J} h_j^2$$
where $h_1, \ldots, h_J \in V$. Such a subspace yields an SDP lift of P of size $\dim V$.

References

[BFPS12] Gábor Braun, Samuel Fiorini, Sebastian Pokutta, and David Steurer. Approximation limits of linear programs (beyond hierarchies). In *IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 480–489, 2012. 4

[BM13] Mark Braverman and Ankur Moitra. An information complexity approach to extended formulations. In *Proceedings of the 45th Annual ACM symposium on Symposium on Theory of Computing*, pages 161–170. ACM, 2013. 4

[BPT13] Grigoriy Blekherman, Pablo A. Parrilo, and Rekha R. Thomas. *Semidefinite optimization and convex algebraic geometry*. SIAM, 2013. 4

[BTN01] Aharon Ben-Tal and Arkadi Nemirovski. On polyhedral approximations of the second-order cone. *Mathematics of Operations Research*, 26(2):193–205, 2001. 6

[CCZ10] Michele Conforti, Gérard Cornuéjols, and Giacomo Zambelli. Extended formulations in combinatorial optimization. *4OR*, 8(1):1–48, 2010. 6

[CLRS13] Siu On Chan, James R. Lee, Prasad Raghavendra, and David Steurer. Approximate constraint satisfaction requires large LP relaxations. In *IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 350–359, 2013. 4

[Faw16] Hamza Fawzi. *Power and limitations of convex formulations via linear and semidefinite programming lifts*. PhD thesis, Massachusetts Institute of Technology, 2016. 1

[FGP+15] Hamza Fawzi, João Gouveia, Pablo A. Parrilo, Richard Z. Robinson, and Rekha R. Thomas. Positive semidefinite rank. *Mathematical Programming*, 153(1):133–177, 2015. 12

[FMP+12] Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans R. Tiwary, and Ronald de Wolf. Linear vs. semidefinite extended formulations: exponential separation and strong lower bounds. In *Proceedings of the 44th Symposium on Theory of Computing*, pages 95–106. ACM, 2012. 4

[FSP16] Hamza Fawzi, James Saunderson, and Pablo A. Parrilo. Sparse sums of squares on finite abelian groups and improved semidefinite lifts. *Mathematical Programming*, 160(1):149–191, 2016. 4

[Goe15] Michel X. Goemans. Smallest compact formulation for the permutahedron. *Mathematical Programming*, 153(1):5–11, 2015. 4, 6

[GPT13] João Gouveia, Pablo A. Parrilo, and Rekha R. Thomas. Lifts of convex sets and cone factorizations. *Mathematics of Operations Research*, 38(2):248–264, 2013. 12, 14

[GT12] João Gouveia and Rekha Thomas. Convex hulls of algebraic sets. In *Handbook on Semidefinite, Conic and Polynomial Optimization*, pages 113–138. Springer, 2012. 4

[HN09] J. William Helton and Jiawang Nie. Sufficient and necessary conditions for semidefinite representability of convex hulls and sets. *SIAM Journal on Optimization*, 20(2):759–791, 2009. 4

[Kai11] V. Kaibel. Extended formulations in combinatorial optimization. *Arxiv preprint arXiv:1104.1023, Appeared in Optima 85 newsletter of the Mathematical Optimization Society*, 2011. 6

[KP11] Volker Kaibel and Kanstantsin Pashkovich. Constructing extended formulations from reflection relations. In *Integer Programming and Combinatoral Optimization*, pages 287–300. Springer, 2011. 4

[KPT12] Volker Kaibel, Kanstantsin Pashkovich, and Dirk Oliver Theis. Symmetry matters for sizes of extended formulations. *SIAM Journal on Discrete Mathematics*, 26(3):1361–1382, 2012. 4

[Las09] Jean-Bernard Lasserre. Convex sets with semidefinite representation. *Mathematical Programming*, 120(2):457–477, 2009. 4

[LRS15] James R. Lee, Prasad Raghavendra, and David Steurer. Lower bounds on the size of semidefinite programming relaxations. In *Proceedings of the 47th Annual Symposium on Theory of Computing (STOC)*, pages 567–576. ACM, 2015. 4

[LW14] Cong Han Lim and Stephen Wright. Beyond the Birkhoff polytope: Convex relaxations for vector permutation problems. In *Advances in Neural Information Processing Systems*, pages 2168–2176, 2014. 5

[Rot14] Thomas Rothvoß. The matching polytope has exponential extension complexity. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 263–272. ACM, 2014. 4

[Sch18] Claus Scheiderer. Spectrahedral shadows. *SIAM Journal on Applied Algebra and Geometry*, 2(1):26–44, 2018. 4

[SPW14] James Saunderson, Pablo A. Parrilo, and Alan S. Willsky. Semidefinite descriptions of the convex hull of rotation matrices. *arXiv preprint arXiv:1403.4914*, 2014. 4

[Yan91] Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences*, 43(3):441–466, 1991. 4, 7, 8