Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 157 (2017)

High Degree Sum of Squares Proofs, Bienstock-Zuckerberg hierarchy and Chvátal-Gomory cuts *

Monaldo Mastrolilli
IDSIA, 6928 Manno, Switzerland monaldo@idsia.ch

Abstract

Chvátal-Gomory (CG) cuts and the Bienstock-Zuckerberg hierarchy capture useful linear programs that the standard bounded degree Lasserre/Sum-of-Squares (SOS) hierarchy fails to capture.

In this paper we present a novel polynomial time SOS hierarchy for 0/1 problems with a custom subspace of high degree polynomials (not the standard subspace of low-degree polynomials). We show that the new SOS hierarchy recovers the Bienstock-Zuckerberg hierarchy. Our result implies a linear program that reproduces the Bienstock-Zuckerberg hierarchy as a polynomial sized, efficiently constructible extended formulation that satisfies all constant pitch inequalities. The construction is also very simple, and it is fully defined by giving the supporting polynomials (one paragraph). Moreover, for a class of polytopes (e.g. set covering and packing problems) it optimizes, up to an arbitrarily small error, over the polytope resulting from any constant rounds of CG cuts.

Arguably, this is the first example where different basis functions can be useful in *asymmetric* situations to obtain a hierarchy of relaxations.

1 Introduction

The Lasserre/Sum-of-Squares (SOS) hierarchy [18, 22, 23, 28] is a systematic procedure for constructing a sequence of increasingly tight semidefinite relaxations. The SOS hierarchy is parameterized by its *level* d, such that the formulation gets tighter as d increases, and a solution of accuracy $\varepsilon > 0$ can be found in time $(mn\log(1/\varepsilon))^{O(d)}$ where n is the number of variables and m the number of constraints in the original problem. It is known that the hierarchy converges to the 0/1 polytope in n levels and captures the convex relaxations used in the best available approximation algorithms for a wide variety of optimization problems (see e.g. [3, 6, 19] and the references therein).

In a recent paper Kurpisz, Leppänen and the author [16] characterize the set of 0/1 integer linear problems that still have an (arbitrarily large) integrality gap at level $n-1$. These problems are the "hardest" for the SOS hierarchy in this sense. In another paper, the same authors [17] consider a problem that is solvable in $O(n\log n)$ time and prove that the integrality gap of the SOS hierarchy is unbounded at level $\Omega(\sqrt{n})$ even after incorporating the objective function as a constraint (a classical trick that sometimes helps to improve the quality of the relaxation). All these "SOS-hard" instances have a "covering nature".

*Supported by the Swiss National Science Foundation project 200020-169022 "Lift and Project Methods for Machine Scheduling Through Theory and Experiments". Preliminary version appeared in IPCO'17 [21].

ISSN 1433-8092

Chvátal-Gomory (CG) rounding is a popular cut generating procedure that is often used in practice (see e.g. [7] and Appendix 7 for a short introduction). There are several prominent examples of CG-cuts in polyhedral combinatorics, including the odd-cycle inequalities of the stable set polytope, the blossom inequalities of the matching polytope, the simple Möbius ladder inequalities of the acyclic subdigraph polytope and the simple comb inequalities of the symmetric traveling salesman polytope, to name a few. Chvátal-Gomory cuts captures useful and efficient linear programs that the bounded degree SOS hierarchy fails to capture. Indeed, the "SOS-hard" instances studied in [16] are the "easiest" for CG cuts, in the sense that they are captured within the *first* CG closure. It is worth noting that it is NP-hard [20] to optimize a linear function over the first CG closure, an interesting contrast to lift-and-project hierarchies (like Sherali-Adams, Lovász-Schrijver, and SOS) where one can optimize in polynomial time for any constant number of levels.

Interestingly, Bienstock and Zuckerberg [5] prove that, in the case of set covering, one can separate over all CG-cuts to an arbitrary fixed precision in polynomial time. The result in [5] is based on another result [4] by the same authors, namely on a (positive semidefinite) lift-and-project operator (which we denote (BZ) herein) that is quite different from the previously proposed operators. This lift-and-project operator generates different variables for different relaxations. They showed that this flexibility can be very useful in attacking relaxations of some set covering problems.

These three methods, (SOS, CG, BZ), are to some extent incomparable, roughly meaning that there are instances where one succeeds while the other fails (see [2] for a comparison between SOS and BZ, the already cited [16] for "easy" cases for CG cuts that are "hard" for SOS, and finally note that clique constraints are "easy" for SOS but "hard" for CG cuts [25], to name a few). In Section 4 we show another simple and basic Chvátal-Gomory cut that is "very hard" for SOS.

One can think of the standard Lasserre/SOS hierarchy at level $O(d)$ as optimizing an objective function over linear functionals that sends n-variate polynomials of degree at most d (over $\mathbb{R}$) to real numbers. The restriction to polynomials of degree d is the standard way (as suggested in [18, 23] and used in most of the applications) to bound the complexity, implying a semidefinite program of size $n^{O(d)}$. However, this is not strictly necessary for getting a polynomial time algorithm and it can be easily extended by considering more general subspaces having a "small" (i.e. polynomially bounded) set of basis functions (see e.g. Chapter 3 in [6] and [9, 11]). This is a less explored direction and it will play a key role in this paper. Indeed, the more general view of the SOS approach has been used so far to exploit very symmetric situations (see e.g. [9, 11, 26]). For symmetric cases the use of a different basis functions has been proved to be very useful.

To the best of author knowledge, in this paper we give the first example where different basis functions can be useful in *asymmetric* situations to obtain a hierarchy of relaxations. More precisely, we show how to reframe the Bienstock-Zuckerberg hierarchy [4] as an augmented version of the SOS hierarchy that uses high degree polynomials (in Section 5 we consider the set cover problem, that is the main known application of the BZ approach, and in Section 6 we sketch the general framework that is based on the set cover case). The resulting high degree SOS approach retains in one single unifying SOS framework the best from the standard bounded degree SOS hierarchy, incorporates the BZ approach and allows to get arbitrary good approximate fixed rank CG cuts for both, set covering and packing problems, in polynomial time (BZ guarantees this only for set covering problems). Moreover, the proposed framework is very simple and, assuming a basic knowledge in SOS machinery (see Section 2), it is fully defined by giving the supporting polynomials (see

Definition 5.1). This is in contrast to the Bienstock-Zuckerberg's hierarchy that requires an elaborate description [4, 29]. Finally, as observed in [1] (see Propositions 25 and 26 in [1]), the performances of the Bienstock-Zuckerberg's hierarchy depend on the presence of redundant constraints.[1] The proposed approach removes these unwanted features.

We emphasize that one can also modify the Sherali-Adam's hierarchy/proof system in the same manner to obtain the covering results, but we decided to take the SDP framework for generality. So the formulation that we are going to describe for the set cover problem is actually an explicit linear program (see Section 5.3) that reproduces the Bienstock-Zuckerberg hierarchy as a polynomial sized, efficiently constructible extended formulation that satisfies all constant pitch inequalities. More details on the Bienstock-Zuckerberg hierarchy are given in Section 6.

In Section 7, we observe that for *both*, set covering and packing problems, the framework used in this paper allows to optimize, up to an arbitrarily small error, over the polytope resulting from any constant rounds of CG cuts.

We have tried to make this article as self-contained as possible and accessible to non-expert readers, providing an introduction to the basic necessary tools and results (see Section 2) and some simple introductory examples. Our main applications are given in Sections 5 and 7. Additional background material, terminology and results can be found in appendix.

Recent developments. Very recently Fiorini et al. [10] claim a new approach to reproduce the Bienstock-Zuckerberg hierarchy. We remark that their framework is weaker than the one presented in this paper, meaning that does not generalize to packing problems (see Section 7). Moreover, their proof is essentially based on similar arguments as used in this paper (formerly appeared in [21]). We give more details in the appendix.

2 SOS-Proofs over the boolean hypercube

Certifying that a polynomial $f(x)$ is non-negative over a semialgebraic set $\mathcal{F}$ is an important problem in optimization, as certificates of non-negativity can often be leveraged into optimization algorithms. For example if $f(x) = f'(x) - \lambda$ and we can certify that $f(x)$ is non-negative over $\mathcal{F}$ then the minimum of $f'(x)$ is not smaller than λ. In this paper we are interested in the case $\mathcal{F}$ is the set of feasible solutions of a 0/1 integer linear program:

$$\mathcal{F} := \{x \in \mathbb{R}^n : x_k^2 - x_k = 0 \quad \forall k \in [n],\ g_i(x) \geq 0 \quad \forall i \in [m]\} \tag{1}$$

where $x_k^2 - x_k = 0$ encodes $x_k \in \{0, 1\}$ and each constraint $g_i(x) \geq 0$ is linear. It is known that the nonnegativity of a polynomial over $\mathcal{F}$ defined in (1) can be certified by showing that $f(x)$ belongs to the cone generated by the nonnegative constraints (i.e. $\mathbf{cone}(g_1, \ldots, g_m)$ see Appendix A) in the quotient ring $\mathbb{R}[\mathbf{x}]/\mathbf{I}(\mathbb{Z}_2^n)$, where $\mathbf{I}(\mathbb{Z}_2^n)$ denotes the vanishing ideal over the boolean hypercube (see Appendix A). We review this derivation from a slightly different perspective by highlighting several aspects that will play a role in all the derivations that we will use. We start with some preliminaries.

Kronecker delta. For any set $Z \subseteq [n]$ and given $I \subseteq Z$ define the *Kronecker delta* function $\delta_I^Z(x)$ by:

$$\delta_I^Z(x) := \prod_{i \in I} x_i \prod_{j \in Z \setminus I} (1 - x_j) \tag{2}$$

[1] I thank Levent Tuncel for pointing me his work [1].

Note that $\sum_{I\subseteq Z}\delta_I^Z(x)=1$, $(\delta_I^Z(x))^2\equiv\delta_I^Z(x) \pmod{\mathbf{I}(\mathbb{Z}_2^n)}$ and $\delta_I^Z(x)\delta_J^Z(x)\equiv 0 \pmod{\mathbf{I}(\mathbb{Z}_2^n)}$ for any $I\neq J$ with $I,J\subseteq Z$. Therefore, it follows that $(\sum_I\delta_I^Z(x))^2 \equiv \sum_I\delta_I^Z(x) \pmod{\mathbf{I}(\mathbb{Z}_2^n)}$. Finally observe that for any linear function $g(x)=\sum_{i\in S}g_ix_i-g_0$ with $S\subseteq[n]$ we have $\delta_I^Z(x)(\sum_{i\in S}g_ix_i-g_0)\equiv\delta_I^Z(x)(\sum_{i\in S\cap I}g_i-g_0+\sum_{S\setminus Z}g_ix_i) \pmod{\mathbf{I}(\mathbb{Z}_2^n)}$. These basic facts will be used several times, in particular over the boolean hypercube we can restrict with no loss to polynomials from $\mathbb{R}[\mathbf{x}]/\mathbf{I}(\mathbb{Z}_2^n)$, i.e. n-degree multilinear polynomials (we use $\mathbb{R}[\mathbf{x}]$ to denote the polynomial ring over the reals in n variables $\mathbb{R}[x_1,\dots,x_n]$ and $\mathbb{R}[\mathbf{x}]_d$ to denote the subspace of $\mathbb{R}[\mathbf{x}]$ of polynomials of degree at most $d\leq n$). So $\delta_I^Z(x)$ is the multilinear representation of $(\delta_I^Z(x))^2$ over the boolean hypercube and we will use them both interchangeably.

Nonnegativity certificate. Given a polynomial $f(x)\in\mathbb{R}[\mathbf{x}]$ that is nonnegative over $\mathcal{F}$, we are interested in certifying this property. We will assume that $f(x)$ is multilinear (for the applications of this paper the reader can assume that $f(x)$ is a linear function). Let $x_I\in\{0,1\}^n$ be the 0/1 solution with $x_i=1$ $\forall i\in I$, and $x_j=0$ $\forall j\in[n]\setminus I$ for any $I\subseteq[n]$. Partition the boolean hypercube into two sets $N^+=\{I\subseteq[n]:f(x_I)\geq 0\}$ and $N^-=\{I\subseteq[n]:f(x_I)<0\}$. If $f(x)$ is nonnegative over $\mathcal{F}$ then for every $I\in N^-$ there exists a violated constraint on x_I, i.e. there is a mapping $h:2^{[n]}\to[m]$ such that $g_{h(I)}(x_I)<0$ for every $I\subseteq[n]$ with $f(x_I)<0$. In the remainder, whenever we use "$\equiv$" assume that the equivalence is modulo the vanishing ideal $\pmod{\mathbf{I}(\mathbb{Z}_2^n)}$ (unless differently defined). Then:

$$
\begin{aligned}
f(x) &= \overbrace{\left(\sum_{I\subseteq[n]}\delta_I^{[n]}(x)\right)}^{=1} f(x) = \overbrace{\sum_{I\in N^+}\delta_I^{[n]}(x)f(x_I)+\sum_{I\in N^-}\delta_I^{[n]}(x)\frac{f(x_I)}{g_{h(I)}(x_I)}g_{h(I)}(x_I)}^{\text{A}} \\
&\equiv \overbrace{\underbrace{\left(\sum_{I\in N^+}\delta_I^{[n]}(x)\sqrt{f(x_I)}\right)^2}_{s_0(x)}+\sum_{I\in N^-}\underbrace{\left(\delta_I^{[n]}(x)\sqrt{\frac{f(x_I)}{g_{h(I)}(x_I)}}\right)^2}_{s_{h(I)}(x)}g_{h(I)}(x)}^{\text{B}}
\end{aligned} \tag{3}
$$

It follows that any nonnegative polynomial over $\mathcal{F}$ can be represented as a polynomial from $\mathbf{cone}(g_1,\dots,g_m)$ in the quotient ring $\mathbb{R}[\mathbf{x}]/\mathbf{I}(\mathbb{Z}_2^n)$. This provides a certificate of its nonnegativity (in the form as described in (4) below). Note that in the above derivation (3) the expression denoted by A is equal to the multilinear representation of B and the highest degree of A is n. Moreover, A is equal to $f(x)$ (recall that $f(x)$ is assumed to be multilinear). In the following we will implicitly use the multilinear representation of the right-hand-side of (4) in the SOS derivations that we will compute, namely we will assume that the elements from $\mathbf{cone}(g_1,\dots,g_m)$ are in the quotient ring $\mathbb{R}[\mathbf{x}]/\mathbf{I}(\mathbb{Z}_2^n)$ and given in the multilinear form.

Proposition 2.1. *Any multilinear polynomial $f\in\mathbb{R}[\mathbf{x}]$ that is non-negative over the semialgebraic set* (1) *has a degree-n* SOS *representation as:*

$$
f(x)\equiv s_0(x)+\sum_{i\in[m]}s_i(x)g_i(x) \pmod{\mathbf{I}(\mathbb{Z}_2^n)} \tag{4}
$$

where $s_i\in\Sigma:=\{s\in\mathbb{R}[\mathbf{x}]:s=\sum_{i=1}^r q_i(x)^2,$ for some $q_1,\dots,q_r\in\mathbb{R}[\mathbf{x}]_n\}$.

These representations can be seen as specific instances of *Positivstellensatz*, a general technique to characterize polynomials that are positive on a semialgebraic set. Computing degree-n SOS representation can be automatized by solving a semidefinite program

(SDP) which is an optimization problem over positive semidefinite (PSD) matrices (see Appendix B.1 for details). However this may take in general exponential time. The "standard" (namely the "most used") way to bound the complexity is to consider the polynomials $q_i \in \mathbb{R}_n[\mathbf{x}]$ used in (4) in the standard monomial basis and to restrict their degree to a constant $d \leq n$. If one restricts the degrees of the polynomials in the certificate to be at most some integer d, it turns out that the positivity certificate is given by a semidefinite program of size $n^{O(d)}$. Clearly this restriction imposes severe restrictions on the kind of proofs that can be obtained. This type of algorithm was proposed first by Shor [28] and the idea was taken further by Parrilo [23, 24] and Lasserre [18]. However, this fact can be easily extended to other subspaces than the standard monomial basis of bounded degree, by considering subspaces having a "small", i.e. polynomially bounded set of basis functions (see e.g. [6, 9]). This is a less explored direction and it will play a key role in this paper. The following introduces this point.

Definition 2.1. *For any fixed subspace $Q \subseteq \mathbb{R}[\mathbf{x}]/\mathbf{I}(\mathbb{Z}_2^n)$, we say that a polynomial $f \in \mathbb{R}[\mathbf{x}]$ that is non-negative over the semialgebraic set* (1) *admits a* Q-SOS representation *(or it is* Q-SOS derivable *and write $\mathcal{F} \vdash_Q f(x) \geq 0$) if*

$$f(x) \equiv s_0(x) + \sum_{i \in [m]} s_i(x) g_i(x) \pmod{\mathbf{I}(\mathbb{Z}_2^n)} \tag{5}$$

where $s_i \in \{s \in \mathbb{R}[\mathbf{x}] : s = \sum_{i=1}^{r} q_i(x)^2,$ for some $q_1, \ldots, q_r \in Q\}$. For a set $\mathcal{S} \subseteq \mathbb{R}_n[\mathbf{x}]$ let $\langle \mathcal{S} \rangle = span(\mathcal{S})$ denote the vector space spanned by $\mathcal{S}$. If $\langle \mathcal{S} \rangle = Q$ then $\mathcal{S}$ is called the Q-SOS spanning set*.*

The existence of a $\langle \mathcal{S} \rangle$-SOS representation can be decided by solving a semidefinite programming feasibility problem whose matrix dimension is bounded by $O(|\mathcal{S}|)$. We refer to [6, 9] and Appendix B.1 for details and an example.

2.1 The dual point of view

Consider the minimization of a given polynomial $p(x)$ over the semialgebraic set (1). Let $\mathcal{G} = \{g_i(x), i \in [m]\}$. For any $\mathcal{S} \subseteq \mathbb{R}_n[\mathbf{x}]$, a relaxation is given by the following conic program:

$$\max\{\gamma : p - \gamma \in \mathbf{cone}_{\langle \mathcal{S} \rangle}(\mathcal{G})\} \tag{6}$$

where $\mathbf{cone}_{\langle \mathcal{S} \rangle}(\mathcal{G}) = \{f(x) : f(x) = s_0(x) + \sum_{g \in \mathcal{G}} s_g(x) g(x)$ where $s_g(x) = \sum_i q_i(x)^2, q_i \in \langle \mathcal{S} \rangle\}$ is the cone of nonnegative polynomials generated by $\langle \mathcal{S} \rangle$ (for $\langle \mathcal{S} \rangle = \mathbb{R}_n[\mathbf{x}]$ then (6) is exact by Proposition 2.1). By definition, the dual of $\mathbf{cone}_{\langle \mathcal{S} \rangle}(\mathcal{G})$ are the linear functionals[2] $\mathbf{cone}^{dual}_{\langle \mathcal{S} \rangle}(\mathcal{G}) = \{l : \langle l, h \rangle \geq 0, \forall h \in \mathbf{cone}_{\langle \mathcal{S} \rangle}(\mathcal{G})\}$ that take nonnegative values on it, namely let $L[h] = \langle l, h \rangle$:

$$L\big[q^2(x)\big] \geq 0, \quad \forall q \in \langle \mathcal{S} \rangle \tag{7}$$

$$L\big[q^2(x) g_i(x)\big] \geq 0, \quad \forall q \in \langle \mathcal{S} \rangle, i = [m] \tag{8}$$

where $q^2(x)$ in (7) and $q^2(x)g_i(x)$ in (8) are given in the multilinear form (so every occurrence of x_j^2 is replaced by x_j for all $j \in [n]$). The dual program of (6) is then given by the following that we call $\text{SOS}_{\langle \mathcal{S} \rangle}(\mathcal{F})$*-relaxation*:

$$\min\{\langle l, p \rangle : l_1 = 1, l \in \mathbf{cone}^{dual}_{\langle \mathcal{S} \rangle}(\mathcal{G})\} \tag{9}$$

[2] Sometimes called *pseudo-expectation*.

The program (9) is actually a semidefinite program whose matrix dimension is bounded by $n^{O(|\mathcal{S}|)}$ (see [6] and Appendix B.2).

When $\langle\mathcal{S}\rangle = \mathbb{R}[\mathbf{x}]_d/\mathbf{I}(\mathbb{Z}_2^n)$, namely $\mathcal{S}$ is the standard monomial basis of degree $\leq d$, then (9) is the (standard) Lasserre/SOS-hierarchy parameterized by the degree d. Note that in order to constraint the dimension of the matrix inequalities to be bounded by $n^{O(d)}$ we can restrict to a different subspace, namely a different basis than the standard basis for polynomials of degree at most d. This is the point of view that is taken in this paper.

In the remainder, we consider some relaxations (9) as given by fixing the spanning polynomials $\mathcal{S}$ (and the set $\mathcal{G}$ of constraints). In proving the properties of (9) we will use the following fact that follows from the definition above.

Proposition 2.2. *Suppose $\mathcal{F} \vdash_{\langle\mathcal{S}\rangle} f(x) \geq 0$ then $L[f(x)] \geq 0$.*

In particular note that if $f(x) = \sum_i a_i x_i - a_0 \geq 0$ is a linear inequality that is $\langle\mathcal{S}\rangle$-SOS derivable then any solution of (9) satisfies the linear inequality $\sum_i a_i L[x_i] - a_0 \geq 0$. It follows that in order to show that any (projected) solution $\{L[x_i] : i \in [n]\}$ of (9) belongs to the polytope $Ax \geq b$ it is sufficient to show that the linear constraints $Ax \geq b$ are $\langle\mathcal{S}\rangle$-SOS derivable. We emphasize that we will use this approach in the remainder of the paper.

3 A family of spanning polynomials

Consider any group G that is acting on monomials in $\mathbb{R}[\mathbf{x}]$ via $gx_i = x_{g(i)}$, for all $g \in G$ and $i \in [n]$. Let $f \in \mathbb{R}[n]$ be a real-valued G-invariant polynomial that is nonnegative over the boolean hypercube. From (3) we have $f(x) \equiv (\sum_{I \in N^+} \delta_I^{[n]}(x)\sqrt{f(x_I)})^2$, and therefore $f(x)$ is congruent $\pmod{\mathbf{I}(\mathbb{Z}_2^n)}$ to the square of a G-invariant polynomial.

Lemma 3.1. *Consider any group G that is acting on monomials in $\mathbb{R}[\mathbf{x}]$ via $gx_i = x_{g(i)}$ for each $g \in G$ and $i \in [n]$. Any real-valued G-invariant polynomial $f \in \mathbb{R}[n]$ that is nonnegative over the boolean hypercube has a degree-n square representation $f(x) \equiv h(x)^2 \pmod{\mathbf{I}(\mathbb{Z}_2^n)}$, for some G-invariant polynomial $h \in \mathbb{R}[x]$.*

Let X be a nonempty set. A permutation σ of X is a bijection $\sigma : X \to X$. The set of all permutations of X is called the *symmetric group* of X and it is denoted by S_X. In the following for any $F \subseteq X$ we will consider the *stabilizer* of F in S_X, namely $stab_{S_X}(F)$ is the subgroup of S_X whose elements are permutations of set X that fix the elements from F. Note that $stab_{S_X}(F)$ is the symmetric subgroup $S_{X\setminus F}$ acting on X and leaving the points in F fixed. The set F is the $S_{X\setminus F}$ group's set of *fixed points* when acting on X.

For the main application of this paper (Section 5) the spanning set is given by products of $S_{X\setminus F}$-invariant polynomials (see Definition 5.1). Note that when $F = \emptyset$ an $S_{X\setminus F}$-invariant polynomial is standardly called a *symmetric* polynomial. Generalizing the latter terminology, we will also use $(X \setminus F)$-*symmetric* polynomial to denote a $S_{X\setminus F}$-invariant polynomial. Observe that *any* polynomial is $(X \setminus F)$-symmetric for some $F \subseteq X$.

From Lemma 3.1 any non-negative $(X\setminus F)$-symmetric polynomial is congruent $\pmod{\mathbf{I}(\mathbb{Z}_2^n)}$ to the square of one $(X\setminus F)$-symmetric polynomial. This simple fact will play a central role in our derivations. In particular the following will be used several times in the following.

Corollary 3.2. *Consider any finite set of polynomials $\mathcal{S} \subseteq \mathbb{R}[\mathbf{x}]/\mathbf{I}(\mathbb{Z}_2^n)$ and let $Q = span(\mathcal{S})$. For any $F \subseteq X \subseteq [n]$, if the ring $(\mathbb{R}[\mathbf{x}]/\mathbf{I}(\mathbb{Z}_2^n))^{S_{X\setminus F}}$ of all $(X \setminus F)$-symmetric polynomials is a subspace of Q then any nonnegative $(X \setminus F)$-symmetric polynomial has a Q-SOS representation.*

A simple counting argument shows the following rough bound on the size of the spanning set of $(X \setminus F)$-symmetric polynomials.

Lemma 3.3. *For any $X \subseteq [n]$, let $Q_{X,t}$ denote the subspace of all $(X \setminus F)$-symmetric polynomials for all $F \subseteq X, |F| \leq t$. There is a spanning set $\mathcal{S}_{X,t}$ such that $Q_{X,t} \subseteq span(\mathcal{S}_{X,t})$ and $|\mathcal{S}_{X,t}| = n^{O(t)}$.*

Proof. For $F \subseteq X \subseteq [n]$, in any multilinear $(X \setminus F)$-symmetric polynomial $g(x) = \sum_{I\subseteq[n]} g_I x^I$ (where $x^I = \prod_{i\in I} x_i$) the coefficients of monomials x^I and x^J are the same as soon as $|I| = |J|$ and $I \cap F = J \cap F$. It follows that we can easily define a basis for $(\mathbb{R}[\mathbf{x}]/\mathbf{I}(\mathbb{Z}_2^n))^{S_{X\setminus F}}$ of size at most $\sum_{s=0}^{n}\sum_{i=0}^{\min\{s,t\}} \binom{t}{i} < n2^t$. We can choose F in $\binom{|X|}{|F|}$ ways, therefore the dimension of the vector space spanned by all the basis of boolean $(X \setminus F)$-symmetric polynomials for all $|F| \leq t$ is at most $\tau = t\binom{n}{t}\sum_{s=0}^{n}\sum_{i=0}^{\min\{s,t\}} \binom{t}{i} < tn2^t\binom{n}{t} = n^{O(t)}$. ◻

Example 3.1. *For any $F \subseteq X \subseteq [n]$ the Kronecker delta $\delta_F^X(x)$ (see (2)) is $(X \setminus F)$-symmetric and $(X\setminus(X\setminus F))$-symmetric. The following two polynomials, $\sum_{F\subseteq X,|F|\geq k} \delta_F^X(x)$ and $\sum_{F\subseteq X,|F|=k} \delta_F^X(x)$, are X-symmetric, for any $k \leq |X|$. Moreover note that the squares of these polynomials are the polynomials themselves on the boolean hypercube. This is the argument that underpins the proof of Corollary 3.2. Actually we can (and we will) restrict ourselves to only those $(X \setminus F)$-symmetric polynomials that come from conical combinations of Kronecker delta functions, since these form a basis.*

4 A simple Chvátal-Gomory cut that is hard for SOS

In this section we provide an introductory educational example where the standard SOS fails and how it can be easily fixed by using high degree polynomials.

The example is motivated by the following situation. Consider the rational polyhedra $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ with $A \in \mathbb{Z}^{m\times n}$ and $b \in \mathbb{Z}^m$. Inequalities of the form $(\lambda^\top A)x \geq \lceil\lambda^\top b\rceil$, with $\lambda \in \mathbb{R}^m$, $\lambda^\top A \in \mathbb{Z}^n$, and $\lambda^\top b \notin \mathbb{Z}$ are commonly referred to Chvátal-Gomory cuts (CG-cuts for short) (see Section 7). It is a natural question to study how many levels (or degree d) of the *standard* Sum-of-Squares hierarchy $\textsc{sos}_{\mathbb{R}_d[\mathbf{x}]}(\mathcal{F})$ are necessary to strengthening $(\lambda^\top A)x \geq \lambda^\top b$ to get $(\lambda^\top A)x \geq \lceil\lambda^\top b\rceil$.

With this aim, consider the following semialgebraic set:

$$\mathcal{F} = \{x \in \mathbb{R}^n : x_k^2 - x_k = 0 \ \forall k \in [n], \sum_{i=1}^{n} x_i \geq b\} \tag{10}$$

where $b \in \mathbb{Q}_+$ is intended to be a positive fractional number. Obviously, any feasible integral solution satisfies $\sum_{i=1}^{n} x_i \geq \lceil b\rceil$ and this is promptly captured by the the first CG closure. In the following (Theorem 4.1) we show that regardless whether b is "small" (i.e. $O(1)$), or "large" (i.e. $\Omega(n)$), the necessary number of levels (or degree) for the standard Sum-of-Squares hierarchy is of linear order $\Omega(n)$.

We remark that Grigoriev, Hirsch, and Pasechnik give in [12] a very interesting and influential result that is related to our Theorem 4.1 below, but *significatively different* in terms of both, lower bounds and techniques. We defer the interested reader to Section 4.1 for a discussion on this point and for a more precise meaning of "significatively different".

In the remainder, we restrict to the cases where the following holds.

$$b := L + \frac{1}{P} \qquad \text{for } L \in \left\{0, 1, \ldots, \left\lceil\frac{n}{2}\right\rceil - 1\right\} \text{ and sufficiently large } P \in \mathbb{N} \tag{11}$$

Theorem 4.1 below (proof in Section C.2) shows that the $\text{SOS}_{\mathbb{R}_d[\mathbf{x}]}(\mathcal{F})$-relaxation requires $d \geq n - L$ for satisfying $\sum_{i=1}^n x_i \geq \left\lceil L + \frac{1}{P} \right\rceil$, for any $L \in \left\{0, 1, \ldots, \left\lceil \frac{n}{2} \right\rceil - 1\right\}$ and sufficiently large $P \in \mathbb{N}$ (that depends on n).

Theorem 4.1. *If* $\mathcal{F} \vdash_{\mathbb{R}_d[\mathbf{x}]} (\sum_{i=1}^n x_i - \lceil b \rceil \geq 0)$ *then* $d \geq n - L$.

The result in Theorem 4.1 is disappointing for at least two reasons: the CG-cut looks pathetically trivial and the proof that $\text{SOS}_{\mathbb{R}_d[\mathbf{x}]}(\mathcal{F})$ fails for small d is relatively complicate (see Section C.2). Clearly it would be sufficient to have in the "bag" Q the symmetric polynomials (so in $\mathcal{S}$ there is the set of the $n+1$ elementary symmetric polynomials) to promptly capture this constraint within Q-SOS in polynomial time. Indeed consider the following equivalences:

$$\sum_{i=1}^n x_i - \lceil b \rceil = \left(\sum_{i=1}^n x_i - \lceil b \rceil\right) \overbrace{\left(\sum_{i=0}^n \underbrace{\sum_{I \subseteq [n]:|I|=i} \delta_I^{[n]}}_{symmetric}\right)}^{=1} \equiv \left(\sum_{i=0}^n (i - \lceil b \rceil) \underbrace{\sum_{I \subseteq [n]:|I|=i} \delta_I^{[n]}}_{symmetric}\right)$$

$$\equiv \underbrace{\left(\sum_{i=\lceil b \rceil}^n \sqrt{i - \lceil b \rceil} \sum_{I \subseteq [n]:|I|=i} \delta_I^{[n]}\right)^2}_{s_0(x)} + \underbrace{\sum_{i=0}^{\lceil b \rceil - 1} \left(\sqrt{\frac{i - \lceil b \rceil}{i - b}} \sum_{I \subseteq [n]:|I|=i} \delta_I^{[n]}\right)^2}_{s_1(x)} \underbrace{\left(\sum_{i=1}^n x_i - b\right)}_{g_1(x)}$$

Note that it has exactly the form in (5), where each $s_i(x)$ is the sum of squares of symmetric polynomials. Therefore $\sum_{i=1}^n x_i - \lceil b \rceil$ is Q-SOS derivable (see also Example B.1). We refer to [9, 11, 26] for other more interesting symmetric situations.

We emphasize that in this paper we show how to handle *asymmetric* situations by exploiting the problem structure, which is our main result.

4.1 On a related result by Grigoriev, Hirsch, and Pasechnik

Grigoriev, Hirsch, and Pasechnik (see Theorem 8.1 in [12]) give a result related to Theorem 4.1 but significatively different as explained in the remainder of this section. In [12], the *symmetric knapsack* is defined as follows:

$$\mathcal{F}' = \{x \in \mathbb{R}^n : x_k^2 - x_k = 0 \ \forall k \in [n], \sum_{i=1}^n x_i = b\} \tag{12}$$

Note that $\mathcal{F}'$ (see (12)) is a more constrained version of set $\mathcal{F}$ (see (10)). The Positivstellensatz Calculus [12] is a proof system for languages consisting of unsolvable systems of polynomial equations (like (12) when b is a non-integral value). A proof in this system consists of polynomials $h_1 \ldots, h_l$ and a derivation of $1 + \sum_j h_j^2 = 0$ from $\mathcal{F}'$ using polynomial calculus rules.

Let δ denote the step function which equals 2 outside the interval $(0, n)$ and $2k+4$ on the intervals $(k, k+1)$ and $(n-k-1, n-k)$ for all integers $0 = k < n/2$. In [12] the following result is proved.

Theorem 4.2. *[12] Any Positivstellensatz calculus refutation of the symmetric knapsack problem* $\mathcal{F}'$ *(see* (12)*) has degree* $\min\{\delta(b), \lceil (n-1)/2 \rceil + 1\}$.

It is easy to observe that any Positivstellensatz Calculus lower bound for the more constrained set $\mathcal{F}'$ (see (12)) gives a SOS lower bound for the set $\mathcal{F}$ (see (10)). However, for $b < n/2$, the bounds given by Theorem 4.2 [12] when applied to set $\mathcal{F}$ are weaker (and also considerably weaker) than the ones provided by Theorem 4.1. For example, for any constant k and $b \in (k, k+1)$ the degree lower bound in Theorem 4.2 is $2k+4 = O(1)$, whereas by Theorem 4.1 the degree lower bound is $n-k$.

Regarding the technique, Theorem 4.1 is proved by building on a result given in [14]. The latter has been shown very powerful in several other situations (see [14, 15] for more examples).

Finally, we remark that studying the number of levels necessary for strengthening *inequalities* (as in Theorem 4.1) is useful for studying the SOS power to strengthening convex combinations of valid covering inequalities, as explained at the beginning of Section 4. Analyzing *equalities* like in (12) is less appropriate in these situations.

5 Set covering

Consider any $m \times n$ 0-1 matrix A, and let $\mathcal{F}_A$ be the feasible region for the 0-1 set covering problem defined by A:

$$\mathcal{F}_A = \{x \in \{0,1\}^n : Ax \geq e\} \tag{13}$$

where e is the vector of 1s. We denote by $A_i \subseteq \{1, \ldots, n\}$ the set of indices of nonzeros in the i-th row of A (namely the *support* of the i-th constraint). By overloading notation, we also use A_i to denote the corresponding set of variables $\{x_j : j \in A_i\}$. We will assume that A is *minimal*, i.e. there is no $i \neq j$ such that $A_i \subseteq A_j$.

We will also use the following notation. For any $T, F \subseteq [n]$ with $T \cap F = \emptyset$, let $\mathcal{F}_{A_{(T,F)}}$ denote the subregion of $\mathcal{F}_A$ where $x_i = 1$, for $i \in T$, and $x_j = 0$, for $j \in F$. Let $A_{(T,F)}$ be the matrix that is obtained from A by removing all the rows where x_i appears for $i \in T$ (these constraints are satisfied when $x_i = 1$ for $i \in T$) and setting to zero the j-th column for $j \in F$. We will assume that $A_{(T,F)}$ is minimal by removing the dominated rows. Therefore, $\mathcal{F}_{A_{(T,F)}} = \{x \in \{0,1\}^n : A_{(T,F)}x \geq e, x_i = 1\ \forall i \in T, x_j = 0\ \forall j \in F\}$ and $\mathcal{F}_{A_{(T,F)}} \subseteq \mathcal{F}_A$.

5.1 The spanning polynomials for the set covering problem

In this section we define the spanning polynomials for the set covering problem. For the sake of simplicity, we will assume that the collection of valid inequalities, i.e. $\{g_i(x) \geq 0, i \in [\ell]\}$, that are defined in the semialgebraic set (1) is given by $Ax \geq e$ *and* the nonnegative constraints $x \geq 0$. The latter is not strictly necessary, since $x_i = x_i^2$ and therefore $x_i \geq 0$, but this will simplify the exposition. We start with a simple structural observation regarding the set covering problem (see e.g. [29] for a proof).

Proposition 5.1. *Consider a set covering problem defined by any $m \times n$ 0-1 matrix B such that no two constraints overlap in any of the variables, namely for any $i, j \in [m]$ with $i \neq j$ we have $B_i \cap B_j = \emptyset$. When this holds then the linear constraints are convex hull defining: $conv(\mathcal{F}_B) = \{x \in [0,1]^n : Bx \geq e\}$.*

Remark 5.1. *From Proposition 5.1 it follows that any valid inequality $a^\top x \geq a_0$ for $\mathcal{F}_B$ is valid also for the feasible region of the linear relaxation $\{x \in [0,1]^n : Bx \geq e\}$, i.e. $a^\top x \geq a_0$ can be derived as a nonnegative linear combination and right-hand-side weakening from $\{x \geq 0, Bx \geq e\}$: $a = \lambda^\top B + \gamma^\top I$ and $a_0 \leq \lambda^\top e$, for some $\lambda, \gamma \geq 0$ and where I denotes the $n \times n$ identity matrix.*

By the previous observation the "interesting" variables are those that appear in more than one constraint. This gives the intuition why the $Q_A(t)$-SOS polynomials that we are going to define are polynomials in these variables.

Definition 5.1. *For any $t \in [n]$ and $\mathcal{C}(t) = \{C : C \subseteq [m] \wedge |C| \le t\}$, let $V_C = \bigcup_{i,j\in C, i\neq j} A_i \cap A_j$ be the set of variables occurring in more than one row with index from $C \in \mathcal{C}(t)$. The subspace of polynomials $Q_A(t)$ is (inductively) defined as the set of all polynomials $p(x) \in \mathbb{R}[\mathbf{x}]$ for which there exists a $C \in \mathcal{C}(t)$ and $I \subseteq C$ with $|I| \le t$ such that $p(x)$ can be written as $p(x) = q(x)r(x)$, where $q(x)$ is $(V_C \setminus I)$-symmetric and, depending on $|I|$, $r(x)$ is either 1 (if $|I| \in \{0, t\}$) or $r(x) \in Q_{A_{(I,V_c\setminus I)}}(t - |I|)$ (else).*

By Lemma B.1 (in appendix) and Lemma 3.3 a $Q_A(t)$-SOS representation can be decided by solving a semidefinite programming feasibility problem of size $n^{O(t^2)}$.

For any given inequality $a^\top x - a_0 \ge 0$ with indices ordered so that $0 < a_1 \le a_2 \le \cdots \le a_h$ and $a_j = 0$ for $j > h$, its *pitch* is the minimum integer $\pi = \pi(a, a_0)$ such that $\sum_{i=1}^{\pi} a_i - a_0 \ge 0$. The definition of pitch was introduced in [4, 29]. The main result of this section is the following.

Theorem 5.2. *Suppose $a^\top x - a_0 \ge 0$ is a valid inequality for $\mathcal{F}_A$ of pitch $\pi = \pi(a, a_0)$ with $a \ge 0$. Then $a^\top x - a_0$ admits a $Q_A(\pi)$-SOS representation.*

Corollary 5.3. *For any $k \ge 1$, any valid solution of the $\text{SOS}_{Q_A(k)}$ relaxation (9) satisfies all the valid inequalities for $\mathcal{F}_A$ of pitch $\le k$.*

Remark 5.2. *Note that for the set-covering problem with a full-circulant constraint matrix (namely $\sum_{j\neq i} x_j \ge 1$ for each $i = 1, \dots, n$) the pitch 2 valid inequality $\sum_{j=1}^n x_j \ge 2$ has rank at least $n - 3$ for a lifting operator stronger than the Sherali-Adams [4] and requires at least $\Omega(\log^{1-\varepsilon} n)$ levels [14] for the standard SOS hierarchy (conjectured to be $n/4$ in [4]). Viceversa, the augmented $\text{SOS}_{Q_A(k)}$ relaxation (9) returns a solution that satisfies all the pitch 2 valid inequalities in polynomial time ($Q_A(2)$, see Definition 5.1, is sufficient for this purpose).*

5.2 Proof of Theorem 5.2

The proof will be by induction on the pitch value. Consider *any* $m \times n$ 0-1 matrix A', and let $\mathcal{F}_{A'}$ be the feasible region for the 0-1 set covering problem defined by A': $\mathcal{F}_{A'} = \{x \in \{0,1\}^n : A'x \ge e\}$. Assume that $a'^\top x - a'_0 \ge 0$ with $a' \ge 0$ is a valid inequality for $\mathcal{F}_{A'}$ of pitch π'. If $\pi' = 0$ we must have $a'_0 \le 0$, so since $a' \ge 0$, $a'^\top x - a'_0$ has a trivial $Q_{A'}(\pi')$-SOS representation as conical combination of x_i, for $i \in [n]$. By induction hypothesis, from now on we will assume that for *any* $m \times n$ 0-1 matrix A' the claim holds for any constraint $a'^\top x - a'_0 \ge 0$ of pitch p, with $0 \le p \le \pi - 1$, that is valid for $\mathcal{F}_{A'}$.

We start describing a key structural property of valid inequalities for set covering that was proved in [4, 29]. We introduce the main property with a simple example.

Example 5.1. *Consider the set cover instance from the full-circulant constraint matrix (namely $\sum_{j\neq i} x_j - 1 \ge 0$ for each $i = 1, \dots, n$) and the pitch 2 valid inequality $\sum_{j=1}^n x_j - 2 \ge 0$. Then we can find two constraints, for example $\sum_{j\neq i} x_j - 1 \ge 0$ for $i \in \{1, 2\}$, such that, let $V = \{3, \dots, n\}$ be the variables in both constraints, then $(\sum_{j=1}^n x_j - 2)_{(\emptyset,V)} \ge 0$ is obtained by summing the two constraints after setting to zero the shared variables in V, i.e. $(\sum_{j\neq 1} x_j - 1)_{(\emptyset,V)} \ge 0$ plus $(\sum_{j\neq 2} x_j - 1)_{(\emptyset,V)} \ge 0$.*

We observe that the statement of Lemma 5.4 below is slightly different from Proposition 4.22 in [29] (or Theorem 6.3 in [4]). The difference is given by Property (17) (see Lemma 5.4). This property is not explicitly given in [4, 29], but it can be derived by their construction as explained in the proof of Lemma 5.4 (see Appendix C.1).

Lemma 5.4. *[4, 29] Suppose $a^\top x - a_0 \geq 0$ is a valid inequality for $\mathcal{F}_A$ with $a \geq 0$. Let supp(a) denote the support of a. Then there is a subset $C = C(a, a_0)$ of the rows of A with $|C| \leq \pi(a, a_0)$, such that*

$$A_i \subseteq supp(a) \quad \forall i \in C \tag{14}$$

$$B_i = A_i - \bigcup_{r \in C - \{i\}} A_r \neq \emptyset \quad \forall i \in C \tag{15}$$

$$(a^\top x - a_0)_{(\emptyset, V)} \geq 0 \text{ is valid for } \mathcal{F}_B = \{x \in \{0,1\}^n : \sum_{j \in B_i} x_j \geq 1, i \in C\} \tag{16}$$

$$\mathcal{F}_{A_{(\emptyset,V)}} \neq \emptyset \tag{17}$$

where $V := \bigcup_{\substack{i,j \in C \\ i \neq j}} A_i \cap A_j$ is the set of variables occurring in more than one row from C.

Consider any valid inequality $a^\top x - a_0 \geq 0$ for $\mathcal{F}_A$ of pitch $\pi = \pi(a, a_0)$ with $a \geq 0$. We show that $a^\top x - a_0$ admits a $Q_A(\pi)$-SOS representation. By Lemma 5.4 there is a subset $C = C(a, a_0)$ of the rows of A that satisfies (14)-(17) where V denotes the set of variables occurring in more than one row of C and $|C| \leq \pi$. The following polynomials (see (2) and Example 3.1) have a $Q_A(\pi)$-SOS representation: δ_J^V, for $J \subseteq V$ with $|J| < \pi$, and $\delta_{\geq \pi}^V := \sum_{J \subseteq V, |J| \geq \pi} \delta_J^V$ (it is zero if $|V| < \pi$). Note that $\sum_{J \subseteq V, |J| < \pi} \delta_J^V + \delta_{\geq \pi}^V = 1$. It follows that

$$
\begin{aligned}
a^\top x - a_0 &= \underbrace{\left(\sum_{J \subseteq V, |J| < \pi} \delta_J^V + \delta_{\geq \pi}^V \right)}_{=1} (a^\top x - a_0) \\
&= \underbrace{\delta_\emptyset^V (a^\top x - a_0)_{(\emptyset, V)}}_{\text{first}} + \underbrace{\left(\sum_{J \subseteq V, 0 < |J| < \pi} \delta_J^V (a^\top x - a_0)_{(J, V \setminus J)} \right)}_{\text{second}} + \underbrace{(\delta_{\geq \pi}^V)(a^\top x - a_0)}_{\text{third}}
\end{aligned} \tag{18}
$$

Therefore, showing that $a^\top x - a_0$ is $Q_A(\pi)$-SOS derivable boils down to prove that each of the summands in (18) is $Q_A(\pi)$-SOS derivable.

Sketch of the Proof: The first summands can be written as conical combination of valid inequalities (by Lemma 5.4); The second summand is a sum of valid smaller pitch inequalities (and therefore by induction can be written as $Q_{A_{(J,V \setminus J)}}(p)$-SOS representation); the last term is always nonnegative because we are considering pitch π, so the value of the inequality is always nonnegative by setting to one at least π variables. In case needed, full details are given below.

5.2.1 More details of the proof

Let's start considering the first summand in (18), namely $\delta_\emptyset^V (a^\top x - a_0)$. By Lemma 5.4, first note that $(a^\top x - a_0)_{(\emptyset, V)} \geq 0$ is valid for $\mathcal{F}_B$ (see (16)). Moreover, no two constraints in $\mathcal{F}_B$ overlap in any of the variables and therefore, by Proposition 5.1, the linear relaxation is convex hull defining: $conv(\mathcal{F}_B) = \{x \in [0,1]^n : \sum_{j\in B_i} x_j \geq 1, i \in C\}$. This means (see Remark 5.1) that $(a^\top x - a_0)_{(\emptyset,V)}$ can be implied by a conical combination of the linear constraints in $conv(\mathcal{F}_B) = \{x \in [0,1]^n : \sum_{j\in B_i} x_j \geq 1, i \in C\}$. Note that these linear constraints are just a subset of the linear constraints from $\{x \in [0,1]^n : Ax \geq e\}$ after setting to zero all the variables from V. It follows that $(a^\top x - a_0)_{(\emptyset,V)} = \big(\lambda^\top(Ax-e) + \gamma^\top x + \mu\big)_{(\emptyset,V)}$ for some $\lambda, \gamma, \mu \geq 0$. For any $x_i \in V$ we have $\delta_\emptyset^V x_i \equiv 0$ (recall that whenever we use "$\equiv$" we assume that the equivalence is $(\text{mod } \mathbf{I}(\mathbb{Z}_2^n))$) and

$$\delta_\emptyset^V(a^\top x - a_0) \equiv \delta_\emptyset^V\left(\lambda^\top(Ax-e) + \gamma^\top x + \mu\right)_{(\emptyset,V)} \equiv \delta_\emptyset^V\left(\lambda^\top(Ax-e) + \gamma^\top x + \mu\right)$$

$$\equiv \sum_{i\in C} \underbrace{(\delta_\emptyset^V \sqrt{\lambda_i})^2}_{s_i(x)} \underbrace{\left(\sum_{j\in A_i} x_j - 1\right)}_{g_i(x)} + \sum_{j\in supp(a)} \underbrace{(\delta_\emptyset^V \sqrt{\gamma_j})^2}_{s_j(x)} \underbrace{x_j}_{g_j(x)} + \underbrace{(\delta_\emptyset^V \sqrt{\mu})^2}_{s_0(x)}$$

Note that the latter has exactly the form in (5), where $\delta_\emptyset^V \in Q_A(\pi)$, and therefore it shows that the first summand $\delta_\emptyset^V(a^\top x - a_0)$ is $Q_A(\pi)$-SOS derivable.

Consider a generic second type summand from (18), i.e. $\delta_J^V(a^\top x - a_0)$ with $J \subseteq V, 0 < |J| < \pi$. Note that $a^\top x - a_0 \geq 0$ is by assumption a valid inequality for any feasible integral solution. By Property (17) we know that by setting to zero all the variables from V we obtain a non-empty subset of feasible integral solutions. It follows that by setting $x_j = 1$, for $j \in J$, and $x_h = 0$, for $h \in V \setminus J$, we obtain a non-empty subset of feasible integral solutions, i.e. $\mathcal{F}_{A_{(J,V\setminus J)}} \neq \emptyset$ and $(a^\top x - a_0)_{(J,V\setminus J)} \geq 0$ is a valid inequality for the solutions in $\mathcal{F}_{A_{(J,V\setminus J)}}$ (since $a^\top x - a_0 \geq 0$ is by assumption a valid inequality for any feasible integral solution). Moreover the pitch p of $(a^\top x - a_0)_{(J,V\setminus J)} \geq 0$ is strictly smaller than π, $0 \leq p \leq \pi - |J|$. It follows, by induction hypothesis that $(a^\top x - a_0)_{(J,V\setminus J)}$ has a $Q_{A_{(J,V\setminus J)}}(p)$-SOS representation, namely $(a^\top x - a_0)_{(J,V\setminus J)} \equiv s'_0(x) + \sum_i s'_i(x) g_i(x)_{(J,V\setminus J)}$ where $s'_i \in \{s' \in \mathbb{R}[\mathbf{x}] : s' = \sum_i q_i(x)^2, q_i \in Q_{A_{(J,V\setminus J)}}(p)\text{-SOS}\}$ and each $g_i(x)_{(J,V\setminus J)} \geq 0$ is a valid linear constraint for $\mathcal{F}_{A_{(J,V\setminus J)}}$, where $g_i(x)_{(J,V\setminus J)}$ is either $(\sum_{j\in A_h} x_j - 1)_{(J,V\setminus J)} \geq 0$ or $(x_j)_{(J,V\setminus J)} \geq 0$ for some $h \in [m], j \in [n]$. Then

$$\begin{aligned}\delta_J^V(a^\top x - a_0) &\equiv \delta_J^V(a^\top x - a_0)_{(J,V\setminus J)} \equiv \delta_J^V\left(s'_0(x) + \sum_i s'_i(x) g_i(x)_{(J,V\setminus J)}\right)\\ &\equiv s'_0(x)\delta_J^V + \sum_i s'_i(x)\delta_J^V g_i(x)\end{aligned}$$

Recall that $0 \leq p \leq \pi - |J|$ and $s'_i(x) = \sum_j q_j(x)^2$ for $q_j \in Q_{A_{(J,V\setminus J)}}(p)$-SOS therefore $\delta_J^V q_j(x)^2 \equiv (\delta_J^V q_j(x))^2$ and $\delta_J^V q_j(x) \in Q_A(\pi)$-SOS (by Definition 5.1). It follows that the second summand is $Q_A(\pi)$-SOS derivable.

Finally, consider the third summand from (18), i.e. $\delta_{\geq\pi}^V(a^\top x - a_0)$. Recall that we are assuming that $0 < a_1 \leq a_2 \leq \cdots \leq a_h$ and $a_j = 0$ for $j > h$ for some $h \in [n]$, so the $supp(a) = \{1, \ldots, h\}$. Moreover the pitch $\pi \leq h$ is the minimum such that $\sum_{i=1}^{\pi} a_i - a_0 \geq 0$. Note that $V \subseteq supp(a)$ and therefore if $\delta_{\geq\pi}^V$ is a non-null polynomial then $|V| \geq \pi$ (we assume this in the following otherwise we are done for this case). Let $a'_i := a_i$ for $i = 1, \ldots, \pi$, $a'_i := a_\pi$ for $i = \pi+1, \ldots, h$ and $a'_i := 0$ for $i \in supp(a) \setminus V$. Note that for any $I \subseteq [\pi]$ and $J_1, J_2 \subseteq V \setminus [\pi]$ with $|J_1| = |J_2| \geq \pi - |I|$ we have for $\ell = 1, 2$

$$\delta^V_{I\cup J_\ell}\left(\sum_{i\in V} a'_i x_i - a_0\right) \equiv \delta^V_{I\cup J_\ell}\left(\sum_{i\in I\cup J_\ell} a'_i - a_0\right) = \delta^V_{I\cup J_\ell}\overbrace{\left(\sum_{i\in I} a'_i + |J_\ell| a_\pi - a_0\right)}^{\geq 0}$$

Note that for any $k = \pi - |I|, \ldots, |V|$ the polynomial $\sum_{\substack{J\subseteq V\setminus[\pi]\\ |J|=k}} \delta^V_{I\cup J}$ is $(V\setminus[\pi])$-symmetric (see Example 3.1) and therefore it belongs to $Q_A(\pi)$. It follows that

$$\delta^V_{\geq\pi}\left(\sum_{i=1}^{h} a_i x_i - a_0\right) = \delta^V_{\geq\pi}\left(\sum_{i\in V} a_i x_i - a_0\right) + \delta^V_{\geq\pi}\left(\sum_{i\in supp(a)\setminus V} a_i x_i\right)$$

$$= \delta^V_{\geq\pi}\left(\sum_{i\in V} a'_i x_i - a_0\right) + \delta^V_{\geq\pi}\left(\sum_{i\in supp(a)} (a_i - a'_i) x_i\right)$$

$$= \overbrace{\sum_{I\subseteq V\cap[\pi]} \sum_{k=\pi-|I|}^{|V|} \sum_{\substack{J\subseteq V\setminus[\pi]\\ |J|=k}} \delta^V_{I\cup J}}^{=\delta^V_{\geq\pi}}\left(\sum_{i\in V} a'_i x_i - a_0\right) + \sum_{i\in supp(a)} ((a_i - a'_i)\delta^V_{\geq\pi}) x_i$$

$$\equiv \sum_{I\subseteq V\cap[\pi]} \sum_{k=\pi-|I|}^{|V|} \left(\sum_{\substack{J\subseteq V\setminus[\pi]\\ |J|=k}} \delta^V_{I\cup J} \overbrace{\left(\sum_{i\in I} a'_i + k a_\pi - a_0\right)}^{\geq 0}\right) + \sum_{i\in supp(a)} (\overbrace{(a_i - a'_i)}^{\geq 0}\delta^V_{\geq\pi}) x_i$$

$$\equiv \underbrace{\sum_{I\subseteq V\cap[\pi]} \sum_{k=\pi-|I|}^{|V|} \left(\overbrace{\sum_{\substack{J\subseteq V\setminus[\pi]\\ |J|=k}} \delta^V_{I\cup J}}^{(V\setminus[\pi])\text{-symmetric}} \sqrt{\sum_{i\in I} a'_i + k a_\pi - a_0}\right)^2}_{s_0(x)} + \sum_{i\in supp(a)} \underbrace{\left(\sqrt{a_i - a'_i}\ \overbrace{\delta^V_{\geq\pi}}^{V\text{-symm.}}\right)^2}_{s_i(x)} \underbrace{x_i}_{g_i(x)}$$

The latter has exactly the form in (5), and each polynomial under the square is from $Q_A(\pi)$, and therefore the third summand $\delta^V_{\geq\pi}(a^\top x - a_0)$ is $Q_A(\pi)$-SOS derivable and the claim follows.

5.3 An explicit compact LP formulation for the BZ hierarchy

As shown in Section 2.1, it turns out that $\text{SOS}_{\langle\mathcal{S}\rangle}$-relaxations are obtained by requiring that the linear functionals of summands of $\langle\mathcal{S}\rangle$-SOS proofs are nonnegative (see (7) and (8)).

By definition of $Q_A(t)$, recall that $q(x)^2 \equiv q(x) \pmod{\mathbf{I}(\mathbb{Z}_2^n)}$ for any $q \in Q_A(t)$. By applying the general arguments given in Section 2.1, note that any summand that appear in the $Q_A(t)$-SOS proof of any bounded pitch t inequality (as explicitly given in Section 5.2) is equal to $g_i(x)q(x)$ or equal to $q(x)$, for some $q(x) \in Q_A(t)$ and $g_i(x) \geq 0$ being any set cover constraint from $Ax \geq e$. It follows that any solution of the following *linear program* satisfies all the pitch t inequalities.

$$L[q(x)] \geq 0, \quad \forall q \in Q_A(t) \tag{19}$$

$$L[q(x)g_i(x)] \geq 0, \quad \forall q \in Q_A(t) \text{ and } g_i(x) \geq 0 \text{ is any constraint from } Ax \geq e \qquad (20)$$

As described in Section 2.1 in a more general setting, the above LP has polynomial size for any fixed pitch.

We remark that a more "explicit" and simplified formulation can be derived, but we decided to leave it as it is to emphasize the use of the high degree polynomials in the SOS framework.

6 The Bienstock-Zuckerberg hierarchy

The Bienstock-Zuckerberg hierarchy (BZ) [4, 29] generalizes the approach for set cover. The full description requires several layers of details and here we sketch only the main points. We refer to the original manuscripts for a more precise and comprehensive description.

Any non-trivial constraint can be rewritten in the set-cover form: $\sum_{i\in I} a_i x_i + \sum_{j\in J} a_j(1-x_j) \geq b$, with all the coefficients a, b nonnegative. Then the BZ hierarchy uses the standard concept of minimal covers[3] (see e.g. [7]): a *minimal cover* is an inclusion-minimal set $C \subseteq supp(a)$ such that $\sum_{j\notin C} a_j < b$ and therefore $\sum_{j\in C} x'_j \geq 1$ is a valid inequality (where $x'_j = x_j$ if $j \in I$ or $x'_j = 1 - x_j$ else). In general, the number of minimal covers can be exponential so the idea in BZ is to generate only the "k-small" ones, which are added to the original relaxation. Here with "k-small" we mean all the valid minimal covers with all the variables from I (or J) but at most k, or at most k from I (or J). These minimal covers can be enumerated in polynomial time for any fixed k. Then the set cover approach is applied to the set cover problem given by the k-small minimal covers. If the minimal covers are polynomially bounded this allows to generate the pitch bounded valid inequalities as for set cover (see the application below). Roughly speaking, the "power" of the BZ approach is given by the presence of the k-small minimal covers, if this set is empty then the hierarchy is not stronger than a variant of the Sherali-Adams hierarchy (see [2]).

The BZ approach can be reframed into the SOS framework by choosing the appropriate spanning polynomials. We omit the complete mapping because this would require the full description of BZ that is quite lengthy. Moreover the currently known most important application of BZ is given by the set cover problem, which has been fully explained in previous sections. By way of example, we show in Section 6.1 that we do not need to explicitly add the k-small minimal covers since they can implied by adding the "right" polynomials. By using the explained ideas, it should be easy to fulfill the missing details.

6.1 k-Small minimal covers

Consider a generic inequality of any given integer problem as written in the covering form, i.e. inequality $g(x) = a^\top x - a_0 \geq 0$ with $a \geq 0$ (here, abusing notation, every variable x_j is either the original one or its negation $1 - x_j$). For each such constraint let $V_a = supp(a)$ be the set of variables in this constraint. Add to the $Q_A(k)$-SOS polynomials the set of all (C)-symmetric polynomials with $C \subseteq V_a$ and $|C| \leq k$. Consider any valid k-small minimal cover of type $\sum_{i\in C} x_i \geq 1$, with $|C| \leq k$ (the other cases are similar). We sketch

[3] More precisely, in [4, 29] a closely related concept that is called *obstruction* is used.

that it is $Q_A(k)$-SOS derivable:

$$\sum_{i\in C} x_i - 1 = \left(\sum_{i\in C} x_i - 1\right)\overbrace{\left(\sum_{i=0}^{|C|} \underbrace{\sum_{I\subseteq C:|I|=i} \delta_I^C}_{symmetric}\right)}^{=1} \equiv \left(\sum_{i=0}^{|C|}(i-1) \underbrace{\sum_{I\subseteq C:|I|=i} \delta_I^C}_{symmetric}\right)$$

$$\equiv \overbrace{\left(\frac{-1}{\sum_{i\in V_a\setminus C} a_i - a_0}\right)}^{>0} \delta_0^C \overbrace{\left(\sum_{i\in V_a\setminus C} a_i + \sum_{i\in C} a_i x_i - a_0\right)}^{<0} + \left(\sum_{i=1}^{|C|}(i-1) \underbrace{\sum_{I\subseteq C:|I|=i} \delta_I^C}_{symmetric}\right)$$

$$\equiv \overbrace{\left(\frac{-1}{\sum_{i\in V_a\setminus C} a_i - a_0}\right)}^{>0} \delta_0^C \left(\sum_{i\in V_a} a_i x_i - a_0 + \sum_{i\in V_a\setminus C} a_i(1-x_i)\right) + \left(\sum_{i=1}^{|C|}(i-1) \underbrace{\sum_{I\subseteq C:|I|=i} \delta_I^C}_{symmetric}\right)$$

$$\equiv \underbrace{\left(\sqrt{\frac{-1}{\sum_{i\in V_a\setminus C} a_i - a_0}}\delta_0^C\right)^2}_{s(x)} \underbrace{\left(\sum_{i\in V_a} a_i x_i - a_0\right)}_{g(x)} + \sum_{i\in V_a\setminus C} \underbrace{\left(\sqrt{\frac{-a_i}{\sum_{i\in V_a\setminus C} a_i - a_0}}\delta_0^C\right)^2}_{s_i(x)} \underbrace{(1-x_i)}_{g_i(x)} +$$

$$+ \underbrace{\left(\sum_{i=1}^{|C|} \sqrt{i-1} \sum_{I\subseteq C:|I|=i} \delta_I^C\right)^2}_{s_0(x)}$$

6.1.1 An application.

As in [4, 29], Theorem 5.2 can be generalized to handle 0/1 integer problems with non-negative constraints having pitch bounded by a constant p. More precisely, consider the feasible region for the 0-1 problem defined by A:

$$\mathcal{F}_A = \{x \in \{0,1\}^n : Ax \geq b\} \tag{21}$$

where $b \in \mathbb{R}_+^m$ and each constraint in $Ax \geq b$ has pitch at most p. (For example any inequality $a^\top x - a_0 \geq 0$ with nonnegative integral coefficients $a_i \in \{0, 1, \dots, p\}$ has pitch at most p.) In this case the number of minimal covers is polynomially bounded. Since the integral polytope defined by using the minimal covers and the integrality constraints coincides with (21) (see e.g. [7]), then we can extend Theorem 5.2 to this more general case.

7 Chvátal-Gomory Cuts

Consider the rational polyhedra $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ with $A \in \mathbb{Z}^{m\times n}$ and $b \in \mathbb{Z}^m$. Inequalities of the form $(\lambda^\top A)x \geq \lceil \lambda^\top b \rceil$, with $\lambda \in \mathbb{R}^m$, $\lambda^\top A \in \mathbb{Z}^n$, and $\lambda^\top b \notin \mathbb{Z}$ are commonly referred to Chvátal-Gomory cuts (CG-cuts for short), see e.g. [7]. CG-cuts are valid for the integer hull, $P_I = conv\{x \in \{0,1\}^n : Ax \geq b\}$, of P.

It is well-known that it is sufficient to consider λ-vectors with small entries:

$$P^{(1)} := \{(\lambda^\top A)x \geq \lceil \lambda^\top b \rceil, \lambda \in [0,1]^m, \lambda^\top A \in \mathbb{Z}^n\} \tag{22}$$

where $P^{(1)}$ is commonly referred to as the *first Chvátal-Gomory (CG) closure*. In particular $P^{(1)}$ is a stronger relaxation of P_I than P, i.e. $P_I \subseteq P^{(1)} \subseteq P$. We can iterate the closure process to obtain the CG closure of $P^{(1)}$. We denote by $P^{(2)}$ this second CG closure. Iteratively, we define the t-th CG closure $P^{(t)}$ of P to be the CG closure of $P^{(t-1)}$, for $t \geq 2$ integer. An inequality that is valid for $P^{(t)}$ but not $P^{(t-1)}$ is said to have *CG-rank t*.

Eisenbrand and Schulz [8] prove that for any polytope P contained in the unit cube $[0,1]^n$, one can choose $t = O(n^2 \log n)$ and obtain the integer hull $P^{(t)} = P_I$. Rothvoss and Sanitá [27] prove that there is a polytope contained in the unit cube whose CG-rank has order n^2, thus showing that the above bound is tight, up to a logarithmic factor.

The CG-cuts that are valid for $P^{(1)}$ and that can be derived by using coefficients in λ of value 0 or 1/2 only are called $\{0, 1/2\}$-cuts. In [20] it is shown that the separation problem for $\{0, 1/2\}$-cuts remains strongly NP-hard even when all integer variables are binary and $P = \{x \in \mathbb{R}^n_+ : Ax \leq \mathbf{1}\}$ with $A \in \{0,1\}^{m\times n}$ and $\mathbf{1}$ denote the all-one vector with m entries. As pointed out in [20], the latter hardness proof can easily be adapted to set partitioning and set covering problems. This result implies that it is NP-hard to optimize a linear function over the first closure $P^{(1)}$. This provides an interesting contrast to lift-and-project hierarchies (like Sherali-Adams, Lovász-Schrijver, and SOS/Lasserre) where one can optimize in polynomial time for any constant number of levels.

For an arbitrary fixed precision $\varepsilon > 0$ and fixed positive integer q, choose π such that $\left(\frac{\pi+1}{\pi}\right)^q \leq 1+\varepsilon$. Bienstock and Zuckerberg (see Lemma 2.1 in [5]) prove that any solution that satisfies the set of valid inequalities of pitch π can be rounded to approximate all the CG-cuts constraint of rank q to precision $\varepsilon > 0$. It follows that the SOS approach with high degree polynomials described in this paper computes fixed rank CG $(1+\varepsilon)$-approximate solutions for any fixed $\varepsilon > 0$ in polynomial time as well (PTAS).

In the next section we present a somehow stronger result for packing problems, meaning that the coefficients of the nonnegative matrix A are not restricted to be 0/1 (or bounded, see Section 6.1) as for the set cover case. It remains an interesting open question to extend the result for set cover to general covering problems, namely for general nonnegative matrix A.

7.1 Approximate fixed-rank CG closure for packing problems

In the following we observe that for the packing problem the standard SOS hierarchy with bounded degree polynomials is sufficient to obtain fixed rank CG $(1-\varepsilon)$-approximate solutions. It follows that the SOS approach can be used for approximating CG cuts of any fixed rank and to any arbitrary precision for both, packing and set covering problems (BZ guarantees this only for set covering problems).

Consider any $m \times n$ nonnegative matrix A, and let $\mathcal{P}$ be the feasible region for the 0-1 set packing problem defined by A: $\mathcal{P} = \{x \in \{0,1\}^n : Ax \leq b\}$ where $b \in \mathbb{R}^m_+$. For an integer $t \geq 0$, denote by $P^{(t)}$ the t-th CG closure and let $cg^{(t)}(c) := \max\{c^\top x : x \in P^{(t)}\}$. Without loss of generality, we will assume that $c \in \mathbb{R}^n_+$ (otherwise it is always optimal to set $x_i = 0$ whenever $c_i \leq 0$).

We can extend the definition of pitch also for packing inequalities as follows. For any given packing inequality $a_0 - a^\top x \geq 0$ with $a_0, a \geq 0$ and indices ordered so that $0 < a_1 \leq a_2 \leq \cdots \leq a_h$ and $a_j = 0$ for $j > h$, its *pitch* is the *maximum* integer $\pi = \pi(a, a_0)$

such that $a_0 - \sum_{i=1}^{\pi} a_i \geq 0$. For example, classical clique inequality $\sum_{i \in clique} x_i \leq 1$ have pitch equal to one.

The following result for packing problems can be seen as the dual of Theorem 5.2 for set cover. It can be easily obtained by using the so called "Decomposition Theorem" due to Karlin, Mathieu, and Nguyen [13]. For completeness, here we sketch a direct simple proof that follows the approach used throughout this paper.

Lemma 7.1. *Suppose $a_0 - a^\top x \geq 0$ is a valid inequality for $\mathcal{P}$ of pitch $\pi = \pi(a, a_0)$ with $a_0, a \geq 0$. Then $a_0 - a^\top x$ admits a $\mathbb{R}_{\pi+1}[\mathbf{x}]$-SOS representation.*[4]

Proof. Let $S = supp(a)$ and $x^I = \prod_{i \in I} x_i$. Note that for any $I \subseteq S$ we have $x^I(a_0 - a^\top x) \equiv x^I(a_0 - \sum_{i \in I} a_i - \sum_{i \notin I} a_i x_i) \pmod{\mathbf{I}(\mathbb{Z}_2^n)}$. Let $F := \{I \subseteq S : (a_0 - \sum_{i \in I} a_i) < 0\}$ and $T := \{J \subseteq S : J \notin F\}$ (and therefore if we set to one all the variables x_i with $i \in I$ for any $I \in F$ then the assumed valid inequality $a_0 - a^\top x \geq 0$ is violated). Let $V := \{x \in \mathbb{R}^n : x^I = 0 \ \forall I \in F, \ x_k^2 - x_k = 0 \ \forall k \in [n]\}$ and note that any feasible integral solution belongs to V. Any δ_J^S is actually equivalent $\pmod{\mathbf{I}(V)}$ to a polynomial $\bar{\delta}_J^S$ of degree at most π (obtained from δ_J^S by zeroing all the monomials x^I with $I \in F$ and therefore at least all the monomials of degree larger than π). Note that $\sum_{I \subseteq [n]} \bar{\delta}_I^S = \sum_{I \subseteq T} \bar{\delta}_I^S = 1$, $(\bar{\delta}_I^S)^2 \equiv \bar{\delta}_I^S \pmod{\mathbf{I}(V)}$ and $\bar{\delta}_I^S(a_0 - a^\top x) \equiv \bar{\delta}_I^S(a_0 - \sum_{i \in I} a_i) \pmod{\mathbf{I}(V)}$. Then

$$a_0 - a^\top x = \Big(a_0 - a^\top x\Big) \overbrace{\left(\sum_{I \subseteq T} \bar{\delta}_I^S\right)}^{=1} \equiv \underbrace{\sum_{I \in T} \overbrace{\left(a_0 - \sum_{i \in I} a_i\right)}^{\geq 0} (\bar{\delta}_I^S)^2}_{s_0(x)} \pmod{\mathbf{I}(V)} \qquad (23)$$

From the above equivalence we see that $a_0 - a^\top x$ can be written $\pmod{\mathbf{I}(V)}$ as a conical combination of squares of polynomials of degree at most π.

By definition of the equivalences $\pmod{\mathbf{I}(V)}$ and $\pmod{\mathbf{I}(\mathbb{Z}_2^n)}$, we can now easily transform the equivalence $\pmod{\mathbf{I}(V)}$ into the equivalence $\pmod{\mathbf{I}(\mathbb{Z}_2^n)}$ as given by (5) by adding some polynomials from $\mathbf{I}(V) \setminus \mathbf{I}(\mathbb{Z}_2^n)$. It is easy to argue that these polynomials have degree $O(\pi)$. (More details will appear in the longer version of this paper.) □

Let $Pr(d)$ denote the set of feasible solutions for $\mathbb{R}_{d+1}[\mathbf{x}]$-SOS projected to the original variables. The following simple result shows that fixed rank CG closures of packing problems can be approximated to any arbitrarily precision in polynomial time by using the SOS hierarchy.

Theorem 7.2. *For each integer $t \geq 0$ and $\varepsilon > 0$ there are integers $d = d(t, \varepsilon)$ such that $\max\{c^\top x : x \in Pr(d)\} \leq (1 + \varepsilon) cg^{(t)}$, for any $c \in \mathbb{R}_+^n$.*

Proof. For any fixed $\varepsilon > 0$ and integer $t \geq 0$ choose $d > 0$ integral large enough that $((d+1)/d)^t \leq 1 + \varepsilon$. Consider the solution $x^{(\ell)}$ obtained by multiplying any given solution $x \in Pr(d)$ by a factor equal to $(\frac{d}{d+1})^\ell$. It follows that $\max\{c^\top x : x \in Pr(d)\}$ is not larger than a factor of $(\frac{d+1}{d})^\ell$ of the value of $x^{(\ell)}$. Now the claim follows by showing that $x^{(t)}$ is feasible for the rank-t CG closure.

The proof is by induction on t. As a base of induction note that when $t = 0$ then clearly $x^{(0)}$ satisfies all the original constraints. Assume now, by induction hypothesis, that the claim is true for any rank equal to $(t - 1)$ with $t \geq 1$ and we need to show that it is valid

[4]This is the standard bounded degree SOS proof system.

also for rank-t. If the pitch of a generic rank-t valid inequality for $P^{(t)}$ is at most d then by Lemma 7.1 it follows that any feasible solution $x \in Pr(d)$ (and therefore $x^{(\ell)}$) satisfies this inequality. Otherwise, consider a generic rank-t valid inequality $\lfloor a_0 \rfloor - a^\top x \geq 0$ of pitch larger than d, where $a_0 - a^\top x \geq 0$ is any valid inequality from the closure $P^{(t-1)}$. By induction hypothesis note that $a_0 - a^\top x^{(t-1)} \geq 0$. Since the pitch is higher than d then $a_0 > d$ (vector a can be assumed, w.l.o.g., to be nonnegative and integral) and therefore $\frac{a_0}{\lfloor a_0 \rfloor} \leq \frac{d+1}{d}$ and by multiplying the solution $x^{(t-1)} \in P^{(t-1)}$ by $d/(d+1)$ we obtain a feasible solution for the rank-t CG closure. □

Open Problems. It would be nice to understand if it is possible (i) to generalize Theorem 5.2 to work with general covering problems, (ii) to get a PTAS to approximate all the CG-cuts constraints for more general problems.

Acknowledgment. This paper is dedicated to Elsa.

References

[1] Y. H. Au and L. Tunçel. A comprehensive analysis of polyhedral lift-and-project methods. *SIAM J. Discrete Math.*, 30(1):411–451, 2016.

[2] Y. H. Au and L. Tunçel. Elementary polytopes with high lift-and-project ranks for strong positive semidefinite operators. *arXiv preprint arXiv:1608.07647*, 2016.

[3] N. Bansal. Hierarchies reading group. http://www.win.tue.nl/ nikhil/hierarchies/index.html.

[4] D. Bienstock and M. Zuckerberg. Subset algebra lift operators for 0-1 integer programming. *SIAM Journal on Optimization*, 15(1):63–95, 2004.

[5] D. Bienstock and M. Zuckerberg. Approximate fixed-rank closures of covering problems. *Math. Program.*, 105(1):9–27, 2006.

[6] G. Blekherman, P. A. Parrilo, and R. R. Thomas. *Semidefinite optimization and convex algebraic geometry*, volume 13. Siam, 2013.

[7] M. Conforti, G. Cornuejols, and G. Zambelli. *Integer Programming*. Springer Publishing Company, Incorporated, 2014.

[8] F. Eisenbrand and A. S. Schulz. Bounds on the chvátal rank of polytopes in the 0/1-cube. *Combinatorica*, 23(2):245–261, 2003.

[9] H. Fawzi, J. Saunderson, and P. A. Parrilo. Equivariant semidefinite lifts and sum-of-squares hierarchies. *SIAM Journal on Optimization*, 25(4):2212–2243, 2015.

[10] S. Fiorini, T. Huynh, and S. Weltge. Strengthening convex relaxations of 0/1-sets using boolean formulas. *https://arxiv.org/abs/1711.01358v1*, November, 2017.

[11] K. Gatermann and P. A. Parrilo. Symmetry groups, semidefinite programs, and sums of squares. *Journal of Pure and Applied Algebra*, 192(1):95–128, 2004.

[12] D. Grigoriev, E. A. Hirsch, and D. V. Pasechnik. Complexity of semialgebraic proofs. *Moscow Mathematical Journal*, 2(4):647–679, 2002.

[13] A. R. Karlin, C. Mathieu, and C. T. Nguyen. Integrality gaps of linear and semi-definite programming relaxations for knapsack. In *Integer Programming and Combinatoral Optimization - 15th International Conference, IPCO*, pages 301–314, 2011.

[14] A. Kurpisz, S. Leppänen, and M. Mastrolilli. Sum-of-squares hierarchy lower bounds for symmetric formulations. In *Integer Programming and Combinatorial Optimization - 18th International Conference, IPCO 2016*, pages 362–374, 2016.

[15] A. Kurpisz, S. Leppänen, and M. Mastrolilli. Tight sum-of-squares lower bounds for binary polynomial optimization problems. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016*, pages 78:1–78:14, 2016.

[16] A. Kurpisz, S. Leppänen, and M. Mastrolilli. On the hardest problem formulations for the 0/1 lasserre hierarchy. *Math. Oper. Res.*, 42(1):135–143, 2017.

[17] A. Kurpisz, S. Leppänen, and M. Mastrolilli. An unbounded sum-of-squares hierarchy integrality gap for a polynomially solvable problem. *Mathematical Programming*, Jan 2017.

[18] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[19] M. Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0-1 programming. *Mathematics of Operations Research*, 28(3):470–496, 2003.

[20] A. N. Letchford, S. Pokutta, and A. S. Schulz. On the membership problem for the {0, 1/2}-closure. *Oper. Res. Lett.*, 39(5):301–304, 2011.

[21] M. Mastrolilli. High degree sum of squares proofs, bienstock-zuckerberg hierarchy and CG cuts. In *Integer Programming and Combinatorial Optimization - 19th International Conference, IPCO 2017, Waterloo, ON, Canada, June 26-28, 2017, Proceedings*, pages 405–416, 2017.

[22] Y. Nesterov. *Global quadratic optimization via conic relaxation*, pages 363–384. Kluwer Academic Publishers, 2000.

[23] P. Parrilo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization.* PhD thesis, California Institute of Technology, 2000.

[24] P. A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Math. Program.*, 96(2):293–320, 2003.

[25] P. Pudlák. Lower bounds for resolution and cutting plane proofs and monotone computations. *J. Symb. Log.*, 62(3):981–998, 1997.

[26] A. Raymond, J. Saunderson, M. Singh, and R. R. Thomas. Symmetric sums of squares over k-subset hypercubes. *arXiv preprint arXiv:1606.05639*, 2016.

[27] T. Rothvoß and L. Sanità. 0/1 polytopes with quadratic chvátal rank. In *Integer Programming and Combinatorial Optimization - 16th International Conference, IPCO*, pages 349–361, 2013.

[28] N. Shor. Class of global minimum bounds of polynomial functions. *Cybernetics and Systems Analysis*, 23(6):731–734, 1987.

[29] M. Zuckerberg. *A set theoretic approach to lifting procedures for 0, 1 integer programming.* PhD thesis, Columbia University, 2004.

A Background material

Definition A.1. *The ideal generated by a finite set of polynomials* $\{f_1, \dots, f_m\}$ *in* $\mathbb{R}[\mathbf{x}]$ *is defined as*

$$\boldsymbol{ideal}(f_1, \dots, f_m) := \{f | f = \sum_{i=1}^{m} t_i f_i \quad t_i \in \mathbb{R}[X]\}$$

Definition A.2. *The set of polynomials that vanish in a given set* $S \subset \mathbb{R}^n$ *is called the* vanishing ideal *of* S *and denoted:*

$$\mathbf{I}(S) := \{f \in \mathbb{R}[\mathbf{x}] : f(a_1, \dots, a_n) = 0 \qquad \forall (a_1, \dots, a_n) \in S\}$$

Definition A.3. *Let $\{f_1, \dots, f_m\}$ in $\mathbb{R}[\mathbf{x}]$ be a finite set of polynomials in $\mathbb{R}[\mathbf{x}]$. Let $\boldsymbol{V}$ be*

$$\boldsymbol{V}(f_1, \dots, f_m) := \{(a_1, \dots, a_n) \in \mathbb{R}^n | f_i(a_1, \dots, a_n) = 0 \quad 1 \leq i \leq m\}$$

We call $\boldsymbol{V}(f_1, \dots, f_m)$ the affine variety *defined by $f_1, \dots, f_m$.*

Definition A.4. *Let I be an ideal, and let $f, g \in \mathbb{R}[\mathbf{x}]$. We say f and g are* congruent modulo I*, written*

$$f \equiv g \pmod{I}$$

if $f - g \in I$. The quotient ring $\mathbb{R}[\mathbf{x}]/I$ is the set of equivalence classes for congruence modulo I.

Quotient rings are particularly useful when considering a polynomial function $p(x)$ over the algebraic variety defined by $\mathbf{V}(f_1, \dots, f_m)$.

Proposition A.1. *If we define the ideal $I = \boldsymbol{ideal}(f_1, \dots, f_m)$, then any polynomial q that is congruent with p modulo I takes exactly the same values in the variety.*

The set of functions that vanish on the boolean hypercube $\mathbb{Z}_2^n$ is the ideal $\mathbf{I}(\mathbb{Z}_2^n)$ in $\mathbb{R}[\mathbf{x}]$ generated by the polynomials $x_i^2 - x_i$ for all $i \in [n]$. In this paper we focus on $\mathbb{R}[\mathbb{Z}_2^n]$, i.e. the set of functions on $\mathbb{Z}_2^n$. This set can be identified with the quotient ring $\mathbb{R}[\mathbf{x}]/\mathbf{I}(\mathbb{Z}_2^n)$. The elements of $\mathbb{R}[\mathbf{x}]/\mathbf{I}(\mathbb{Z}_2^n)$ are in bijection with the square-free polynomials in $\mathbb{R}[\mathbf{x}]$, namely, those polynomials in which every monomial is square-free or multilinear. As a vector space, it will be convenient to identify $\mathbb{R}[\mathbb{Z}_2^n]$ with the set of all square-free polynomials in $\mathbb{R}[\mathbf{x}]$.

A multivariate polynomial is a *sum of squares* (SOS) if it can be written as the sum of squares of some other polynomials. Formally, we have the following.

Definition A.5. *A polynomial $p \in \mathbb{R}[\mathbf{x}]$ is a* sum of squares (SOS) *if there exists $q_1, \dots, q_m \in \mathbb{R}[\mathbf{x}]$ such that $p = \sum_{i=1}^m q_i^2$.*

It quickly follows from its definition that SOS polynomials are invariant under nonnegative scalings and convex combinations; i.e., it is a convex cone. If a feasible set is defined by a system of inequalities $\{g_1(x) \geq 0, \dots, g_\ell(x) \geq 0\}$ then we can define the following nonnegative region on the feasible set.

Definition A.6. *Given a set of multivariate polynomials $\{g_1, \dots, g_\ell\}$, let*

$$\boldsymbol{cone}(g_1, \dots, g_\ell) := \{g | g = s_0 + \sum_{i=1}^{\ell} s_i g_i\}$$

where each $s_i \in \mathbb{R}[\mathbf{x}]$ is a SOS.

These algebraic objects will be used for deriving *valid inequalities*, which are logical consequences of the given constraints. Note that by construction, every polynomial in $\mathbf{ideal}\,(f_i)$ vanishes in the solution set of $f_i(x) = 0$. Similarly, every element of $\mathbf{cone}(g_i)$ is clearly nonnegative on the feasible set of $g_i(x) \geq 0$.

Notice that as geometric objects, ideals are affine sets, and cones are closed under convex combinations and nonnegative scalings (i.e., they are actually cones in the convex geometry sense). These convexity properties, coupled with the relationships between SDP and SOS, will be key for our developments.

B Sum of squares over the boolean hypercube

B.1 The complexity of computing Q-SOS representations

Lemma B.1. *Consider any finite set of polynomials $\mathcal{S} \subseteq \mathbb{R}[\mathbf{x}]/\mathbf{I}(\mathbb{Z}_2^n)$ with $|\mathcal{S}| = n^{O(d)}$ and let $Q = \langle \mathcal{S} \rangle$. Then the existence of a Q-SOS representation can be decided by solving a semidefinite programming feasibility problem. The dimension of the matrix inequality is bounded by $n^{O(d)}$.*

In the following we sketch the proof of the above lemma. If f is Q-SOS, then a SOS certificate for it can be found by solving semidefinite programs (see e.g. [6] for more details and Example B.1 below). Indeed, by simply expanding the right-hand side of (4) as expressed in multilinear form, and matching the coefficient of x, we obtain a system of linear equalities with the coefficients of polynomials $s_i(x)$ as variables. We need now to impose that each $s_i(x)$ is a SOS. Let us review this. We start considering the case that the polynomials are expressed in the standard monomial basis. Let $\mathcal{P}([n])$ denote the collection of all subsets of $[n]$ and let $N = |\mathcal{P}([n])|$. Then any polynomial can be seen as a vector in $\mathbb{R}^{\mathcal{P}(N)}$ and let x denote the vector of all possible multilinear monomials of degree at most n (so x is a vector with N entries).

Lemma B.2. *Let $s(x) \in \mathbb{R}[\mathbf{x}]$. The following statements are equivalent:*

1. *$s(x)$ has a representation as a sum of squares in $\mathbb{R}[\mathbf{x}]$.*
2. *There is a matrix W such that $s(x) = x^\top W x$ with $W \succeq 0$ and x is the vector of different multilinear monomials.*

Proof. The matrix W is PSD if and only if there is a factorization $W = V^\top V$. If this holds then $s(x) = x^\top W x = x^\top V^\top V x = (Vx)^\top (Vx) = \sum_i ((Vx)_i)^2$ is a SOS. Vice versa, if $s(x) = \sum_i ((Vx)_i)^2$ then going backward in the previous equality the claim follows. $\square$

By using the previous lemma it follows that $s(x)$ is a SOS if and only if there is a symmetric matrix W (known as the Gram matrix of the SOS representation) that satisfies: $s(x) = x^\top W x,\ W \succeq 0$. Notice that the latter is a semidefinite program, since $s(x) = x^\top W x$ is affine in the matrix W, and thus the set of possible Gram matrices W is given exactly by the intersection of an affine subspace and the cone of positive semidefinite matrices.

Consider any finite set of polynomials $\mathcal{S} \subseteq \mathbb{R}[\mathbf{x}]/\mathbf{I}(\mathbb{Z}_2^n)$ with $|\mathcal{S}| = n^{O(d)}$ and let $Q = \langle \mathcal{S} \rangle$ (for any positive constant d). Let S be the matrix having as columns the spanning set $\mathcal{S}$. It follows that for any vector $q \in Q$ there is a vector $u \in \mathbb{R}^{|\mathcal{S}|}$ such that $q = Su$.

Since we are assuming that $s(x)$ is Q-SOS then $s(x) = \sum_i ((Vx)_i)^2$ and each $(Vx)_i$ belongs to Q and therefore there exists a $u_i \in \mathbb{R}^{|\mathcal{S}|}$ such that $(Su_i)^\top x = (Vx)_i$. Let U denote the matrix whose columns are the u_i, then we have the following: $\sum_i ((Vx)_i)^2 = x^\top S(UU^\top)S^\top x$. Polynomials are expressed in the new basis $S^\top x$ (this basis is in general not isomorphic to the standard monomial basis of degree d) and the complexity is given by the size of the matrix $UU^\top$, i.e. $n^{O(d)}$.

By the above discussion it follows that a Q-SOS proof (i.e. a proof where the nonnegative can be proved only for polynomials that admit a Q-SOS representation) is completely specified by giving the set of spanning polynomials.

Example B.1. *Consider the following set $\mathcal{F} = \{x \in \mathbb{R}^2 : x_1 - x_1^2 = x_2 - x_2^2 = 0, x_1 + x_2 - \varepsilon \geq 0\}$ where $\varepsilon \in (0,1)$. We want to show that the valid inequality $x_1 + x_2 - 1 \geq 0$ admits* *a $\langle\mathcal{S}\rangle$-SOS representation, where $\mathcal{S}$ is the set of the elementary symmetric polynomials in 2 variables, i.e $\mathcal{S} = \{1, x_1 + x_2, x_1x_2\}$ and therefore $\langle\mathcal{S}\rangle$ is the ring of all symmetric polynomials. Let $x = [1, x_1, x_2, x_1x_2]^\top$, the matrix S is equal to*

$S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ *and the new basis is* $S^\top x = [1, x_1 + x_2, x_1x_2]^\top$.

We want to show that $\mathcal{F} \vdash_{\langle\mathcal{S}\rangle} x_1 + x_2 - 1 \geq 0$ therefore we need to show that

$$x_1 + x_2 - 1 \equiv s_0(x) + s_1(x)(x_1 + x_2 - \varepsilon) \pmod{\mathbf{I}(\mathbb{Z}_2^n)} \tag{24}$$

where $s_0, s_1 \in \{s \in \mathbb{R}[\mathbf{x}] : s = \sum_i q_i(x)^2, q_i \in \langle\mathcal{S}\rangle\}$. By Lemma B.2 there are two PSD matrices W_0 and W_1 such that $s_0(x) = x^\top W_0 x$ and $s_1(x) = x^\top W_1 x$ with the additional constraint on the structure of W_0 and W_1 given by the restriction that $q_i \in \langle\mathcal{S}\rangle$. Let us first perform the change of basis $\sigma_i = (S^\top x)_i$ for $i = 0, 1, 2$. So the new variables are $\sigma_0 = 1$, $\sigma_1 = x_1 + x_2$ and $\sigma_2 = x_1x_2$ and the corresponding vector form $\sigma = [1, \sigma_1, \sigma_2]^\top$. Note that in the new basis $\sigma_1^2 \equiv \sigma_1 + 2\sigma_2 \pmod{\mathbf{I}(\mathbb{Z}_2^n)}$, $\sigma_2^2 \equiv \sigma_2 \pmod{\mathbf{I}(\mathbb{Z}_2^n)}$ and $\sigma_1\sigma_2 \equiv 2\sigma_2 \pmod{\mathbf{I}(\mathbb{Z}_2^n)}$ which correspond to the multilinear forms in the new basis. By rephrasing our goal in the new basis, we need to show that

$$\sigma_1 - 1 \equiv \sigma^\top T_0 \sigma + (\sigma^\top T_1 \sigma)(\sigma_1 - \varepsilon) \pmod{\mathbf{I}(\mathbb{Z}_2^n)} \tag{25}$$

for some PSD matrices T_0, T_1 with $T_i = \begin{bmatrix} t_{i00} & t_{i01} & t_{i02} \\ t_{i01} & t_{i11} & t_{i12} \\ t_{i02} & t_{i12} & t_{i22} \end{bmatrix}$ for $i = 0, 1$. By writing (25) *in the multilinear form, our goal is to prove that there are two PSD matrices T_0, T_1 such that the following is satisfied:*

$$\sigma_1 - 1 = \underbrace{t_{000} - \varepsilon t_{100}}_{\alpha} + \underbrace{(t_{011} + 2t_{001} + t_{100} + (t_{111} + 2t_{101})(1 - \varepsilon))}_{\beta}\sigma_1 +$$

$$\underbrace{(2t_{011} + t_{022} + 2t_{002} + 4t_{012} + 6t_{111} + 4t_{101} + 2t_{122} + 4t_{102} + 8t_{112} - \varepsilon(2t_{111} + t_{122} + 2t_{102} + 4t_{112})}_{\gamma}\sigma_2$$

So the solution of the following SDP= $\{\alpha = -1, \beta = 1, \gamma = 0, T_0 \succeq 0, T_1 \succeq 0\}$ gives the desired $\langle\mathcal{S}\rangle$-SOS representation. By choosing $T_0 = [0, 0, 1]^\top[0, 0, 1]$ and $T_1 = \frac{1}{\varepsilon}[1, -1, 1]^\top[1, -1, 1]$ the SDP is satisfied.

B.2 The complexity of $\langle\mathcal{S}\rangle$-SOS relaxation (9)

Lemma B.3. *Consider any $\mathcal{S} \subseteq \mathbb{R}[\mathbf{x}]/\mathbf{I}(\mathbb{Z}_2^n)$. Then the $\langle\mathcal{S}\rangle$-SOS relaxation* (9) *is a semidefinite programming with $m + 1$ matrix inequalities. The dimension of each matrix inequality is bounded by $|\mathcal{S}|$.*

We sketch the proof of the above lemma. The positivity condition (7) corresponds to the fact that, when viewed as a matrix, $L[\cdot]$ is positive semidefinite. Indeed, consider the (exponentially large) *moment matrix* $M \in \mathbb{R}^{\mathcal{P}([n])} \times \mathbb{R}^{\mathcal{P}([n])}$ where the generic entry $M_{I,J} = y_{I\cup J} = L\left[\prod_{i\in I\cup J} x_i\right]$, with $I, J \in \mathcal{P}([n])$ and $y_\emptyset = L[1]$. Then any multilinear polynomial $q \in \langle\mathcal{S}\rangle$ can be written as $q(x) = \sum_{I\subseteq[n]} q_I \prod_{i\in I} x_i$ for some q_I, as written in the standard monomial basis. Note that we are using and restricting to multilinear polynomials. Let q denote the vector representation of $q(x)$, i.e. the column vector of the coefficients of $q(x)$, then condition (7) is equivalent to require $q^\top M q \geq 0 \ \forall\, q \in \langle\mathcal{S}\rangle$.

We now consider a change of basis. Let S denote the matrix whose columns are the vectors from $\mathcal{S}$. The vector representation of $q \in \langle \mathcal{S} \rangle$ belongs to the vector space of matrix S, namely $q = Su$, for some vector $u \in \mathbb{R}^{|\mathcal{S}|}$. It follows that if we restrict to polynomials $q \in \langle \mathcal{S} \rangle$ then condition (7) is equivalent to require

$$q^\top M q = u^\top \underbrace{\left(S^\top M S\right)}_{M_{\langle \mathcal{S} \rangle}} u \geq 0 \qquad \forall\, u \in \mathbb{R}^{|\mathcal{S}|}$$

and therefore condition (7) is equivalent to the PSD condition $M_{\langle \mathcal{S} \rangle} \succeq 0$. A similar argument holds for the other condition (8). It follows that the restriction to polynomials from $\langle \mathcal{S} \rangle$ implies that the computation of a solution to (7)-(8) boils down to solving a semidefinite programming problem where the dimension of each matrix inequality is bounded by $|\mathcal{S}|$.

Note that when $\mathcal{S}$ is the set of monomials of degree at most d, then $M_{\langle \mathcal{S} \rangle}$ is equal to the familiar *truncated moment matrix* (see e.g. [19]) with rows and columns indexed by sets of size at most d. More in general, the polynomials u are expressed in the new basis $\sigma_1, \ldots, \sigma_t$, where $\sigma_i = (S^\top x)_i$ and $t = |\mathcal{S}|$ (here x denotes the vector of all multilinear monomials and the column vector of S are the representation of the vectors from $\mathcal{S}$ according to the standard monomial basis).

C Omitted proofs

C.1 Proof of Lemma 5.4

The proof will be by induction on $\pi = \pi(a, a_0)$. If $\pi = 0$ it follows that $\mathcal{F}_B = \{x : x \in \{0,1\}^n\}$ and $V = \emptyset$. As a pitch zero constraint we must have $a_0 \leq 0$, so since $a \geq 0$, $a^\top x - a_0 \geq 0$ is indeed valid for $\mathcal{F}_B = \{x : x \in \{0,1\}^n\}$ and for $\mathcal{F}_{A_{(\emptyset,\emptyset)}} = \mathcal{F}_A$ $(\neq \emptyset)$.

Assume now that the claim holds for all valid inequalities of pitch p, $p \leq \pi - 1 \geq 0$, and consider a valid inequality $a^\top x - a_0 \geq 0$ of pitch π. Note that there must be some $A_v \subseteq supp(a)$ for $v \in [m]$ or else we could set $x_j = 0$ for all $j \in supp(a)$, and $x_j = 1$ everywhere else, and thereby satisfy every constraint and nevertheless have $a^\top x = 0$ (so contradicting the hypothesis that $a^\top x - a_0 \geq 0$ is a valid inequality of pitch $\pi \geq 1$). Choose $A_v \subseteq supp(a)$. Note that we are assuming, w.l.o.g., that A is minimal, so there is no A_i, with $i \in [m]$ and $i \neq v$, that is a proper subset of A_v. Let $v(1) \in A_v$ be the index of the minimum coefficient $a_j : j \in A_v$, where a_j is the coefficient of variable x_j in the valid inequality $a^\top x - a_0 \geq 0$.

We first obtain a strengthen by setting to zero all the variables from V_v, where V_v are all the variables from all A_i, with $i \neq v$, that appear in $A_v - \{v(1)\}$, i.e. $V_v := (A_v - \{v(1)\}) \bigcap (\cup_{i \neq v} A_i)$. Consider $\mathcal{F}_{A_{(\emptyset,V_v)}}$ and note that $\mathcal{F}_{A_{(\emptyset,V_v)}} \neq \emptyset$ because by assumption no $A_j \subset A_v$ and therefore $(a^\top x - a_0)_{(\emptyset,V_v)} \geq 0$ is a valid inequality for $\mathcal{F}_{A_{(\emptyset,V_v)}}$. Set $x_{v(1)} = 1$ in $(a^\top x - a_0)_{(\emptyset,V_v)} \geq 0$ to get $(a^\top x - a_0)_{(\{v(1)\},V_v)} \geq 0$ which is a valid inequality for $\mathcal{F}_{A_{(\emptyset,V_v)}}$. Note that the pitch p of $(a^\top x - a_0)_{(\{v(1)\},V_v)}$ is such that $p \leq \pi - 1$ and therefore, by induction hypothesis, it satisfies the properties of the claim when we consider $(a^\top x - a_0)_{(\{v(1)\},V_v)} \geq 0$ as valid inequality for $\mathcal{F}_{A_{(\emptyset,V_v)}}$. Let a' be the vector that is obtained from a by setting to zero all the coefficients from $V_v \cup \{v\}$ and let $a'_0 := a_0 - a_{v(1)}$, so $(a^\top x - a_0)_{(\{v(1)\},V_v)} = a'^\top x - a'_0$. By the induction hypothesis there must be a subset C' of the rows from $A' := A_{(\emptyset,V_v)}$ such that $|C'| \leq p$ and

$$A'_i \subseteq supp(a') \quad \forall i \in C' \tag{26}$$

$$B_i' = A_i' - \bigcup_{r \in C' - \{i\}} A_r' \neq \emptyset \quad \forall i \in C' \tag{27}$$

$$(a'^\top x - a_0')_{(\emptyset, V')} \geq 0 \text{ is valid for } \mathcal{F}_{B'} = \{x \in \{0,1\}^n : \sum_{j \in B_i'} x_j \geq 1, i \in C'\} \tag{28}$$

$$\mathcal{F}_{A'_{(\emptyset, V')}} \neq \emptyset \tag{29}$$

where V' is the set of variables occurring in more than one row from C'.

Define the collection $C := \{v\} \cup C'$ and therefore Condition (14) is satisfied for this collection by construction.

Define B_i for $i \in C$ as in the statement of the claim. Clearly $B_i = B_i'$ for $i \in C'$ as these sets had their indices that overlap with $A_v - \{v(1)\}$ removed, and they never overlapped $v(1)$ (because they belong to $supp(a')$). Moreover, $v(1) \in B_v$, so $B_v \neq \emptyset$, and Condition (15) is satisfied as well.

Suppose now that we are given an arbitrary $x \in \{0,1\}^n$ that satisfies $\mathcal{F}_B$. Consider that we must have $x_j = 1$ for some $j \in B_v$, that $a_j \geq a_{v(1)}$, and that, since all B_i for $i \in C$ are disjoint, if we define x' to be the same as x but with $x_j = 0$, then x' still satisfy all $\sum_{x \in B_i} x \geq 1$ for $i \in C'$. Thus by induction $a'^\top x' \geq a_0 - a_{v(1)}$ which implies that $a^\top x = a'^\top x' + a_j x_j = a'^\top x' + a_j \geq a_0 - a_{v(1)} + a_j \geq a_0$. This proves Property (16).

To prove Property (17) we show that we can set to zero all the overlapping variables from the rows in C, namely the variables from V and still get a non empty set of integral solutions, i.e. $\mathcal{F}_{A_{(\emptyset,V)}} \neq \emptyset$. Indeed, by the induction hypothesis we have that $\mathcal{F}_{A'_{(\emptyset,V')}} \neq \emptyset$, where $A'_{(\emptyset,V')} = A_{(\emptyset, V_v \cup V')}$. Therefore $\mathcal{F}_{A_{(\emptyset,V)}} \neq \emptyset$ because $V \subseteq V_v \cup V'$.

C.2 Proof of Theorem 4.1

Before proving the bound given in Theorem 4.1 on the number of levels for our simple example we need some preliminaries. In particular we first introduce the SOS hierarchy in matrix form that is more convenient for proving lower bounds. In the following we assume that the SOS hierarchy is the "standard" one, namely the one that follows by considering the subspace of bounded degree polynomials as functional basis.

C.2.1 The Sum-of-Squares hierarchy in matrix form

Consider the SOS hierarchy for approximating the convex hull of the semialgebraic set

$$P = \{x \in \{0,1\}^n \mid g_\ell(x) \geq 0, \forall \ell \in [p]\} \tag{30}$$

where $g_\ell(x)$ are linear constraints and p a positive integer. The form of the SOS hierarchy we use here is equivalent to the one introduced before and follows from applying a change of basis to the dual certificate of the refutation of the proof system (see [14] for the details on the change of basis). We use this change of basis in order to obtain a useful decomposition of the moment matrices as a sum of rank one matrices of special kind.

For any $I \subseteq N = \{1, \ldots, n\}$, let x_I denote the 0/1 solution obtained by setting $x_i = 1$ for $i \in I$, and $x_i = 0$ for $i \in N \setminus I$. For a function $f : \{0,1\}^n \to \mathbb{R}$, we denote by $f(x_I)$ the value of the function evaluated at x_I. In the SOS hierarchy defined below there is a variable y_I^N that can be interpreted as the "relaxed" indicator variable for the solution x_I. We point out that in this formulation of the hierarchy the number of variables $\{y_I^N : I \subseteq N\}$ is exponential in n, but this is not a problem in our context since we are interested in proving lower and upper bounds rather than solving an optimization problem.

Let $\mathcal{P}_t(N)$ be the collection of subsets of N of size at most $t \in \mathbb{N}$. For every $I \subseteq N$, the q-zeta vector $Z_I \in \mathbb{R}^{\mathcal{P}_q(N)}$ is a 0/1 vector with J-th entry ($|J| \leq q$) equal to 1 if and only if $J \subseteq I$.[5] Note that $Z_I Z_I^\top$ is a rank one matrix and the matrices considered in Definition C.1 are linear combinations of these rank one matrices.

Definition C.1. *The t-th round SoS hierarchy relaxation for the set P as given in* (30)*, denoted by* $\mathrm{SoS}_t(P)$*, is the set of variables* $\{y_I^N \in \mathbb{R} : \forall I \subseteq N\}$ *that satisfy*

$$\sum_{I \subseteq N} y_I^N = 1, \tag{31}$$

$$\sum_{I \subseteq N} y_I^N Z_I Z_I^\top \succeq 0, \text{ where } Z_I \in \mathbb{R}^{\mathcal{P}_{t+1}(N)} \tag{32}$$

$$\sum_{I \subseteq N} g_\ell(x_I) y_I^N Z_I Z_I^\top \succeq 0, \ \forall \ell \in [p], \text{ where } Z_I \in \mathbb{R}^{\mathcal{P}_t(N)} \tag{33}$$

It is straightforward to see that the SoS hierarchy formulation given in Definition C.1 is a relaxation of the integral polytope. Indeed consider any feasible integral solution $x_I \in P$ and set $y_I^N = 1$ and the other variables to zero. This solution clearly satisfies (31) and (32) because the rank one matrix $Z_I Z_I^\top$ is positive semidefinite (PSD), and (33) since $x_I \in P$.

For a set $Q \subseteq [0,1]^n$, we define the projection from $\mathrm{SoS}_t(Q)$ to $\mathbb{R}^n$ as $x_i = \sum_{i \in I \subseteq N} y_I^N$ for each $i \in \{1, ..., n\}$. The *SoS rank* of Q, $\rho(Q)$, is the smallest t such that $\mathrm{SoS}_t(Q)$ projects exactly to the convex hull of $Q \cap \{0,1\}^n$.

C.2.2 Using symmetry to simplify the PSDness conditions

In this section we present a theorem given in [14] that can be used to simplify the PSDness conditions (32) and (33) when the problem formulation is very symmetric. More precisely, the theorem can be applied whenever the solutions and constraints are symmetric in the sense that $w_I^N = w_J^N$ whenever $|I| = |J|$ where w_I^N is understood to denote either y_I^N or $g_\ell(x_I) y_I^N$. In what follows we denote by $\mathbb{R}[x]$ the ring of polynomials with real coefficients and by $\mathbb{R}[x]_d$ the polynomials in $\mathbb{R}[x]$ with degree less or equal to d.

Theorem C.1 ([14])**.** *For any* $t \in \{1, \ldots, n\}$*, let* $\mathcal{S}_t$ *be the set of univariate polynomials* $G_h(k) \in \mathbb{R}[k]$*, for* $h \in \{0, \ldots, t\}$*, that satisfy the following conditions:*

$$G_h(k) \in \mathbb{R}[k]_{2t} \tag{34}$$

$$G_h(k) = 0 \qquad \text{for } k \in \{0, \ldots, h-1\} \cup \{n-h+1, \ldots, n\}, \text{ when } h \geq 1 \tag{35}$$

$$G_h(k) \geq 0 \qquad \text{for } k \in [h-1, n-h+1] \tag{36}$$

For any fixed set of values $\{w_k^N \in \mathbb{R} : k = 0, \ldots, n\}$*, if the following holds*

$$\sum_{k=h}^{n-h} \binom{n}{k} w_k^N G_h(k) \geq 0 \qquad \forall G_h(k) \in \mathcal{S}_t \tag{37}$$

then

$$\sum_{k=0}^{n} w_k^N \sum_{\substack{I \subseteq N \\ |I|=k}} Z_I Z_I^\top \succeq 0$$

where $Z_I \in \mathbb{R}^{\mathcal{P}_t(N)}$.

[5] In order to keep the notation simple, we do not emphasize the parameter q as the dimension of the vectors should be clear from the context.

Note that polynomial $G_h(k)$ in (36) is nonnegative in a *real interval*, and in (35) it is zero over a *set of integers*. Moreover, constraints (37) are trivially satisfied for $h > \lfloor n/2 \rfloor$.

C.2.3 The simple example proof

The single constraint of the simple example can be rewritten, w.l.o.g., as follows:

$$g(x) = \sum_{i=1}^{n} x_i - L + 1 - \frac{1}{P} \geq 0$$

where L and P are positive integers. Clearly any integral $\{0,1\}$-solution requires to set to one at least L variables.

Let (LP) be the polytope $\{x \in [0,1]^n : g(x) \geq 0\}$. The SoS *rank* is the minimal number of levels needed to obtain the integer hull (IP) of (LP).

In the following we will restrict the analysis to $L \leq \lceil n/2 \rceil$. Consider any solution that satisfies the following conditions:

$$\begin{cases} y_k^N = 0 & \text{for } k \leq L-2 \\ y_k^N > 0 & \text{for } k \geq L-1 \\ \sum_{k=0}^{n} y_k^N \binom{n}{k} = 1 \end{cases} \tag{38}$$

Note that in (38) we do not impose any restriction on the exact value of the positive probabilities. The value of the suggested solution is $\sum_{k=L-1}^{n} \binom{n}{k} y_k^N k$. By choosing P sufficiently large we will show that almost all the probability mass (but an arbitrarily small part) can be assigned to y_{L-1}^N, resulting therefore into an objective function value equal to $L-1+\varepsilon$, (for any $\varepsilon > 0$) and an integrality gap of $\frac{L}{L-1+\varepsilon}$.

Lemma C.2. *For $L \leq \lceil n/2 \rceil$ and a suitable large value of P that depends on n the* SoS *rank for (LP) is at least $n - L + 1$.*

Proof. For any solution that satisfies (38) there is a unique nonpositive term in conditions (37), namely $z_{L-1}^N G_h(L-1) = y_{L-1}^N(-1/P)G_h(L-1) = -\varepsilon G_h(L-1)$ (for some $\varepsilon = y_{L-1}^N/P > 0$), where we use the following notation $z_k^N = y_k^N g(k)$ (with $g(k)$ denoting the value of the constraint $g(x)$ when exactly k variables are set to one).

If we chose h such that $L-1 = n-h$ then we would have that $z_k^N G_h(k)$ is equal to zero for all $k \neq n-h$, and by choosing $G_h(k)$ such that $G_h(L-1) > 0$ we would have that (37) is never satisfied. To avoid this problem we assume that $L-1 \leq n-h-1$ and since $h \leq \lfloor n/2 \rfloor$, the claim holds when $L \leq n - \lfloor n/2 \rfloor = \lceil n/2 \rceil$.

According to Theorem C.1 and (38) note that

- $G_h(k)$ has $2t$ roots.
- $G_h(k)$ has at least $h - 1 + 1 + n - (n-h+1) + 1 = 2h$ roots outside the (open) interval $(h-1, \ldots, n-h+1)$.
- $G_h(k)$ has at most $2(t-h)$ roots within the (open) interval $(h-1, \ldots, n-h+1)$. Moreover $G_h(k) \geq 0$ for any $k \in (h-1, \ldots, n-h+1)$ and therefore the at most $2(t-h)$ roots that are within the (open) interval $(h-1, \ldots, n-h+1)$ must appear in pairs. It follows that $G_h(k)$ has at most $t-h$ different roots within the (open) interval $(h-1, \ldots, n-h+1)$.

Consider any h such that $h \leq L-1 \leq n-h-1$ (if $L-1 \leq h-1$ then (37) is trivially satisfied). Note that there are $n-h-L+1$ terms $z_k^N > 0$ for $k \in \{L, \ldots, n-h\}$ (note that $L \leq n-h$ by assumption, so set $\{L, \ldots, n-h\}$ is never empty). From the above arguments we know that $G_h(k)$ has at most $t-h$ different roots within the (open) interval $(h-1, \ldots, n-h+1)$. So if $t-h$ is strictly smaller than the number $n-h-L+1$ of terms $z_k^N > 0$ (with $k \in \{L, \ldots, n-h\}$) then it exists a $k^* \in \{L, \ldots, n-h\}$ that is not a root for $G_h(k)$ and such that $z_{k^*}^N \binom{n}{k^*} G_h(k^*) > 0$ (recall that $G_h(k) \geq 0$ within the considered interval which implies that $G_h(k^*) > 0$). The latter condition is satisfied when $t-h \leq n-h-L$, namely when $t \leq n-L$. It follows that if $t \leq n-L$ then there exists a $k^* \in \{L, \ldots, n-h\}$ such that $z_{k^*}^N \binom{n}{k^*} G_h(k^*) > 0$. Moreover, let $r_1, \ldots, r_{2t}$ be the roots of $G_h(x)$. It is easy to see that $k^* \in \{L, \ldots, n-h\}$ can be chosen such that the following two conditions are both satisfied:

$$|k^* - r_i| \geq 1/2 \quad \text{for every } i \in [2t] \tag{39}$$

$$z_{k^*}^N \binom{n}{k^*} G_h(k^*) > 0 \tag{40}$$

Let j^* such that $k^* = L-1+j^*$, where $j^* \in \{1, \ldots, n-h-L+1\}$. The claim follows by showing how to choose P such that:

$$z_{L-1+j^*}^N \binom{n}{L-1+j^*} G_h(L-1+j^*) > \frac{y_{L-1}^N}{P} \binom{n}{L-1} G_h(L-1)$$

From (40) the above condition is equivalent to satisfy the following

$$z_{L-1+j^*}^N > \frac{y_{L-1}^N}{P} \frac{\binom{n}{L-1}}{\binom{n}{L-1+j^*}} \frac{G_h(L-1)}{G_h(L-1+j^*)} \tag{41}$$

Clearly, the interesting cases are when $G_h(L-1) > 0$. By the latter, (39) and (40), we have that:

$$\frac{G_h(L-1)}{G_h(L-1+j^*)} = \prod_{i=1}^{2t} \frac{|L-1-r_i|}{|L-1+j^*-r_i|} \leq \prod_{i=1}^{2t} \left(1 + \frac{j^*}{|L-1+j^*-r_i|}\right) \leq \prod_{i=1}^{2t} (1+2j^*) \tag{42}$$

By (42), if the following is satisfied then (41) holds.

$$z_{L-1+j^*}^N > \frac{y_{L-1}^N}{P} \frac{\binom{n}{L-1}}{\binom{n}{L-1+j^*}} (1+2j^*)^{2t} \tag{43}$$

Then it is sufficient to choose P such that

$$P \geq 2 \frac{y_{L-1}^N}{y_{L-1+j^*}^N} \frac{\binom{n}{L-1}}{\binom{n}{L-1+j^*}} \frac{(1+2j^*)^{2t}}{j^*}$$

Note that the right-hand-side of the above inequality is bounded by a function of n.

□

D On a very recent claim by Fiorini et al. [10]

We describe the approach suggested in [10] for the 0/1 set cover problem which is also the main application advertised in the abstract. We observe in the following that their approach is essentially based on similar arguments as in this paper (formerly appeared in [21]) but specialized for a weaker framework that does not generalize to packing problems (see Section 7.1). We sketch this for pitch 2 in the following. The generalization to any pitch is straightforward.

Let A be the $m \times n$ set cover matrix defined as in (21) and let A_{ij} denote the (i, j)-entry of A. By overloading notation, we will interchangeably use A_i to denote the i-th row of A and its support. In [10], they consider the canonical monotone formula for set cover:

$$\phi := \bigwedge_{i=1}^{m} \bigvee_{A_{ij=1}} x_j \tag{44}$$

Starting with any convex set $Q \subseteq [0, 1]^n$ containing $\mathcal{F}_A$ (see (21)) the improved relaxation is obtained by recursively "feeding" Q into the formula ϕ, denoted by $\phi(Q)$ and defined as follows:

$$\phi(Q) := \bigcap_{i=1}^{m} \operatorname{conv}\left(\bigcup_{A_{ij=1}} \{x \in Q : x_j = 1\} \right) \tag{45}$$

By starting with $Q := [0, 1]^n$ it is easy to see that $\phi([0, 1]^n) = \{x \in [0, 1]^n : Ax \geq e\}$. This is also the base of induction in the proof of Lemma 5.4 in this paper. So their approach obtains, after the first application, the starting linear program relaxation that corresponds to all pitch one inequalities (also used in (20)). Now let $Q := \phi([0, 1]^n)$ and let's analyze the second application, namely $\phi(Q) = \phi^2([0, 1]^n)$:

$$\phi(Q) := \bigcap_{i=1}^{m} \operatorname{conv} \underbrace{\left(\bigcup_{A_{ij=1}} \{x \in [0, 1]^n : Ax \geq e, x_j = 1\} \right)}_{U_i} \tag{46}$$

It can be easily observed that the relaxation given by (46) is obtained by considering the "interaction" of the i-th pitch 1 constraint (for any $i \in [m]$, see the outer intersection) with any other constraint $h \in [m]$ from $Ax \geq e$. The "interaction" is given by the common variables, denoted by $A_i \cap A_h$ in this paper, otherwise (i.e. $j \notin A_h$) setting $x_j = 1$ does not effect the corresponding constraint $A_h x \geq 1$. These are exactly the variables considered in V_C with $C = \{i, h\}$ (see Definition 5.1, V_C will be used to define the subspace of polynomials $Q_A(t)$).

Lemma 5.4 gives a property of these interactions that are used for proving that these pairs of interactions are sufficient to show pitch 2 inequalities. Higher pitches use recursive polynomials which correspond to recursive application of ϕ by considering triplets for pitch 3 and so on, as in this paper.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il