Approximate Extended Formulations: First Impossibility Results

Gábor Braun[1], Samuel Fiorini[2], Sebastian Pokutta[3], and David Steurer[4]

[1]Universität Leipzig, Institut für Informatik, PF 100920, 04009 Leipzig, Germany.
Email: gabor.braun@informatik.uni-leipzig.de
[2]Department of Mathematics, Université libre de Bruxelles CP 216, Bd. du Triomphe, 1050 Brussels, Belgium. *Email:* sfiorini@ulb.ac.be
[3]Friedrich-Alexander-Universität Erlangen-Nürnberg, Department of Mathematics, Office 03.045, Cauerstrasse 11, 91058 Erlangen, Germany. *Email:* sebastian.pokutta@math.uni-erlangen.de
[4]Department of Computer Science, Cornell University, Ithaca, NY 14853, United States.
Email: dsteurer@cs.princeton.edu

March 22, 2012

Abstract

The unique disjointness matrix is a partial matrix with rows and columns corresponding to subsets of a finite set of size n. The entries corresponding to pairs of disjoint sets equal 1, those corresponding to pairs of sets intersecting in exactly one element equal 0. The other entries are not defined. A famous lemma due to Razborov [Razborov, 1992] implies that the nondeterministic communication complexity of any matrix containing the unique disjointness matrix is $\Omega(n)$. In particular, the nonnegative rank of (any matrix containing) the unique disjointness matrix is $2^{\Omega(n)}$. Recent work on the fundamental limits of linear programming [Fiorini et al., 2012] has shown that the unique disjointness matrix can be embedded in the slack matrix of various polytopes such as the cut, stable set and TSP polytopes. Via Yannakakis's factorization theorem [Yannakakis, 1991], this implied super-polynomial lower bounds on the number of inequalities in any linear program expressing any of these polytopes.

We extend Razborov's lemma to deal with perturbations of the unique disjointness matrix. In particular, we show that the nonnegative rank of the unique disjointness matrix remains $2^{\Omega(n)}$ even if a constant is added to all its entries. Moreover, the nonnegative rank of the unique disjointness matrix remains super-polynomial as long as the perturbation is $O(n^{1/3-\varepsilon})$. We obtain from this implies a strong, unconditional polyhedral inapproximability result for a problem that we call the universal stable set problem: this problem has a polynomial size n-approximate extended formulation, but no polynomial size $O(n^{1/3-\varepsilon})$-approximate extended formulation. Finally, we prove strong limitations on how well polynomial-size linear programs can approximate semidefinite programs.

1 Introduction

1.1 Context

What has to go in: definition of EF, extension, unique disjointness problem, statement of Razborov's lemma. Definition of what we call a problem. Universal stable set problem. The necessity of taking a pair of polytopes rather than just one polytope.

$c^\intercal x$.

1.2 Previous Work

The proof of Fiorini et al. [2012] in a nutshell. Works on approximate extended formulations, starting with Arora-Bollobás-Lovász-(Tourlakis). Lots of bibliography here. In the linear world Sherali-Adams, Lovász Schrijver. Construction of approximate extended formulations by, e.g., Van Vyve and Wolsey. Cite open question in the survey of Conforti et al.?

1.3 Contribution

- first matrices with provably high nonnegative rank, where there are no zeroes;
- a framework for proving unconditional polyhedral inapproximability results (not conditioned on the fact that P is not NP), and first such result.

1.4 Outline

2 Pushing Razborov's Lemma Further

In this section, we establish a generalization of a famous lemma due to Razborov in the context of the disjointness problem (see Razborov [1992] or Kushilevitz and Nisan [1997, Lemma 4.49] for the original version). This generalization is well suited for obtaining lower bounds on the nonnegative rank of certain matrices related to the communication matrix of the disjointness problem.

Lemma 1. *For every $0 < p < 1$, $n \geq 3$ and $1 \leq \ell \leq (n+1)/4$ let μ be the following probability distribution: of random subsets a and b of size ℓ of $[n]$. We flip a biased coin. With probability p, we choose (a, b) uniformly among the pairs of subsets intersecting in exactly one element. With probability $1 - p$, we choose (a, b) uniformly among the pairs of disjoint subsets. Let the corresponding events be $A = \{a \cap b = \emptyset\}$ and $B = \{|a \cap b| = 1\}$.*

Now for every sequence of non-negative functions $f_1, g_1, \ldots, f_r, g_r$ defined on the subsets of $[n]$, we introduce a random variable $X := \sum_{i=1}^{r} f_i(a) g_i(b)$. Then for every $0 < \alpha < 1$

$$\mathbb{E}\left[XI_B\right] \geqslant \alpha \frac{p}{1-p} \mathbb{E}\left[XI_A\right] - rp \left\|XI_A\right\|_\infty 2^{-\frac{(1-\sqrt[3]{\alpha})^3}{16 \ln 2}\ell + O(\log \ell)}, \tag{1}$$

where $O(\log \ell)$ does not depend on n, α and p. Here I_A and I_B are the indicator of the events A and B.

For $r = 1$ and X the characteristic function of the rectangle $R = C \times D$ we recover a strengthened version of Razborov's Lemma.

We now turn on to prove Lemma 1. The proof is mostly a repetition of the version in [Kushilevitz and Nisan, 1997, Lemma 4.49] with improved constants and adapted to random variables. For completeness, we include the whole proof with a different presentation. It also corrects a minor mistake (namely, in the middle of Page 62, $\mathrm{Row}_1(T) \geq \mathrm{Row}_0(T)/3 - 2^{-\delta n}$ and $\mathrm{Col}_1(T) \geq \mathrm{Col}_0(T)/3 - 2^{-\delta n}$ do not imply $\mathrm{Row}_1(T)\,\mathrm{Col}_1(T) \geq (\mathrm{Row}_0(T)/3 - 2^{-\delta n})(\mathrm{Col}_0(T)/3 - 2^{-\delta n})$ when the factors on the right-hand side are negative).

Extending the previous notation I_A and I_B, we will write I_C for the indicator of C for every event C.

Proof of Lemma 1. In the following, all expectations are taken over a and b if not indicated otherwise. First we show that it suffices to consider the case $n = 4\ell - 1$. In general $n = 4\ell - 1 + k$ for some $k \geq 0$. We reconstruct μ as follows. Let H be a uniformly chosen random subset of size $4\ell - 1$ of $[n]$. Let us choose a and b as subsets of H with the distribution as in the Lemma. By reasons of symmetry, this gives the distribution μ of a, b. The claim of the Lemma applied to the set H gives

$$\mathbb{E}\left[XI_B \mid H\right] \geqslant \alpha \frac{p}{1-p} \mathbb{E}\left[XI_A \mid H\right] - rp \left\| XI_A \restriction H \right\|_\infty 2^{-\frac{(1-\sqrt[3]{\alpha})^3}{16 \ln 2}\ell + O(\log \ell)},$$

with $X \restriction H \coloneqq XI_{\{(a,b) \mid a,b \subseteq H\}}$. Taking expected value of both sides establishes the Lemma. Hence it is enough to prove the lemma for $n = 4\ell - 1$

Second, we redefine the distribution μ in an alternative fashion. Let $q = \sqrt{p}$, and T be a uniformly chosen partition of $[n]$ into 2 subsets T_1, T_2 with $2\ell - 1$ elements and a singleton $\{i\}$. Given T we choose a and b independently as subsets with ℓ elements. We flip a biased coin to decide whether i is an element of a. With probability q, we select a as a uniform random subset of $T_1 \cup \{i\}$ of size ℓ containing $\{i\}$. With probability $1 - q$, we choose a as a uniform random subset of T_1 of size ℓ. We choose b similarly by using T_2 instead of T_1. By reasons of symmetry, this distribution is exactly μ. Note that a and b intersect if and only if they both contain i, therefore

$$\mu(B) = q^2 = p, \qquad \mu(A) = 1 - q^2 = 1 - p. \tag{2}$$

We are ready to prove equation (1). It is easy to reduce the statement to the $r = 1$ case. When established for $r = 1$ we obtain the general case by adding up (1) with $f_i(a)g_i(b)$ instead of X, and then use the estimate $\|f_i(a)g_i(b) \cdot I_A\|_\infty \leq \|X \cdot I_A\|_\infty$. Hence we shall assume $r = 1$ and suppress the subscript 1, i.e., $f = f_1$, $g = g_1$ and $X = f(a)g(b)$.

Now $\mathbb{E}\left[f(a)g(b) \cdot I_A\right]$ and $\mathbb{E}\left[f(a)g(b) \cdot I_B\right]$ are conveniently expressed using the following values, which do not depend on p:

$$\mathrm{Row}_0(T) \coloneqq \mathbb{E}\left[f(a) \mid T, i \notin a\right], \qquad \mathrm{Row}_1(T) \coloneqq \mathbb{E}\left[f(a) \mid T, i \in a\right], \tag{3}$$

$$\mathrm{Col}_0(T) \coloneqq \mathbb{E}\left[g(b) \mid T, i \notin b\right], \qquad \mathrm{Col}_1(T) \coloneqq \mathbb{E}\left[g(b) \mid T, i \in b\right]. \tag{4}$$

First we note that $\mathrm{Col}_0(T)$ depends only on T_2, and similarly $\mathrm{Row}_0(T)$ depends only on T_1. Moreover, the conditional distribution of a is the same for condition T_2, condition $T_2 \wedge i \in a$ and condition $T_2 \wedge i \notin a$: namely, the uniform distribution of ℓ-element subsets of $[n] \setminus T_2$. Therefore

$$\mathbb{E}\left[\mathrm{Row}_0(T) \mid T_2\right] = \mathbb{E}\left[\mathrm{Row}_1(T) \mid T_2\right], \tag{5}$$

$$\mathbb{E}\left[\mathrm{Col}_0(T) \mid T_1\right] = \mathbb{E}\left[\mathrm{Col}_1(T) \mid T_1\right], \tag{6}$$

especially

$$\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)\right] = \mathbb{E}\left[\mathrm{Row}_1(T)\,\mathrm{Col}_0(T)\right] = \mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_1(T)\right]. \tag{7}$$

These lead to (c.f., [Kushilevitz and Nisan, 1997, Claim 4.52])

$$\begin{aligned}
\mathbb{E}\left[f(a)g(b)I_A\right] = {} & (1-q)^2\, \mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)\right] + q(1-q)\, \mathbb{E}\left[\mathrm{Row}_1(T)\,\mathrm{Col}_0(T)\right] \\
& + (1-q)q\, \mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_1(T)\right] \\
= {} & (1-q^2)\, \mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)\right],
\end{aligned} \tag{8}$$

$$\mathbb{E}\left[f(a)g(b)I_B\right] = q^2\, \mathbb{E}\left[\mathrm{Row}_1(T)\,\mathrm{Col}_1(T)\right]. \tag{9}$$

Hence the claimed (1) for $f(a)g(b)$ reduces to

$$\mathbb{E}\left[\mathrm{Row}_1(T)\,\mathrm{Col}_1(T)\right] \geq \alpha\,\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)\right] - 2^{-\frac{(1-\sqrt[3]{\alpha})^3}{16\ln 2}\ell + O(\log \ell)} \tag{10}$$

We will now prove equation (10). For this we will decompose the above terms based on properties of the partition and we will bound each part separately. Note that, at this point, p is essentially eliminated from the claim. However, for convenience we set $q = 1/2$ (i.e., $p = 1/4$), which gives a nice interpretation of $\mathrm{Row}_0(T) + \mathrm{Row}_1(T)$ and $\mathrm{Col}_0(T) + \mathrm{Col}_1(T)$:

$$\mathbb{E}\left[f(a) \mid T\right] = \frac{\mathrm{Row}_0(T) + \mathrm{Row}_1(T)}{2},$$
$$\mathbb{E}\left[g(b) \mid T\right] = \frac{\mathrm{Col}_0(T) + \mathrm{Col}_1(T)}{2}.$$

These values depend only on T_2 and T_1, respectively.

Let $\beta < 1$ and $\delta > 0$ be constants. (We will see later how to choose β and δ. Essentially δ is the coefficient of ℓ in the exponent.) Let $\mathrm{Bad}_a(T)$ denote the event that the partition T is *a-bad*, i.e.,

$$\mathrm{Row}_1(T) \leq \beta\,\mathrm{Row}_0(T) \quad \text{and} \quad \mathbb{E}\left[f(a) \mid T\right] > 2^{-\delta\ell-1}\left\| f \restriction ([n] \setminus T_2)\right\|_\infty. \tag{11}$$

We will now show that $\mathbb{P}\left[\mathrm{Bad}_a(T) \mid T_2\right]$ is small when δ is small which is the main part of the later error estimation. To prove this for a fixed T_2, as $\mathbb{E}\left[f(a) \mid T\right]$ depends only on T_2, we may assume $\mathbb{E}\left[f(a) \mid T\right] > 2^{-\delta\ell-1}\left\| f \restriction ([n] \setminus T_2)\right\|_\infty$. In particular $\mathbb{E}\left[f(a) \mid T\right]$ is positive.

Note that

$$\mathbb{E}\left[f(a) \mid T_2\right] = \mathbb{E}\left[f(a) \mid T\right] = \frac{1}{\binom{2\ell}{\ell}} \sum_{\substack{x \subseteq [n] \setminus T_2 \\ |x| = \ell}} f(x),$$

hence we can define s as a random ℓ-element subset of $[n] \setminus T_2$ with distribution

$$\mathbb{P}\left[s = x \mid T_2\right] = \frac{f(x)}{\binom{2\ell}{\ell}\,\mathbb{E}\left[f(a) \mid T_2\right]} \leq \frac{2^{\delta\ell+1}}{\binom{2\ell}{\ell}}$$

Then

$$\mathrm{Row}_0(T) = \frac{\mathbb{E}\left[f(a) I_{i \notin a} \mid T\right]}{\mathbb{P}\left[i \notin a \mid T\right]} = 2\,\mathbb{E}\left[f(a) I_{i \notin a} \mid T\right] = 2\frac{\mathbb{P}\left[i \notin s \mid T\right]}{\mathbb{E}\left[f(a) \mid T_2\right]}, \tag{12}$$

$$\mathrm{Row}_1(T) = \frac{\mathbb{E}\left[f(a) I_{i \in a} \mid T\right]}{\mathbb{P}\left[i \in a \mid T\right]} = 2\,\mathbb{E}\left[f(a) I_{i \in a} \mid T\right] = 2\frac{\mathbb{P}\left[i \in s \mid T\right]}{\mathbb{E}\left[f(a) \mid T_2\right]}. \tag{13}$$

In particular, if T is a-bad, i.e., $\mathrm{Row}_1(T) \leq \beta\,\mathrm{Row}_0(T)$, then $\mathbb{P}\left[i \in s \mid T\right] \leq \beta/(1+\beta)$.

We now estimate the entropy of s. On the one hand, we clearly have

$$\begin{aligned} H\left(s \mid T_2\right) &= \sum_x \mathbb{P}\left[s = x \mid T_2\right] \log \frac{1}{\mathbb{P}\left[s = x \mid T_2\right]} \\ &\geq \sum_x \mathbb{P}\left[s = x \mid T_2\right] \log \frac{\binom{2\ell}{\ell}}{2^{\delta\ell+1}} = \log \frac{\binom{2\ell}{\ell}}{2^{\delta\ell+1}} = 2\ell\left(1 - \frac{\delta}{2} - O\left(\frac{\log \ell}{\ell}\right)\right). \end{aligned} \tag{14}$$

On the other hand, we obtain

$$
\begin{aligned}
H\left(s \mid T_2\right) &\leq \sum_{j\in[n]\setminus T_2} H\left(I_{j\in s} \mid T_2\right) \\
&\leq 2\ell\left(\mathbb{P}\left[\mathrm{Bad}_a(T) \mid T_2\right] \cdot H\left(\frac{\beta}{1+\beta}\right) + \left(1-\mathbb{P}\left[\mathrm{Bad}_a(T) \mid T_2\right]\right) \cdot 1\right) \\
&= 2\ell\left(1-\mathbb{P}\left[\mathrm{Bad}_a(T) \mid T_2\right]\left(1-H\left(\frac{\beta}{1+\beta}\right)\right)\right). \qquad (15)
\end{aligned}
$$

This implies

$$
\mathbb{P}\left[\mathrm{Bad}_a(T) \mid T_2\right] \leq \frac{\frac{\delta}{2}+O\left(\frac{\log \ell}{\ell}\right)}{1-H\left(\beta/(1+\beta)\right)} =: \gamma, \qquad (16)
$$

which is small when δ is small with respect to β. Note that γ does not depend on T_2. (This is an improvement to [Kushilevitz and Nisan, 1997, Claim 4.50].)

As a consequence, we derive

$$
\mathbb{E}\left[\mathrm{Row}_0(T) I_{\mathrm{Bad}_a(T)} \,\middle|\, T_2\right] \leq \mathbb{P}\left[\mathrm{Bad}_a(T) \mid T_2\right]\left(\mathrm{Row}_0(T)+\mathrm{Row}_1(T)\right) \leq 2\gamma\, \mathbb{E}\left[\mathrm{Row}_0(T) \mid T_2\right], \qquad (17)
$$

hence (c.f., [Kushilevitz and Nisan, 1997, Claim 4.51]) with $\mathrm{Col}_0(T)$ being fixed when conditioning over T_2

$$
\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) I_{\mathrm{Bad}_a(T)} \,\middle|\, T_2\right] \leq 2\gamma\, \mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) \mid T_2\right], \qquad (18)
$$

$$
\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) I_{\mathrm{Bad}_a(T)}\right] \leq 2\gamma\, \mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)\right]. \qquad (19)
$$

Similarly

$$
\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) I_{\mathrm{Bad}_b(T)}\right] \leq 2\gamma\, \mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)\right]. \qquad (20)
$$

Let $\mathrm{small}(T)$ be the event that either $\mathbb{E}\left[f(a) \mid T\right] \leq 2^{-\delta\ell-1}\left\|f \restriction ([n]\setminus T_2)\right\|_\infty$ or $\mathbb{E}\left[g(b) \mid T\right] \leq 2^{-\delta\ell-1}\left\|g \restriction ([n]\setminus T_1)\right\|_\infty$ occurs. Obviously, the first inequality implies

$$
\begin{aligned}
\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) &\leq 2\,\mathbb{E}\left[f(a) \mid T\right]\mathbb{E}\left[g(b) \mid T, i \notin b\right] \\
&\leq 2^{-\delta\ell}\left\|f \restriction ([n]\setminus T_2)\right\|_\infty \left\|g \restriction T_2\right\|_\infty \leq 2^{-\delta\ell}\left\|f(a)g(b)\cdot I_A\right\|_\infty,
\end{aligned}
$$

as f and g are only considered on disjoint sets here. An analogous argument results in the same estimation when $\mathbb{E}\left[g(b) \mid T\right] \leq 2^{-\delta\ell-1}\cdot\left\|g \restriction ([n]\setminus T_1)\right\|_\infty$. Thus

$$
\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) I_{\mathrm{small}(T)}\right] \leq \left\|f(a)g(b)\cdot I_A\right\|_\infty 2^{-\delta\ell} \qquad (21)
$$

Finally, let $\mathrm{good}(T)$ be the event $\mathrm{Row}_1(T) \geq \beta\,\mathrm{Row}_0(T)$ and $\mathrm{Col}_1(T) \geq \beta\,\mathrm{Col}_0(T)$, so

$$
\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) I_{\mathrm{good}(T)}\right] \leq \frac{1}{\beta^2}\,\mathbb{E}\left[\mathrm{Row}_1(T)\,\mathrm{Col}_1(T)\right]. \qquad (22)
$$

Adding (19), (20), (21) and (22), we get

$$
\begin{aligned}
\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)\right] &\leq \mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) I_{\mathrm{good}(T)}\right] + \mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) I_{\mathrm{small}(T)}\right] \\
&\quad + \mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) I_{\mathrm{Bad}_a(T)}\right] + \mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) I_{\mathrm{Bad}_b(T)}\right] \\
&\leq \frac{1}{\beta^2}\,\mathbb{E}\left[\mathrm{Row}_1(T)\,\mathrm{Col}_1(T)\right] + \left\|f(a)g(b)\cdot I_A\right\|_\infty 2^{-\delta\ell} + 4\gamma\,\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)\right]. \qquad (23)
\end{aligned}
$$

We conclude

$$\mathbb{E}\left[\mathrm{Row}_1(T)\,\mathrm{Col}_1(T)\right] \geq \beta^2\left((1-4\gamma)\,\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)\right] - \|f(a)g(b)\cdot I_A\|_\infty 2^{-\delta\ell}\right)$$
$$> \beta^2(1-4\gamma)\,\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)\right] - \|f(a)g(b)\cdot I_A\|_\infty 2^{-\delta\ell}. \quad (24)$$

We now choose the constants β and δ to reduce this to (10). Therefore we require $\alpha = \beta^2(1-4\gamma)$, which expresses δ in terms of α and β (see equation (16)):

$$\delta + O\left(\frac{\log\ell}{\ell}\right) = \frac{(1-\alpha/\beta^2)(1-H\left(\beta/(1+\beta)\right))}{2}. \quad (25)$$

To estimate this expression, we use the Taylor expansion of the binary entropy function at $1/2$:

$$H(x) = 1 - (2/\ln 2)\left(x-\frac{1}{2}\right)^2 - \frac{\xi - 1/2}{12(\xi(1-\xi))^2\ln 2}\left(x-\frac{1}{2}\right)^3 \quad (26)$$

The last term is non-negative, hence

$$1 - H(x) \geq (2/\ln 2)\left(x-\frac{1}{2}\right)^2.$$

Combining it with (25) we obtain

$$\delta + O\left(\frac{\log\ell}{\ell}\right) \geq \frac{(1-\alpha/\beta^2)(\frac{2}{\ln 2}(\beta/(1+\beta)-1/2)^2}{2} = \frac{(1-\alpha/\beta^2)(1-\beta)^2}{4(1+\beta)^2\ln 2} \geq \frac{(1-\alpha/\beta^2)(1-\beta)^2}{16\ln 2}.$$

As easily seen, the denominator attains its maximum at $\beta = \sqrt[3]{\alpha}$, hence we make this choice and obtain:

$$\delta \geq \frac{(1-\sqrt[3]{\alpha})^3}{16\ln 2} - O\left(\frac{\log\ell}{\ell}\right). \quad (27)$$

This and $\alpha = \beta^2(1-4\gamma)$ reduce (24) to (10). □

3 Approximate Extended Formulations

3.1 The Basics

We begin with a lemma that characterizes the extension complexity of any polytope sandwiched between two given polytopes. The lemma generalizes Yannakakis' factorization theorem, which concerns the case $P = Q$.

Lemma 2. *Let $P := \mathrm{conv}\left(\{v_j \mid j \in [n]\}\right)$ and $Q := \{x \in \mathbb{R}^d \mid A_i x \leqslant b_i,\ i \in [m]\}$ be polytopes such that $P \subseteq Q$. Then for every polytope K with $P \subseteq K \subseteq Q$ it holds*

$$\mathrm{xc}(K) \geqslant \mathrm{rank}_+(S^{P,Q})$$

with $S^{P,Q}_{ij} = b_i - A_i v_j$ with $i \in [m]$ and $j \in [n]$. Moreover, there exists a polytope K such that $P \subseteq K \subseteq Q$ and

$$\mathrm{xc}(K) \leqslant \mathrm{rank}_+(S^{P,Q}).$$

Proof (sketch). There exists a slack matrix S^K of K that has $S^{P,Q}$ as a submatrix. The first part of the lemma then follows from the two following facts: (i) the extension complexity of K is the nonnegative rank of *any* of its slack matrices, thus $\mathrm{xc}(K) = \mathrm{rank}_+(S^K)$; (ii) the nonnegative rank of submatrix $S^{P,Q}$ is at most the nonnegative rank of matrix S^K, thus $\mathrm{rank}_+(S^K) \geqslant \mathrm{rank}_+(S^{P,Q})$. Combining both inequalities, we find that $\mathrm{xc}(K) \geqslant \mathrm{rank}_+(S^{P,Q})$.

For the second part of the lemma, consider any rank-r nonnegative factorization $S^{P,Q} = TU$ with $r = \mathrm{rank}_+(S^{P,Q})$ and the polytope $L = \{(x,y) \in \mathbb{R}^{d+r} \mid Ax + Ty = b,\ y \geqslant 0\}$. Then, as can be easily verified, defining K to be the projection of L to the x-space gives the desired polytope. □

Consider polytopes $P, Q \subseteq \mathbb{R}^d$ such that $0 \in P \subseteq Q$, and $\rho \geqslant 1$. Then $\rho Q := \{\rho x \mid x \in Q\}$. A polytope $L \subseteq \mathbb{R}^e$ such that there exists a linear projection $\pi : \mathbb{R}^e \to \mathbb{R}^d$ with

$$P \subseteq \pi(L) \subseteq \rho Q$$

is called a *ρ-approximate extended formulation* of P w.r.t. Q. The *size* of L is defined as the number of facets of L. We define $\mathrm{xc}_\rho(P,Q)$ as the minimum size of a ρ-approximate extended formulation of P w.r.t. Q. Then, we have the following straightforward characterization.

Lemma 3. *Let $P, Q \subseteq \mathbb{R}^d$ be polytopes with $0 \in P \subseteq Q$. Then $L \subseteq \mathbb{R}^e$ together with $\pi : \mathbb{R}^e \to \mathbb{R}^d$ define a ρ-approximate extended formulation of P w.r.t. Q if and only if*

$$\max\{f(x) \mid x \in P\} \leqslant \max\{f(\pi(y)) \mid y \in L\} \leqslant \rho \max\{f(x) \mid x \in Q\}$$

for every linear objective function f. In the second inequality, it suffices to restrict to the functions f which yield facets of Q.

Intuitively, the vertices of P encode the feasible solutions of the problem we are considering and the facets of Q encode the admissible linear objective functions. Often, it will be the case that Q is defined by considering some set of $\mathcal{F}$ of objective functions and letting Q be the set of points $x \in \mathbb{R}^d$ such that $f(x) \leqslant \max\{f(x) \mid x \in P\}$ for all $f \in \mathcal{F}$. Notice that some care has to be taken in the choice of $\mathcal{F}$ to ensure that the resulting Q is indeed a polytope.

Now suppose again that $P = \mathrm{conv}\,(v_j \mid j \in [n])$ and $Q = \{x \in \mathbb{R}^d \mid A_i x \leqslant b_i,\ i \in [m]\}$. Then $\rho Q = \{x \in \mathbb{R}^d \mid A_i x \leqslant \rho b_i,\ i \in [m]\}$ and the slack matrix of the pair $P, \rho Q$ is related to the slack matrix of the pair P, Q in the following way:

$$S^{P,\rho Q}_{ij} = \rho b_i - A_i v_j = (\rho - 1) b_i + b_i - A_i v_j = S^{P,Q}_{ij} + (\rho - 1) b_i.$$

From Lemma 2, we obtain the following result.

Theorem 4. *Let P, Q be polytopes such that $0 \in P \subseteq Q \subseteq \mathbb{R}^d$, and let $\rho \geqslant 1$. Consider any slack matrix $S^{P,Q}$ for the pair $P, \rho Q$ and the corresponding slack matrix $S^{P,\rho Q}$ for the pair $P, \rho Q$. Then we have*

$$\mathrm{xc}_\rho(P,Q) = \mathrm{rank}_+(S^{P,\rho Q}).$$

3.2 The Case of the Correlation and Cut Polytopes

Notice that $\mathrm{CUT}(n)$ is linearly equivalent to $\mathrm{COR}(n-1)$, thus much of what holds for one polytope in terms of approximate extended formulations also holds for the other. We begin with a negative result that explains why it is sometimes necessary to consider pairs of polytopes P, Q rather than having $P = Q$ always. Intuitively, the problem comes from the fact that, because 0 is a vertex of the cut polytope, every approximate extended formulation necessarily "captures" all facets of the cut polytope incident to 0. But these facets define the cut cone, which has high extension complexity. (Although we have not defined formally the extension complexity for cones, the reader should have no difficulty figuring out the definition because it is very similar to that for polytopes.)

Theorem 5. *For every $\rho \geqslant 1$, every ρ-approximate extended formulation of the cut polytope* $\mathrm{CUT}(n)$ *has $2^{\Omega(\sqrt{n})}$ size (here $P = Q = \mathrm{CUT}(n)$). More precisely, disregarding the value of $\rho \geqslant 1$, we have* $\mathrm{xc}_\rho(\mathrm{CUT}(n), \mathrm{CUT}(n)) = 2^{\Omega(\sqrt{n})}$.

Proof (sketch). Let $L := \{(x,y) \mid Ex + Fy \leqslant g\}$ denote a minimum size ρ-approximate extended formulation of $\mathrm{CUT}(n)$. Then $L' := \{(x,y,\lambda) \mid Ex + Fy \leqslant \lambda g, \lambda \geqslant \mathbf{0}, x \leqslant \mathbf{1}\}$ is an extended formulation of the *multicut polytope* $\mathrm{MULTICUT}(n)$. But then the size of L' is at least $2^{\Omega(\sqrt{n})}$, and so is the size of L. This is due to the fact that, for every graph G with $n-1$ vertices, there exists a face of $\mathrm{MULTICUT}(n)$ that projects to $\mathrm{STAB}(G)$. Thus we have $\mathrm{xc}(\mathrm{MULTICUT}(n)) = 2^{\Omega(\sqrt{n})}$. ☐

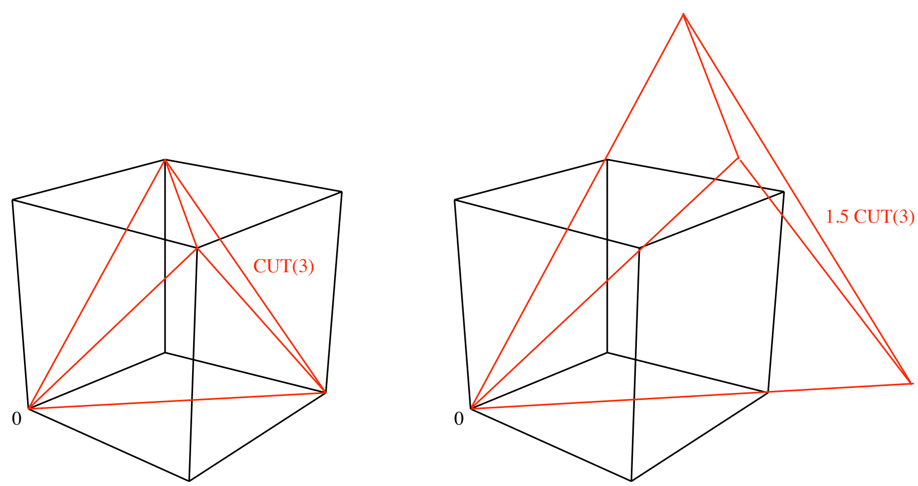


Figure 1: $\mathrm{CUT}(3)$ and a dilate $\rho\,\mathrm{CUT}(3)$ for $\rho = 3/2$.

Let $\rho \geqslant 1$. (The actual value of ρ will be set later.) Now let $P := \mathrm{COR}(n)$ and

$$Q := \{x \in \mathbb{R}^{n\times n} \mid \langle 2\,\mathrm{diag}(a) - aa^T, x\rangle \leqslant 1,\ a \in \{0,1\}^n\}.$$

Then $P \subseteq Q$. Denoting by $M = M(\rho)$ the slack matrix of the pair $P, \rho Q$, we have

$$M_{ab} = S_{ab}^{P,\rho Q} = (1 - a^T b)^2 + \rho - 1.$$

We would like to prove that the nonnegative rank of M is large, at least when ρ is chosen in a certain range, which via Theorem 5 implies that every polytope sandwiched between $P = \mathrm{COR}(n)$ and ρQ has large extension complexity, that is, $\mathrm{COR}(n)$ has no polynomial-size ρ-approximate extended formulation w.r.t. Q. The lower bound on the nonnegative rank of M follows directly from Lemma 1.

Theorem 6. *Let $\rho \geqslant 1$, $P = \mathrm{COR}(n)$ and $Q = \{x \in \mathbb{R}^{n\times n} \mid \langle 2\,\mathrm{diag}(a) - aa^T, x\rangle \leqslant 1,\ a \in \{0,1\}^n\}$. Then*

(i) If ρ is a fixed constant, then $\mathrm{xc}_\rho(P,Q) = 2^{\Omega(n)}$.

(ii) If $\rho = O(n^\beta)$ for some constant $\beta < 1/3$ then $\mathrm{xc}_\rho(P,Q) = \Omega(n^{-\beta})\, 2^{\Omega(n^{1-3\beta})}$.

Proof. Regarding the $2^n \times 2^n$ matrix $M = M(\rho)$ as a random variable over $2^{[n]} \times 2^{[n]}$, we apply Lemma 1 to $X := M$.

Assume that M has a rank-r nonnegative factorization. This means that X can be written as $X(a,b) = \sum_{i=1}^r f_i(a) g_i(b)$ where f_i and g_i are nonnegative functions defined over $[2^n]$ and $i = 1, \ldots, r$. Note that $MI_A = \rho I_A$ and $MI_B = (\rho - 1) I_B$. Therefore, Equation (1) reduces to

$$p(\rho - 1) \geqslant \alpha \frac{p}{1-p}(1-p) \cdot \rho - rp \cdot \rho \cdot 2^{-\frac{(1-\sqrt[3]{\alpha})^3}{64 \ln 2} n + O(\log n)}$$

which gives the lower bound

$$r \geqslant \left(\alpha - 1 + \frac{1}{\rho}\right) 2^{\frac{(1-\sqrt[3]{\alpha})^3}{64 \ln 2} n - O(\log n)}.$$

If ρ is constant, this last expression is $2^{\Omega(n)}$ provided α is chosen sufficiently close to 1. This proves part (i) of the theorem.

If $\rho \leqslant C n^\beta$ for some positive constant C, then we take $\alpha = 1 - \frac{1}{2C} n^{-\beta}$. Thus

$$\alpha - 1 + \frac{1}{\rho} \geqslant \frac{1}{C} n^{-\beta} = \Omega(n^{-\beta})$$

and

$$1 - \sqrt[3]{\alpha} \approx \frac{1-\alpha}{3} = \frac{1}{6C} n^{-\beta} = \Omega(n^{-\beta}).$$

This leads to the lower bound

$$r \geqslant \Omega(n^{-\beta}) 2^{\Omega(n^{1-3\beta})}$$

claimed in part (ii) of the theorem. □

3.3 Polyhedral Inapproximability of the Universal Stable Set Problem

The class of admissible objective functions encoded by the polytope Q considered in the previous section is not algorithmically interesting, because optimizing $f(x) = \langle 2\,\mathrm{diag}(a) - aa^T, x\rangle$ for $a \in \{0,1\}^n$ over $P = \mathrm{COR}(n)$ is trivial (if $a \neq \mathbf{0}$, the optimum value is always 1 and $x = e_i e_i^T$ is an optimum solution whenever $a_i \neq 0$, here e_i is the ith unit vector). We now define a larger polytope $\tilde{Q}$ that is more interesting algorithmically.

Given a graph G with vertex set $[n]$ and adjacency matrix A, we define $f_G(x) = \langle I - A, x\rangle$. Then $\max\{f_G(x) \mid x \in \mathrm{COR}(n)\} = \alpha(G)$, the stability number of G. We define $\tilde{Q}$ as follows:

$$\tilde{Q} := \{x \in \mathbb{R}^{n \times n} \mid f_G(x) \leqslant \alpha(G),\ G \text{ graph with vertex set } [n]\}.$$

Optimizing an admissible objective function $f_G(x)$ over $P = \mathrm{COR}(n)$ amounts to finding a maximum stable set of G. We say that the pair $P, \tilde{Q}$ defines the *universal stable set problem*. Clearly, any ρ-approximate extended formulation of P w.r.t. $\tilde{Q}$ is a ρ-approximate extended formulation of P w.r.t. Q, because $\rho Q \subseteq \rho \tilde{Q}$ for every $\rho \geqslant 1$. We obtain the following result.

Theorem 7. *The universal stable problem admits a polynomial size n-approximate extended formulation but no polynomial size $n^{1/3-\epsilon}$-approximate extended formulation, for $\epsilon > 0$.*

Proof. For the first part, simply take L to be the 0/1-cube of appropriate dimension (the projection is the identity map). The second part follows directly from Theorem 6. □

4 SOCP Extended Formulations

We will now use the approximation of the SOCP cone given in Ben-Tal and Nemirovski [2001] to show that every SOCP-extended formulation of the the correlation polytope has to have super-polynomial size.

In Ben-Tal and Nemirovski [2001] the following result was proven:

Theorem 8 (SOCP approximation)**.** *Let*

$$P = \left\{ x \in \mathbb{R}^n \;\middle|\; Ax \leq b, \|x^1\|_2 \leq x_0^1, \ldots, \|x^\ell\|_2 \leq x_0^\ell \right\}$$

where the x_0^i, x^i are parts of the x-variables, and $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $\bigcup_{i \in [\ell]} (\operatorname{supp}(x^i) \cup \{x_0^i\}) = [n]$. Let $c = \sum_{i \in [\ell]} \dim(x^i)$, then for any $\epsilon > 0$, there exists a polytope

$$Q = \left\{ (x, u)^T \in \mathbb{R}^r \;\middle|\; W(x, u)^T = d, (x, u)^T \geq 0 \right\}$$

with $W \in \mathbb{R}^{k \times r}$ and $d \in \mathbb{R}^k$ such that

1. *$k, r \leq O(1)(n + m + c) \ln(\frac{1}{\epsilon})$;*
2. *$P \subseteq \pi(Q) \subseteq (1 + \epsilon)_{SOCP} P$ with*

$$(1 + \epsilon)_{SOCP} P \coloneqq \left\{ x \in \mathbb{R}^n \;\middle|\; Ax \leq b, \|x^1\|_2 \leq (1 + \epsilon) x_0^1, \ldots, \|x^\ell\|_2 \leq (1 + \epsilon) x_0^\ell \right\},$$

where π is the orthogonal projection on the variables x.

We refer to the term $n + m + c$ as the *size of the SOCP-formulation*. In the following we will consider a special subclass of SOCPs. Let QP_0 denote those SOCPs S with fixed $x_0^i = r^i$ and $0 \in S$ (after substituting $x_0^i = r^i$). The fixing of the x_0^i is effectively achieved by inequalities in the system $Ax \leq b$ and can be substituted out subsequently. Note that we assume 0 being in the feasible region as solely requiring a *preimage* of 0 in the feasible region is not sufficient (see Example 1). We will first establish a simple lemma for QP_0.

Lemma 9. *Let S be a QP_0 and π be a linear map. Then for any $\tau \geq 0$ it holds*

$$\pi\left((1 + \tau)_{SOCP} S\right) \subseteq (1 + \tau)\pi(S) = \pi\left((1 + \tau)S\right).$$

Proof. We have

$$(1 + \tau)_{SOCP} S = \left\{ x \in \mathbb{R}^n \;\middle|\; Ax \leq b, \|x^1\|_2 \leq (1 + \tau) r^1, \ldots, \|x^\ell\|_2 \leq (1 + \tau) r^\ell \right\}$$

and

$$(1 + \tau) S = \left\{ x \in \mathbb{R}^n \;\middle|\; Ax \leq (1 + \tau) b, \|x^1\|_2 \leq (1 + \tau) r^1, \ldots, \|x^\ell\|_2 \leq (1 + \tau) r^\ell \right\}.$$

Let $z \in \pi\left((1 + \tau)_{SOCP} S\right)$. Then there exists $x_0 \in (1 + \tau)_{SOCP} S$ with $\pi(x_0) = z$. Clearly, x_0 satisfies the quadratic constraints of $(1 + \tau)S$ and thus it remains to show that $Ax_0 \leq (1 + \tau)b$ holds. For this observe that $(1 + \tau)x_0$ satisfies $A(1 + \tau)x_0 \leq (1 + \tau)b$ as this is equivalent to $Ax_0 \leq b$ and $x_0 \in (1 + \tau)_{SOCP} S$. Also, $A0 \leq b$ holds as $0 \in S$ and so $A0 \leq (1 + \tau)b$ as well. We write $x_0 = \frac{1}{1+\tau}(1 + \tau)x_0 + (1 - \frac{1}{1+\tau})0$ and so $Ax_0 \leq (1 + \tau)b$ follows. □

Using Theorem 8 we obtain the following theorem

Theorem 10. *Let P be a polytope with $0 \in P$ and $\mathrm{xc}_{\epsilon(n)}(P) \geq f(n)$ for some functions $f(n)$ and $\epsilon(n) > 0$. Then*

$$\mathrm{xc}_{QP0}(P) \geq f(n)/\ln \epsilon(n).$$

Proof. Let S be an QP_0-extended formulation of size r of $P \subseteq \mathbb{R}^n$, i.e., there exists a linear map π such that $\pi(S) = P$.

By Theorem 8 for any $\tau > 0$ there exists a τ-approximation L_τ of S and a linear map ψ such that $S \subseteq \psi(L_\tau) \subseteq (1+\tau)_{SOCP}S$. We obtain

$$P \subseteq \pi\psi(L_\tau) \subseteq \pi((1+\tau)_{SOCP}S)$$

For $\tau = \epsilon(n)$ we have

$$\pi((1+\epsilon(n))_{SOCP}S) \subseteq \pi((1+\epsilon(n))S) = (1+\epsilon(n))\pi(S) = (1+\epsilon(n))P$$

by Lemma 9. Therefore it follows that

$$P \subseteq \pi\psi(L_{\epsilon(n)}) \subseteq (1+\epsilon(n))P,$$

and so $L_{\epsilon(n)}$ is an $\epsilon(n)$-approximate extended formulation of P. Thus it has at least $f(n)$ inequalities by assumption. Moreover, by Theorem 8 we know that the number of inequalities of $L_{\epsilon(n)}$ is at most $O(1)r\ln(1/\epsilon(n))$. Thus in order to have $O(1)r\ln(1/\epsilon(n)) \geq f(n)$ we need $r \geq f(n)/\ln(1/\epsilon(n))$.

□

Combining Theorem 6 and Theorem 10 we obtain the following corollary:

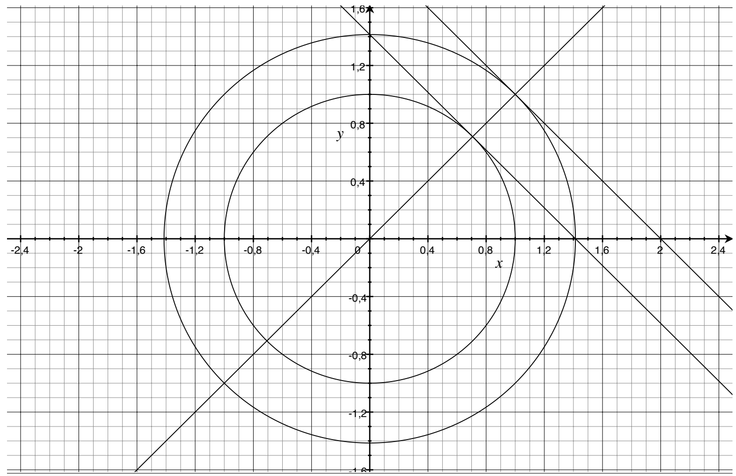
Corollary 11. *Every QP_0-extended formulation of the correlation polytope is of size $2^{\Omega(n)}$.*

Proof. We know that $\mathrm{xc}_{1/50}(COR(n)) = 2^{\Omega(n)}$. Therefore the result follows with Theorem 10. □

The following example shows that we indeed need that $0 \in S$ and it is not sufficient to have $z \in S$ with $\pi(z) = 0$. Therefore we need to restrict the class we consider: whereas there is always $z \in S$ with $\pi(z) = 0$ as we assume throughout that the polytope $P = \pi(S)$ we project satisfies $0 \in P$ it is not necessarily true that $0 \in S$ and therefore we have to restrict the considered class.

Example 1. We consider $S = \left\{ x \in R^2 \;\middle|\; \|(x,y)\|_2 \leq 1, x+y = \sqrt{2} \right\}$ and $\pi = x - y$. Observe that $S = \left\{ (\frac{1}{2}\sqrt{2}, \frac{1}{2}\sqrt{2}) \right\}$ and $\pi(z) = 0$ with $z := (\frac{1}{2}\sqrt{2}, \frac{1}{2}\sqrt{2})$. We have $\pi(S) = \{0\}$.

Now consider $(\sqrt{2})_{SOCP}S = \mathrm{conv}\left(\left\{ (\sqrt{2}, 0), (0, \sqrt{2}) \right\}\right)$ on the one hand and $(\sqrt{2})S = \{(1,1)\}$ on the other hand. Clearly, $(\sqrt{2})_{SOCP}S \not\subseteq (\sqrt{2})S$ and moreover, $\pi((\sqrt{2})_{SOCP}S) \not\subseteq (\sqrt{2})\pi(S)$ as $(\sqrt{2})\pi(S) = \{0\}$ but $\pi((\sqrt{2})_{SOCP}S) = [-\sqrt{2}, \sqrt{2}]$.



Remark 1. It is unlikely that we can do better than QP_0 due to the Glineur arguments. @Sebastian: check out the paper and make this more precise.

5 Approximating Spectrahedra

*** section needs to hardened considerably ***

In this section we will show that there exists a spectrahedron S with $0 \in S$ and an $\epsilon > 0$ such that any polytope P approximately S in the sense that $S \subseteq P \subseteq (1+\epsilon)S$ has exponential extension complexity. For this we combine results from Section 3.2 with previous results.

In a first step we show that there exists a spectrahedron S with the properties $\mathrm{COR}(n) \subseteq S \subseteq Q$ where Q is the polyhedron defined by the inequalities given by M with small extension complexity.

Lemma 12 (Existence of spectrahedron). *Let $n \in \mathbb{N}$. Then there exists a spectrahedron $S \subseteq \mathbb{R}^{n(n+1)/2}$ with*

$$\mathrm{COR}(n) \subseteq S \subseteq Q(n),$$

and S has small SDP-extension complexity, where $Q(n) := \left\{ x \in \mathbb{R}^{n(n+1)/2} \;\middle|\; Ax \leq b \right\}$ is the polyhedron given by the matrix $M(n)$.

Proof. As shown in Fiorini et al. [2012], there exists a PSD-factorization of $M = TU$ with $T^i, U_j \in \mathbb{S}_+^{n+1}$. We define

$$K := \left\{ (x,z) \in \mathbb{R}^{n(n+1)/2+n(n+1)/2} \;\middle|\; Ax + Tz = b, z \in \mathbb{S}^{n+1} \right\}$$

and $S = \mathrm{proj}_x(K)$. First observe that $S \subseteq Q(n)$: as $T^i \in \mathbb{S}^{n+1}$ we have $Tz \geq 0$ and thus $Ax \leq b$ holds for $x \in S$.

In order to show that $\mathrm{COR}(n) \subseteq S$ let $\mathrm{COR}(n) = \mathrm{conv}\left(\left\{ v_j \;\middle|\; v_j \in \mathrm{vert}(P) \right\}\right)$. Recall that the factorization of the matrix $M(n)$ is with respect to the vertices of $\mathrm{COR}(n)$ and the inequalities defining $Q(n)$. Therefore, for each $v_j \in \mathrm{vert}(P)$ we pick U_j from the factorization and hence the point $(v_j, U_j) \in K$ which implies $v_j \in S$. It follows $\mathrm{COR}(n) \subseteq S$. □

We can now combine Lemma 12 with Corollary 6 to obtain the following inapproximability theorem for the SDP cone.

*** we can do much better now, as no constant factor is possible!! ***

Theorem 13 (Inapproximability of the SDP-cone)**.** *There exists an $\epsilon > 0$ (in fact $\epsilon = \frac{1}{50}$) such that for $n \in \mathbb{N}$ there exists a spectrahedron $S \subseteq \mathbb{R}^{n(n+1/2)}$ with small extension complexity such that any polyhedral approximation P better than $1+\epsilon$ in the metric sense, i.e.,*

$$S \subseteq P \subseteq (1+\epsilon)S$$

has exponential extension complexity, i.e., $\mathrm{xc}(P) = 2^{\Omega(n)}$.

Proof. By Lemma 5 we know that there exists a spectrahedron S with $\mathrm{COR}(n) \subseteq S \subseteq Q(n)$ where $Q(n)$ is the polyhedron from Lemma 12. As $0 \in S$ this implies in particular:

$$\mathrm{COR}(n) \subseteq S \subseteq (1+\epsilon)S \subseteq (1+\epsilon)Q(n).$$

If now $S \subseteq P \subseteq (1+\epsilon)S$ then in particular $\mathrm{COR}(n) \subseteq P \subseteq (1+\epsilon)Q(n)$ and by Corollary 6 and the preceding construction we know that if we choose $\epsilon = \frac{1}{50}$, then $\mathrm{xc}(P) = 2^{\Omega(n)}$. This completes the proof. □

Remark 2. *** preliminary *** While the statement in Theorem 13 is about the spectrahedron S (S could *potentially* have large size in the projected space) which arises as a projection of K as defined in Lemma 12 the same holds true for K: if K would have a small polyhedral approximate extension complexity, the so would S by concatenation of the projection maps and linearity of the operators. *** we need to double-check this statement, for this we again need that $0 \in K$ ***

6 Open Questions

6.1 The matching polytope

Let $G = (V, E)$ be a graph. The *matching polytope* for G is given by the following inequalities

$$\begin{aligned} x(e) &\geq 0 && e \in E \\ x(\delta(v)) &= 1 && v \in V \\ x(\delta(U)) &\geq 1 && U \subseteq V, |U| = 2k+1 \geq 3, k \in \mathbb{N} \end{aligned}$$

We are interested in the extension complexity of the matching polytope for $G = K_n$. The slack matrix is essentially generated by the last inequalities (those of the third type).

6.2 Reduction mechanisms

Question 14. *Can we establish some general conditions or framework to carry over NP-hardness reduction (used typically in optimization) to reductions for extension complexity via being a face and/or an extended formulation.*

Question 15. *Can we also find (maybe easier) reductions via the communication complexity route, i.e, by runnning a protocol as a subprotocol as opposed to the current reduction mechanism via being a face and/or an extended formulation.*

Question 16. *Provide a direct reduction from $COR(\cdot)$ to the TSP polytope, i.e.,* without *using the stable set polytope as an intermediate result.*

Question 17. *Provide more reductions for different families of polytope to establish lower bounds on their extension complexity, such as*

1. ...

Question 18. *Can we find a hierarchy of error vs. number of inequalities? E.g., for the cut polytope we would like to better a stronger inapproximability if we allow for super-polynomial formulations.*

6.3 Different cones to consider

Question 19 (Implications for the SOCP cone)**.** *Can we say something about the SOCP cone and extended formulations arising from it? The SOCP cone is generated by all points $x = (x_0, x_1, \dots, x_n)$ so that $\|(x_1, \dots, x_n)\|_2 \leq x_0$ holds. This cone lies between LP-cone $(x_0, \dots, x_n) \geq 0$ and SDP (more on this below) and therefore should effectively correspond to some communication between classical and quantum or some restriction of quantum communication.*

a) The SOCP cone has a natural SDP formulation, i.e., it is a sub-cone. Given a constraint $\|x\|_2 \leq y$ (as above) then this is equivalent to require the matrix $A = \begin{pmatrix} yI & x \\ x^T & y \end{pmatrix}$ to be psd, i.e., $A \succeq 0$; I is the identity matrix of suitable dimension.

b) It is known that the SOCP cone can be arbitrarily well approximated (in an efficient manner!) using linear inequalities (see Ben-Tal and Nemirovski [2001]).

c) There is a natural hierarchy: linear programs can be written as second order cone programs (SOCP) and those in turn can be written as SDPs.

Now, b) together with the rather restrictive nature of the matrices in a) might indicate that we can 'simulate' the SOCP-communication (whatever restricted form of quantum communication it is) with classical communication with a loss of a small factor. The result in b) only states that this is true in an approximate sense but maybe we can show that whenever the SOCP formulation is exact (i.e., it projects to a given polytope), then there exists a linear extended formulation that is not much larger. The rationelle is that at some point we would approximate below the encoding length of the vertices of our polytope and therefore it should be exact.

Question 20 (Larger cones - Serge's question)**.** *There has been increasing (if specialised) interest in the physics community in looking at theories different from classical and quantum. These can be formulated in terms of cones, see arXiv:1012.1215 section 2 for an introduction.*

In particular going back to polytopes, any factorisation in terms of cones can be interpreted in this language as a one way communication complexity problem in which one communicates members of the cone, and measurements are defined as members of the dual cone.

It seems to me a very interesting line of enquiry to try to understand the communication complexity of such theories. However obviously one must put some restrictions on the cones, otherwise they may contain so much structure that solving the problem becomes trivial.

In the mentioned paper, there is some discussion about how to put a tensor product structure on cones, which is probably relevant.

6.4 Other questions

Question 21. *Can we find an approximate extended formulation for the matching polytope by somehow sampling only $O(\log n)$ vertices/edges. @Sam, can you make this more precise?*

Question 22. *A question from Sam: In Theorem 6 we could go up to $n^{1/3}$ inapprox — can we do better with sets n^{1-eps}? It seems that our previous discussion suggest no.*

The exponent in Razborov's lemma must be $O(\ell)$ for subsets a and b of size ℓ to be correct for rectangles $\binom{T_1}{\ell} \times \binom{T_2}{\ell}$ where T_1, T_2 form a partition of $[n]$ into sets of size $n/2$. Here $\binom{S}{\ell}$ is the set of all subsets of S of size ℓ, and the function X is the characteristic function of $\binom{T_1}{\ell} \times \binom{T_2}{\ell}$, ie, the product of the characteristic *functions of* $\binom{T_1}{\ell}$ *and* $\binom{T_2}{\ell}$*. Then*

$$\mathbb{E}\left[XI_B\right] = 0, \quad r = 1, \quad \|X\|_\infty = 1 \tag{28}$$

$$\mathbb{E}\left[XI_A\right] = \frac{\binom{n/2}{\ell}^2}{\binom{n}{\ell}\cdot\binom{n-\ell}{\ell}} = \frac{\binom{n-2\ell}{n/2-\ell}}{\binom{n}{n/2}} = \Theta\left(2^{-2\ell}\sqrt{\frac{n}{n-\ell}}\right). \tag{29}$$

Now the inequality (1) *can be only true with an* $O(\ell)$ *exponent in the last term.*

Question 23. *About the question: "Can the inequality be reversed?", I have some doubts. Indeed, assume that one can prove* $\mu(A \cap R) \geqslant \tau\mu(B \cap R) - f(n)$ *for all rectangles* R*, where* τ *is a constant and* $f(n)$ *is a function. There are* n *rectangles* $R_1, \ldots, R_n$ *which do no contain any pair in* A *and cover* B*. Simply,* $R_i := \{(x,y) \mid i \in x \cap y\}$*. Then at least one of these rectangles has measure* $\mu(B)/n$*. We would get then* $\mu(B) \leqslant \tau^{-1}n \cdot f(n)$ *and the only way to keep* $\mu(B)$ *constant is by letting* $f(n) = \Omega(1/n)$*, that is* $1/f(n) = O(n)$*, which would provide poor lower bounds.*

Question 24. *In order to tackle approximate extended formulations, can we use a reduction via protocols by showing that by having a protocol that computes an "approximate slack matrix" in expectation, we can construct a protocol that computes the exact slack matrix in expectation or a protocol that computes the support of the (original, non-approximate) slack matrix in a non-deterministic fashion.*

Remark 3. @Sam: add remark about the submissive from your email at 2/10.

References

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

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

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

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

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