Mathematical Programming (2020) 182:369–397
https://doi.org/10.1007/s10107-019-01398-9

FULL LENGTH PAPER

Series A

Sum-of-squares hierarchy lower bounds for symmetric formulations

Adam Kurpisz[1] · **Samuli Leppänen**[2] · **Monaldo Mastrolilli**[2]

Received: 9 January 2018 / Accepted: 22 April 2019 / Published online: 22 May 2019

Abstract
We introduce a method for proving Sum-of-Squares (SoS)/Lasserre hierarchy lower bounds when the initial problem formulation exhibits a high degree of symmetry. Our main technical theorem allows us to reduce the study of the positive semidefiniteness to the analysis of "well-behaved" univariate polynomial inequalities. We illustrate the technique on two problems, one unconstrained and the other with constraints. More precisely, we give a short elementary proof of Grigoriev/Laurent lower bound for finding the integer cut polytope of the complete graph. We also show that the SoS hierarchy requires a non-constant number of rounds to improve the initial integrality gap of 2 for the Min-Knapsack linear program strengthened with cover inequalities.

Keywords Sum-of-squares hierarchy · Integrality gap

Mathematics Subject Classification 90C05 · 90C22

1 Introduction

Proving lower bounds for the Sum-of-Squares (SoS)/Lasserre hierarchy [28,36] has attracted notable attention in the theoretical computer science community during the last decade, see e.g. [7,14,15,20,21,30,31,34,39,41]. This is partly because the hierarchy captures many of the best known approximation algorithms based on semidefinite programming (SDP) for several natural 0/1 optimization problems (see [31] for a recent

This is the full version of the paper that was presented at IPCO 2016. Supported by the Swiss National Science Foundation project 200020-169022 "Lift and Project Methods for Machine Scheduling Through Theory and Experiments" and by the project PZ00P2_174117 "Theory and Applications of Linear and Semidefinite Relaxations for Combinatorial Optimization Problems".

✉ Adam Kurpisz
kurpisz.adam@gmail.com

1 Department of Mathematics, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland

2 IDSIA, 6928 Manno, Switzerland

result). Indeed, it can be argued that the SoS hierarchy is the strongest candidate to be the "optimal" meta-algorithm predicted by the Unique Games Conjecture (UGC) [24,37]. On the other hand, the hierarchy is also one of the best known candidates for refuting the conjecture since it is still conceivable that one could show that the SoS hierarchy achieves better approximation guarantees than the UGC predicts (see [6] for discussion). Despite the interest in the algorithm and due to the many technical challenges presented by semidefinite programming, only relatively few techniques are known for proving lower bounds for the hierarchy. In particular, several integrality gap results follow from applying gadget reductions to the few known original lower bound constructions.

Indeed, many of the known lower bounds for the SoS hierarchy originated in the works of Grigoriev [20,21].[1] We defer the formal definition of the hierarchy for later and only point out that solving the hierarchy after t rounds takes $n^{O(t)}$ time. In [21] Grigoriev showed that random 3Xor or 3Sat instances cannot be solved even by $\Omega(n)$ rounds of the SoS hierarchy (some of these results were later independently rediscovered by Schoenebeck [39]). Lower bounds, such as those of [7,41] rely on [21,39] combined with gadget reductions. Another important lower bound was also given by Grigoriev [20] for the Knapsack problem (a simplified proof can be found in [22]), showing that the SoS hierarchy cannot prove within $\lfloor n/2 \rfloor$ rounds that the polytope $\{x \in [0, 1]^n : \sum_{i=1}^{n} x_i = n/2\}$ contains no integer point when n is odd. Using essentially the same construction as in [22], Laurent [30] independently showed that $\lfloor \frac{n}{2} \rfloor$ rounds are not enough for finding the integer cut polytope of the complete graph with n nodes, where n is odd (this result was recently shown to be tight in [17]).[2] By using several new ideas and techniques, but a similar starting point as in [22,30], Meka, Potechin and Wigderson [34] were able to show a lower bound of $\Omega(\log^{1/2} n)$ for the Planted-Clique problem. Common to the works [22,30,34] is that the matrix involved in the analysis has a large kernel, and they prove that a principal submatrix is positive definite by applying the theory of association schemes [18]. It is also interesting to point out that for the class of Max- CSPs, Lee, Raghavendra and Steurer [31] proved that the SoS relaxation yields the "optimal" approximation, meaning that SDPs of polynomial-size are equivalent in power to those arising from $O(1)$ rounds of the SoS relaxations. Then, by appealing to the result by Grigoriev/Laurent [20,30] they showed an exponential lower bound on the size of SDP formulations for the integer cut polytope. For different techniques to obtain lower bounds, we refer for example to the recent papers [5,25,26] (see also Sect. 5.3) and the survey [15] for an overview of previous results.

In this paper we introduce a method for proving SoS hierarchy lower bounds when the initial problem formulation exhibits a high degree of symmetry. Our main technical theorem (Theorem 2) allows us to reduce the study of the positive semidefiniteness to the analysis of "well-behaved" univariate polynomial inequalities. The theorem

[1] More precisely, Grigoriev considers the positivstellensatz proof system, which is the dual of the SoS hierarchy considered in this paper. For brevity, we will use SoS hierarchy/proof system interchangeably as is customary in theoretical computer science literature. In optimization context the moment matrix formulation considered in this paper is usually called Lasserre hierarchy.

[2] The algebraic description of the two problem instances of Knapsack and Max- Cut in the complete graph, considered respectively in [20,22] and in [30], are essentially the same, and we will use Max- Cut to refer to both.

applies whenever the solution and constraints are *symmetric*, informally meaning that all subsets of the variables of equal cardinality play the same role in the formulation (see Sect. 3 for the formal definition). For example, the solution in [20,22,30] for Max- Cut is symmetric in this sense.

We note that exploiting symmetry reduces the number of variables involved in the analysis, and different ways of utilizing symmetry have been widely used in the past for proving integrality gaps for different hierarchies, see for example [9,19,21,23,26,40]. An interesting difference of our approach from others is that we establish several lower bounds without fully identifying the formula of eigenvectors. More specifically, the common task in this context is to identify the spectral structure to get a simple diagonalized form. In the previous papers the moment matrices belong to the Bose-Mesner algebra of a well-studied association scheme, and hence one can use the existing theory. In this paper, instead of identifying the spectral structure completely, we identify only possible forms and propose to test all the possible candidates. This is in fact an important point, since the approach may be extended even if the underlying symmetry is imperfect or its spectral property is not well understood.

The proof of Theorem 2 is obtained by a sequence of elementary operations, as opposed to notions such as big kernel in the matrix form, the use of interlacing eigenvalues, the machinery of association schemes and various results about hyper-geometric series as in [20,22,30]. Thus Theorem 2 applies to the whole class of symmetric solutions, even when several conditions and machinery exploited in [20,22,30] cannot be directly applied. For example the kernel dimension, which was one of the important key property used to prove the results in [20,22,30], depends on the particular solution that is used and it is not a general property of the class of symmetric solutions. The solutions for two problems considered in this paper have completely different kernel sizes of the analyzed matrices, one large and the other zero.

We demonstrate the technique with two illustrative and complementary applications. First, we show that the analysis of the lower bound for Max- Cut in [20,22,30] simplifies to few elementary calculations once the main theorem is in place. This result is partially motivated by the open question posed by O'Donnell [35] of finding a simpler proof for Grigoriev's lower bound for the Knapsack problem.

As a second application we consider a constrained problem. We show that after $\Omega(\log^{1-\epsilon} n)$ levels the SoS hierarchy does not improve the integrality gap of 2 for the Min- knapsack linear program formulation strengthened with *cover inequalities* [12] introduced by Wolsey [42]. Adding cover inequalities is currently the most successful approach for capacitated covering problems of this type [1–3,11,13]. Moreover, this result partially answers the question raised in [8] for the rank of the certain set-covering polytope. Our result is the first SoS lower bound for formulations with cover inequalities. In this application we demonstrate that our technique can also be used for suggesting the solution and for analyzing its feasibility.

Finally we point it out that the same analysis can be used to provide a non trivial lower bound to an open question raised by Laurent [29] regarding the Lasserre rank of the knapsack problem (see Sect. 5.3 for a discussion).

Related results Recently an independent argument by Blekherman of similar type was presented in [32, Theorem 2.4]. The argument works on the dual of the Lasserre/SoS hierarchy namely the SoS certificate for nonnegativity of a polynomial over the boolean hypercube.

In the unconstrained boolean hypercube setting, the SoS certifies the nonnegativity over the boolean hypercube of the polynomial by writing it as a SoS polynomial. In the constrained case, when the polynomial is nonnegative over subset of the boolean hypercube, the certificate writes the polynomial as a summation of SoS polynomials multiplied by the constraints defining the nonnegativity set.

Blekherman proved the following theorem:

Theorem 1 (Blekherman: [32] Theorem 2.4) *Let $f : \{0,1\}^n \to \mathbb{R}_+$ be a symmetric non-negative real-valued boolean function and $\tilde{f}$ a univariate polynomial such that $f(x_1,\ldots,x_n) = \tilde{f}(x_1+\cdots+x_n)$. If f can be written as the sum of squares of n-variate polynomials of degree $d \leq n/2$, then we can write*

$$\begin{aligned}\tilde{f}(z) &= q_d(z) + z(n-z)q_{d-1}(z) + z(z-1)(n-z)(n-1-z)q_{d-2}(z) + \cdots \\ &\quad + z(z-1)\cdots(z-d+1)(n-z)(n-1-z)\cdots(n-d+1-z)q_0(z) \quad (1)\end{aligned}$$

where each $q_t(z)$ is a univariate SoS polynomial with $deg(q_t) \leq 2t$.

For the unconstrained boolean hypercube the Blekherman theorem allows to significantly reduces the complexity of writing a symmetric polynomial as a SoS polynomial since it reduces the search space to univariate polynomials only. This is of similar flavor as our result on the dual side (Theorem 2) which says that for symmetric problem formulations the search space can be reduced to specific univariate polynomials (see Definition 5). Indeed, by Theorem 3.72 in [10] one can check that the RHS of the SoS representation of univariate polynomial $\tilde{f}$ defined in (1) can be written as a combination of polynomials $G(x)$ in Definition 5 and vice versa.

In [32] Blekherman's theorem is used to reprove the Grigoriev's Max- Cut lower bound result. However, in our paper not only we obtain a simpler proof of Grigoriev's lower bound but our theorem generalizes to constrained hypercube setting, which is used for our main new application, the Min Knapsack problem with covering inequalities.

For now it is unclear if Blekherman's theorem can provide similar results for the constrained boolean hypercube. Indeed, as pointed out in [32] [Sect. 1.3.2 and Future work: Sect. 5] "Kurpisz et al. [...showed] a more general theorem that reduces the analysis of dual certificates for very symmetric SoS proof systems (such as for knapsack) to the analysis of univariate polynomials [...] Similar results may be obtainable using Blekherman's theorem " but no formal proof appeared so far.

Finally we would like to point out that Theorems 2 and 6 recently have been furthered applied to obtain a tight SoS lower bounds for the family of the unconstrained binary polynomial optimization problems and to disprove the conjecture by Laurent, who considered the linear representation of a set with no integral points and conjectured that the SoS/Lasserre rank for the problem is $n-1$ (see [27,29]).

2 The SoS hierarchy

In this section we recall the usual definition of the SoS/Lasserre hierarchy [28] using moment matrices. We then perform a change of basis and obtain an alternative definition for the hierarchy, which is useful for deriving our results later. The SDP hierarchy that we discuss here is the dual certificate of a refutation of the positivstellensatz proof system. For further information about the connection to the proof system see [34], and about the SoS hierarchy in matrix form [29].

In our setting we restrict ourselves to problems with $0/1$-variables and linear constraints. More precisely, we consider the following general optimization problem $\mathbb{P}$: given a multilinear polynomial $f : \{0, 1\}^n \to \mathbb{R}$

$$\mathbb{P} : \min\{f(x)|x \in \{0, 1\}^n \cap K\}, \tag{2}$$

where K is a polytope defined by m linear inequalities $g_\ell(x) \geq 0$ for $\ell \in [m]$. Many basic optimization problems are special cases of $\mathbb{P}$. For example, any k-ary boolean constraint satisfaction problem, such as Max- Cut, is captured by (2) where a degree k function $f(x)$ counts the number of satisfied constraints, and no linear constraints $g_\ell(x) \geq 0$ are present. Also any $0/1$ integer linear program is a special case of (2), where $f(x)$ is a linear function.

Let $\mathbb{R}[x]$ denote the ring of real polynomials, and $\mathbb{R}[x]_d$ the subset of $\mathbb{R}[x]$ of polynomials with degree less or equal to d. Lasserre [28] proposed a hierarchy of SDP relaxations parameterized by an integer r,

$$\min_L\{L(f)|L(1) = 1, \quad L(x^2 - x) = 0 \text{ and } L(u^2), \quad L(u^2 g_\ell) \geq 0, \quad \forall u \in \mathbb{R}[x]_r\}, \tag{3}$$

where $L : \mathbb{R}[X]_{2r} \to \mathbb{R}$ is a linear map.[3] Note that (3) is a relaxation since one can take L to be the evaluation map $f \to f(x^*)$ for any optimal solution x^*.

Relaxation (3) can be equivalently formulated in terms of *moment matrices* [28]. In the context of this paper, this matrix point of view is more convenient to use and it is described below. In our notation we mainly follow the survey of Laurent [29] (see also [38]).

Variables and moment matrix Let N denote the set $\{1, \ldots, n\}$. The collection of all subsets of N is denoted by $\mathcal{P}(N)$. For any integer $t \geq 0$, let $\mathcal{P}_t(N)$ denote the collection of subsets of N having cardinality at most t. Let $y \in \mathbb{R}^{\mathcal{P}(N)}$. For any nonnegative integer $t \leq n$, let $M_t(y)$ denote the matrix with (I, J)-entry $y_{I\cup J}$ for all $I, J \in \mathcal{P}_t(N)$. Matrix $M_t(y)$ is termed in the following as the *t-moment matrix* of y. For a linear function $g(x) = \sum_{i=1}^n g_i \cdot x_i + g_0$, we define $g * y$ as a vector, often called *shift operator*, where the Ith entry is $(g * y)_I = \sum_{i=1}^n g_i y_{I\cup\{i\}} + g_0 y_I$. Let f denote the vector of coefficients of polynomial $f(x)$ (where f_I is the coefficient of monomial $\Pi_{i\in I} x_i$ in $f(x)$).

[3] In [4], $L(p)$ is written $\tilde{\mathbb{E}}[p]$ and called the "pseudo-expectation" of p.

Definition 1 The tth round SoS (or Lasserre) relaxation of problem (2), denoted as $\text{SoS}_t(\mathbb{P})$, is the following

$$\text{SoS}_t(\mathbb{P}) : \min_y \left\{ \sum_{I \subseteq N} f_I y_I | y \in \mathbb{R}^{\mathcal{P}_{2t+2d}(N)} \text{ and } y \in \mathbb{M} \right\}, \tag{4}$$

where $\mathbb{M}$ is the set of vectors $y \in \mathbb{R}^{\mathcal{P}_{2t+2d}(N)}$ that satisfy the following PSD conditions

$$y_{\varnothing} = 1, \tag{5}$$

$$M_{t+d}(y) \succeq 0, \tag{6}$$

$$M_t(g_\ell * y) \succeq 0 \qquad \ell \in [m] \tag{7}$$

where $d = 0$ if $m = 0$ (no linear constraints), otherwise $d = 1$.

Change of variables A solution of the SoS hierarchy as defined in Definition 1 is given by a vector $y \in \mathbb{R}^{\mathcal{P}_{2t+2d}(N)}$. Next we show we can make a change of basis and replace the variables y_I with other variables y_I^N that are indexed by *all* the subsets of N. Variable y_I^N can be interpreted as the "relaxed" indicator variable for the integral solution x_I, i.e. the 0/1 solution obtained by setting $x_i = 1$ for $i \in I$, and $x_i = 0$ for $i \in N \setminus I$. We use this change of basis in order to obtain a useful decomposition of the moment matrix as a sum of rank one matrices of special kind.

In deriving the change of basis it is not necessary to distinguish between the moment matrices of the variables and constraints, since the matrices are structurally similar in the sense that the (I, J)-entry of any moment matrix depends only on the set $I \cup J$. Therefore, in the derivation we denote a generic vector by $w \in \mathbb{R}^{\mathcal{P}_{2q}(N)}$, where q is either t or $t + 1$.

Before introducing the variables let us start by giving a definition.

Definition 2 Given a vector $w \in \mathbb{R}^{\mathcal{P}(N)}$, define a vector $w^N \in \mathbb{R}^{\mathcal{P}(N)}$ entry-wise by

$$w_I^N := \sum_{H \subseteq N \setminus I} (-1)^{|H|} w_{I \cup H}. \tag{8}$$

Lemma 1 *Let w and w^N be as in Definition 2. Then, the vector w is given entry-wise by*

$$w_I = \sum_{I \subseteq H \subseteq N} w_H^N, \tag{9}$$

for each $I \subseteq N$.

Proof The proof follows from a direct application of the inclusion–exclusion principle. □

By Lemma 1, we can relate the vectors w and w^N in matrix form by using $w = Tw^N$, where T is square matrix of a suitable size that directly follows from (9). Moreover, T is invertible and upper triangular. It is easy to observe that if we "truncate" vector w to be a vector in $\mathbb{R}^{\mathcal{P}_{2q}(N)}$, then we can "truncate" matrix T (by taking the upper left $|\mathcal{P}_{2q}(N)| \times |\mathcal{P}_{2q}(N)|$ matrix) and "truncate" vector w^I such that $w = Tw^N$ still holds. Note that the truncated T is again invertible and upper triangular. (More precisely, truncating means considering only the entries w_I and w_I^N with $|I| \leq 2q$ or, alternatively, setting to zero all the others.)

The new variables are defined by the (truncated) vector $w^N \in \mathbb{R}^{\mathcal{P}_{2q}(N)}$ and they are related to the previous variables in the (truncated) vector $w \in \mathbb{R}^{\mathcal{P}_{2q}(N)}$ by the (truncated) matrix T as explained above.

Definition 3 For every $I \subseteq N$, the q-zeta vector $Z_I^q \in \mathbb{R}^{\mathcal{P}_q(N)}$ is defined entry-wise for each $J \in \mathcal{P}_q(N)$ as follows

$$[Z_I^q]_J = \begin{cases} 1, & \text{if } J \subseteq I, \\ 0, & \text{otherwise.} \end{cases}$$

Lemma 2 *Let $w \in \mathbb{R}^{\mathcal{P}_{2q}(N)}$, and $M \in \mathbb{R}^{\mathcal{P}_q(N)\times\mathcal{P}_q(N)}$ such that $M_{I,J} = w_{I\cup J}$. Then*

$$M = \sum_{H\subseteq N} w_H^N Z_H^q (Z_H^q)^\top.$$

Proof Since $M_{I,J} = w_{I\cup J}$, we have by the change of variables (9) that

$$M_{I,J} = \sum_{I\cup J\subseteq H\subseteq N} w_H^N = \sum_{H\subseteq N} \chi_{I\cup J}(H) w_H^N,$$

where $\chi_I(H)$ is the 0-1 indicator function such that $\chi_I(H) = 1$ if and only if $I \subseteq H$. On the other hand, $[Z_H^q(Z_H^q)^\top]_{I,J} = [Z_H^q]_I[Z_H^q]_J = 1$ if $I\cup J \subseteq H$, and 0 otherwise. Therefore $[Z_H^q(Z_H^q)^\top]_{I,J} = \chi_{I\cup J}(H)$. □

By the previous lemma it follows that given a solution by using variables $\{w_I^N\}$ we can obtain a solution with variables $\{w_I : |I| \leq 2t\}$. Vice versa, given any assignment of variables in $\{w_I : |I| \leq 2t\}$ we can find an assignment of variables in $\{w_I^N\}$ such that $M_{I,J} = w_{I\cup J}$ and $M = \sum_{H\subseteq N} w_H^N Z_H^q(Z_H^q)^\top$. Indeed, set $w_I^N = 0$ for every I such that $|I| > 2t$. For the remaining ones note that for $|I| \leq 2t$ the square matrix corresponding to the equalities $w_I = \sum_{I\subseteq H\subseteq N} w_H^N$ is invertible since it is upper triangular.

Another way of saying the above is noting that the change of basis is given by (a principal submatrix) of the *zeta matrix* (see e.g. [33]), which is the matrix that has the vectors Z_I^n as its columns and is thus upper triangular and invertible.

Lemma 3 [29] *Given $y \in \mathbb{R}^{\mathcal{P}_{2t+2}(N)}$, for the vector $z_I = \sum_{i=1}^n A_{\ell i} y_{I\cup\{i\}} - b_\ell y_I$ we have*

$$z_I^N = g_\ell(x_I) y_I^N, \tag{10}$$

where $g_\ell(x_I) = \sum_{i=1}^N A_{\ell i} x_i - b_\ell$ *is a linear function corresponding to the constraint* ℓ*, evaluated at* x_I *such that* $x_i = 1$ *if* $i \in I$ *and* $x_i = 0$ *otherwise.*

Proof We need to show that this choice of z_I^N yields $z_I = \sum_{I \subseteq H \subseteq N} z_H^N$. We plug in (10) to obtain

$$\begin{aligned}
\sum_{I\subseteq H\subseteq N} z_H^N &= \sum_{I\subseteq H\subseteq N} g_\ell(x_H) y_H^N = \sum_{I\subseteq H\subseteq N} \left[\sum_{i=1}^n A_{\ell i} x_i - b_\ell\right]_{x=x_H} y_H^N \\
&= \sum_{I\subseteq H\subseteq N} \left(\sum_{i=1}^n [A_{\ell i} x_i]_{x=x_H}\, y_H^N - b_\ell y_H^N\right) \\
&= \sum_{I\subseteq H\subseteq N} \sum_{i=1}^n [A_{\ell i} x_i]_{x=x_H}\, y_H^N - b_\ell y_I.
\end{aligned}$$

Here the term $[A_{\ell i} x_i]_{x=x_H}\, y_H^N$ is $A_{\ell i} y_H^N$ if $i \in H$ and 0 otherwise. Taking this into account and changing the order of the sums, the above becomes

$$\sum_{i=1}^n \sum_{I\cup\{i\}\subseteq H\subseteq N} A_{\ell i} y_H^N - b_\ell y_I = \sum_{i=1}^n A_{\ell i} y_{I\cup\{i\}} - b_\ell y_I,$$

which proves the claim. □

Finally, the change of basis modifies the objective function as follows:

Lemma 4 *Let* f *denote the vector of coefficients of polynomial* $f(x)$ *of* (2)*. Then the objective value of the solution* y *is given by*

$$\sum_{I\subseteq N} f_I y_I = \sum_{I\subseteq N} f(x_I) y_I^N.$$

Proof Similar as the proof of Lemmas 2 and 3. □

The above discussion together with the observation that $y_\emptyset = 1$ implies that $\sum_{J\subseteq N} y_J^N = 1$. Thus, we have justified the following definition:

Definition 4 The tth round SoS relaxation of problem (2), denoted as $\mathrm{SoS}_t(\mathbb{P})$, can be expressed as

$$\mathrm{SoS}_t(\mathbb{P}) : \min_{y^N} \left\{ \sum_{I\subseteq N} f(x_I) y_I^N \,|\, y^N \in \mathbb{R}^{\mathcal{P}_n(N)} \text{ and } y^N \in \mathbb{M}' \right\}, \tag{11}$$

where $\mathbb{M}'$ is the set of vectors $y^N \in \mathbb{R}^{\mathcal{P}_n(N)}$ that satisfy the following PSD conditions

$$\sum_{I\subseteq N} y_I^N = 1, \tag{12}$$

$$\sum_{I \subseteq N} y_I^N Z_I^{t+d} (Z_I^{t+d})^\top \succeq 0, \tag{13}$$

$$\sum_{I \subseteq N} g_\ell(x_I) y_I^N Z_I^t (Z_I^t)^\top \succeq 0, \quad \forall \ell \in [m]. \tag{14}$$

where $d = 0$ if $m = 0$ (no linear constraints), otherwise $d = 1$.

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 bounds rather than solving the optimization problem. It is straightforward to see that the SoS hierarchy formulation of Definition 4 is a relaxation of the integral polytope. Indeed consider any feasible integral solution $x_I \in K$ and set $y_I^N = 1$ and the other variables to zero. This solution clearly satisfies Condition (12), Condition (13) because the rank one matrix $Z_I^{t+d}(Z_I^{t+d})^\top$ is positive semidefinite (PSD), and Condition (14) since $x_I \in K$.

3 The main technical theorem

The main result of this paper (see Theorem 2 below) allows us to reduce the study of the positive semidefiniteness for matrices (13) and (14) to the analysis of "well-behaved" univariate polynomial inequalities. It can be applied whenever the solutions and constraints are *symmetric*, namely they are invariant under all permutations π of the set N: $z_I^N = z_{\pi(I)}^N$ for all $I \subseteq N$ (equivalently when $z_I^N = z_J^N$ whenever $|I| = |J|$),[4] where z_I^N is understood to denote either y_I^N or $g_\ell(x_I) y_I^N$. For example, the solution for Max- Cut considered by Grigoriev [20] and Laurent [30] belongs to this class.

Theorem 2 is somewhat technical, and should be understood on high-level as "if a certain inequality [see (18)] holds for every univariate polynomial of a special form (denoted by $G(x)$), then the related moment matrix is PSD". More precisely, the polynomials of "special" form are as follows. Here the parameter q is connected to the SoS round t and is intended to be either t or $t + 1$.

Definition 5 For any $q \in \{1, \dots, n\}$, let $\mathcal{S}_q$ be the set of all polynomials $G(x) \in \mathbb{R}[x]$ with $\deg(G) \leq 2q$, such that

$$\exists P \in \mathbb{R}[x]_q \ \text{ such that } \ G(x) = P(x)^2 \tag{15}$$

or for some $h \in \{1, \dots, q\}$, the following hold:

$$G(x) = 0, \quad \text{for } \ x \in \{0, \dots, h-1\} \cup \{n-h+1, \dots, n\} \tag{16}$$

$$G(x) \geq 0, \quad \text{for } \ x \in [h-1, \dots, n-h+1]. \tag{17}$$

[4] We define the set-valued permutation by $\pi(I) = \{\pi(i) \mid i \in I\}$.

Note that the polynomial $G(x)$ in (16) is required to be zero for a *finite set of integers*, and in (17) nonnegative on a *real interval*. Furthermore, observe that any $G(x) \in \mathcal{S}_q$ is nonnegative on the interval $[q-1, n-q+1]$.

Theorem 2 *For any fixed set of values $\{z_k^N \in \mathbb{R} : k = 0, \ldots, n\}$, if the following holds*

$$\sum_{k=0}^{n} z_k^N \binom{n}{k} G(k) \geq 0 \quad \forall G \in \mathcal{S}_q, \tag{18}$$

then the matrix

$$\sum_{k=0}^{n} z_k^N \sum_{\substack{I \subseteq N \\ |I|=k}} Z_I^q (Z_I^q)^\top \tag{19}$$

is positive-semidefinite.

Observe that if $q > \lfloor n/2 \rfloor$, constraints (18) are trivially satisfied for polynomials $G(x) \in \mathcal{S}_q$ with $h > \lfloor n/2 \rfloor$.

Theorem 2 is a actually a corollary of a technical theorem that is not strictly necessary for the applications of this paper, and therefore deferred to a later section (see Theorem 6 in Sect. 6). The proof (given in Sect. 6) is obtained by exploiting the high symmetry of the eigenvectors of the matrix appearing in (19). Condition (18) corresponds to the requirement that the Rayleigh quotient being non-negative restricted to some highly symmetric vectors (which we show are the only ones we need to consider).

4 Max-Cut for the complete graph

In the Max- Cut problem, we are given an undirected graph $G = (V, E)$ with set of vertices V and edges E, and we wish to partition the vertices in two sets such that the number of edges with end points in different sets of the partition (a cut) is maximized. For the complete graph with n vertices, consider any solution with ω vertices on one side and the remaining $n - \omega$ on the other side of the partition. This gives a cut of value $\omega(n - \omega)$.

In order to show a lower bound, we consider a *real valued* $\omega \geq 0$ and choose it such that the cut given by the SoS hierarchy has superoptimal value. In particular, when n is odd and for *any* $\omega \leq n/2$, Grigoriev [20] and Laurent [30] considered the following solution (reformulated in the basis considered in Definition 4, see "Appendix A"):

$$y_I^N = (n+1)\binom{\omega}{n+1} \frac{(-1)^{n-|I|}}{\omega - |I|} \quad \forall I \subseteq N. \tag{20}$$

Notice here that for any real $\omega \geq 0$, $\binom{\omega}{n+1}$ is defined as $\frac{\omega(\omega-1)\cdots(\omega-n+1)}{n!}$. It can be checked that the SoS hierarchy attains the value $\omega(n - \omega)$ using solution (20) (see

"Appendix A"). The papers [20,30] then show that (20) is a feasible solution for the SoS hierarchy for *any* $\omega \leq n/2$ up to round $t \leq \lfloor \omega \rfloor$. In particular for $\omega = n/2$ the cut value of the SoS relaxation is strictly larger than the value of the optimal integral cut (i.e. $\lfloor \frac{n}{2} \rfloor (\lfloor \frac{n}{2} \rfloor + 1)$), showing therefore an integrality gap at round $\lfloor n/2 \rfloor$.

We note that the formula for the solution (20) is essentially implied by the requirement of having exactly ω vertices (perhaps fractionally) on one side of the partition (see [20,30,34] for more details) and the core of the analysis in [20,30] is in showing that (20) is a feasible solution for the SoS hierarchy.

By taking advantage of Theorem 2, the proof that (20) is a feasible solution for the SoS relaxation becomes a short calculation.

Theorem 3 *The vector y^N given by (20) is a feasible solution to the SoS hierarchy up to round $t \leq \lfloor \omega \rfloor$.*

Proof In order to simplify notation, we write $z_i^N = y_I^N$ for each I such that $|I| = i \in \mathbb{N}$, since y^N in (20) depends only on the size of the set I. To show the feasibility, we have to show that the matrix $\sum_{k=0}^{n} z_k^N \sum_{\substack{I \subseteq N \\ |I|=k}} Z_I^t (Z_I^t)^\top$ is PSD. Theorem 2 implies that in order to achieve this, we only need to show that for any polynomial $G(x) \in \mathcal{S}_t$ we have

$$\sum_{k=0}^{n} \binom{n}{k} z_k^N G(k) \geq 0.$$

By the polynomial remainder theorem $G(k) = (\omega - k)Q(k) + G(\omega)$, where $Q(k)$ is a unique polynomial of degree at most $2t - 1$. It follows that

$$\sum_{k=0}^{n} \binom{n}{k} z_k^N G(k) = \underbrace{\sum_{k=0}^{n} \binom{n}{k} z_k^N (\omega - k) Q(k)}_{=0} + G(\omega) \underbrace{\sum_{k=0}^{n} \binom{n}{k} z_k^N}_{=1} = G(\omega)$$

since $\sum_{k=0}^{n} (-1)^k \binom{n}{k} Q(k) = 0$ for any polynomial of degree at most $n - 1$.[5]

Now the feasibility of (20) follows by Theorem 2 condition (17), since we have that $G(\omega) \geq 0$ whenever $t \leq \omega$ for $\omega \leq n/2$. □

Notice that the proof only uses the fact that $G \in \mathcal{S}_t$ in controlling the degree of G (and thus the polynomial Q), and in order to have $G(\omega) \geq 0$. On the other hand, the proof works for any sequence of numbers $z_k^N \in \mathbb{R}$, $k = 0, \ldots, n$ (with perhaps small modifications) as long as the sequence alternates in sign (i.e., contains the term $(-1)^k$) and simplifies to a polynomial in k as happened here when the denominator $\omega - k$ cancelled out.

[5] A quick calculation reveals that $(1 - x)^k = \sum_{k=0}^{n} (-1)^k \binom{n}{k} x^k$. Taking the jth derivative with $j < n$ on both sides, setting $x = 1$ and simplifying yields $\sum_{k=0}^{n} (-1)^k \binom{n}{k} k(k-1) \cdots (k - j + 1) = 0$. Using this derivation one can show inductively that $\sum_{k=0}^{n} (-1)^k \binom{n}{k} k^j = 0$ for every $0 \leq j < n$, and by taking linear combinations of such expressions one obtains that $\sum_{k=0}^{n} (-1) \binom{n}{k} Q(k) = 0$ for any polynomial Q of degree at most $n - 1$.

5 Min-Knapsack with cover inequalities

The MIN- KNAPSACK problem is defined as follows: we have n items with costs c_i and profits p_i, and we want to choose a subset of items such that the sum of the costs of the selected items is minimized and the sum of the profits is at least a given demand P. Formally, this can be formulated as an integer program (IP) $\min\{\sum_{j=1}^{n} c_j x_j : \sum_{j=1}^{n} p_j x_j \geq P, x \in \{0,1\}^n\}$. It is easy to see that the natural linear program (LP), obtained by relaxing $x \in \{0,1\}^n$ to $x \in [0,1]^n$ in (IP), has an unbounded integrality gap. This can be seen by considering the instance

$$
\begin{aligned}
(LP') \min \sum_{j=1}^{n} x_j \quad \text{s.t.} \sum_{j=1}^{n} x_j &\geq \frac{1}{n} \\
0 \leq x_j \leq 1 \quad &\text{for each} \quad j \in N
\end{aligned}
$$

Indeed, the optimal integral solution has the value 1, whereas $x_1 = 1/n$, $x_2 = x_3 = \cdots = x_n = 0$ is a feasible solution to (LP') with objective value $1/n$.

One way to strengthen (LP) is by adding the *Knapsack Cover* (KC) inequalities introduced by Wolsey [42] (see also [12]). Then, the arbitrarily large integrality gap of (LP) reduces to 2 (and it is tight [12]). The KC constraints applied to a linear constraint $\sum_{j=1}^{n} p_j x_j \geq P$ are

$$
\sum_{j \notin A} p_j^A x_j \geq P - p(A) \quad \text{for all} \quad A \subseteq N \quad \text{such that} \quad \sum_{i \in A} p_i < P,
$$

where $p(A) = \sum_{i \in A} p_i$ and $p_j^A = \min\left\{p_j, P - p(A)\right\}$. Note that these constraints are valid constraints for integral solutions. Indeed, in the "integral world" if a set A of items is picked we still need to cover $P - p(A)$; the remaining profits are "trimmed" to be at most $P - p(A)$ and this again does not remove any feasible integral solution.

It is easy to see that when the KC inequalities are applied to (LP'), the integrality gap vanishes. However, for an example where the KC constraints do not improve the starting linear program, consider the instance [12]

$$
(LP'') \min \sum_{j=1}^{n} x_j \quad \text{s.t.} \sum_{j=1}^{n} x_j \geq 1 + \frac{1}{n-1} \quad 0 \leq x_j \leq 1 \text{ for each } j \in N
$$

The KC inequalities for the inequality $\sum_{j=1}^{n} x_j \geq 1 + \frac{1}{n-1}$ are defined for $A = \emptyset$ and $A = \{i\}$ for $i \in N$, in other cases $\sum_{i \in A} p_i = |A| > 1 + 1/(n-1)$. For the case $A = \emptyset$ we get the original constraint, and for $A = \{i\}$ we have that $p(A) = 1$ and $p_j^A = \min\left\{1, 1 + \frac{1}{n-1} - 1\right\} = \frac{1}{n-1}$. Therefore, we obtain the strengthened formulation

$$(LP^+)\ \min \sum_{j=1}^{n} x_j \quad \text{s.t.} \sum_{j=1}^{n} x_j \geq 1 + 1/(n-1) \tag{21}$$

$$\sum_{j \in N \setminus \{i\}} x_j \geq 1 \quad \forall i \in N$$

$$0 \leq x_j \leq 1 \quad \text{for each} \quad j \in N \tag{22}$$

Note that the solution $x_j = 1/(n-1)$ is a valid fractional solution of value $1+1/(n-1)$ whereas the optimal integral solution has value 2. In [26], the authors of this paper show that the SoS hierarchy does not improve the unbounded integrality gap of (LP') even after $n-1$ rounds. Therefore it is interesting to ask if the SoS hierarchy can improve the integrality gap of the stronger formulation (LP^+). In this section we answer this question with a negative by showing that $\text{SoS}_t(LP^+)$, with t arbitrarily close to a logarithmic function of n, admits the same integrality gap as the initial linear program (LP^+) relaxation.

The considered instance of the Min- Knapsack with KC inequalities can be seen as an instance of the Set- Covering problem. Indeed, this instance was considered in [8] [note that constraint (21) is a conical combination of the constraints (22)]. In [8] an open question was raised asking what is the rank of this polytope, conjecturing that the rank is at least $n/4$, based on the numerical experiments. Our next Theorem 4 supports this conjecture by proving that the rank is at least $\log^{1-\delta}(n)$ for any $\delta > 0$. Moreover it shows that at this level the integrality gap is still asymptotically close to 2.

Theorem 4 *For any $\delta > 0$ and sufficiently large n', let $t = \lfloor \log^{1-\delta} n' \rfloor$ and $n = \lfloor \frac{n'}{t} \rfloor t$. Then there exists an $\epsilon = o(t^{-1})$ such that the following solution is feasible for* $\text{SoS}_t(LP^+)$ *with integrality gap of* $2 - o(1)$

$$y_I^N = \binom{n}{|I|}^{-1} \cdot \begin{cases} \frac{(1+\epsilon)n}{(n-1)\lfloor \log n \rfloor} & \text{for } |I| = \lfloor \log n \rfloor \\ \frac{\epsilon t}{jn} & \text{for } |I| = j\frac{n}{t} \text{ and } j \in [t] \\ 1 - \sum_{\emptyset \neq I \subseteq N} y_I^N & \text{for } I = \emptyset \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

The solution y^N in Theorem 4 is symmetric. Therefore, for notational convenience we denote by $z_k^N = y_I^N$ for every I such that $|I| = k$.

5.1 Overview of the proof

In this section we describe the intuition behind obtaining Theorem 4. An integrality gap proof for the SoS hierarchy can in general be thought of having two steps: first, choosing a solution to the hierarchy that attains a superoptimal value, and second showing that this solution is feasible for the hierarchy. We take advantage of Theorem 2 in both steps. Before going into the technical details, we describe the overview of our integrality gap construction while keeping the discussion informal.

Choosing the solution. We now give a brief explanation of how the solution y_I^N was found. Due to the symmetry in the problem formulation, it is natural to make the symmetry assumption $y_I^N = y_J^N = z_k^N$ for each I, J such that $|I| = |J| = k$. Furthermore, for every $I \subseteq N$ we set $y_I^N \geq 0$ in order to satisfy (13) for free. The objective function of the SoS relaxation of the problem is $\sum_{I \subseteq N} y_I^N |I|$, and thus in order to have an integrality gap (i.e., a small objective function value), we would like to have $y_0^N \approx 1$ forcing the other variables to be small due to (12).

We then show that satisfying (14) for the constraint (22) follows if it holds

$$\sum_{k=0}^{n} \binom{n}{k} z_k^N (k-1) \prod_{i=1}^{t} (k - r_i)^2 \geq 0, \tag{24}$$

for every choice of t real variables r_i. We get this condition by observing similarities in the structure of the constraints and applying Theorem 2, then expressing the polynomial in root form.[6] The only negative term in the sum corresponds to z_0^N. Then, it is clear that we need at least $t+1$ non-zero variables z_k^N, otherwise the roots r_i can be set such that the positive terms in (24) vanish and the inequality is not satisfied. Therefore, we choose exactly $t+1$ of the z_k^N to be strictly positive (and the rest 0 excluding z_0^N), and we distribute them "far away" from each other, so that no root can be placed such that the coefficient of two positive terms become small. To take this idea further, for one "very small" k' (logarithmic in n), we set $z_{k'}^N$ positive and space out the rest evenly.

Proving that the solution is feasible. We show that (24) holds for all possible r_i with our chosen solution by analyzing two cases. In the first case we assume that all of the roots r_i are larger than $\log^3 n$. Then, we show that the "small" point k' we chose is enough to satisfy the condition. In the complement case, we assume that there is at least one root r_i that is smaller than $\log^3 n$. It follows that one of the evenly spaced points is "far" from any remaining root, and can be used to show that the condition is satisfied.

5.2 Proof of Theorem 4

We start by proving the claimed integrality gap. The defined solution has an objective value that is arbitrarily close to 1 whereas the optimal integral value is 2. Indeed, due to symmetry the value of the solution at $x_I \in \{0, 1\}^n$ is $|I|$, and thus he objective value of the relaxation is (see Lemma 4)

$$\sum_{I \subseteq N} y_I^N |I| = \frac{n}{n-1}(1+\varepsilon) + \varepsilon t \overset{n\to\infty}{=} 1.$$

Note that we have $\varepsilon = o(t^{-1})$ and $t = \lfloor \log^{1-\delta} n' \rfloor$ for a constant $\delta > 0$, so that $\varepsilon t \overset{n\to\infty}{=} 0$.

The remaining part of the Theorem 4 follows by showing that the suggested solution satisfies (12), (13) and (14). Note that (12) is immediately satisfied by the definition

[6] We show that the roots r_i can be assumed to be real numbers.

of variables $\{y_I^N\}$, and (13) is satisfied since $y_I^N \geq 0$ and the rank one matrix $Z_I^t(Z_I^t)^\top$ is positive semidefinite for every $I \subseteq N$. It remains to prove that the condition (14) is also satisfied for all the constraints (21) and (22). In fact, by summing up the constraints (22) and dividing by $n-1$, we obtain the constraint (21). Hence the constraint (21) is redundant and we only need to worry about the constraint (22). Note that the constraint (22) is not symmetric (one variable is missing and sets of variables of the same size do not play the same role with respect to this constraint). However, the following lemma shows how to solve this issue by reducing to the form (19) of Theorem 2.

Lemma 5 *Assume that the solution y_I^N satisfies*

$$\sum_{\emptyset \neq I \subseteq N} y_I^N \, (|I|-2) \, Z_I^t(Z_I^t)^\top \succeq Z_\emptyset^t(Z_\emptyset^t)^\top. \tag{25}$$

Then the SoS constraint (14) holds for the linear function $\sum_{j\in N\setminus\{i\}} x_j - 1$ for each $i \in N$.

Proof We show that (25) implies that (14) is satisfied for the cover constraint $\sum_{\substack{j=1\\ j\neq i}}^n x_j \geq 1$. For this constraint Condition (14) can be written as

$$\begin{aligned}
&\sum_{\substack{i\notin I\subseteq N\\ I\neq\emptyset}} y_I^N(|I|-1)Z_I^t(Z_I^t)^\top + \sum_{i\in I\subseteq N} y_I^N(|I|-2)Z_I^t(Z_I^t)^\top - y_\emptyset^N Z_\emptyset^t(Z_\emptyset^t)^\top \\
&\succeq \sum_{\substack{I\subseteq N\\ I\neq\emptyset}} y_I^N(|I|-2)Z_I^t(Z_I^t)^\top - y_\emptyset^N Z_\emptyset^t(Z_\emptyset^t)^\top \succeq 0.
\end{aligned}$$

which is also implied if (25) is satisfied, as $y_\emptyset^N \leq 1$. In the last step we used implicitly the fact that $y_I^N = 0$ for I such that $|I| = 1$ when n is large. $\square$

We take advantage of the symmetry in the solution y_I^N and simplify (25) such that Theorem 2 can be applied. Indeed, Theorem 2 implies that if

$$\sum_{k=1}^n z_k^N (k-2)\binom{n}{k} G(k) \geq G(0), \tag{26}$$

for every $G \in \mathcal{S}_t$, then (25) holds. We observe that if $G(0) = 0$, then (26) is satisfied, since the left hand side of the inequality is non-negative (recall that the solution in Theorem 4 is such that for large n, $y_1^N = 0$). Therefore, the only nontrivial $G(x) \in \mathcal{S}_t$ we need to consider are such that $G(x) = P^2(x)$ as in (15), and in order to complete the proof of Theorem 4 it is enough to show that the following is satisfied

$$\sum_{k=1}^n z_k^N (k-2)\binom{n}{k} P^2(k) \geq P^2(0) \quad \forall P : \deg(P) \leq t. \tag{27}$$

The following lemma further reduces the interesting cases.

Lemma 6 *In order to prove that Solution* (23) *satisfies* (27) *it is sufficient to prove that* (23) *satisfies* (27) *for polynomials* $P(x)$ *with the following properties:*

(a) *The degree of* $P(x)$ *is exactly* t.
(b) *All the roots* $r_1, \ldots, r_t$ *of* $P(x) = 0$ *are in the range,* $1 \leq r_j \leq n$ *for all* $j = 1, \ldots, t$,
(c) *All the roots* $r_1, \ldots, r_t$ *of* $P(x) = 0$ *are real numbers,*

Notice that (a) of Lemma 6 implies that (27) is equivalent to

$$\sum_{k=1}^{n} z_k^N \binom{n}{k} (k-2) \prod_{j=1}^{t} \left(\frac{r_j - k}{r_j}\right)^2 \geq 1. \tag{28}$$

In particular the case $r_i = 0$ for some i is redundant, since in that case (27) is immediately satisfied.

Proof (a) Let $P(k)$ be the univariate polynomial with degree $s < t$ with all real roots. Let $P'(k)$ be the polynomial of degree t with all real roots such that $r'_j = r_j$, $j \leq s$ and $r'_j = n$ for $s < j \leq t$
For any $k \in N$, we have

$$1 \geq \left(\frac{n-k}{n}\right)^2.$$

Hence,

$$\left(\frac{r_1 - k}{r_1}\right)^2 \cdots \left(\frac{r_s - k}{r_s}\right)^2 \geq \left(\frac{r_1 - k}{r_1}\right)^2 \cdots \left(\frac{r_s - k}{r_s}\right)^2 \left(\frac{n-k}{n}\right)^{2(t-s)},$$

and finally

$$\sum_{k=1}^{n} z_k^N \binom{n}{k} (k-2) \prod_{j=1}^{t} \left(\frac{r_j - k}{r_j}\right)^2 \geq \sum_{k=1}^{n} z_k^N \binom{n}{k} (k-2) \prod_{j=1}^{t} \left(\frac{r'_j - k}{r'_j}\right)^2.$$

(b) For $1 \leq a \leq n$ the function $(1 - a/x)^2$ is increasing in a negative orthant with the limit in $-\infty$ equal to 1. Moreover, it is decreasing when $0 < x < a$, increasing when $x > a$, converging to the value 1 in ∞. Therefore, for any $r \in (-\infty, 1) \cup (n, \infty)$ we can find an $r' \in [1, n]$ such that for every $k \in N$, $k \leq n$,

$$\left(\frac{r-k}{r}\right)^2 \geq \left(\frac{r'-k}{r'}\right)^2. \tag{29}$$

Hence, if $P(x)$ is a univariate polynomial with roots $r_1, \ldots, r_t$ such that $r_i \in (-\infty, 1) \cup (n, \infty)$ and $H(x)$ a univariate polynomial with roots $r'_1, \ldots, r'_t$ such that $r_i' \in [1, n]$ is chosen so that (29) holds for $r = r_i$ and $r' = r_i'$, for $j \neq i$ we have $r_j' = r_j$, then

$$\sum_{k=1}^{n} z_k^N \binom{n}{k} (k-2) \prod_{j=1}^{t} \left(\frac{r_j - k}{r_j}\right)^2 \geq \sum_{k=1}^{n} z_k^N \binom{n}{k} (k-2) \prod_{j=1}^{t} \left(\frac{r_j' - k}{r_j'}\right)^2.$$

Thus $\sum_{k=1}^{n} z_k^N (k-2) \binom{n}{k} P(k)^2 \geq \sum_{k=1}^{n} z_k^N (k-2) \binom{n}{k} H(k)^2$ and we conclude that we do not need to consider polynomials with roots outside the interval $[1, n]$ in our analysis.

(c) Let $P(k)$ be the univariate polynomial with $2q$ complex roots (complex roots appear in conjugate pairs) i.e. $r_{2j-1} = a_j + b_j\mathbf{i}$, $r_{2j} = a_j - b_j\mathbf{i}$ for $j = 1, \ldots, q$ and the rest real roots. Let $P'(k)$ be the polynomial with all real roots such that $r_{2j-1}' = r_{2j}' = \sqrt{a_j^2 + b_j^2}$ for $j = 1, \ldots, q$ and $r_j' = r_j$, $j > 2q$.

For any $k \in N$ and $j \in [t]$, a simple calculation shows that

$$\left(\frac{r_{2j-1} - k}{r_{2j-1}}\right)^2 \left(\frac{r_{2j} - k}{r_{2j}}\right)^2 \geq \left(\frac{r_{2j-1}' - k}{r_{2j-1}'}\right)^2 \left(\frac{r_{2j}' - k}{r_{2j}'}\right)^2.$$

Hence,

$$\sum_{k=1}^{n} z_k^N \binom{n}{k} (k-2) \prod_{j=1}^{t} \left(\frac{r_j - k}{r_j}\right)^2 \geq \sum_{k=1}^{n} z_k^N \binom{n}{k} (k-2) \prod_{j=1}^{t} \left(\frac{r_j' - k}{r_j'}\right)^2,$$

and thus it suffices to show (27) for polynomials with real roots.

□

Next we show that the vector (23) satisfies (27) for polynomials P with properties (a)-(c) from Lemma 6. Our analysis also implies that there exists an $\varepsilon = o(t^{-1})$ as claimed. We express the univariate polynomial P using its roots $r_1, \ldots, r_t$, so that (27) becomes

$$\sum_{k=1}^{n} \binom{n}{k} z_k^N (k-2) \prod_{i=1}^{t} (r_i - k)^2 \geq \prod_{i=1}^{t} r_i^2. \tag{30}$$

To show that (30) is satisfied we separate two cases: when all of the roots of the polynomial are greater or equal to a fixed threshold $\alpha = \log^3 n$ and when at least one root is smaller than this threshold. In order to simplify the computations we denote $\beta = \lfloor \log n \rfloor$.

1. $r_j \geq \alpha$ for all j. It is sufficient to show that the left-hand side term in (30) corresponding to $k = \beta$ satisfies

$$\binom{n}{\beta} y_\beta^N \, (\beta - 2) \prod_{i=1}^{t} (r_i - \beta)^2 \geq \prod_{i=1}^{t} r_i^2.$$

Replacing the variables with the values and using $\binom{n}{\beta} \geq 1$ we get

$$\frac{n}{n-1} \frac{1+\varepsilon}{\beta} \, (\beta - 2) \prod_{i=1}^{t} (r_i - \beta)^2 \geq \prod_{i=1}^{t} r_i^2$$
$$\Longleftrightarrow 1 + \varepsilon \geq \prod_{i=1}^{t} \left(\frac{r_i}{r_i - \beta} \right)^2 \frac{1}{1 - 2\beta^{-1}} \frac{n-1}{n}.$$

By Lemma 6 and assumption, all roots r_j satisfy $\alpha \leq r_j \leq n$. Since $\frac{r_i}{r_i - \beta} \leq \frac{\alpha}{\alpha - \beta}$ and $\frac{n-1}{n} < 1$, it is sufficient that it holds $1 + \varepsilon \geq \frac{1}{1-2\beta^{-1}} \left(\frac{\alpha}{\alpha - \beta} \right)^{2t}$.

2. There is at least one root r_j such that $r_j < \alpha$. It can be shown by straightforward induction on the number of roots that if for at least one j, $r_j < \alpha$, then there exists a point $u = l\frac{n}{t}$, $l = 1, \ldots, t$ such that $\binom{n}{u} y_u > 0$ and $|u - r_i| \geq \frac{n}{2t}$ for all $i = 1, \ldots, t$. Let u be such a point. It is sufficient to show that we can satisfy $\binom{n}{u} y_u^N \, (u-2) \prod_{i=1}^{t} (r_i - u)^2 \geq \prod_{i=1}^{t} r_i^2$.
   We have $\binom{n}{u} y_u^N = \frac{\varepsilon}{u}$ and the estimates $u - 2 \geq \frac{u}{2}$, $(r_i - u)^2 \geq \frac{n^2}{(2t)^2}$, $\prod_{i=1}^{t} r_i \leq n^{t-1}\alpha$. Substituting these we get the condition $\frac{\varepsilon}{2} \left(\frac{n}{2t} \right)^{2t} \geq n^{2t-2} \alpha^2$ which gives us the requirement that $\varepsilon \geq \frac{2\alpha^2}{n^2} (2t)^{2t}$.

These two cases suggest that we fix ε as

$$\varepsilon = \max \left\{ \frac{1}{1 - 2\beta^{-1}} \left(1 - \frac{\beta}{\alpha} \right)^{-2t} - 1, \; \frac{2\alpha^2}{n^2} (2t)^{2t} \right\}.$$

The proof has now been reduced to showing that with this choice of ε we have $\varepsilon t \to 0$, i.e., $\varepsilon = o(t^{-1})$. Assume $\varepsilon = \frac{1}{1-2\beta^{-1}} (1 - \frac{\beta}{\alpha})^{-2t} - 1$. Then $\varepsilon t = t (\frac{1}{1-2\beta^{-1}} (1 - \frac{\beta}{\alpha})^{-2t} - 1) \leq t (\frac{1}{1-2\beta^{-1}} e^{4t\frac{\beta}{\alpha}} - 1)$, when $\beta / \alpha \leq 1/2$, using the estimate $1 - x \geq e^{-2x} \Rightarrow (1-x)^{-2t} \leq e^{4xt}$ which holds when $x \leq 1/2$. Furthermore, the same estimate yields $e^x - 1 \leq 2x$ when $x \leq 1/2$. Hence, we have the bound

$$\varepsilon t \leq t \frac{1}{1 - 2\beta^{-1}} \cdot 8t \frac{\beta}{\alpha} + t \left(\frac{1}{1 - 2\beta^{-1}} - 1 \right) = \frac{8}{1 - 2\beta^{-1}} \cdot t^2 \frac{\beta}{\alpha} + \frac{2t\beta^{-1}}{1 - 2\beta^{-1}}.$$

The right-hand side goes to 0 if $\frac{t^2 \beta}{\alpha} \to 0$ and $\frac{t}{\beta} \to 0$ as $n \to \infty$. This is clearly the case for $t \leq \log^{1-\delta} n$, for any $\delta > 0$.

Next, assume $\varepsilon = \frac{2\alpha^2}{n^2} (2t)^{2t}$. Then $\varepsilon t = t \frac{2\alpha^2}{n^2} (2t)^{2t}$, which immediately yields the condition on α and t that we need $\frac{t\alpha^2}{n^2} (2t)^{2t} \to 0$ as $n \to \infty$. Substituting $\alpha = \log^3 n$ and $t = \log^{1-\delta} n$, for any $\delta > 0$, allows us to write this as

$\frac{t\alpha^2}{n^2}(2t)^{2t} = \log^{1-\delta} n \frac{\log^6 n}{n^2}(2\log^{1-\delta} n)^{2\log^{1-\delta} n}$. By a change of variables of the form $w = \log^{1-\delta} n$ we get

$$\frac{w^{2w+\frac{7-\delta}{1-\delta}}2^{2w}}{e^{2w^{\frac{1}{1-\delta}}}} \leq \frac{w^{4w+\frac{7-\delta}{1-\delta}}}{e^{2w^{\frac{1}{1-\delta}}}} = \frac{e^{(4w+\frac{7-\delta}{1-\delta})\log w}}{e^{2w^{\frac{1}{1-\delta}}}} = e^{(4w+\frac{7-\delta}{1-\delta})\log w - 2w^{\frac{1}{1-\delta}}},$$

which tends to 0 as $n \to \infty$.

5.3 Further results

In a recent paper [25] the authors characterize the class of $0/1$ integer linear programming problems that are *maximally hard* for the SoS hierarchy. Here, maximally hard means those problems that still have an integrality gap even after $n-1$ rounds of the SoS hierarchy.[7] An illustrative natural member of this class is given by the simple LP relaxation for the Min Knapsack problem, i.e.

$$(LP)\ \min\left\{\sum_{j=1}^{n} x_j : \sum_{j=1}^{n} x_j \geq \frac{1}{P}, x \in [0,1]^n\right\}.$$

In [25] it is shown that at level $n-1$ the integrality gap is k, for any $k \geq 2$ if and only if $P = \Theta(k)\cdot 2^{2n}$. A natural question is to understand if the SoS hierarchy is able to reduce the gap when P is "small".

This problem, for $P = 2$, was considered by Cook and Dash [16] as an example where the Lovasz-Schrijver hierarchy rank is n. Laurent [29] showed that the Sherali-Adams hierarchy rank is also equal to n and raised the open question to find the rank for the Lasserre hierarchy. She also showed that when $n = 2$, the Lasserre relaxation has an integrality gap at level 1, but leaves open whether or not this happens at level $n-1$ for a general n. In [25] it is ruled out the possibility that the Lasserre/SoS rank is n for $n \geq 3$.

The following lemma provides a feasible solution for $\text{SoS}_t(LP)$ with integrality gap arbitrarily close to P when $t = \Omega(\log^{1-\varepsilon} n)$ and for any $P > 1$. The proof is omitted since it is similar to the proof of Theorem 4.

Theorem 5 *For any $\delta > 0$ and sufficiently large n', let $t = \lfloor \log^{1-\delta} n' \rfloor$, $n = \lfloor \frac{n'}{t} \rfloor t$ and $\epsilon = o(t^{-1})$. Then the following solution is feasible for $\text{SoS}_t(LP^+)$ with integrality gap arbitrarily close to P.*

$$y_I^N = \binom{n}{|I|}^{-1} \cdot \begin{cases} \frac{(1+\epsilon)}{P\lfloor \log n \rfloor} & \text{for } |I| = \lfloor \log n \rfloor \\ \frac{\epsilon t}{jn} & \text{for } |I| = j\frac{n}{t} \text{ and } j \in [t] \\ 1 - \sum_{\emptyset \neq I \subseteq N} y_I^N & \text{for } I = \emptyset \\ 0 & \text{otherwise} \end{cases} \qquad (31)$$

[7] Recall that at level n the integrality gap vanishes.

6 Proof of Theorem 2

Theorem 2 is actually a corollary of a stronger statement we prove in this section and that provides *necessary and sufficient* conditions for the matrix (19) being positive-semidefinite. This statement is given in Theorem 6, and it uses a special family of polynomials that we introduce first. We remind the reader here that the falling factorial is defined as $x^{\underline{m}} = x(x-1)\cdots(x-m+1)$ (with the convention that $x^{\underline{0}} = 1$).

Definition 6 For any $h \in \{0,\ldots,t\}$ and scalars $\alpha_{ij} \in \mathbb{R}$ (for $0 \le i \le t$ and $0 \le j \le \min\{h, i\}$), let $G_h(x) \in \mathbb{R}[x]$ be a univariate polynomial defined as follows

$$G_h(x) = \sum_{r=0}^{h} \binom{h}{r} h_r(x) \left(\sum_{j=0}^{h} \binom{r}{j} p_j(x-r) \right)^2 \tag{32}$$

where $h_r(x) = x^{\underline{r}} \cdot (n-x)^{\underline{h-r}}$ and $p_j(x-r) = \sum_{i=0}^{t-j} \alpha_{i+j,j} \binom{x-r}{i}$.

We remark that the polynomials $G_h(x)$ of Definition 6 are in $\mathcal{S}_t$ of Definition 5 and thus satisfy the conditions of Theorem 2, as shown in Lemma 14 later.

Theorem 6 *Let $z_k^N \in \mathbb{R}$ for $k \in \{0,\ldots,n\}$. Then for any $t \in \{1,\ldots,n\}$ the matrix*

$$\sum_{k=0}^{n} z_k^N \sum_{\substack{I \subseteq N \\ |I|=k}} Z_I^t (Z_I^t)^\top \tag{33}$$

is positive-semidefinite if and only if

$$\sum_{k=0}^{n} z_k^N \binom{n}{k} G_h(k) \ge 0 \quad \textit{for} \quad h \in \{0,\ldots,t\}, \tag{34}$$

for every univariate polynomial $G_h(x) \in \mathbb{R}[x]$ of degree at most $2t$ as defined in Definition 6.

By Lemma 14, Theorem 2 is a straightforward corollary of Theorem 6.

6.1 Proof of Theorem 6

We study when the matrix $M = \sum_{k=0}^{n} z_k \sum_{I \subseteq N, |I|=k} Z_I^t (Z_I^t)^\top$, where is positive-semidefinite. Theorem 6 allows us to reduce the condition $M \succeq 0$ to inequalities of the form $\sum_{k=0}^{n} \binom{n}{k} z_k p(k) \ge 0$, where $p(k)$ is a univariate polynomial of degree $2t$ with some additional remarkable properties.

A basic key idea that is used to obtain such a characterization is that the eigenvectors of M are "very well" structured. This structure is used to get $p(k)$ with the claimed properties.

The structure of the eigenvectors. Let Π denote the group of all permutations of the set N, i.e. the *symmetric* group. Let P_π be the permutation matrix of size $\mathcal{P}_t(N) \times \mathcal{P}_t(N)$ corresponding to any permutation π of set N, i.e. for any vector v we have $[P_\pi v]_I = v_{\pi(I)}$ for any $I \in \mathcal{P}_t(N)$ (see Footnote 4). Note that $P_\pi^{-1} = P_\pi^\top$.

Lemma 7 *For every $\pi \in \Pi$ we have $P_\pi^\top M P_\pi = M$ or, equivalently, M and P_π commute $M P_\pi = P_\pi M$.*

Proof We have that $[P_\pi^\top M P_\pi]_{I,J} = M_{\pi(I),\pi(J)} = M_{I,J}$, where the first equality follows from the fact that the conjugation by P_π permutes the rows and columns of M by π. The second equality holds, since the entries $M_{I,J}$ depend only on $|I \cup J|$ and we have that $|\pi(I) \cup \pi(J)| = |I \cup J|$. □

Corollary 1 *If $w \in \mathbb{R}^{\mathcal{P}_t(N)}$ is an eigenvector of M then $v = P_\pi w$ is also an eigenvector of M for any $\pi \in \Pi$.*

Proof By the assumption, $Mw = \lambda w$ and by Lemma 7, $Mv = M(P_\pi w) = P_\pi M w = \lambda v$. □

Corollary 1 allows us to consider eigenvectors that are symmetric with respect to the action of permutation groups.

Definition 7 Let $w \in \mathbb{R}^{\mathcal{P}_t(N)}$ and G a group of permutations that act on w element-wise by $[P_\pi w]_I = w_{\pi(I)}$. Then the *symmetrization of w w.r.t. G* is the vector

$$u = \sum_{\pi \in G} P_\pi w.$$

Notice that the symmetrization of an arbitrary vector can be the zero vector. Hence, by Corollary 1 the symmetrization of an eigenvector of the moment matrix M is either also an eigenvector corresponding to the same eigenvalue, or the zero vector. An important property of the symmetrization u of a vector w is that it is *invariant* under the action of the corresponding permutation group, i.e., $P_\pi u = u$ for each permutation π in the group.

Definition 8 For any $H \subseteq N$, we denote by Π_H the *stabilizer* of H, i.e.,$\pi \in \Pi_H \Leftrightarrow \pi(H) = H$. We consider $\Pi_\emptyset = \Pi$.

Note that the definition is equivalent to saying that $\pi \in \Pi_H$ if and only if $\pi(i) \in H$ for every $i \in H$ and $\pi(i) \notin H$ for every $i \notin H$.

We want to avoid the situation where the symmetrization of the eigenvector w is the zero vector by choosing the group of symmetrization. Hence, we choose a subset $H \subseteq N$ of the smallest cardinality such that $u = \sum_{\pi \in \Pi_H} P_\pi w \neq 0$. Such a set H always exists, since $w_I \neq 0$ implies that $\sum_{\pi \in \Pi_I} P_\pi w \neq 0$, and thus one can choose $H = I$ of the smallest cardinality. The choice of H is not unique, but we can always assume that it is the subset of the first $h = |H|$ elements from N, i.e. $H = \{1, \ldots, h\}$. Indeed, if it is not the case, there exists a permutation $\pi \in \Pi$ that maps H to the subset of the first $|H|$ elements from N and such that $P_\pi w$ is an eigenvector of M

by Corollary 1. Now it holds that $u \neq 0$ and the vector $u/\|u\|$ is a unit eigenvector corresponding to the same eigenvalue as w and is invariant under the action of Π_H. We summarize the discussion so far in the following lemma.

Lemma 8 *Let $w \in \mathbb{R}^{\mathcal{P}_t(N)}$ be a unit eigenvector of M corresponding to eigenvalue λ, and H be the smallest subset of N such that $u = \sum_{\pi \in \Pi_H} P_\pi w \neq 0$. Then $u/\|u\|$ is also a unit eigenvector of M corresponding to eigenvalue λ.*

The following lemma shows the structure of eigenvectors obtained by symmetrizing arbitrary eigenvectors.

Lemma 9 *Let $u = \sum_{\pi \in \Pi_H} P_\pi w$. Then the vector u is invariant under the permutations of Π_H, namely $u_I = u_{\sigma(I)}$ for each $\sigma \in \Pi_H$. Equivalently, $u_I = u_J$ for all $|I| = |J|$ such that $|I \cap H| = |J \cap H|$.*

Proof We have $u_I = \left[\sum_{\pi \in \Pi_H} P_\pi w\right]_I = \sum_{\pi \in \Pi_H} [P_\pi w]_I = \sum_{\pi \in \Pi_H} w_{\pi(I)} = \sum_{\pi \in \Pi_H} w_{\pi(\sigma(I))} = u_{\pi(I)}$ where the last but one equality follows since permutations are bijective. The claim follows by observing that for all $|I| = |J|$ such that $|I \cap H| = |J \cap H|$ there exists $\pi \in \Pi_H$ such that $\pi(I) = J$. □

Lemma 9 and the arguments above imply Lemma 10.

Lemma 10 *Let $h \in \{0, \ldots, t\}$. For any vector u_h that is invariant under the action of the group Π_H with $H = \{1, \ldots, h\}$, we have*

$$u_h = \sum_{i=0}^{t} \sum_{j=0}^{\min\{h,i\}} \alpha_{i,j} b_{i,j}, \tag{35}$$

where $\alpha_{i,j} \in \mathbb{R}$ and $b_{i,j} \in \mathbb{R}^{\mathcal{P}_t(N)}$ such that $[b_{i,j}]_Q = 1$ if $|Q| = i$ and $|Q \cap H| = j$, $[b_{i,j}]_Q = 0$ otherwise.

What we have achieved now is that instead of considering arbitrary eigenvectors of the moment matrix M, we only need to consider symmetrized vectors of the form (35). More precisely, positive semidefiniteness of M is equivalent to saying that $w^\top M w \succeq 0$ for each eigenvector w of M. Using symmetrization and Lemma 10, the PSDness of M follows by ensuring that for any $h = 0, 1, \ldots, t$ we have $u_h^\top M u_h \geq 0$, i.e.

$$u_h^\top M u_h = \sum_{k=0}^{n} z_k \sum_{\substack{I \subseteq N \\ |I|=k}} \left(u_h^\top Z_I^t\right)^2 = \sum_{k=0}^{n} z_k \sum_{\substack{I \subseteq N \\ |I|=k}} \left(\sum_{i=0}^{t} \sum_{j=0}^{\min\{h,i\}} \alpha_{i,j} b_{i,j}^\top Z_I^t\right)^2 \geq 0.$$

We can now connect the polynomial G_h of Definition 6 to the moment matrix M. Note that we do not require here that the vectors u_h are symmetrizations of eigenvectors of M, but any vector that is invariant under the action of the group Π_H will do. When considering the PSDness of M in Theorem 6, we only use the fact that *any* eigenvector of M can be symmetrized and thus we can obtain a corresponding symmetric vector u_h.

Lemma 11 *Let u_h be a vector of the form (35), and G_h the polynomial of Definition 6. Then*

$$\sum_{\substack{I \subseteq N \\ |I|=k}} \left(u_h^\top Z_I^t\right)^2 = \binom{n}{k} \frac{1}{n^{\underline{h}}} G_h(k).$$

In particular the numbers $\alpha_{i,j} \in \mathbb{R}$ in (35) and in Definition 6 are the same. It follows that

$$u_h^\top M u_h = \sum_{k=0}^{n} z_k \binom{n}{k} \frac{1}{n^{\underline{h}}} G_h(k).$$

Proof We start by noting that for every $i = 0, \ldots, t$, $j = 0, \ldots, |H|$ we have[8]

$$b_{i,j}^\top Z_I^t = \binom{|I \cap H|}{j} \binom{|I \setminus H|}{i-j}.$$

Indeed

$$b_{i,j}^\top Z_I^t = \sum_{Q \in \mathcal{P}_t(N)} (b_{i,j})_Q (Z_I^t)_Q = \sum_{\substack{Q \subseteq I, |Q|=i \\ |Q \cap H|=j}} 1 = \binom{|I \cap H|}{j} \binom{|I \setminus H|}{i-j}$$

It follows that we have

$$\begin{aligned}
\sum_{\substack{I \subseteq N \\ |I|=k}} \left(u_h^\top Z_I^t\right)^2 &= \sum_{\substack{I \subseteq N \\ |I|=k}} \left(\sum_{i=0}^{t} \sum_{j=0}^{|H|} \alpha_{i,j} b_{i,j}^\top Z_I^t\right)^2 \\
&= \sum_{\substack{I \subseteq N \\ |I|=k}} \left(\sum_{i=0}^{t} \sum_{j=0}^{|H|} \alpha_{i,j} \binom{|I \cap H|}{j} \binom{|I \setminus H|}{i-j}\right)^2
\end{aligned}$$

Splitting the sum over I by considering the intersections $I \cap H$ of sizes $r = 0, \ldots, |H|$, we have

$$\begin{aligned}
&\sum_{r=0}^{|H|} \sum_{\substack{|I|=k \\ |I \cap H|=r}} \left(\sum_{i=0}^{t} \sum_{j=0}^{|H|} \alpha_{i,j} \binom{r}{j} \binom{k-r}{i-j}\right)^2 \\
&= \sum_{r=0}^{|H|} \binom{|H|}{r} \binom{n-|H|}{k-r} \left(\sum_{j=0}^{|H|} \binom{r}{j} \sum_{i=0}^{t} \alpha_{i,j} \binom{k-r}{i-j}\right)^2.
\end{aligned}$$

[8] Recall that $\binom{n}{-k} = \binom{n}{n+k} = 0$ for any positive integer k.

Finally, we shift the sum over i by j and thus justify the equality

$$\sum_{\substack{I\subseteq N\\|I|=k}}\left(u_h^\top Z_I^t\right)^2=\sum_{r=0}^{|H|}\binom{|H|}{r}\binom{n-|H|}{k-r}\left(\sum_{j=0}^{|H|}\binom{r}{j}\sum_{i=0}^{t-j}\alpha_{i+j,j}\binom{k-r}{i}\right)^2.$$

Now, the sum over i is a Newton polynomial that we denote by $p_j(k-r)=\sum_{i=0}^{t-j}\alpha_{i+j,j}\binom{k-r}{i}$. Note that by definition, here $\deg(p)\leq t-j$. Furthermore, observe that

$$\binom{n-|H|}{k-r}=\binom{n}{k}\frac{1}{n^{\underline{|H|}}}k^{\underline{r}}\cdot(n-k)^{\underline{|H|-r}},$$

and writing $h_r(k)=k^{\underline{r}}\cdot(n-k)^{\underline{|H|-r}}$ yields the claim. □

6.2 Properties of the univariate polynomials G_h

In this section we provide some properties of the polynomials G_h. In particular, we show that the polynomials G_h of Definition 6 are contained in the set $\mathcal{S}_t$ of Definition 5, by showing that the degree of G_h is bounded and proving that G_h satisfies (16) and (17). Once we have bounded the degree of G_h, it is straightforward to verify that G_0 satisfies condition (15).

We begin by recalling Vandermonde's identity and proving a useful lemma.

Lemma 12 (Vandermonde's identity) *Let m, n and r be nonnegative integers. Then*

$$\binom{m+n}{h}=\sum_{r=0}^{h}\binom{m}{r}\binom{n}{h-r}.$$

Lemma 13 *Let h, i, j, m, n, p be nonnegative integers and*

$$C(x)=\sum_{r=0}^{h}\binom{h}{r}x^{\underline{r}}(n-x)^{\underline{h-r}}f(r),$$

with

$$f(r)=\binom{r}{i}\binom{r}{j}\binom{h-r}{m}\binom{h-r}{p}.$$

Then the $C(x)$ is a polynomial in x of degree at most the degree of $f(r)$ as a polynomial in r.

Proof Notice that the degree of $f(r)$ is $i+j+m+p$. Recall that the forward difference of function $g(X)$ with respect to variable X is a finite difference defined by $\Delta_X[g(X)] = g(X+1) - g(X)$. Higher order differences are obtained by repeated operations of the forward difference operator, and we use $\Delta_X^{\ell}[g(X)]_{X=b}$ to denote the ℓth forward difference evaluated at $X = b$. The following identity can be proven with a simple induction argument

$$\Delta_X^d[(x+X)^{\underline{r+d}}] = (x+X)^{\underline{r}}(r+d)^{\underline{d}}.$$

First note that the polynomials $P_d(r) = (r+1)^{\overline{d}} = (r+d)^{\underline{d}}$ with $0 \le d \le \delta$ form a basis of the space of real polynomials of degree at most δ. Hence the claim follows by showing that the degree of the following $C'(x)$ (as a polynomial in x) is at most the degree of $P_d(r)$ (as a polynomial in r)

$$\begin{aligned} C'(x) &= \sum_{r=0}^{h} \binom{h}{r} (n-x)^{\underline{h-r}} \cdot x^{\underline{r}} (r+d)^{\underline{d}} \\ &= \sum_{r=0}^{h} \binom{h}{r} (n-x)^{\underline{h-r}} \cdot \Delta_X^d \left[(x+X)^{\underline{r+d}} \right]_{X=0} \\ &= \Delta_X^d \left[\sum_{r=0}^{h} \binom{h}{r} (n-x)^{\underline{h-r}} (x+X)^{\underline{r+d}} \right]_{X=0}, \end{aligned}$$

where at the last step we use the linearity of the forward difference operator. We can further simplify the above to

$$\begin{aligned} &\Delta_X^d \left[(x+X)^{\underline{d}} \sum_{r=0}^{h} \binom{h}{r} (n-x)^{\underline{h-r}} (x+X-d)^{\underline{r}} \right]_{X=0} \\ &= \Delta_X^d \left[(x+X)^{\underline{d}} \sum_{r=0}^{h} \binom{h}{r} r!(h-r)! \binom{n-x}{h-r} \binom{x+X-d}{r} \right]_{X=0} \end{aligned}$$

and now using Vandermonde's identity (Lemma 12) obtain from (36) that

$$\begin{aligned} C'(x) &= \Delta_X^d \left[(x+X)^{\underline{d}} h! \binom{n-x+x+X-d}{h} \right]_{X=0} \\ &= \Delta_X^d \left[(x+X)^{\underline{d}} (n+X-d)^{\underline{h}} \right]_{X=0}. \end{aligned}$$

The claim follows by observing that the forward difference operator does not increase the degree of its argument and therefore $C'(x)$ has degree at most d. $\square$

Lemma 14 *For any $h \in \{0, \ldots, t\}$, the polynomials $G_h(x)$ as defined in Definition 6 have the following properties:*

(a) *$G_h(x)$ is a univariate polynomial of degree at most $2t$,*
(b) *$G_h(x) \ge 0$ for $x \in [h-1, n-h+1]$*
(c) *$G_h(x) = 0$ for every $x \in \{0, \ldots, h-1\} \cup \{n-h+1, \ldots, n\}$.*

Proof of (a).

$$
\begin{aligned}
G_h(k) &= \sum_{r=0}^{h} \binom{h}{r} h_r(k) \left(\sum_{j=0}^{h} \binom{r}{j} p_j(k-r) \right)^2 \\
&= \sum_{r=0}^{h} \binom{h}{r} h_r(k) \left(\sum_{i=0}^{h} \sum_{j=0}^{h} \binom{r}{i} \binom{r}{j} p_i(k-r) p_j(k-r) \right) \\
&= \sum_{r=0}^{h} \binom{h}{r} h_r(k) \left(\sum_{i=0}^{h} \sum_{j=0}^{h} \binom{r}{i} \binom{r}{j} \left(\sum_{a=0}^{t-i} \alpha_{a+i,i} \binom{k-r}{a} \right) \left(\sum_{b=0}^{t-j} \alpha_{b+j,j} \binom{k-r}{b} \right) \right) \\
&= \sum_{i=0}^{h} \sum_{j=0}^{h} \sum_{a=0}^{t-i} \sum_{b=0}^{t-j} \alpha_{a+i,i} \alpha_{b+j,j} \\
&\quad \cdot \sum_{q=0}^{a} \sum_{s=0}^{b} \underbrace{\binom{k-h}{q} \binom{k-h}{s} \left(\sum_{r=0}^{h} \binom{h}{r} h_r(k) \binom{r}{i} \binom{r}{j} \binom{h-r}{a-q} \binom{h-r}{b-s} \right)}_{B(k)}
\end{aligned}
$$

Here in the last equality we wrote $\binom{k-r}{a} = \sum_{q=0}^{a} \binom{k-h}{q}\binom{h-r}{a-q}$ by Vandermonde's identity (Lemma 12). We prove the theorem by showing that $B(k)$ considered as a polynomial in k has degree bounded above by $2t$.

$$
B(k) = \binom{k-h}{q} \binom{k-h}{s} \underbrace{\left(\sum_{r=0}^{h} \binom{h}{r} k^{\underline{r}} (n-k)^{\underline{h-r}} \overbrace{\binom{r}{i} \binom{r}{j} \binom{h-r}{a-q} \binom{h-r}{b-s}}^{f(r)} \right)}_{C(k)}
$$

By Lemma 13, the degree of $C(k)$ is at most $i + j + a - q + b - s$ and thus the degree of $B(k)$ is at most $i + j + a + b \leq 2t$

Proof of (b). Let $k \in [h-1, n-h+1]$. We have

$$
G_h(k) = \sum_{r=0}^{h} \binom{h}{r} h_r(k) \left(\sum_{j=0}^{h} \binom{r}{j} p_j(k-r) \right)^2,
$$

where $h_r(k) = k^{\underline{r}} \cdot (n-k)^{\underline{h-r}} \geq 0$ for each $r = 0, \dots, h$. Therefore $G_h(k)$ is a sum of non-negative numbers $\left(\sum_{j=0}^{h} \binom{r}{j} p_j(k-r) \right)^2$ weighted by positive coefficients $h_r(k)$.

Proof of (c). From Lemma 11 we have that

$$
\frac{1}{n^{\underline{h}}} \binom{n}{k} G_h(k) = \sum_{\substack{I \subseteq N \\ |I|=k}} \left(u^\top Z_I^t \right)^2.
$$

Therefore $G_h(k) = 0$ for $k \in \{0, \ldots, h-1\} \cup \{n-h+1, \ldots, n\}$ if we can show that $u^\top Z_Q^t = 0$ for all $Q \subseteq N$ such that $|Q| = k$.

With this aim, we start noting that for every set $S \subseteq N$ we have that the permutation group Π_S is the same as $\Pi_{N\setminus S}$. Moreover if $|S| < h$ then $\sum_{\pi \in \Pi_S} P_\pi u = 0$, otherwise we would obtain a set S smaller than H with $\sum_{\pi \in \Pi_S} P_\pi u \neq 0$ (contradicting our assumption that H is a set with the smallest cardinality having $\sum_{\pi \in \Pi_H} P_\pi u \neq 0$).

Now consider any set I such that $I \subseteq Q$ with $Q \in \{S, N\setminus S\}$ and $|S| < h$. By the previous observations it follows that $[\sum_{\pi \in \Pi_Q} P_\pi u]_I = \sum_{\pi \in \Pi_Q} P_\pi u_I = \sum_{\pi \in \Pi_Q} u_{\pi(I)} = 0$. Note that since $I \subseteq Q$ the set $\{\pi(I) : \pi \in \Pi_Q\}$ is equal to $\{J : J \subseteq Q, |J| = |I|\}$, since Π_Q is the permutation group that maps any element I from Q to any other element from Q of the same size. It follows that $\sum_{\pi \in \Pi_Q} u_{\pi(I)} = \sum_{J \subseteq Q, |J|=|I|} u_J = 0$. Using the latter we get

$$u^\top Z_Q^t = \sum_{J \subseteq Q} u_J = \sum_{i=0}^{|Q|} \sum_{J \subseteq Q, |J|=i} u_J = 0,$$

proving the claim.

Acknowledgements The authors would like to express their gratitude to Ola Svensson for helpful discussions and ideas regarding this paper.

Appendix

A Omitted calculations for Max-Cut

Grigoriev [20] and Laurent [30] proved that the following solution is feasible for any $\omega \leq n/2$ up to round $t \leq \lfloor \omega \rfloor$ for the SoS hierarchy given in Definition 1:

$$y_I = \frac{\binom{\omega}{|I|}}{\binom{n}{|I|}} \qquad \forall I \subseteq N : |I| \leq 2t. \tag{36}$$

For a graph $G = (V, E)$, the objective function of the Max- Cut problem in the usual formulation for 0/1 variables is $\sum_{(i,j)\in E} (x_i - x_j)^2$. The cut value of the SoS relaxation of Definition 1 is then $\sum_{(i,j)\in E} y_{\{i\}} + y_{\{j\}} - 2y_{\{i,j\}}$. When G is the complete graph and the solution to the SoS relaxation is taken to be (36), the cut value of the SoS relaxation is

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} y_{\{i\}} + y_{\{j\}} - 2y_{\{i,j\}} = 2\binom{n}{2} \left(\frac{\binom{\omega}{1}}{\binom{n}{1}} - \frac{\binom{\omega}{2}}{\binom{n}{2}} \right) = \omega(n - \omega).$$

It is straightforward to check that for odd n and $\omega = n/2$ this value is larger than the maximum cut in the complete graph of n vertices.

Change of basis. Using the change of basis of Lemma 1, solution $\{y_I\}$ is equivalent to solution $\{y_I^N\}$:

$$y_I^N = \sum_{H \subseteq N \setminus I} (-1)^{|H|} y_{I \cup H} = \sum_{h=0}^{n-|I|} \binom{n-|I|}{h} (-1)^h \frac{\binom{\omega}{|I|+h}}{\binom{n}{|I|+h}}$$
$$= y_I \binom{\omega - |I| - 1}{n - |I|} (-1)^{n-|I|} = (n+1) \binom{\omega}{n+1} \frac{(-1)^{n-|I|}}{\omega - |I|} \quad (37)$$

where we use the identity $\sum_{\omega=0}^{m} (-1)^{\omega} \binom{n}{\omega} = (-1)^m \binom{n-1}{m}$.

References

1. Bansal, N., Buchbinder, N., Naor, J.: Randomized competitive algorithms for generalized caching. In: STOC, pp. 235–244 (2008)
2. Bansal, N., Gupta, A., Krishnaswamy, R.: A constant factor approximation algorithm for generalized min-sum set cover. In: SODA, pp. 1539–1545 (2010)
3. Bansal, N., Pruhs, K.: The geometry of scheduling. In: FOCS, pp. 407–414 (2010)
4. Barak, B., Brandão, F.G.S.L., Harrow, A.W., Kelner, J.A., Steurer, D., Zhou, Y.: Hypercontractivity, sum-of-squares proofs, and their applications. In: STOC, pp. 307–326 (2012)
5. Barak, B., Chan, S.O., Kothari, P.: Sum of squares lower bounds from pairwise independence. In: STOC (2015). arXiv:1501.00734
6. Barak, B., Steurer, D.: Sum-of-squares proofs and the quest toward optimal algorithms. Electron. Colloq. Comput. Complex. (ECCC) **21**, 59 (2014)
7. Bhaskara, A., Charikar, M., Vijayaraghavan, A., Guruswami, V., Zhou, Y.: Polynomial integrality gaps for strong SDP relaxations of densest k-subgraph. In: SODA, pp. 388–405 (2012)
8. Bienstock, D., Zuckerberg, M.: Subset algebra lift operators for 0–1 integer programming. SIAM J. Optim. **15**(1), 63–95 (2004)
9. Blekherman, G., ao Gouveia, J., Pfeiffer, J.: Sums of squares on the hypercube. CoRR arXiv:1402.4199 (2014)
10. Blekherman, G., Parrilo, P., Thomas, R.: Semidefinite optimization and convex algebraic geometry. Society for Industrial and Applied Mathematics, Philadelphia, PA (2012). https://doi.org/10.1137/1.9781611972290
11. Carnes, T., Shmoys, D.B.: Primal-dual schema for capacitated covering problems. In: IPCO, pp. 288–302 (2008)
12. Carr, R.D., Fleischer, L., Leung, V.J., Phillips, C.A.: Strengthening integrality gaps for capacitated network design and covering problems. In: SODA, pp. 106–115 (2000)
13. Chakrabarty, D., Grant, E., Könemann, J.: On column-restricted and priority covering integer programs. In: IPCO, pp. 355–368 (2010)
14. Cheung, K.K.H.: Computation of the Lasserre ranks of some polytopes. Math. Oper. Res. **32**(1), 88–94 (2007)
15. Chlamtac, E., Tulsiani, M.: Convex relaxations and integrality gaps. In: Anjos, M., Lasserre, J. (eds.) Handbook on Semidefinite, Conic and Polynomial Optimization, vol. 166, pp. 139–169. Springer, Berlin (2011)
16. Cook, W., Dash, S.: On the matrix-cut rank of polyhedra. Math. Oper. Res. **26**(1), 19–30 (2001)
17. Fawzi, H., Saunderson, J., Parrilo, P.: Sparse sum-of-squares certificates on finite Abelian groups. CoRR. arXiv:1503.01207 (2015)
18. Godsil, C.: Association schemes (2010). Lecture Notes Available at http://quoll.uwaterloo.ca/mine/Notes/assoc2.pdf
19. Goemans, M.X., Tunçel, L.: When does the positive semidefiniteness constraint help in lifting procedures? Math. Oper. Res. **26**(4), 796–815 (2001). https://doi.org/10.1287/moor.26.4.796.10012
20. Grigoriev, D.: Complexity of positivstellensatz proofs for the knapsack. Comput. Complex. **10**(2), 139–154 (2001)
21. Grigoriev, D.: Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. Theor. Comput. Sci. **259**(1–2), 613–622 (2001)
22. Grigoriev, D., Hirsch, E.A., Pasechnik, D.V.: Complexity of semi-algebraic proofs. In: STACS, pp. 419–430 (2002)

23. Hong, S., Tunçel, L.: Unification of lower-bound analyses of the lift-and-project rank of combinatorial optimization polyhedra. Discrete Appl. Math. **156**(1), 25–41 (2008). https://doi.org/10.1016/j.dam.2007.07.021
24. Khot, S.: On the power of unique 2-prover 1-round games. In: STOC, pp. 767–775 (2002). https://doi.org/10.1145/509907.510017
25. Kurpisz, A., Leppänen, S., Mastrolilli, M.: A Lasserre lower bound for the min-sum single machine scheduling problem. In: ESA, pp. 853–864 (2015)
26. Kurpisz, A., Leppänen, S., Mastrolilli, M.: On the hardest problem formulations for the 0/1 lasserre hierarchy. In: ICALP, pp. 872–885 (2015)
27. Kurpisz, A., Leppänen, S., Mastrolilli, M.: Tight sum-of-squares lower bounds for binary polynomial optimization problems. pp. 78:1–78:14 (2016). https://doi.org/10.4230/LIPIcs.ICALP.2016.78
28. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. SIAM J. Optim. **11**(3), 796–817 (2001)
29. Laurent, M.: A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0–1 programming. Math. Oper. Res. **28**(3), 470–496 (2003)
30. Laurent, M.: Lower bound for the number of iterations in semidefinite hierarchies for the cut polytope. Math. Oper. Res. **28**(4), 871–883 (2003). https://doi.org/10.1287/moor.28.4.871.20508
31. Lee, J.R., Raghavendra, P., Steurer, D.: Lower bounds on the size of semidefinite programming relaxations. In: STOC, pp. 567–576 (2015). https://doi.org/10.1145/2746539.2746599
32. Lee, T., Prakash, A., de Wolf, R., Yuen, H.: On the sum-of-squares degree of symmetric quadratic functions. In: 31st Conference on Computational Complexity, CCC 2016, Tokyo, Japan, pp. 17:1–17:31 (2016). https://doi.org/10.4230/LIPIcs.CCC.2016.17
33. Lovász, L., Schrijver, A.: Cones of matrices and set-functions and 0–1 optimization. SIAM J. Optim. **1**(2), 166–190 (1991). https://doi.org/10.1137/0801013
34. Meka, R., Potechin, A., Wigderson, A.: Sum-of-squares lower bounds for planted clique. In: STOC, pp. 87–96 (2015). https://doi.org/10.1145/2746539.2746600
35. O'Donnell, R.: Approximability and proof complexity (2013). Talk at ELC Tokyo. Slides Available at http://www.cs.cmu.edu/odonnell/slides/approx-proof-cxty.pps
36. Parrilo, P.: Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. Ph.D. thesis, California Institute of Technology, California (2000)
37. Raghavendra, P.: Optimal algorithms and inapproximability results for every CSP? In: STOC, pp. 245–254 (2008). https://doi.org/10.1145/1374376.1374414
38. Rothvoß, T.: The lasserre hierarchy in approximation algorithms (2013). In: Lecture Notes for the MAPSP 2013—Tutorial
39. Schoenebeck, G.: Linear level Lasserre lower bounds for certain k-csps. In: FOCS, pp. 593–602 (2008)
40. Stephen, T., Tunçel, L.: On a representation of the matching polytope via semidefinite liftings. Math. Oper. Res. **24**(1), 1–7 (1999)
41. Tulsiani, M.: CSP gaps and reductions in the Lasserre hierarchy. In: STOC, pp. 303–312 (2009)
42. Wolsey, L.A.: Facets for a linear inequality in 0–1 variables. Math. Program. **8**, 168–175 (1975)