On the existence of 0/1 polytopes with high semidefinite extension complexity*

Jop Briët[1†], Daniel Dadush[1], and Sebastian Pokutta[2‡]

[1] New York University, Courant Institute of Mathematical Sciences, New York, NY, USA. {jop.briet,dadush}@cims.nyu.edu
[2] Georgia Institute of Technology, H. Milton Stewart School of Industrial and Systems Engineering, Atlanta, GA, USA. sebastian.pokutta@isye.gatech.edu

Abstract. In Rothvoß [2011] it was shown that there exists a 0/1 polytope (a polytope whose vertices are in $\{0,1\}^n$) such that any higher-dimensional polytope projecting to it must have $2^{\Omega(n)}$ facets, i.e., its linear extension complexity is exponential. The question whether there exists a 0/1 polytope with high PSD extension complexity was left open. We answer this question in the affirmative by showing that there is a 0/1 polytope such that any spectrahedron projecting to it must be the intersection of a semidefinite cone of dimension $2^{\Omega(n)}$ and an affine space. Our proof relies on a new technique to rescale semidefinite factorizations.

1 Introduction

The subject of lower bounds on the size of extended formulations has recently regained a lot of attention. This is due to several reasons. First of all, essentially all NP-Hard problems in combinatorial optimization can be expressed as linear optimization over an appropriate convex hull of integer points. Indeed, many past (erroneous) approaches for proving that P=NP have proceeded by attempting to give polynomial sized linear extended formulations for hard convex hulls (convex hull of TSP tours, indicators of cuts in a graph, etc.). Recent breakthroughs Fiorini et al. [2012], Braun et al. [2012] have unconditionally ruled out such approaches for the TSP and Correlation polytope, complementing the classic result of Yannakakis [1991] which gave lower bounds for symmetric extended formulations. Furthermore, even for polytopes over which optimization is in P, it is very natural to ask what the "optimal" representation of the polytope is. From this perspective, the smallest extended formulation represents the "description complexity" of the polytope in terms of a linear or semidefinite program.

A *(linear) extension* of a polytope $P \subseteq \mathbb{R}^n$ is another polytope $Q \subseteq \mathbb{R}^d$, so that there exists a linear projection π with $\pi(Q) = P$. The *extension complexity*

† J.B. was supported by a Rubicon grant from the Netherlands Organisation for Scientific Research (NWO).

‡ Research reported in this paper was partially supported by NSF grant CMMI-1300144.

* *Keywords and phrases:* semidefinite extended formulations, extended formulations, extension complexity

of a polytope is the minimum number of facets in any of its extensions. The linear extension complexity of P can be thought of as the inherent complexity of expressing P with linear inequalities. Note that in many cases it is possible to save an exponential number of inequalities by writing the polytope in higher-dimensional space. Well-known examples include the regular polygon, see Ben-Tal and Nemirovski [2001] and Fiorini et al. [2011] or the permutahedron, see Goemans [2009]. A *(linear) extended formulation* is simply a normalized way of expressing an extension as an intersection of the nonnegative cone with an affine space; in fact we will use these notions in an interchangeable fashion. In the seminal work of Yannakakis [1988] a fundamental link between the extension complexity of a polytope and the nonnegative rank of an associated matrix, the so called *slack matrix*, was established and it is precisely this link that provided all known strong lower bounds. It states that the nonnegative rank of any slack matrix is equal to the extension complexity of the polytope.

As shown in Fiorini et al. [2012] and Gouveia et al. [2011] the above readily generalizes to semidefinite extended formulations. Let $P \subseteq \mathbb{R}^n$ be a polytope. Then a *semidefinite extension* of P is a spectrahedron $Q \subseteq \mathbb{R}^d$ so that there exists a linear map π with $\pi(Q) = P$. While the projection of a polyhedron is polyhedral, it is open which convex sets can be obtained as projections of spectrahedra. We can again normalize the representation by considering Q as the intersection of an affine space with the cone of positive semidefinite (PSD) matrices. The *semidefinite extension complexity* is then defined as the smallest r for which there exists an affine space such that its intersection with the cone of $r \times r$ PSD matrices projects to P. We thus ask for the smallest representation of P as a projection of a spectrahedron. In both the linear and the semidefinite case, one can think of the extension complexity as the minimum size of the cone needed to represent P. Yannakakis's theorem can be generalized to this case, as was done in Fiorini et al. [2012] and Gouveia et al. [2011], and it asserts that the semidefinite extension complexity of a polytope is equal to the semidefinite rank (see Definition 3) of any of its slack matrices.

An important fact in the study of extended formulations is that the encoding length of the coefficients is disregarded, i.e., we only measure the dimension of the required cone. Furthermore, a lower bound on the extension complexity of a polytope does not imply that building a separation oracle for the polytope is computationally hard. Indeed, as recently shown in **?**, the perfect matching polytope has exponential extension complexity, while the associated separation problem (which allows us to compute min-cost perfect matchings) is in P. Thus standard complexity theoretic assumptions and limitations do not apply. In fact one of the main features of extended formulations is that they *unconditionally* provide lower bounds for the size of linear and semidefinite programs *independent of P vs. NP*.

The first natural class of polytopes with high linear extension complexity comes from the work of Rothvoß [2011]. Rothvoß showed that “random” 0/1 polytopes have exponential linear extension complexity via an elegant counting argument. Given that SDP relaxations are often far more powerful than LP relaxations, an important open question is whether random 0/1 polytopes also have high PSD extension complexity.

1.1 Related work

The basis for the study of linear and semidefinite extended formulations is the work of Yannakakis (see Yannakakis [1988] and Yannakakis [1991]). The existence of a 0/1 polytope with exponential extension complexity was shown in Rothvoß [2011] which in turn was inspired by Shannon [1949]. The first explicit example, answering a long standing open problem of Yannakakis, was provided in Fiorini et al. [2012] which, together with Gouveia et al. [2011], also lay the foundation for the study of extended formulations over general closed convex cones. In Fiorini et al. [2012] it was also shown that there exist matrices with large nonnegative rank but small semidefinite rank, indicating that semidefinite extended formulations can be exponentially stronger than linear ones, however falling short of giving an explicit proof. They thereby separated the expressive power of linear programs from those of semidefinite programs and raised the question:

Does every *0/1 polytope have an efficient semidefinite lift?*

Other related work includes Braun et al. [2012], where the authors study approximate extended formulations and provide examples of spectrahedra that cannot be approximated well by linear programs with a polynomial number of inequalities as well as improvements thereof by Braverman and Moitra [2012]. Faenza et al. [2011] proved equivalence of extended formulations to communication complexity. Recently there has been also significant progress in terms of lower bounding the linear extension complexity of polytopes by means of information theory, see Braverman and Moitra [2012] and Braun and Pokutta [2013]. Similar techniques are not known for the semidefinite case.

1.2 Contribution

We answer the above question in the negative, i.e., we show the existence of a 0/1 polytope with exponential semidefinite extension complexity. In particular, we show that the counting argument of Rothvoß [2011] extends to the PSD setting.

The main challenge when moving to the PSD setting, is that the largest value occurring in the slack matrix does not easily translate to a bound on the largest values occurring in the factorizations. Obtaining such a bound is crucial for the counting argument to carry over.

Our main technical contribution is a new rescaling technique for semidefinite factorizations of slack matrices. In particular, we show that any rank-r semidefinite factorization of a slack matrix with maximum entry size Δ can be "rescaled" to a semidefinite factorization where each factor has operator norm at most $\sqrt{r\Delta}$ (see Theorem 6). Here our proof proceeds by a variational argument and relies on John's theorem on ellipsoidal approximation of convex bodies John [1948].

We note that in the linear case proving such a result is far simpler, here the only required observation is that after independent nonnegative scalings of the coordinates a nonnegative vector remains nonnegative. However, one cannot in general independently scale the entries of a PSD matrix while maintaining the PSD property.

Using our rescaling lemma, the existence proof of the 0/1 polytopes with high semidefinite extension complexity follows in a similar fashion to the linear case as presented in Rothvoß [2011]. In addition to our main result, we show the existence of a polygon with d integral vertices and semidefinite extension complexity $\Omega((\frac{d}{\log d})^{\frac{1}{4}})$. The argument follows similarly to Fiorini et al. [2011] adapting Rothvoß [2011].

1.3 Outline

In Section 2 we provide basic results and notions. We then present the rescaling technique in Section 3 which is at the core of our existence proof. In Section 4 we establish the existence of 0/1 polytopes with subexponential semidefinite extension complexity and we conclude with some final remarks in Section 6.

2 Preliminaries

Let $[n] \coloneqq \{1, \dots, n\}$. In the following we will consider semidefinite extended formulations. We refer the interested reader to Fiorini et al. [2012] and Braun et al. [2012] for a broader overview and proofs.

Let $B_2^n \subseteq \mathbb{R}^n$ denote the n-dimensional Euclidean ball, and let $S^{n-1} = \partial B_2^n$ denote the Euclidean sphere in $\mathbb{R}^n$. We denote by $\mathbb{S}_+^n$ the set of $n \times n$ PSD matrices which form a (non-polyhedral) convex cone. Note that $M \in \mathbb{S}_+^n$ if and only if M is symmetric ($M^\intercal = M$) and

$$x^\intercal M x \geq 0 \quad \forall x \in \mathbb{R}^n.$$

Equivalently, $M \in \mathbb{S}_+^n$ iff M is symmetric and has nonnegative eigenvalues. For a linear subspace $W \subseteq \mathbb{R}^n$, let $\dim(W)$ denote its dimension, $W^\perp$ its orthogonal complement, and $P_W : \mathbb{R}^n \to \mathbb{R}^n$ the orthogonal projection onto W. Note that as a matrix $P_W \in \mathbb{S}_+^n$ and $P_W^2 = P_W$. For a matrix $A \in \mathbb{R}^{n \times m}$, let $\mathrm{Im}(A)$ denote its image or column span, and let $\mathrm{Ker}(A)$ denote its kernel. For a symmetric matrix $A \in \mathbb{R}^{n \times n}$, we have that $\mathrm{Im}(A) = \mathrm{Ker}(A)^\perp$. If $A \in \mathbb{S}_+^n$, we have that $x \in \mathrm{Ker}(A) \Leftrightarrow x^\intercal A x = 0$. We define the pseudo-inverse A^+ of a symmetric matrix A to be the unique matrix satisfying $A^+ A = A A^+ = P_W$, where $W = \mathrm{Im}(A)$. If A has spectral decomposition $A = \sum_{i=1}^k \lambda_i v_i v_i^\intercal$, $v_1, \dots, v_k$ orthonormal, then $A^+ = \sum_{i=1}^k \lambda_i^{-1} v_i v_i^\intercal$.

For matrices $A, B \in \mathbb{S}_+^n$, we have that $\mathrm{Im}(A+B) = \mathrm{Im}(A) + \mathrm{Im}(B)$ and that $\mathrm{Ker}(A+B) = \mathrm{Ker}(A) \cap \mathrm{Ker}(B)$. We denote the trace of $A \in \mathbb{S}_+^n$ by $\mathrm{Tr}[A] = \sum_{i=1}^n A_{ii}$. For a pair of equally-sized matrices A, B we let $\langle A, B \rangle = \mathrm{Tr}[A^\intercal B]$

denote their trace inner product and let $\|A\|_F = \sqrt{\langle A, A\rangle}$ denote the Frobenius norm of A. We denote the operator norm of a matrix $M \in \mathbb{R}^{m\times n}$ by

$$\|M\| = \sup_{\|x\|_2=1} \|Mx\|_2 .$$

If M is square and symmetric ($M^\top = M$), then $\|M\| = \sup_{\|x\|_2=1} |x^\top M x|$, in which case $\|M\|$ denotes the largest eigenvalue of M in absolute value. Lastly, if $M \in \mathbb{S}^n_+$ then $\|M\| = \sup_{\|x\|_2=1} x^\top M x$ by nonnegativity of the inner expression.

For every positive integer ℓ and any ℓ-tuple of matrices $\mathbf{M} = (M_1, \dots, M_\ell)$ we define

$$\|\mathbf{M}\|_\infty = \max\{\|M_i\| \,\big|\, i \in [\ell]\}.$$

Definition 1 (Semidefinite extended formulation) *Let $K \subseteq \mathbb{R}^n$ be a convex set. A* semidefinite extended formulation *(semidefinite EF) of K is a system consisting of a positive integer r, an index set I and a set of triples $(a_i, U_i, b_i)_{i\in I} \subseteq \mathbb{R}^n \times \mathbb{S}^r_+ \times \mathbb{R}$ such that*

$$K = \{x \in \mathbb{R}^n \,\big|\, \exists Y \in \mathbb{S}^r_+ :\ a_i^\top x + \langle U_i, Y\rangle = b_i\ \forall i \in I\}.$$

The size *of a semidefinite EF is the size r of the positive semidefinite matrices U_i. The* semidefinite extension complexity *of K, denoted $\mathrm{xc}_{SDP}(K)$, is the minimum size of a semidefinite EF of K.*

In order to characterize the semidefinite extension complexity of a polytope $P \subseteq [0,1]^n$ we will need the concept of a slack matrix.

Definition 2 (Slack matrix) *Let $P \subseteq [0,1]^n$ be a polytope, I, J be finite sets, $\mathcal{A} = (a_i, b_i)_{i\in I} \subseteq \mathbb{R}^n \times \mathbb{R}$ be a set of pairs and let $\mathcal{X} = (x_j)_{j\in J} \subseteq \mathbb{R}^n$ be a set of points, such that*

$$P = \{x \in \mathbb{R}^n \,\big|\, a_i^\top x \le b_i\ \forall i \in I\} = \mathrm{conv}\,(\mathcal{X})\,.$$

Then, the slack matrix *of P associated with $(\mathcal{A}, \mathcal{X})$ is given by $S_{ij} = b_i - a_i^\top x_j$.*

Finally, the definition of a semidefinite factorization is as follows.

Definition 3 (Semidefinite factorization) *Let I, J be finite sets, $S \in \mathbb{R}^{I\times J}_+$ be a nonnegative matrix and r be a positive integer. Then, a* rank-r semidefinite factorization *of S is a set of pairs $(U_i, V^j)_{(i,j)\in I\times J} \subseteq \mathbb{S}^r_+ \times \mathbb{S}^r_+$ such that*

$$S_{ij} = \langle U_i, V^j\rangle$$

for every $(i,j) \in I \times J$. The semidefinite rank *of S, denoted $\mathrm{rank}_{\mathrm{PSD}}(S)$, is the minimum r such that there exists a rank r semidefinite factorization of S.*

Using the above notions the semidefinite extension complexity of a polytope can be characterized by the semidefinite rank of any of its slack matrices, which is a generalization of Yannakakis's factorization theorem (Yannakakis [1988] and Yannakakis [1991]) established in Fiorini et al. [2012] and Gouveia et al. [2011].

Theorem 4 (Yannakakis's Factorization Theorem for SDPs). *Let $P \subseteq [0,1]^n$ be a polytope and $\mathcal{A} = (a_i, b_i)_{i \in I}$ and $\mathcal{X} = (x_j)_{j \in J}$ be as in Definition 2. Let S be the slack matrix of P associated with $(\mathcal{A}, \mathcal{X})$. Then, S has a rank-r semidefinite factorization if and only if P has a semidefinite EF of size r. That is,* $\operatorname{rank}_{\mathrm{PSD}}(S) = \operatorname{xc}_{SDP}(P)$.

Moreover, if $(U_i, V^j)_{(i,j) \in I \times J} \subseteq \mathbb{S}^r_+ \times \mathbb{S}^r_+$ is a factorization of S, then

$$P = \{x \in \mathbb{R}^n \big| \, \exists Y \in \mathbb{S}^r_+ : \, a_i^\intercal x + \langle U_i, Y \rangle = b_i \; \forall i \in I\}$$

and the pairs $(x_j, V^j)_{j \in J}$ satisfy $a_i^\intercal x_j + \langle U_i, V^j \rangle = b_i$ for every $i \in I$.

In particular, the extension complexity is independent of the choice of the slack matrix and the semidefinite rank of all slack matrices of P is identical.

The following well-known theorem due to John [1948] lies at the core of our rescaling argument. We state a version that is suitable for the later application. Recall that B_2^n denotes the n-dimensional Euclidean unit ball. A *probability vector* is a vector $p \in \mathbb{R}^n_+$ such that $p(1) + p(2) + \cdots + p(n) = 1$. For a convex set $K \subseteq \mathbb{R}^n$, we let $\operatorname{aff}(K)$ denote the affine hull of K, the smallest affine space containing K. We let $\dim(K)$ denote the linear dimension of the affine hull of K. Last, we let $\operatorname{relbd}(K)$ denote the relative boundary of K, i.e., the topological boundary of K with respect to its affine hull $\operatorname{aff}(K)$.

Theorem 5 (John [1948]). *Let $K \subseteq \mathbb{R}^n$ be a centrally symmetric convex set with* $\dim(K) = k$. *Let $T \in \mathbb{R}^{n \times k}$ be such that $E = T \cdot B_2^k = \{Tx \big| \, \|x\| \leq 1\}$ is the smallest volume ellipsoid containing K. Then, there exist a finite set of points $\mathcal{Z} \subseteq \operatorname{relbd}(K) \cap \operatorname{relbd}(E)$ and a probability vector $p \in \mathbb{R}^{\mathcal{Z}}_+$ such that*

$$\sum_{z \in \mathcal{Z}} p(z) \, zz^\intercal = \frac{1}{k} TT^\intercal.$$

For a real-valued function $f : \mathbb{R} \to \mathbb{R}$, we denote its right-sided derivative at $a \in \mathbb{R}$ by

$$\frac{d_+}{dx} f|_{x=a} = \lim_{\varepsilon \to 0^+} \frac{f(a+\varepsilon) - f(a)}{\varepsilon}.$$

We will need the following lemma; for a general theory on perturbations on linear operators we refer the reader to **?**.

Lemma 1. *Let r be a positive integer, $X \in \mathbb{S}^r_+$ be a non-zero positive semidefinite matrix. Let $\lambda_1 = \|X\|$ and W denote the λ_1-eigenspace of X. Then for $Z \in \mathbb{R}^{r \times r}$ symmetric,*

$$\frac{d_+}{d\varepsilon} \, \|X + \varepsilon Z\| \bigg|_{\varepsilon = 0} = \max_{\substack{w \in W \\ \|w\|_2 = 1}} w^\intercal Z w$$

Proof: Observe that

$$\|X + \varepsilon Z\| \geq \max_{\|w\|_2 = 1, w \in W} w^\intercal (X + \varepsilon Z) w = \max_{\|w\|_2 = 1, w \in W} \underbrace{w^\intercal X w}_{= \lambda_1} + \varepsilon w^\intercal Z w$$
$$= \lambda_1 + \varepsilon \cdot \max_{\|w\|_2 = 1, w \in W} w^T Z w.$$

It therefore suffices to show that $\|X + \varepsilon Z\|$ cannot exceed the lower bound by more than $o(\varepsilon)$.

Let u be an arbitrary vector with $\|u\|_2 = 1$ and write $u = u_1 + u_2$ with $u_1 \in W$ and $u_2 \in W^\perp$, where the latter is the orthogonal complement of W. Clearly, $\|u_1\|_2^2 + \|u_2\|_2^2 = 1$. Further let $\Delta := \lambda_1 - \lambda_2$ where λ_2 is the second largest Eigenvalue of X and for readibility let $\lambda_1(Z \upharpoonright W) := \max_{\|w\|_2=1, w\in W} w^\top Z w$. We estimate

$$\begin{aligned}
u^\top (X + \varepsilon Z)u &= u_1^\top X u_1 + u_2^\top X u_2 + \varepsilon(u_1^\top Z u_1 + u_1^\top Z u_2 + u_2^\top Z u_1 + u_2^\top Z u_2) \\
&\le \lambda_1 \|u_1\|_2^2 + \lambda_2 \|u_2\|_2^2 + \varepsilon\lambda_1(Z \upharpoonright W) + 3\varepsilon \|Z\| \|u_2\|_2 \\
&= \lambda_1 + \varepsilon\lambda_1(Z \upharpoonright W) + (3\varepsilon \|Z\| - \Delta \|u_2\|_2) \|u_2\|_2 \\
&= \lambda_1 + \epsilon\lambda_1(Z \upharpoonright W) - (\sqrt{\Delta} \|u_2\|_2 - 3\varepsilon \|Z\| / \sqrt{4\Delta})^2 + 9\varepsilon^2 \|Z\|^2 / (4\Delta) \\
&\le \lambda_1 + \epsilon\lambda_1(Z \upharpoonright W) + 9\varepsilon^2 \|Z\|^2 / (4\Delta),
\end{aligned}$$

which finishes the proof. □

We record the following corollary of Lemma 1 for later use. Recall that for a square matrix X, its *exponential* is given by

$$e^X = \sum_{k=0}^{\infty} \frac{1}{k!} X^k = I + X + \frac{1}{2} X^2 + \cdots .$$

Corollary 1. *Let r be a positive integer, $X \in \mathbb{S}_+^r$ be a non-zero positive semidefinite matrices. Let $\lambda_1 = \|X\|$ and W denote the λ_1-eigenspace of X. Then for $Z \in \mathbb{R}^{r\times r}$ symmetric,*

$$\frac{d_+}{d\varepsilon} \left\| e^{\varepsilon Z} X e^{\varepsilon Z} \right\| \Big|_{\varepsilon=0} = 2\lambda_1 \max_{\substack{w\in W \\ \|w\|_2=1}} w^\top Z w$$

Proof: Let us write $e^{\varepsilon Z} = \sum_{k=0}^{\infty} \frac{\varepsilon^k Z^k}{k!} = I + \varepsilon Z + \varepsilon^2 R_\varepsilon$, where $R_\varepsilon = \sum_{k=2}^{\infty} \frac{\varepsilon^{k-2} Z^k}{k!}$. For $\varepsilon < 1/(2\|Z\|)$, by the triangle inequality

$$\|R_\varepsilon\| \le \sum_{k=2}^{\infty} \frac{\varepsilon^{k-2} \|Z\|^k}{k!} \le \frac{\|Z\|^2}{2} \sum_{k=0}^{\infty} (\varepsilon \|Z\|)^k = \frac{\|Z\|^2}{2(1 - \varepsilon \|Z\|)} \le \|Z\|^2$$

From here we see that

$$e^{\varepsilon Z} X e^{\varepsilon Z} = (I + \varepsilon Z + \varepsilon^2 R_\varepsilon) X (I + \varepsilon Z + \varepsilon^2 R_\varepsilon) = X + \varepsilon(ZX + XZ) + \varepsilon^2 (ZXR_\varepsilon + R_\varepsilon XZ + R_\varepsilon X R_\varepsilon)$$

Let $R'_\varepsilon = ZXR_\varepsilon + R_\varepsilon XZ + R_\varepsilon X R_\varepsilon$. Again by the triangle inequality, we have that

$$\|R'_\varepsilon\| \le 2\|Z\| \|X\| \|R_\varepsilon\| + \|R_\varepsilon\|^2 \|X\| \le 2\|Z\|^3 \|X\| + \|Z\|^4 \|X\| = O(1),$$

for ε small enough. Therefore, we have that

$$\begin{aligned}
\left\| e^{\varepsilon Z} X e^{\varepsilon Z} \right\| &= \left\| X + \varepsilon(XZ + ZX) + \varepsilon^2 R'_\varepsilon \right\| = \|X + \varepsilon(XZ + ZX)\| \pm O(\varepsilon^2 \|R'_\varepsilon\|) \\
&= \|X + \varepsilon(XZ + ZX)\| \pm O(\varepsilon^2).
\end{aligned}$$

Since $XZ + ZX$ is symmetric and $X \in \mathbb{S}_+^r$ and non-zero, by Lemma 1 we have that

$$\begin{aligned}\|X + \varepsilon(XZ + ZX)\| &= \lambda_1 + \varepsilon(\max_{\substack{w \in W \\ \|w\|_2 = 1}} w^\mathsf{T}(XZ + ZX)w) \pm O(\varepsilon^2) \\ &= \lambda_1 + \varepsilon\lambda_1(\max_{\substack{w \in W \\ \|w\|_2 = 1}} w^\mathsf{T}(Z + Z)w) \pm O(\varepsilon^2) \\ &= \lambda_1 + 2\lambda_1\varepsilon(\max_{\substack{w \in W \\ \|w\|_2 = 1}} w^\mathsf{T}Zw) \pm O(\varepsilon^2)\end{aligned}$$

Putting it all together, we get that

$$\left\|e^{\varepsilon Z} X e^{\varepsilon Z}\right\| = \|X + \varepsilon(XZ + ZX)\| + O(\varepsilon^2) = \lambda_1 + 2\lambda_1\varepsilon(\max_{\substack{w \in W \\ \|w\|_2 = 1}} w^\mathsf{T}Zw) \pm O(\varepsilon^2)$$

as needed. □

3 Rescaling semidefinite factorizations

A crucial point will be the rescaling of a semidefinite factorization of a nonnegative matrix M. In the case of linear extended formulations an upper bound of Δ on the largest entry of a slack matrix S implies the existence of a minimal nonnegative factorization $S = UV$ where the entries of U, V are bounded by $\sqrt{\Delta}$. This ensures that the approximation of the extended formulation can be captured by means of a polynomial-size (in Δ) grid. In the linear case, we note that any factorization $S = UV$ can be rescaled by a nonnegative diagonal matrix D where $S = (UD)(D^{-1}U)$ and the factorization $(UD, D^{-1}V)$ has entries bounded by $\sqrt{\Delta}$. However, such a rescaling relies crucially on the fact that after independent nonnegative scalings of the coordinates a nonnegative vector remains nonnegative. However, in the PSD setting, it is not true that the PSD property is preserved after independent nonnegative scalings of the matrix entries. We circumvent this issue by showing that a restricted class of transformations, i.e. the symmetries of the semidefinite cone, suffice to rescale any PSD factorization such that the largest eigenvalue occurring in the factorization is bounded in terms of the maximum entry in M and the rank of the factorization.

Theorem 6 (Rescaling semidefinite factorizations). *Let Δ be a positive real number, I, J be finite sets, $M \in [0, \Delta]^{I \times J}$ be a nonnegative matrix with a rank r semidefinite factorization factorization $(\mathbf{U}, \mathbf{V})$, $\mathbf{U} = (U_i)_{i \in I}$, $\mathbf{V} = (V^j)_{j \in J}$, satisfying $M_{ij} = \operatorname{Tr}\left[U_i V^j\right]$, $i \in I, j \in J$. Then there exists $A \in \mathbb{S}_+^r$ such that $A\mathbf{U}A = (AU_iA)_{i \in I}$, $A^+\mathbf{V}A^+ = (A^+V^jA^+)_{j \in J}$ is a semidefinite factorization of M satisfying*

$$\begin{aligned}\|A\mathbf{U}A\|_\infty &= \max_{i \in I} \|AU_iA\| \leq \sqrt{r\Delta} \\ \left\|A^+\mathbf{V}A^+\right\|_\infty &= \max_{j \in J} \left\|A^+V^jA^+\right\| \leq \sqrt{r\Delta}\ .\end{aligned}$$

Proof: Let $\bar{U} = \sum_{i \in I} U_i/|I|$, $\bar{V} = \sum_{j \in J} V^j/|J|$. Let $W_1 = \mathrm{Im}(\bar{U})$, $W_2 = \mathrm{Im}(\bar{V})$, $W = P_{W_1}(W_2)$ and $d = \dim(W)$. Let $O \in \mathbb{R}^{r \times d}$ denote an orthonormal basis matrix for W, that is $\mathrm{Im}(O) = W$, $OO^\mathsf{T} = P_W$, and $O^\mathsf{T}O = I_d$ (the $d \times d$ identity).

As a first step, we preprocess the factorization to make it full dimensional (i.e., by reducing the ambient dimension).

Claim. $(O^\mathsf{T}\mathbf{U}O, O^\mathsf{T}\mathbf{V}O)$ is a semidefinite factorization of M. Furthermore, $O^\mathsf{T}\bar{U}O$ and $O^\mathsf{T}\bar{V}O$ are $d \times d$ nonsingular matrices.

Proof of claim: If $T \in \mathbb{S}^r_+$ then for any matrix $A \in \mathbb{R}^{r \times d}$, we have that $A^\mathsf{T}TA \in \mathbb{S}^d_+$. Hence $O^\mathsf{T}U_iO, O^\mathsf{T}V^jO \in \mathbb{S}^d_+$, for all $i \in I, j \in J$. To show that the new matrices factorize M, it suffices to show that $M_{ij} = \mathrm{Tr}\left[O^\mathsf{T}U_iOO^\mathsf{T}V^jO\right]$ for all $i \in I, j \in J$. We examine spectral decompositions of U_i and V_j,

$$U_i = \sum_{k=1}^{r} \lambda_k u_k u_k^\mathsf{T} \quad \text{and} \quad V^j = \sum_{k=1}^{r} \gamma_k v_k v_k^\mathsf{T}.$$

For $k \in [r]$, we have that $u_k \in \mathrm{Im}(U_i) \subseteq \sum_{i \in I} \mathrm{Im}(U_i) = \mathrm{Im}(\bar{U}) = W_1$. Similarly for $l \in [r]$, $v_l \in \mathrm{Im}(V^j) \subseteq \mathrm{Im}(\bar{V}) = W_2$. Given the previous containment, remembering that $P_{W_1}(W_2) = W$, for $k, l \in [r]$ we have that

$$\langle u_k, v_l \rangle = \langle P_{W_1} u_k, v_l \rangle = \langle u_k, P_{W_1} v_l \rangle = \langle u_k, P_W v_l \rangle = \langle P_W u_k, P_W v_l \rangle = \langle O^\mathsf{T} u_k, O^\mathsf{T} v_l \rangle,$$

since O is an orthonormal basis matrix for W. The trace inner product can now be analyzed as follows

$$\begin{aligned}\mathrm{Tr}\left[U_i V^j\right] &= \sum_{1 \le k,l \le r} \lambda_k \gamma_l \langle u_k, v_l \rangle^2 = \sum_{1 \le k,l \le r} \lambda_k \gamma_l \langle O^\mathsf{T} u_k, O^\mathsf{T} v_l \rangle^2 \\ &= \mathrm{Tr}\left[(O^\mathsf{T}U_iO)(O^\mathsf{T}V^jO)\right].\end{aligned}$$

Hence $(O^\mathsf{T}\mathbf{U}O, O^\mathsf{T}\mathbf{V}O)$ is a semidefinite factorization of M as needed.

For the furthermore, we must show that the matrices $O^\mathsf{T}\bar{U}O$ and $O^\mathsf{T}\bar{V}O$ have trivial kernels. By construction $\mathrm{Ker}(\bar{U}) = W_1^\perp$, $\mathrm{Ker}(\bar{V}) = W_2^\perp$, and

$$W = P_{W_1}(W_2) = (W_2 + W_1^\perp) \cap W_1.$$

From here, we have that

$$\dim(\mathrm{Ker}(O^\mathsf{T}\bar{U}O)) = \dim(\mathrm{Ker}(\bar{U}O)) = \dim(W_1^\perp \cap W) \le \dim(W_1^\perp \cap W_1) = \dim(\{0\}) = 0.$$

Next, we have that

$$\begin{aligned}\dim(\mathrm{Ker}(O^\mathsf{T}\bar{V}O)) &= \dim(\mathrm{Ker}(\bar{V}O)) = \dim(W_2^\perp \cap W) = \dim((W_2^\perp \cap W_1) \cap (W_2 + W_1^\perp)) \\ &= \dim((W_2 + W_1^\perp)^\perp \cap (W_2 + W_1^\perp)) = \dim(\{0\}) = 0,\end{aligned}$$

as needed. ♦

We will now examine factorizations of the form $(AO^\intercal \mathbf{U} OA, A^{-1}O^\intercal \mathbf{V} OA^{-1})$, for $A \in \mathbb{S}_+^d$ nonsingular. To see that this yields a factorization, note that

$$\begin{aligned}
&\operatorname{Tr}\left[AO^\intercal U_i OAA^{-1}O^\intercal V^j OA^{-1}\right] = \operatorname{Tr}\left[AO^\intercal U_i OO^\intercal V^j OA^{-1}\right] \\
&= \operatorname{Tr}\left[O^\intercal U_i OO^\intercal V^j OA^{-1}A\right] = \operatorname{Tr}\left[O^\intercal U_i OO^\intercal V^j O\right] = M_{ij},
\end{aligned}$$

where the last inequality follows from Claim 3. To prove the theorem, it suffices to construct a nonsingular matrix $A \in \mathbb{S}_+^d$ such that

$$\left\|AO^\intercal \mathbf{U} OA\right\|_\infty \leq \sqrt{d\Delta} \quad \text{and} \quad \left\|A^{-1}O^\intercal \mathbf{V} OA^{-1}\right\|_\infty \leq \sqrt{d\Delta}.$$

Given such an A, we can recover the rescaling matrix claimed in the theorem using $OAO^\intercal$, where $(OAO^\intercal)^+ = OA^{-1}O^\intercal$. It is easy to check that this lifting is valid and preserves the maximum eigenvalues of the factorization matrices.

Given the above reduction, we may now assume that $d = r$ and that $\bar{U}$, $\bar{V}$ are nonsingular. We define the following potential function over factorizations,

$$\Phi_M(\mathbf{U}, \mathbf{V}) = \|\mathbf{U}\|_\infty \cdot \|\mathbf{V}\|_\infty .$$

We now examine the optimization problem

$$\inf_{\substack{A \in \mathbb{S}_+^r \\ A \text{ nonsingular}}} \Phi_M(A\mathbf{U}A, A^{-1}\mathbf{V}A^{-1}) \ . \tag{1}$$

For any nonsingular $T \in \mathbb{R}^{r\times r}$, $(T\mathbf{U}T^\intercal, T^{-\intercal}\mathbf{V}T^{-1})$ is a valid PSD factorization of M. Without loss of generality we can require T to be PSD as above, since T can be always be expressed as $T = OA$, where O is orthogonal and $A \in \mathbb{S}_+^r$. Here it is easy to check that substituting A for T does not change the Φ_M value of the factorization.

Recall that the goal is to construct a nonsingular $A \in \mathbb{S}_+^r$ such that

$$\|A\mathbf{U}A\|_\infty \leq \sqrt{r\Delta} \quad \text{and} \quad \left\|A^{-1}\mathbf{V}A^{-1}\right\|_\infty \leq \sqrt{r\Delta}\ .$$

For any scalar $s > 0$, we see that

$$\|sA\mathbf{U}sA\|_\infty = s^2 \|A\mathbf{U}A\|_\infty \quad \text{and} \quad \left\|(sA)^{-1}\mathbf{V}(sA)^{-1}\right\|_\infty = \left\|A^{-1}\mathbf{V}A^{-1}\right\|_\infty / s^2.$$

Given this, if $\Phi_M(A\mathbf{U}A, A^{-1}\mathbf{V}A^{-1}) \leq \mu^2$ then setting

$$s = \Phi_M(A\mathbf{U}A, A^{-1}\mathbf{V}A^{-1})^{1/4} / \|A\mathbf{U}A\|_\infty^{1/2} ,$$

we get that

$$\|sA\mathbf{U}sA\|_\infty = \left\|(sA)^{-1}\mathbf{V}(sA)^{-1}\right\|_\infty = \Phi_M(A\mathbf{U}A, A^{-1}\mathbf{V}A^{-1})^{1/2} \leq \mu.$$

Hence it suffices to show that the infimum value for (1) is less than or equal to $r\Delta$. We claim that this infimum is attained. Since the objective functions is clearly continuous in A, it suffices to show that the infimum can be taken over a compact subset of $\mathbb{S}_+^r$. Let $\tau = \Phi_M(\mathbf{U}, \mathbf{V})$, and $\sigma > 0$ be the largest value such that $\bar{U} \succeq \sigma I_r$, $\bar{V} \succeq \sigma I_r$. Note that $\sigma > 0$ exists since $\bar{U}$,$\bar{V}$ are nonsingular $r \times r$ PSD matrices.

Claim. Let $A \in \mathbb{S}_+^r$ nonsingular. If $\Phi_M(A\mathbf{U}A, A^{-1}\mathbf{V}A^{-1}) \leq \tau$, then there exists $s > 0$ such that $I_r \preceq sA \preceq (\tau/\sigma^2)I_r$.

PROOF OF CLAIM: We examine the spectral decomposition of $A = \sum_{i=1}^r \lambda_i v_i v_i^\mathsf{T}$, where $v_1, \ldots, v_r$ form an orthornomal basis of $\mathbb{R}^r$ and $\lambda_1 \geq \cdots \geq \lambda_r \geq 0$. Note that $A^{-1} = \sum_{i=1}^r \lambda_i^{-1} v_i v_i^\mathsf{T}$. Here $\lambda_r > 0$ since A is nonsingular. Since multiplying A by a positive scalar does not change the potential Φ_M, we may rescale A such that $\lambda_r = 1$. Since $\lambda_r I_r \preceq A \preceq \lambda_1 I_r$, and $\lambda_r = 1$, we must now show that $\lambda_1 \leq \tau/\sigma^2$.

We lower bound $\Phi(A)$ in terms of λ_1. Firstly, note that

$$\begin{aligned}\|A\mathbf{U}A\|_\infty &= \max_{i\in I} \|AU_iA\| \geq \max_{i\in I} v_1^\mathsf{T} AU_iAv_1 = \lambda_1 \max_{i\in I} v_1^\mathsf{T} U_i v_1 \\ &\geq \lambda_1 \frac{1}{|I|}\sum_{i\in I} v_1^\mathsf{T} U_i v_1 = \lambda_1 v_1^\mathsf{T} \bar{U} v_1 \geq \sigma\lambda_1.\end{aligned}$$

Next, we have that

$$\begin{aligned}\left\|A^{-1}\mathbf{V}A^{-1}\right\|_\infty &= \max_{j\in J} \left\|A^{-1}V^jA^{-1}\right\| \geq \max_{j\in J} v_r^\mathsf{T} A^{-1}V^jA^{-1}v_r = \lambda_r^{-1} \max_{j\in J} v_r^\mathsf{T} V^j v_r \\ &= \max_{j\in J} v_r^\mathsf{T} V^j v_r \geq \frac{1}{|J|}\sum_{j\in J} v_r^\mathsf{T} V^j v_r = v_r^\mathsf{T} \bar{V} v_r \geq \sigma.\end{aligned}$$

Therefore

$$\tau \geq \|A\mathbf{U}A\|_\infty \left\|A^{-1}\mathbf{V}A^{-1}\right\|_\infty \geq \lambda_1\sigma^2 \Rightarrow \lambda_1 \leq \tau/\sigma^2,$$

as needed. ♦

From the above claim, and our assumption that $\Phi_M(\mathbf{U}, \mathbf{V}) = \tau$, we get that

$$\inf_{\substack{A\in\mathbb{S}_+^r \\ A \text{ nonsingular}}} \Phi_M(A\mathbf{U}A, A^{-1}\mathbf{V}A^{-1}) = \inf_{\substack{A\in\mathbb{S}_+^r \\ I_r\preceq A\preceq(\tau/\sigma^2)I_r}} \Phi_M(A\mathbf{U}A, A^{-1}\mathbf{V}A^{-1}).$$

Since the infimum on the right hand side is taken on a compact set, the infimum is attained as claimed. Let μ^2 denote the infimum value. Letting $(\widetilde{\mathbf{U}}, \widetilde{\mathbf{V}}) = (A\mathbf{U}A, A^{-1}\mathbf{V}A^{-1})$, for the appropriate matrix $A \in \mathbb{S}_+^r$, we can assume

$$\left\|\widetilde{\mathbf{U}}\right\|_\infty = \left\|\widetilde{\mathbf{V}}\right\|_\infty = \Phi_M(\widetilde{\mathbf{U}}, \widetilde{\mathbf{V}})^{1/2} = \mu.$$

We shall now analyze how Φ_M behaves under small perturbations of the minimizer $(\widetilde{\mathbf{U}}, \widetilde{\mathbf{V}})$. Our goal is to obtain a contradiction by assuming that $\mu^2 > \Delta r + \tau$ for some $\tau > 0$. To this end we bound the value of Φ_M at infinitesimal perturbations of the point $(\widetilde{\mathbf{U}}, \widetilde{\mathbf{V}})$. For a symmetric matrix Z and parameter $\varepsilon > 0$ the type of perturbations we consider are those defined by the invertible matrix $e^{-\varepsilon Z}$, which will take the role of the matrix A above. Notice that if Z is symmetric, then so is $e^{-\varepsilon Z}$. We show that there exists a matrix Z such that for every $U \in \{\widetilde{U}_i \big| i \in I\}$ such that $\|U\| = \mu$, we have

$$\left\|e^{-\varepsilon Z} U e^{-\varepsilon Z}\right\| \leq \mu - \frac{2\mu}{r}\varepsilon + O(\varepsilon^2), \tag{2}$$

while at the same time for every $V \in \{\widetilde{V}^j \big| j \in J\}$ such that $\|V\| = \mu$, we have

$$\left\|e^{\varepsilon Z} V e^{\varepsilon Z}\right\| \leq \mu + \frac{2\Delta}{\mu}\varepsilon + O(\varepsilon^2). \tag{3}$$

This implies that there is a point $(\mathbf{U}', \mathbf{V}')$ in the neighborhood of the minimizer $(\widetilde{\mathbf{U}}, \widetilde{\mathbf{V}})$ where

$$\begin{aligned}
\Phi_M(\mathbf{U}', \mathbf{V}') &\leq \Big(\mu - \frac{2\mu}{r}\varepsilon + O(\varepsilon^2)\Big) \cdot \Big(\mu + \frac{2\Delta}{\mu}\varepsilon + O(\varepsilon^2)\Big) \\
&= \mu^2 - 2\Big(\frac{\mu^2}{r} - \Delta\Big)\varepsilon + O(\varepsilon^2) \\
&< \mu^2 - \frac{2\tau}{r}\varepsilon + O(\varepsilon^2),
\end{aligned}$$

where the last inequality follows from our assumption that $\mu^2 > \Delta r + \tau$. Thus, for small enough $\varepsilon > 0$, we have $\Phi_M(\mathbf{U}', \mathbf{V}') < \mu^2$, a contradiction to the minimality of μ. It suffices to consider the factorization matrices with the largest eigenvalues as small perturbations cannot change the eigenvalue structure. Hence, to prove the theorem we need to show the existence of such a matrix Z.

Let $\mathcal{Z} \subseteq S^{r-1}$ be a finite set of unit vectors such that every $z \in \mathcal{Z}$ is a μ-eigenvector of at least one of the matrices $\widetilde{U}_i$ for $i \in I$. Let $p \in \mathbb{R}_+^{\mathcal{Z}}$ be a probability vector (i.e., $\sum_{z \in \mathcal{Z}} p(z) = 1$) and define the symmetric matrix

$$Z = \sum_{z \in \mathcal{Z}} p(z)\, zz^{\mathsf{T}}. \tag{4}$$

Claim. Let $V \in \{\widetilde{V}^j \big| j \in J\}$ be one of the factorization matrices such that $\|V\| = \mu$. Then,

$$\frac{d_+}{d\varepsilon} \left\|e^{\varepsilon Z} V e^{\varepsilon Z}\right\| \bigg|_{\varepsilon = 0} \leq \frac{2\Delta}{\mu}. \tag{5}$$

Proof of claim: Let $\mathcal{V} \subseteq S^{r-1}$ be the set of eigenvectors of V that have eigenvalue μ. Then, Corollary 1 gives

$$\frac{d_+}{d\varepsilon} \left\|e^{\varepsilon Z} V e^{\varepsilon Z}\right\| \bigg|_{\varepsilon = 0} = 2\mu \max_{v \in \mathcal{V}} v^{\mathsf{T}} Z v = 2\mu \max_{v \in \mathcal{V}} \sum_{z \in \mathcal{Z}} p(z)(z^{\mathsf{T}} v)^2 \tag{6}$$

We show that for any $z \in \mathcal{Z}$ and $v \in \mathcal{V}$, we have $(z^{\mathsf{T}} v)^2 \leq \Delta/\mu^2$. The claim then follows from (6) since p is a probability vector. Let us fix vectors $z \in \mathcal{Z}$ and $v \in \mathcal{V}$ and let $U \in \{\widetilde{U}_i \big| i \in I\}$ be a factorization matrix such that z is a

μ-eigenvector of U. Recall that the matrices U and V are part of a semidefinite factorization of the matrix M and that we assumed the entries of M to have value at most Δ. Hence, $\mathrm{Tr}[U^\mathsf{T} V] \le \Delta$. We now argue that $\mu^2(z^\mathsf{T} v) \le \mathrm{Tr}[U^\mathsf{T} V]$. Let $U = \sum_{k\in[r]} \lambda_k u_k u_k^\mathsf{T}$ and $V = \sum_{\ell\in[r]} \gamma_\ell v_\ell v_\ell^\mathsf{T}$ be spectral decompositions of U and V, respectively, such that $u_1 = z$ and $v_1 = v$. The λ_k and γ_ℓ are nonnegative (as U, V are PSD) and $\lambda_1 = \gamma_1 = \mu$. Hence, expanding the trace inner product

$$\mathrm{Tr}[U^\mathsf{T} V] = \sum_{k,\ell\in[r]} \lambda_k \gamma_\ell (u_k^\mathsf{T} v_\ell)^2, \tag{7}$$

we get that the terms on the right-hand side of (7) are nonnegative and that the sum in (7) is at least $\lambda_1\gamma_1(u_1^\mathsf{T} v_1)^2 = \mu^2(z^\mathsf{T} v)^2$. Putting these observations together we conclude that $\mu^2(z^\mathsf{T} v)^2 \le \mathrm{Tr}[U^\mathsf{T} V] \le \Delta$, which proves the claim. ♦

Claim. There exists a choice of unit vectors $\mathcal{Z}$ and probabilities p such that the following holds. Let $I' = \{i \in I \,|\, \|\widetilde{U}_i\| = \mu\}$. Then, for Z as in (4) we have

$$\frac{d_+}{d\varepsilon} \left\| e^{-\varepsilon Z} \widetilde{U}_i e^{-\varepsilon Z} \right\| \Big|_{\varepsilon=0} \le -\frac{2\mu}{r} \quad \forall i \in I'. \tag{8}$$

Proof of claim: For every $i \in I'$, let $\mathcal{U}_i \subseteq \mathbb{R}^r$ be the vector space spanned by the μ-eigenvectors of $\widetilde{U}_i$. Define the convex set $K = \mathrm{conv}\left(\bigcup_{i\in I'} (\mathcal{U}_i \cap B_2^r)\right)$. Notice that K is centrally symmetric. Let $k = \dim(K)$, and let $T \in \mathbb{R}^{r\times k}$ denote a linear transformation such that that $E = TB_2^k$ is the smallest volume ellipsoid containing K. By John's Theorem, there exists a finite set $\mathcal{Z} \subseteq \mathrm{relbd}(K) \cap \mathrm{relbd}(E)$ and a probability vector $p \in \mathbb{R}_+^{\mathcal{Z}}$ such that

$$Z = \sum_{z\in\mathcal{Z}} p(z)\, zz^\mathsf{T} = \frac{1}{k} TT^\mathsf{T}. \tag{9}$$

Notice that each $z \in \mathcal{Z}$ must be an extreme point of K (as it is one for E) and the set of extreme points of K is exactly $\bigcup_{i\in I'} (\mathcal{U}_i \cap S^{r-1})$. Hence, each $z \in \mathcal{Z}$ is a unit vector and at the same time a μ-eigenvector of some $\widetilde{U}_i$, $i \in I'$.

For $i \in I'$, by Corollary 1 and (9) we have that

$$\begin{aligned}
\frac{d_+}{d\varepsilon} \left\| e^{-\varepsilon Z} \widetilde{U}_i e^{-\varepsilon Z} \right\| \Big|_{\varepsilon=0} &= 2\mu \max\{u^\mathsf{T}(-Z)u \,|\, u \in \mathcal{U}_i \cap S^{r-1}\} \\
&= -2\mu \min\{u^\mathsf{T} Z u \,|\, u \in \mathcal{U}_i \cap S^{r-1}\} \\
&= -\frac{2\mu}{k} \min\{u^\mathsf{T} TT^\mathsf{T} u \,|\, u \in \mathcal{U}_i \cap S^{r-1}\} \\
&\le -\frac{2\mu}{r} \min\{\|T^\mathsf{T} u\|_2^2 \,|\, u \in \mathcal{U}_i \cap S^{r-1}\}.
\end{aligned}$$

Since $E \supseteq K \supseteq (\mathcal{U}_i \cap S^{r-1})$, for any $u \in \mathcal{U}_i \cap S^{r-1}$, we have

$$\|T^\mathsf{T} u\|_2 = \sup_{x\in E} x^\mathsf{T} u \ge \sup_{y\in K} y^\mathsf{T} u \ge u^\mathsf{T} u = 1 \text{ as needed.}$$

♦

Notice that the first claim implies (3) and the second claim implies (2). Hence, our assumption $\mu^2 > \Delta r + \tau$ contradicts that μ is the minimum value of Φ_M. $\square$

4 0/1 polytopes with high semidefinite xc

The lower bound estimation will crucially rely on the fact that any 0/1 polytope in the n-dimensional unit cube can be written as a linear system of inequalities $Ax \leq b$ with integral coefficients where the largest coefficient is bounded by $(\sqrt{n+1})^{n+1} \leq 2^{n \log(n)}$, see e.g., [**?**, Corollary 26]. Using Theorem 6 the proof follows along the lines of Rothvoß [2011]; for simplicity and exposition we chose a compatible notation. We use different estimation however and we need to invoke Theorem 6. In the following let $\mathbb{S}^r_+(\alpha) = \{X \in \mathbb{S}^r_+ \mid \|X\| \leq \alpha\}$.

Lemma 2 (Rounding lemma). *For a positive integer n set $\Delta := (n+1)^{(n+1)/2}$. Let $\mathcal{X} \subseteq \{0,1\}^n$ be a nonempty set, let $r := \mathrm{xc}_{SDP}(\mathrm{conv}(\mathcal{X}))$ and let $\delta \leq \left(16r^3(n+r^2)\right)^{-1}$. Then, for every $i \in [n+r^2]$ there exist:*

1. *an integer vector $a_i \in \mathbb{Z}^n$ such that $\|a_i\|_\infty \leq \Delta$,*
2. *an integer b_i such that $|b_i| \leq \Delta$,*
3. *a matrix $U_i \in \mathbb{S}^r_+(\sqrt{r\Delta})$ whose entries are integer multiples of δ/Δ and have absolute value at most $8r^{3/2}\Delta$, such that*

$$\mathcal{X} = \left\{x \in \{0,1\}^n \,\middle|\, \exists Y \in \mathbb{S}^r_+(\sqrt{r\Delta}) : \left|b_i - a_i^\intercal x - \langle Y, U_i\rangle\right| \leq \frac{1}{4(n+r^2)} \forall i \in [n+r^2]\right\}.$$

Proof: For some index set I let $\mathcal{A} = (a_i, b_i)_{i \in I} \subseteq \mathbb{Z}^n \times \mathbb{Z}$ be a non-redundant description of $\mathrm{conv}(\mathcal{X})$ (i.e., $|I|$ is minimal) such that for every $i \in I$, we have $\|a_i\|_\infty \leq \Delta$ and $|b_i| \leq \Delta$. Let J be an index set for $\mathcal{X} = (x_j)_{j \in J}$ and let $S \in \mathbb{Z}^{I \times J}_{\geq 0}$ be the slack matrix of $\mathrm{conv}(\mathcal{X})$ associated with the pair $(\mathcal{A}, \mathcal{X})$. The largest entry of the slack matrix is at most Δ. By Yannakakis's Theorem (Theorem 4) there exists a semidefinite factorization $(U_i, V^j)_{(i,j) \in I \times J} \subseteq \mathbb{S}^r_+ \times \mathbb{S}^r_+$ of S such that

$$\mathrm{conv}(\mathcal{X}) = \{x \in \mathbb{R}^n \,|\, \exists Y \in \mathbb{S}^r_+ : \, a_i^\intercal x + \langle U_i, Y\rangle = b_i \;\; \forall i \in I\}.$$

By Theorem 6 we may assume that $\|U_i\| \leq \sqrt{r\Delta}$ for every $i \in I$ and $\left\|V^j\right\| \leq \sqrt{r\Delta}$ for every $j \in J$. We will now pick a subsystem of maximum volume. For a linearly independent set of vectors $x_1, \ldots, x_k \in \mathbb{R}^n$, we let $\mathrm{vol}(\{x_1, \ldots, x_k\})$ denote the k-dimensional parallelepiped volume

$$\mathrm{vol}\left(\sum_{i=1}^k a_i x_i \,\middle|\, a_1, \ldots, a_k \in [0,1]\right) = \det((x_i^\intercal x_j)_{ij})^{\frac{1}{2}}.$$

If the vectors are dependent, then by convention the volume is zero. Let $\mathcal{W} = \mathrm{span}\left\{(a_i, U_i) \,\middle|\, i \in I\right\}$ and let $I' \subseteq I$ be a subset of size $|I'| = \dim(\mathcal{W})$ such that $\mathrm{vol}\left(\{(a_i, U_i) \,|\, i \in I'\}\right)$ is maximized. Note that $|I'| \leq n + r^2$.

For any positive semidefinite matrix $U \in \mathbb{S}^r_+$ with spectral decomposition

$$U = \sum_{k\in[r]} \lambda_k\, u_k u_k^\intercal, \quad \text{we let} \quad \bar{U} = \sum_{k\in[r]} \bar{\lambda}_k\, \bar{u}_k \bar{u}_k^\intercal$$

be the matrix where for every $k \in [r]$, the value of $\bar{\lambda}_k$ is the nearest integer multiple of δ/Δ to λ_k and $\bar{u}_k$ is the vector we get by rounding each of the entries of u_k to the nearest integer multiple of δ/Δ. Since each u_k is a unit vector, the matrices $u_k u_k^\intercal$ have entries in $[-1, 1]$ and it follows that U has entries in $r\,\|U\|\,[-1, 1]$. Similarly, since each $\bar{u}_k$ has entries in $(1 + \delta/\Delta)[-1, 1]$ each of the matrices $\bar{u}_k \bar{u}_k^\intercal$ has entries in $(1 + \delta/\Delta)^2[-1, 1]$, and it follows that $\bar{U}$ has entries in $r(\|U\| + \delta/\Delta)(1 + \delta/\Delta)^2[-1, 1]$. In particular, for every $i \in I'$, the entries of $\bar{U}_i$ are bounded in absolute value by

$$r\big(\,\|U_i\| + \delta/\Delta\big)(1 + \delta/\Delta)^2 \le r(\sqrt{r\Delta} + \delta/\Delta)(1 + \delta/\Delta)^2 \le 8r^{3/2}\sqrt{\Delta}.$$

We use the following simple claim.

Claim. Let U and $\bar{U}$ be as above. Then, $\|\bar{U} - U\|_2 \le 4\delta r^2/\sqrt{\Delta}$

Proof of claim: By the triangle inequality we have

$$\begin{aligned}
\big\|\bar{U} - U\big\|_F &= \Bigg\|\sum_{k\in[r]} \bar{\lambda}_k\, \bar{u}_k\bar{u}_k^\intercal - \lambda_k\, u_k u_k^\intercal\Bigg\|_F \\
&\le r \max_{k\in[r]} \big\|\bar{\lambda}_k\, \bar{u}_k\bar{u}_k^\intercal - \lambda_k\, u_k u_k^\intercal\big\|_F \\
&= r \max_{k\in[r]} \big\|(\bar{\lambda}_k - \lambda_k)\, \bar{u}_k\bar{u}_k^\intercal - \lambda_k(u_k u_k^\intercal - \bar{u}_k\bar{u}_k^\intercal)\big\|_F \\
&\le r \max_{k\in[r]} \frac{\delta}{\Delta}\, \big\|\bar{u}_k\bar{u}_k^\intercal\big\|_F + \sqrt{r\Delta}\,\big\|u_k u_k^\intercal - \bar{u}_k\bar{u}_k^\intercal\big\|_F \\
&= r \max_{k\in[r]} \frac{\delta}{\Delta} \bar{u}_k^\intercal \bar{u}_k + \sqrt{r\Delta}\,\big\|(u_k - \bar{u}_k)u_k^\intercal - \bar{u}_k(\bar{u}_k^\intercal - u_k^\intercal)\big\|_F \\
&\le r \max_{k\in[r]} \frac{\delta}{\Delta}\Big(1 + \frac{\delta}{\Delta}\sqrt{r}\Big)^2 + \sqrt{r\Delta}\Big(\,\|u_k - \bar{u}_k\|_F + \|\bar{u}_k\|_F\,\|u_k - \bar{u}_k\|_F\,\Big) \\
&\le r\frac{\delta}{\Delta}\Big(1 + \frac{\delta}{\Delta}\sqrt{r}\Big)^2 + r\sqrt{r\Delta}\Big(\frac{\delta}{\Delta}\sqrt{r} + \big(1 + \frac{\delta}{\Delta}\sqrt{r}\big)\frac{\delta}{\Delta}\sqrt{r}\Big) \\
&\le r \cdot 4\delta r/\sqrt{\Delta}.
\end{aligned}$$

The claim now follows from the fact that $\delta\sqrt{r}/\Delta < 1$. ♦

Define the set

$$\bar{\mathcal{X}} = \Big\{x \in \{0, 1\}^n \,\Big|\, \exists Y \in \mathbb{S}^r_+(\sqrt{r\Delta}) : \,\big|b_i - a_i^\intercal x - \langle \bar{U}_i, Y\rangle\big| \le \frac{1}{4(n + r^2)} \;\; \forall i \in I'\Big\}.$$

We claim that $\bar{\mathcal{X}} = \mathcal{X}$, which will complete the proof.

We will first show that $\mathcal{X} \subseteq \bar{\mathcal{X}}$. To this end, fix an index $j \in J$. By Theorem 4 we can pick $Y = V^j \in \mathbb{S}^r_+$ such that $a_i^\intercal x_j + \langle U_i, Y\rangle = b_i$ for every $i \in I'$. Moreover, $\|Y\| = \left\|V^j\right\| \le \sqrt{r\Delta}$. This implies that for every $i \in I'$, we have

$$\left|b_i - a_i^\intercal x_j - \langle \bar{U}_i, Y\rangle\right| = \big|\underbrace{b_i - a_i^\intercal x_j - \langle U_i, Y\rangle}_{0} + \langle \bar{U}_i - U_i, Y\rangle\big|$$
$$\le \left\|\bar{U}_i - U_i\right\|_F \|Y\|_F \le 4\delta r^3,$$

where the second line follows from the Cauchy-Schwarz inequality, the above claim, and $\|Y\|_F \le \sqrt{r}\,\|Y\| \le r\sqrt{\Delta}$. Now, since $4\delta r^3 \le 4r^3/(16r^3(n+r^2)) = 1/(4(n+r^2))$ we conclude that $x_j \in \bar{\mathcal{X}}$ and hence $\mathcal{X} \subseteq \bar{\mathcal{X}}$.

It remains to show that $\bar{\mathcal{X}} \subseteq \mathcal{X}$. For this we show that whenever $x \in \{0,1\}^n$ is such that $x \notin \mathcal{X}$ it follows that $x \notin \bar{\mathcal{X}}$. To this end, fix an $x \in \{0,1\}^n$ such that $x \notin \mathcal{X}$. Clearly $x \notin \operatorname{conv}(\mathcal{X})$ and hence, there must be an $i^* \in I$ such that $a_{i^*}^\intercal x > b_{i^*}$. Since x, a_{i^*} and b_{i^*} are integral we must in fact have $a_{i^*}^\intercal x \ge b_{i^*} + 1$. We express this violation in terms of the above selected subsystem corresponding to the set I'.

There exist unique multipliers $\nu \in \mathbb{R}^{I'}$ such that $\left(a_{i^*}, U_{i^*}\right) = \sum_{i \in I'} \nu_i (a_i, U_i)$. Observe that this implies that $\sum_{i\in I'} \nu_i b_i = b_{i^*}$; otherwise it would be impossible for $a_i^\intercal x + \langle U_i, Y\rangle = b_i$ to hold for every $i \in I$ and hence we would have $\mathcal{X} = \emptyset$ (which we assumed is not the case).

Using the fact that the chosen subsystem I' is volume maximizing and using Cramer's rule,

$$|\nu_i| = \frac{\operatorname{vol}\left(\{(a_t, U_t) \mid t \in I' \setminus \{i\} \cup \{i^*\}\}\right)}{\operatorname{vol}\left(\{(a_t, U_t) \mid t \in I'\}\right)} \le 1.$$

For any $Y \in \mathbb{S}^r_+(\sqrt{r\Delta})$ using $\langle U_{i^*}, Y\rangle \ge 0$ it follows thus

$$1 \le \left|a_{i^*}^\intercal x - b_{i^*} + \langle U_{i^*}, Y\rangle\right| = \left|\sum_{i\in I'} \nu_i (a_i^\intercal x - b_i + \langle U_i, Y\rangle)\right|$$
$$\le \sum_{i \in I'} |\nu_i| \left|a_i^\intercal x - b_i + \langle U_i, Y\rangle\right| \le (n + r^2) \max_{i\in I'} \left|a_i^\intercal x - b_i + \langle U_i, Y\rangle\right|.$$

Using a similar estimation as above, for every $i \in I'$, we have

$$\left|a_i^\intercal x - b_i + \langle U_i, Y\rangle\right| = |a_i^\intercal x - b_i + \langle \bar{U}_i, Y\rangle + \langle U_i - \bar{U}_i, Y\rangle|$$
$$\le |a_i^\intercal x - b_i + \langle \bar{U}_i, Y\rangle| + |\langle U_i - \bar{U}_i, Y\rangle|$$
$$\le |a_i^\intercal x - b_i + \langle \bar{U}_i, Y\rangle| + \frac{1}{4(n+r^2)}.$$

Combining this with $1 \le (n+r^2)\max_{i\in I'} \left|a_i^\intercal x - b_i + \langle U_i, Y\rangle\right|$ we obtain

$$\frac{1}{2(n+r^2)} \le \frac{1}{n+r^2} - \frac{1}{4(n+r^2)} \le \max_{i\in I'} \left|a_i^\intercal x - b_i + \langle \bar{U}_i, Y\rangle\right|,$$

and so $x \notin Y$.

Via padding with empty rows we can ensure that $|I'| = n + r^2$ as claimed. □

Using Lemma 2 we can establish the existence of 0/1 polytopes that do not admit any small semidefinite extended formulation following the proof of [Rothvoß, 2011, Theorem 4].

Theorem 7. *For any $n \in \mathbb{N}$ there exists $\mathcal{X} \subseteq \{0,1\}^n$ such that*

$$\mathrm{xc}_{SDP}(\mathrm{conv}\,(\mathcal{X})) = \Omega\left(\frac{2^{n/4}}{(n \log n)^{1/4}}\right).$$

Proof: Let $R := R(n) := \max_{\mathcal{X} \subseteq \{0,1\}^n} \mathrm{xc}_{\mathrm{SDP}}(\mathrm{conv}\,(\mathcal{X}))$ and suppose that $R(n) \leq 2^n$; otherwise the statement is trivial. The construction of Lemma 2 induces an injective map from $\mathcal{X} \subseteq \{0,1\}^n$ to systems $(a_i, U_i, b_i)_{i \in [n+r^2]}$ as the set $\mathcal{X}$ can be reconstructed from the system. Also, adding zero rows and columns to A, U and zero rows to b does not affect this property. Thus without loss of generality we assume that A is a $(n+R^2) \times n$ matrix, U is a $(n+R^2) \times R^2$ matrix (using $\frac{R(R+1)}{2} \leq R^2$). Furthermore, by Lemma 2, every value in U has absolute value at most Δ and can be chosen to be a multiple of $(16R^3(n+R^2))^{-1}\Delta^{-1}$. Thus each entry can take at most $3(16R^3(n+R^2))\Delta \cdot \Delta = \Delta^{2+o(1)}$ values, since $R \leq 2^n$ and $\Delta \geq n^{n/2}$. Furthermore, the entries of A, b are integral and have absolute value at most Δ, and hence each entry can take at most $3\Delta \leq \Delta^{2+o(1)}$ different values.

We shall now assume that $R \geq n$ (this will be justified by the lower bound on R later). By injectivity we cannot have more sets than distinct systems, i.e.

$$2^{2^n} - 1 \leq \Delta^{(2+o(1))(n+R^2+1)(n+R^2)} = \Delta^{(2+o(1))R^4} = 2^{(2+o(1))n \log n R^4}.$$

Hence for n large enough, $R \geq \frac{2^{n/4}}{(3n \log n)^{1/4}}$ as needed. □

5 On the semidefinite xc of polygons

In an analogous fashion to Fiorini et al. [2011] we can use a slightly adapted version of Theorem 2 to show the existence of a polygon with d integral vertices with semidefinite extension complexity $\Omega((\frac{d}{\log d})^{\frac{1}{4}})$. For this we change Theorem 2 to work for arbitrary polytopes with bounded vertex coordinates; the proof is almost identical to Theorem 2 and follows with the analogous changes as in Fiorini et al. [2011].

Lemma 3 (Generalized rounding lemma). *Let $n, N \geq 2$ be a positive integer and set $\Delta := ((n+1)N)^{2n}$. Let $\mathcal{V} \subseteq \mathbb{Z}^n \cap [-N,N]^n$ be a nonempty and convex independent set and $\mathcal{X} := \mathrm{conv}\,(\mathcal{V}) \cap \mathbb{Z}^n$. With $r := \mathrm{xc}_{SDP}(\mathrm{conv}\,(\mathcal{X}))$ and $\delta \leq \left(16r^3(n+r^2)\right)^{-1}$, for every $i \in [n+r^2]$ there exist:*

1. *an integer vector $a_i \in \mathbb{Z}^n$ such that $\|a_i\|_\infty \leq \Delta$,*
2. *an integer b_i such that $|b_i| \leq \Delta$,*
3. *a matrix $U_i \in \mathbb{S}^r_+(\sqrt{r\Delta})$ whose entries are integer multiples of δ/Δ and have absolute value at most $8r^{3/2}\Delta$, such that*

$$\mathcal{X} = \Big\{ x \in \mathbb{Z}^n \,\Big|\, \exists Y \in \mathbb{S}_+^r(\sqrt{r\Delta}) : \big| b_i - a_i^\intercal x - \langle Y, U_i \rangle \big| \leq \frac{1}{4(n+r^2)} \;\; \forall i \in [n+r^2] \Big\}.$$

Proof: By, e.g., [**?**, Lemma D.4.1] it follows that P has a non-redundant description with integral coefficients of largest absolute value of at most $((n+1)N)^n$. Thus the maximal entry occurring in the slack matrix is $((n+1)N)^{2n} = \Delta$. The proof follows now with a similar argument as in Theorem 2. □

We are ready to prove the existence of a polygon with d vertices, with integral coefficients, so that its semidefinite extension complexity is $\Omega((\frac{d}{\log d})^{\frac{1}{4}})$.

Theorem 8 (Integral polygon with high semidefinite xc). *For every $d \geq 3$, there exists a d-gon P with vertices in $[2d] \times [4d^2]$ and $\mathrm{xc}_{SDP}(P) = \Omega((\frac{d}{\log d})^{\frac{1}{4}})$.*

Proof: The proof is identical to the one is Fiorini et al. [2011] except for adjusting parameters as follows. The set $Z \coloneqq \{(z, z^2) \mid z \in [2d]\}$ is convex independent, thus every subset $X \subseteq Z$ of size $|X| = d$ yields a different convex d-gon. Let $R \coloneqq \max\{\mathrm{xc}_{\mathrm{SDP}} \operatorname{conv}(X) \mid X \subseteq Z, |X| = d\}$.

As in the proof of Theorem 7, we need to count the number of systems (which the above set of polygons map to in an injective manner). Using $\Delta = (12d^2)^2$, $n = 2$, $N = 4d^2$ by Lemma 3 it follows easily that each entry in the system can take at most cd^{14} different values. Without loss of generality, by padding with zeros, we assume that the system given by Lemma 3 has the following dimensions: the A, b part from (1.) and (2.), where A is formed by the rows a_i, is a $(3+R^2) \times 3$ matrix and U from (3.), formed by the U_i read as rows vectors, is a $(3+R^2) \times R^2$ matrix. We estimate

$$2^d \leq (cd^{14})^{(3+R^2)^2} \leq 2^{c' \cdot R^4 \cdot \log d}$$

and hence $R \geq c'(\frac{d}{\log d})^{\frac{1}{4}}$ for some constant $c' > 0$ follows. □

6 Final remarks

Most of the questions and complexity theoretic considerations in Rothvoß [2011] as well as the approximation theorem carry over immediately to our setting and the proofs follow similarly. For example, in analogy to [Rothvoß, 2011, Theorem 6], an approximation theorem for 0/1 polytopes can be derived showing that every semidefinite extended formulation for a 0/1 polytope can be approximated arbitrarily well by one with coefficients of bounded size.

The following important problems remain open:

Problem 1. Does the CUT polytope have high semidefinite extension complexity. We highly suspect that the answer is in the affirmative, similar to the linear case. However the partial slack matrix analyzed in Fiorini et al. [2012] to establish the lower bound for linear EFs has an efficient semidefinite factorization. In fact, it was precisely this fact that established the separation between semidefinite EFs and linear EFs in Braun et al. [2012].

Problem 2. Is there an information theoretic framework for lower bounding semidefinite rank similar to the framework laid out in Braverman and Moitra [2012], Braun and Pokutta [2013] for nonnegative rank?

Problem 3. As asked in Fiorini et al. [2011], we can ask similarly for semidefinite EFs: is the provided lower bound for the semidefinite extension complexity of polygons tight?

Acknowledgements

We are indebted to the anonymous referees for their remarks and the shortening of the proof of Lemma 1.

Bibliography

A. Ben-Tal and A. Nemirovski. On polyhedral approximations of the second-order cone. *Math. Oper. Res.*, 26:193–205, 2001. doi: 10.1287/moor.26.2.193.10561.

G. Braun and S. Pokutta. Common information and unique disjointness. *submitted*, 2013.

G. Braun, S. Fiorini, S. Pokutta, and D. Steurer. Approximation Limits of Linear Programs (Beyond Hierarchies). In *53rd IEEE Symp. on Foundations of Computer Science (FOCS 2012)*, pages 480–489, 2012. ISBN 978-1-4673-4383-1. doi: 10.1109/FOCS.2012.10.

M. Braverman and A. Moitra. An information complexity approach to extended formulations. *Electronic Colloquium on Computational Complexity (ECCC)*, 19(131), 2012.

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

S. Fiorini, T. Rothvoß, and H. Tiwary. Extended formulations for polygons, 2011. arXiv:1107.0371.

S. Fiorini, S. Massar, S. Pokutta, H. R. Tiwary, and R. de Wolf. Linear vs. Semidefinite Extended Formulations: Exponential Separation and Strong Lower Bounds. *Proceedings of STOC 2012*, 2012.

M. X. Goemans. Smallest compact formulation for the permutahedron. Manuscript, 2009.

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

F. John. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays presented to R. Courant on his 60th Birthday*, pages 187–204, 1948.

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

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

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

M. Yannakakis. Expressing combinatorial optimization problems by linear programs. *J. Comput. System Sci.*, 43(3):441–466, 1991. doi: 10.1016/0022-0000(91)90024-Y.

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