

An unbounded Sum-of-Squares hierarchy integrality gap for a polynomially solvable problem

Adam Kurpisz¹ · Samuli Leppänen¹ ·
Monaldo Mastrolilli¹

Received: 27 November 2015 / Accepted: 15 December 2016 / Published online: 4 January 2017
© Springer-Verlag Berlin Heidelberg and Mathematical Optimization Society 2017

Abstract In this paper we study the complexity of the Min-sum single machine scheduling problem under algorithms from the Sum-of-Squares/Lasserre hierarchy. We prove the first lower bound for this model by showing that the integrality gap is unbounded at level $\Omega(\sqrt{n})$ even for a variant of the problem that is solvable in $O(n \log n)$ time by the Moore–Hodgson algorithm, namely Min-number of late jobs. We consider a natural formulation that incorporates the objective function as a constraint and prove the result by partially diagonalizing the matrix associated with the relaxation and exploiting this characterization. To the best of our knowledge, our result provides the first example where the Sum-of-Squares hierarchy exhibits an unbounded integrality gap for a polynomially solvable problem after non-constant number of levels.

Keywords Sum-of-Squares hierarchy · Integrality gap · Min-number of late jobs

Mathematics Subject Classification 90C05 · 90C22

Supported by the Swiss National Science Foundation project 200020-144491/1 “Approximation Algorithms for Machine Scheduling Through Theory and Experiments”.
A preliminary version of this paper appeared in 23rd European Symposium on Algorithms—ESA 2015.

✉ Monaldo Mastrolilli
monaldo@idsia.ch

Adam Kurpisz
adam@idsia.ch

Samuli Leppänen
samuli@idsia.ch

¹ Dalle Molle Institute for Artificial Intelligence Research (IDSIA), 6928 Manno, Switzerland

1 Introduction

The MIN-SUM SINGLE MACHINE scheduling problem (often denoted $1||\sum f_j$) is defined by a set of n jobs to be scheduled on a single machine. Each job has an integral processing time, and there is a monotone function $f_j(C_j)$ specifying the cost incurred when the job j is completed at a particular time C_j ; the goal is to minimize $\sum f_j(C_j)$. A natural special case of this problem is given by the weighted version of the MIN-NUMBER OF LATE JOBS (denoted $1||\sum w_j U_j$), with $f_j(C_j) = w_j$ if $C_j > d_j$, and 0 otherwise, where $w_j \geq 0$, $d_j > 0$ are the specific cost and due date of the job j respectively. This problem is known to be NP-complete [12]. However, the unweighted version, where $w_j = 1$ for every j , can be solved in $O(n \log n)$ time by the Moore–Hodgson algorithm [21].

The first constant approximation algorithm for $1||\sum f_j$ was obtained by Bansal and Pruhs [1], who considered an even more general scheduling problem. Their 16-approximation has been recently improved to $4 + \epsilon$ by Mestre and Verschae [20] based on a primal-dual algorithm by Cheung and Shmoys [6].

A particular difficulty in approximating this problem lies in the fact that the ratio (*integrality gap*) between the optimal IP solution to the optimal solution of “natural” LPs can be arbitrarily large, since the MIN-KNAPSACK LP is a common special case. Thus, in [1, 6] the authors strengthen natural time-indexed LP relaxations by adding (exponentially many) *Knapsack-Cover* (KC) inequalities introduced by Wolsey [29] (see also [4]) that have been proved to be a useful tool to address capacitated covering problems.

One source of improvements could be the use of semidefinite relaxations such as the powerful Sum-of-Squares (SoS)/Lasserre hierarchy [16, 23, 27] (we defer the definition and related results to Sect. 2). Indeed, it is known [11] that for MIN-KNAPSACK the SoS hierarchy relaxation, when the objective function is incorporated as a constraint in the natural LP, reduces the gap to $(1 + \epsilon)$ at level $O(1/\epsilon)$, for any $\epsilon > 0$ (in fact, the same holds even for the weaker Sherali–Adams hierarchy). In light of this observation, it is therefore tempting to understand whether the SoS hierarchy can replace the use of exponentially many KC inequalities to get a better approximation for the problem $1||\sum f_j$.¹

In this paper we study the complexity of the MIN-SUM SINGLE MACHINE scheduling problem under algorithms from the SoS hierarchy. Our contribution is two-fold. We provide a novel technique that is interesting on its own for analysing integrality gaps for the SoS hierarchy. We then use this technique to prove the first SoS hierarchy lower bound for this scheduling problem by showing that the integrality gap is unbounded at level $\Omega(\sqrt{n})$ even for the unweighted MIN-NUMBER OF LATE JOBS problem, a variant of the problem that admits an $O(n \log n)$ time algorithm [21]. This result is one of the few known examples where the SoS hierarchy requires a non-constant number of levels to exactly solve a problem that admits a polynomial time algorithm, and to the best of our knowledge the only one where integrality gap remains

¹ Note that in order to claim that one can optimize over the SoS hierarchy in polynomial time, one needs to assume that the number of constraints of the starting LP is polynomial in the number of variables (see the discussion in [17]).

unbounded. Another well-known such example is the MATCHING problem, where the SoS hierarchy is known to exhibit a vanishing gap at $\Omega(n)$ levels (Corollary 2 in [10]).

Our lower bound (Theorem 4.1) is obtained by formulating the hierarchy as a sum of (exponentially many) rank-one matrices (Sect. 2) and, for every constraint, by choosing a dedicated collection (Sect. 3) of rank-one matrices whose sum can be shown to be positive definite by diagonalizing it; it is then sufficient to compare its smallest eigenvalue to the smallest eigenvalue of the remaining part of the sum of the rank-one matrices. Furthermore, we complement the result by proving a tight characterization of the considered instance (Theorem 4.2) by analysing the sign of the Rayleigh quotient.

Finally, we show a different use of the above technique to prove that the class of unconstrained k ($\leq n$) degree 0/1 n -variate polynomial optimization problems cannot be solved exactly within $k - 1$ levels of the SoS hierarchy relaxation. We do this by exhibiting for each k a 0/1 polynomial optimization problem of degree k with an integrality gap. We include this proof in the paper to demonstrate the developed technique, although the result has been recently improved. Indeed, after submitting the present paper for reviewing, the performance of the SoS hierarchy on unconstrained 0/1 polynomial optimization problems has been fully characterized in this sense. Sakaue et al. [25] showed that $\lceil \frac{n+k-1}{2} \rceil$ levels are enough to exactly solve polynomial optimization problems of degree k in n variables, generalizing the result of Fawzi et al. [8]. Moreover, using a technique different from the current work, the authors of this paper proved that the bound of Sakaue et al. is tight for even k and odd n [13].

2 The Sum-of-Squares hierarchy

In this section we provide a formal definition of the SoS hierarchy [16] together with a brief overview of the literature. We refer the reader to “Appendix” for an extended discussion of the form of the hierarchy used here.

Related work. The SoS hierarchy [16,23,27] is a systematic procedure to strengthen a relaxation for an optimization problem by constructing a sequence of increasingly tight formulations, obtained by adding additional variables and positive semidefiniteness (PSD) constraints. The hierarchy is parameterized by its level t , such that the formulation gets tighter as t increases, and the size of the resulting semidefinite program is $mn^{O(t)}$ where n is the number of variables and m the number of constraints in the original problem. This approach captures the convex relaxations used in the best available approximation algorithms for a wide variety of optimization problems. Due to space restrictions, we refer the reader to [7,17,22,24] and the references therein.

The limitations of the SoS hierarchy have also been studied, but not many techniques for proving lower bounds are known. Most of the known lower bounds for the hierarchy originated in the works of Grigoriev [9,10] (also independently rediscovered later by Schoenebeck [26]). In [10] it is shown that random 3XOR or 3SAT instances with n variables cannot be solved by even $\Omega(n)$ rounds of SoS hierarchy. Lower bounds, such as those of [3,28] rely on [10,26] plus gadget reductions. For different techniques to obtain lower bounds see [2,5,14,15,18,19].

Notation and the formal definition. In the context of this paper, it is convenient to define the hierarchy in an equivalent form that follows easily from “standard” definitions (see e.g. [17]) after a change of variables.

For the applications that we have in mind, we restrict our discussion to optimization problems with n 0/1-variables and m linear constraints. We denote $K = \{x \in \mathbb{R}^n \mid g_\ell(x) \geq 0, \forall \ell \in [m]\}$ to be the feasible set of the linear relaxation. We are interested in approximating the convex hull of the integral points in K . We refer to the ℓ -th linear constraint evaluated at the set $I \subseteq N$ ($x_i = 1$ for $i \in I$, and $x_i = 0$ for $i \notin I$) as $g_\ell(x_I)$. Let $N = \{1, \dots, n\}$. For each integral solution x_I , where $I \subseteq N$, 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 .

For a set $I \subseteq N$ and fixed integer t , let $\mathcal{P}_t(I)$ denote the set of the subsets of I of size at most t . Define d -zeta vectors: $Z_I \in \mathbb{R}^{\mathcal{P}_d(N)}$ for every $I \subseteq N$, such that for each $|J| \leq d$, $[Z_I]_J = \begin{cases} 1, & \text{if } J \subseteq I \\ 0, & \text{otherwise} \end{cases}$. In order to keep the notation simple, we do not emphasize the parameter d as the dimension of the vectors should be clear from the context (we can think of the parameter d as either t or $t + 1$).

Definition 2.1 The SoS hierarchy relaxation at level t for the set K , denoted by $\text{SoS}_t(K)$, is given by the set of values $y_I^N \in \mathbb{R}$ for $I \subseteq N$ that satisfy

$$\sum_{I \subseteq N} y_I^N = 1, \quad (1)$$

$$\sum_{I \subseteq N} y_I^N Z_I Z_I^\top \geq 0, \text{ where } Z_I \in \mathbb{R}^{\mathcal{P}_{t+1}(N)} \quad (2)$$

$$\sum_{I \subseteq N} g_\ell(x_I) y_I^N Z_I Z_I^\top \geq 0, \forall \ell \in [m], \text{ where } Z_I \in \mathbb{R}^{\mathcal{P}_t(N)} \quad (3)$$

Notice that the formulation of the SoS hierarchy given in Definition 2.1 has exponentially many variables y_I^N , due to the change of variables. This is not a problem for our purposes, since we are interested in showing an integrality gap rather than solving an optimization problem. Furthermore, it is straightforward to see that SoS hierarchy 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 (1), Condition (2) because the rank one matrix $Z_I Z_I^\top$ is PSD, and Condition (3) since $x_I \in K$.

3 Partial diagonalization

In this section we describe how to partially diagonalize the matrices associated to SoS hierarchy. This will be used in the proofs of Theorems 4.1 and 4.2.

Below we denote by w_I^N either y_I^N or $y_I^N g_\ell(x_I)$. The following simple observation describes a congruent transformation (\cong) to obtain a partial diagonalization of the matrices used in Definition 2.1. We will use this partial diagonalization in our bound derivation.

Lemma 3.1 *Let $\mathcal{C} \subseteq \mathcal{P}_n(N)$ be a collection of size $|\mathcal{P}_d(N)|$ (where d is either t or $t + 1$). If \mathcal{C} is such that the matrix Z with columns Z_I for every $I \in \mathcal{C}$ is invertible, then*

$$\sum_{I \subseteq N} w_I^N Z_I Z_I^\top \cong D + \sum_{I \in \mathcal{P}_n(N) \setminus \mathcal{C}} w_I^N Z^{-1} Z_I (Z^{-1} Z_I)^\top$$

where D is a diagonal matrix with entries w_I^N , for $I \in \mathcal{C}$.

Proof It is sufficient to note that $\sum_{I \in \mathcal{C}} w_I^N Z_I Z_I^\top = Z D Z^\top$. □

Since congruent transformations are known to preserve the sign of the eigenvalues, the above lemma in principle gives us a technique to check whether or not (2) and (3) are satisfied: show that the sum of the smallest diagonal element of D and the smallest eigenvalue of the matrix $\sum_{I \in \mathcal{P}_n(N) \setminus \mathcal{C}} w_I^N Z^{-1} Z_I (Z^{-1} Z_I)^\top$ is non-negative. In what follows we introduce a method to select the collection \mathcal{C} such that the matrix Z is invertible.

Let Z_d denote the matrix with columns $[Z_d]_I = Z_I$ indexed by sets $I \subseteq N$ of size at most d . The matrix Z_d is invertible as it is upper triangular with ones on the diagonal. It is straightforward to check that the inverse Z_d^{-1} is given by $[Z_d^{-1}]_{I,J} = (-1)^{|J \setminus I|}$ if $I \subseteq J$ and 0 otherwise (see e.g. [17]).

In Lemma 3.1 we require a collection \mathcal{C} such that the matrix, whose columns are the zeta vectors corresponding to elements in \mathcal{C} , is invertible. The above indicates that if we take \mathcal{C} to be the set of subsets of N with size less or equal to d , then this requirement is satisfied. We can think that the matrix Z_d contains as columns the zeta vectors corresponding to the set \emptyset and all the symmetric differences of the set \emptyset with sets of size at most d . The observation allows us to generalize this notion: fix a set $S \subseteq N$, and define \mathcal{C} to contain all the sets $S \oplus I$ for $|I| \leq d$ (here \oplus denotes the symmetric difference). More formally, consider the following $|\mathcal{P}_d(N)| \times |\mathcal{P}_d(N)|$ matrix $Z_{d,S}$, whose generic entry $I, J \subseteq \mathcal{P}_d(N)$ is

$$[Z_{d,S}]_{I,J} = \begin{cases} 1 & \text{if } I \subseteq J \oplus S, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Note that $Z_{d,\emptyset} = Z_d$. In order to apply Lemma 3.1, we show that $Z_{d,S}$ is invertible.

Lemma 3.2 *Let $A_{d,S}$ be a $|\mathcal{P}_d(N)| \times |\mathcal{P}_d(N)|$ matrix defined as*

$$[A_{d,S}]_{I,K} = \begin{cases} (-1)^{|K \cap S|} & \text{if } (I \setminus S) \subseteq K \subseteq I \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Then $Z_{d,S}^{-1} = Z_d^{-1} A_{d,S}$.

Proof The claim follows by proving that $Z_d = A_{d,S} Z_{d,S}$. The generic entry (I, J) of $A_{d,S} Z_{d,S}$ is

$$[A_{d,S}Z_{d,S}]_{I,J} = \sum_{K \in \mathcal{P}_d(N)} [A_{d,S}]_{I,K} [Z_{d,S}]_{K,J} = \sum_{\substack{K \in \mathcal{P}_d(N) \\ (I \setminus S) \subseteq K \subseteq I \\ K \subseteq J \oplus S}} (-1)^{|K \cap S|}$$

We first note that unless $I \subseteq J \cup S$, the sum is over an empty set, and thus zero. Indeed, assume there exists an element $a \in I, a \notin J \cup S$. Then, since $I \setminus S \subseteq K$, we require that $a \in K$. On the other hand, $K \subseteq S \oplus J$ implies that $a \in S \oplus J$, which contradicts the assumption on a , and hence no such K exists.

Since $K \subseteq (J \oplus S) \cap I$, and $I \setminus S \subseteq J$, we can partition K in the form $K = (I \setminus S) \cup H$, where H is any subset of $I \cap (S \setminus J)$. Indeed, it is easy to see that such a K satisfies the conditions of the sum, and that no other choice is possible. Then, the sum becomes of the form

$$[A_{d,S}Z_{d,S}]_{I,J} = \sum_{i=0}^m (-1)^i \binom{m}{i}$$

where m is the size of the set $S \cap (I \setminus J)$. Therefore, the sum equals 1 if $m = 0$ and 0 otherwise. It follows that $[A_{d,S}Z_{d,S}]_{I,J} = 1$ if and only if $I \subseteq J$, and 0 otherwise. \square

We also give a closed form of the elements of the matrix $Z_{d,S}^{-1}$.

Lemma 3.3 *For each $I, J \subseteq \mathcal{P}_d(N)$ the generic entry (I, J) of $Z_{d,S}^{-1}$ is*

$$[Z_{d,S}^{-1}]_{I,J} = (-1)^{|J \cap S| + |J \setminus I|} \begin{cases} (-1)^{d - |I \cup J|} \binom{|S \setminus (I \cup J)| - 1}{d - |I \cap J|}, & \text{if } I \setminus S \subseteq J \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

In particular $\left| [Z_{d,S}^{-1}]_{I,J} \right| \leq |S|^{O(d)}$.

Proof From Lemma 3.2 we know that $Z_{d,S}^{-1} = Z_d^{-1} A_{d,S}$, thus

$$[Z_{d,S}^{-1}]_{I,J} = \sum_{K \in \mathcal{P}_d(N)} [Z_d^{-1}]_{I,K} [A_{d,S}]_{K,J} = \sum_{\substack{K \in \mathcal{P}_d(N) \\ I \subseteq K \\ K \setminus S \subseteq J \subseteq K}} (-1)^{|K \setminus I| + |J \cap S|}$$

First, note that $K \setminus S \subseteq J$ implies that $K \subseteq J \cup S$. This with $I \subseteq K$ implies in particular that the sum has no terms unless $I \subseteq J \cup S$. Next, we see that $I \cup J \subseteq K$, so we can write $K = I \cup J \cup H$ for some set H disjoint from $I \cup J$. Using the first observation we get that $H \subseteq S \setminus (I \cup J)$. Since $K \in \mathcal{P}_d(N)$, we thus require that $H \in \mathcal{P}_{d - |I \cup J|}(S \setminus (I \cup J))$. The above sum then becomes

$$[Z_{d,S}^{-1}]_{I,J} = (-1)^{|J \cap S| + |J \setminus I|} \sum_{\substack{H \in \mathcal{P}_{d - |I \cup J|}(S \setminus (I \cup J)) \\ I \subseteq J \cup S}} (-1)^{|H|}$$

This simplifies to

$$\left[Z_{d,S}^{-1} \right]_{I,J} = (-1)^{|J \cap S| + |J \setminus I|} \begin{cases} (-1)^{d - |I \cup J|} \binom{|S \setminus (I \cup J)| - 1}{d - |I \cap J|}, & \text{if } I \subseteq J \cup S \\ 0, & \text{otherwise} \end{cases}$$

□

4 A lower bound for Min-number of late jobs

We consider the single machine scheduling problem to minimize the number of late jobs: we are given a set of n jobs, each with a processing time $p_j > 0$, and a due date $d_j > 0$. We have to sequence the jobs on a single machine such that no two jobs overlap. For each job j that is not completed by its due date, we pay the cost w_j .

4.1 The starting linear program

Our result is based on the following “natural” linear programming (LP) relaxation that is a special case of the LPs used in [1,6] (therefore our gap result also holds if we apply those LP formulations). For each job we introduce a variable $x_j \in [0, 1]$ with the intended (integral) meaning that $x_j = 1$ if and only if the job j completes after its deadline. Then, for any time $s \in \{d_1, \dots, d_n\}$, the sum of the processing times of the jobs with deadlines less than s , and that complete before s , must satisfy $\sum_{j:d_j \leq s} (1 - x_j) p_j \leq s$. The latter constraint can be rewritten as a capacitated covering constraint, $\sum_{j:d_j \leq t} x_j p_j \geq D_t$, where $D_s := \sum_{j:d_j \leq s} p_j - s$ represents the demand at time s . The goal is to minimize $\sum_j x_j$.

The feasible sets of our LP relaxation and the relaxation used in [1,6] both consist of convex combinations of fractional schedules that have no idle periods and that process at most one job at a time. However, the LP of [1,6] uses time-indexed variables z_{jt} with the intended (integral) meaning that $z_{jt} = 1$ if and only if the job j completes at time t . This allows the formulation to model more general problems than just the MIN-NUMBER OF LATE JOBS. In the case of MIN-NUMBER OF LATE JOBS one can show that a feasible solution to the formulation used in this paper can be mapped to a feasible solution to the time-indexed formulation by assigning $x_j = \sum_{t > d_j} z_{j,t}$ and resolving the ambiguity in the z_{ij} variables by ordering according to the deadlines.

4.2 The integrality gap instance

Consider the following instance with $n = m^2$ jobs of unit costs. The jobs are partitioned into m blocks N_1, N_2, \dots, N_m , each with m jobs. For $i \in [m]$, the jobs belonging to block N_i have the same processing time P^i , for $P > 1$, and the same deadline $d_i = m \sum_{j=1}^i P^j - \sum_{j=1}^i P^{j-1}$. Then the demand at time d_i is $D_i = \sum_{j=1}^i P^{j-1}$. For any $t \geq 0$, let T be the smallest value that makes $\text{SOS}_t(LP(T))$ feasible, where $LP(T)$ is defined as follows for the variables $x_{ij} \in [0, 1]$ for the j -th job from the i -th block, where $i, j \in [m]$:

$$LP(T) \sum_{i=1}^m \sum_{j=1}^m x_{ij} \leq T, \quad (7a)$$

$$\sum_{i=1}^{\ell} \sum_{j=1}^m x_{ij} \cdot P^i \geq D_{\ell}, \quad \text{for } \ell \in [m] \quad (7b)$$

Note that, for any feasible *integral* solution for $LP(T)$, the smallest T (i.e. the optimal integral value) can be obtained by selecting one job for each block, so the smallest T for integral solutions is $m = \sqrt{n}$. The *integrality gap* of $\text{SOS}_t(LP(T))$ (or $LP(T)$) is defined as the ratio between \sqrt{n} (i.e. the optimal integral value) and the smallest T that makes $\text{SOS}_t(LP(T))$ (or $LP(T)$) feasible. It is easy to check that $LP(T)$ has an integrality gap P for any $P \geq 1$: for $T = \sqrt{n}/P$, a feasible fractional solution for $LP(T)$ exists by setting $x_{ij} = \frac{1}{\sqrt{n}P}$.

4.3 Proof of integrality gap for $\text{SOS}_t(LP(T))$

Theorem 4.1 *For any $k \geq 1$ and n such that $t = \frac{\sqrt{n}}{2k} - \frac{1}{2} \in \mathbb{N}$, the following solution is feasible for $\text{SOS}_t(LP(\sqrt{n}/k))$*

$$y_I^N = \begin{cases} \alpha, & \forall I \in \mathcal{P}_{2t+1}(N) \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where $\alpha > 0$ is such that $\sum_{I \subseteq N} y_I^N = 1$ and the parameter P is large enough.

Proof We need to show that the solution (8) satisfies the feasibility conditions (1)–(2) for the variables and the condition (3) for every constraint. The condition (1) is satisfied by definition of the solution, and (2) becomes a sum of positive semidefinite matrices $Z_I Z_I^\top$ with non-negative weights y_I^N , so it is satisfied as well.

It remains to show that (3) is satisfied for both (7a) and (7b). Consider the Eq. (7a) first, and let $g(x_I) = T - \sum_{i,j} x_{ij}$ be the value of the constraint when the decision variables are $x_{ij} = 1$ whenever $(i, j) \in I$, and 0 otherwise.² Now for every $I \subseteq N$, it holds $g(x_I) y_I^N \geq 0$, as we have $y_I^N = 0$ for every I containing more than $2t + 1 = \frac{\sqrt{n}}{k} = T$ elements. Hence the sum in (3) is again a sum of positive semidefinite matrices with non-negative weights, and the condition is satisfied.

Finally, consider the ℓ -th constraint of the form (7b), and let $g_{\ell}(x_I) = \sum_{i=1}^{\ell} \sum_{j=1}^m x_{ij} \cdot P^i - D_{\ell}$. In order to prove that (3) is satisfied, we apply Lemma 3.1 with the following collection of subsets of N : $\mathcal{C} = \{I \oplus S \mid I \subseteq [m], |I| \leq t\}$, where we take $S = \{(\ell, j) \mid j \in [t + 1]\}$. Now, S is a set of $t + 1$ jobs from the ℓ th block. Since any $J \in \mathcal{C}$ is of the form $J = I \oplus S$ where $|I| \leq t$, we have that J always contains at least one job from the block ℓ . Therefore, any allocation x_J with $J \in \mathcal{C}$ satisfies the constraint g_{ℓ} .

² Strictly speaking $I \subseteq N$ is a set of numbers, so we associate to each pair i, j a number via the one-to-one mapping $(i - 1)m + j$. Hence, to keep the notation simple, we here understand $(i, j) \in I$ to mean $(i - 1)m + j \in I$.

By Lemma 3.2, the matrix $Z_{t,S}$ is invertible and by Lemma 3.1 we have for (3) that $\sum_{I \subseteq N} g_\ell(x_I) y_I^N Z_I Z_I^\top \cong D + \sum_{I \in \mathcal{P}_n(N) \setminus \mathcal{C}} g_\ell(x_I) y_I^N Z_{t,S}^{-1} Z_I (Z_{t,S}^{-1} Z_I)^\top$, where D is a diagonal matrix with elements $g_\ell(x_I) y_I^N$ for each $I \in \mathcal{C}$. We prove that the latter is positive semidefinite by analysing its smallest eigenvalue λ_{\min} . Writing $R_I = Z_{t,S}^{-1} Z_I (Z_{t,S}^{-1} Z_I)^\top$, we have by Weyl's inequality

$$\lambda_{\min} \left(D + \sum_{I \in \mathcal{P}_n(N) \setminus \mathcal{C}} g_\ell(x_I) y_I^N R_I \right) \geq \lambda_{\min}(D) + \lambda_{\min} \left(\sum_{I \in \mathcal{P}_n(N) \setminus \mathcal{C}} g_\ell(x_I) y_I^N R_I \right) \tag{9}$$

Since D is a diagonal matrix with entries $g_\ell(x_I) y_I^N$ for $I \in \mathcal{C}$, and for every $I \in \mathcal{C}$ the constraint $g_\ell(x_I)$ is satisfied, we have $\lambda_{\min}(D) \geq \alpha \left(P^\ell - D_\ell \right) = \alpha \left(P^\ell - \frac{P^\ell - 1}{P - 1} \right)$.

On the other hand for every $I \subseteq N$, $g_\ell(x_I) \geq -\sum_{j=1}^\ell P^{j-1} = -\frac{P^\ell - 1}{P - 1}$. The nonzero eigenvalue of the rank one matrix R_I is $\left(Z_{t,S}^{-1} Z_I \right)^\top Z_{t,S}^{-1} Z_I \leq |\mathcal{P}_t(N)|^3 t^{O(t)} = n^{O(\sqrt{n})}$. This is because by Lemma 3.3, for every $I, J \in \mathcal{P}_t(N)$, $\left| [Z_{t,S}^{-1}]_{I,J} \right| \leq t^{O(t)}$, for $|S| = t + 1$. Furthermore, $[Z_I]_J \in \{0, 1\}$ and thus $Z_{t,S}^{-1} Z_I$ is a vector of dimension $|\mathcal{P}_t(N)|$ with entries bounded by $|\mathcal{P}_t(N)| t^{O(t)}$. This in turn gives that $\left(Z_{t,S}^{-1} Z_I \right)^\top Z_{t,S}^{-1} Z_I \leq |\mathcal{P}_t(N)|^3 t^{O(t)}$ which, by the fact that $t \leq \sqrt{n}$, $|\mathcal{P}_t(N)| \leq n^{O(\sqrt{n})}$, is equal to $n^{O(\sqrt{n})}$. Finally, since $1/2^n \leq \alpha \leq 1$, for $P \geq 2$ we get

$$\begin{aligned} \lambda_{\min} \left(D + \sum_{I \in \mathcal{P}_n(N) \setminus \mathcal{C}} g_\ell(x_I) y_I^N R_I \right) &\geq \alpha \left(P^k - \frac{P^k - 1}{P - 1} \right) - \alpha \frac{P^k - 1}{P - 1} 2^n n^{O(\sqrt{n})} \\ &\geq \alpha P^k - \alpha \frac{P^k - 1}{P - 1} \left(1 + 2^n n^{O(\sqrt{n})} \right) \\ &\geq \frac{P^k}{2^n} - \frac{P^k}{P/2} 2^{n+1} n^{O(\sqrt{n})} \geq P^{k-1} \left(\frac{P}{2^n} - 2^{n+2} n^{O(\sqrt{n})} \right) \end{aligned}$$

which is non-negative for $P = n^{C(\sqrt{n})}$, for C being a sufficiently large constant. \square

Note that the level t cannot be easily increased in the above proof to meet for example the level of Theorem 4.2. This is because we require that the solutions y_I^N in (8) and (10) assign zero for sets of size greater than \sqrt{n}/k . Requiring this ensures that satisfying (3) for (7a) can be done easily (second paragraph of the above proof). Moreover, if the level t is increased in Theorem 4.1 without changing the solution (8), our construction would yield subsets of size greater than \sqrt{n}/k in \mathcal{C} . This in turn would produce zero diagonal entries in the diagonal matrix (9) implying $\lambda_{\min}(D) = 0$.

Theorem 4.1 states that the SoS hierarchy has an arbitrarily large integrality gap k even at level $t = \frac{\sqrt{n}}{2k} - \frac{1}{2}$. In the following we provide a tight analysis characterization for this instance, namely we prove that the SoS hierarchy admits an

arbitrarily large gap k even at level $t = \frac{\sqrt{n}}{k} - 1$. Note that at the next level, namely $t + 1 = \sqrt{n}/k$, $\text{SOS}_{t+1}(LP(\sqrt{n}/k))$ has no feasible solution for $k > 1$, since the constraint (7b) implies that any feasible solution for $\text{SOS}_{t+1}(LP(\sqrt{n}/k))$ has $y_I^N = 0$ for all $|I| > \sqrt{n}/k$. This in turn implies, with Lemma 3.1 for $\mathcal{C} = \mathcal{P}_t(N)$, that $\sum_{I \subseteq N} g_\ell(x_I) y_I^N Z_I Z_I^\top \cong D_\ell$, where D_ℓ is a diagonal matrix with entries $g_\ell(x_I) y_I^N$, for every $|I| \leq t$, there exists ℓ such that $g_\ell(x_I) < 0$ which, in any feasible solution implies $y_I^N = 0$, contradicting (1). Therefore, our analysis gives a tight characterization of the integrality gap threshold phenomenon. The claimed tight bound is obtained by utilizing a more involved analysis of the sign of the Rayleigh quotient for the almost diagonal matrix characterization of the SoS hierarchy.

Theorem 4.2 *For any $k \geq 1$ and n such that $t = \frac{\sqrt{n}}{k} - 1 \in \mathbb{N}$, the following solution is feasible for $\text{SOS}_t(LP(\sqrt{n}/k))$*

$$y_I^N = \begin{cases} \alpha, & \forall I \in \mathcal{P}_{t+1}(N) \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

where $\alpha > 0$ is such that $\sum_{I \subseteq N} y_I^N = 1$ and the parameter P is large enough.

Proof The solution satisfies the conditions (1), (2) and (3) for (7a) by the same argument as in the proof of Theorem 4.1.

We prove that the solution satisfies the condition (3) for any constraint ℓ of the form (7b). Since $M \geq 0$ if and only if $v^\top M v \geq 0$, for every unit vector v , by Lemma 3.1 (for the collection $\mathcal{C} = \mathcal{P}_t(N)$) and using the solution (10) we can transform (3) to the following semi-infinite system of linear inequalities (recall that $g_\ell(x_I) = \sum_{i=1}^I \sum_{j=1}^m x_{ij} \cdot P^i - D_\ell$):

$$\sum_{I \in \mathcal{P}_t(N)} g_\ell(x_I) v_I^2 + \sum_{J \subseteq N: |J|=t+1} \left(\sum_{\substack{I \in \mathcal{P}_t(N) \\ I \subset J}} v_I (-1)^{|I|} \right)^2 g_\ell(x_J) \geq 0, \quad \forall v \in \mathbb{S}^{|\mathcal{P}_t(N)|-1} \tag{11}$$

where $\mathbb{S}^{|\mathcal{P}_t(N)|-1}$ is the set of unit vectors in $\mathbb{R}^{|\mathcal{P}_t(N)|}$. Now consider the ℓ -th covering constraint $g_\ell(x) \geq 0$ of the form (7b) and the corresponding semi-infinite set of linear inequalities (11). Then consider the following partition of $\mathcal{P}_{t+1}(N)$: $A = \{I \in \mathcal{P}_{t+1}(N) : I \cap N_\ell \neq \emptyset\}$ and $B = \{I \in \mathcal{P}_{t+1}(N) : I \cap N_\ell = \emptyset\}$.

Note that A corresponds to the assignments that are guaranteed to satisfy the constraint ℓ . More precisely, for $S \in A$ we have $g_\ell(x_S) \geq \left(P^\ell - \sum_{j=1}^\ell P^{j-1} \right) = P^\ell \left(1 - \frac{P^\ell - 1}{P^\ell(P-1)} \right) \geq P^\ell \left(1 - \frac{1}{P-1} \right)$, and for $S \in B$ we have $g_\ell(x_S) \geq -\sum_{j=1}^\ell P^{j-1} \geq P^\ell \left(-\frac{1}{P-1} \right)$. Since $P > 0$, by scaling $g_\ell(x) \geq 0$ [see (7b)] by P^ℓ , we will assume, w.l.o.g., that

$$g_\ell(x_S) \geq \begin{cases} 1 - \frac{1}{P-1}, & \text{if } S \in A \\ -\frac{1}{P-1}, & \text{if } S \in B \end{cases}$$

Note that, since v is a unit vector, we have $v_I^2 \leq 1$, and for any $J \subseteq N$ such that $|J| = t + 1$, the coefficient of $g_\ell(x_J)$ is bounded by $\left(\sum_{I \subseteq J} v_I (-1)^{|I|}\right)^2 \leq 2^{O(t)}$. For all unit vectors v , let β denote the smallest possible total sum of the negative terms in (11) (these are those related to $g_\ell(x_I)$ for $I \in B$). Note that, $\beta \geq -\frac{|B|2^{O(t)}}{P} = -\frac{n^{O(t)}}{P}$, since $|B| \leq \mathcal{P}_t(N) \leq n^{O(t)}$.

In the following, we show that, for sufficiently large P , the claimed solution satisfies (11). We prove this by contradiction.

Assume that there exists a unit vector v such that (11) is not satisfied by the solution. We start by observing that under the previous assumption the following holds $v_I^2 = \frac{n^{O(t)}}{P}$ for all $I \in A \cap \mathcal{P}_t(N)$. If not, we would have an $I \in A \cap \mathcal{P}_t(N)$ such that $v_I^2 g_\ell(x_I) \geq -\beta$ contradicting the assumption that (11) is not satisfied. We claim that under the contradiction assumption, the previous bound on v_I^2 can be generalized to $v_I^2 = \frac{n^{O(t^2)}}{P}$ for any $I \in \mathcal{P}_t(N)$. Then, by choosing P such that $v_I^2 < 1/n^{2t}$, for $I \in \mathcal{P}_t(N)$, we have $\sum_{I \in \mathcal{P}_t(N)} v_I^2 < 1$, which contradicts the assumption that v is a unit vector.

The claim follows by showing that $\forall I \in B \cap \mathcal{P}_t(N)$ it holds $v_I^2 \leq n^{O(t^2)}/P$. The proof is by induction on the size of I for any $I \in B \cap \mathcal{P}_t(N)$.

Consider the empty set, since $\emptyset \in B \cap \mathcal{P}_t(N)$. We show that $v_\emptyset^2 = n^{O(t)}/P$. With this aim, consider any $J \subseteq N_\ell$ with $|J| = t + 1$. Note that $J \in A$, so $g_\ell(x_J) \geq 1 - 1/(P - 1)$ and its coefficient $u_J^2 = \left(\sum_{I \subseteq J} v_I (-1)^{|I|}\right)^2$ is the square of the sum of v_\emptyset and other terms v_I , all with $I \in A \cap \mathcal{P}_t(N)$. Ignoring all the other positive terms apart from the one corresponding to J in (11), evaluating the sum of all the negative terms as β and using a loose bound $g_\ell(x_J) \geq 1/2$ for large P , we obtain the following bound b_0

$$|v_\emptyset| \leq \sqrt{-2\beta} + \sum_{\emptyset \neq I \subseteq J} |v_I| \leq b_0 = O\left(\sqrt{-\beta} + 2^{O(t)} \frac{n^{O(t)}}{\sqrt{P}}\right) = \frac{n^{O(t)}}{\sqrt{P}} \tag{12}$$

which implies that $v_\emptyset^2 = n^{O(t)}/P$.

Similarly as before, consider any singleton set $\{i\}$ with $\{i\} \in B \cap \mathcal{P}_t(N)$ and any $J \subseteq N_\ell$ with $|J| = t$. Note that $J \in A$, $g_\ell(x_J) \geq 1 - 1/(P - 1)$ and its coefficient $u_J^2 = \left(\sum_{I \subseteq J \cup \{i\}} v_I (-1)^{|I|}\right)^2$ is the square of the sum of $v_{\{i\}}$, v_\emptyset and other terms v_I , with $I \subseteq J$ and therefore $v_I^2 = \frac{n^{O(t)}}{P}$. Moreover, again note that u_J^2 is smaller than $-\beta$ [otherwise (11) is satisfied]. Therefore, for any singleton set $\{i\} \in B \cap \mathcal{P}_t(N)$, we have that $|v_{\{i\}}| \leq |v_\emptyset| + \sqrt{-2\beta} + \sum_{\emptyset \neq I \subseteq J} |v_I| \leq 2b_0$.

Generalizing by induction, consider any set $S \in B \cap \mathcal{P}_t(N)$ and any $J \subseteq N_\ell$ with $|J| = t + 1 - |S|$. We claim that $|v_{|S|}| \leq b_{|S|}$ where

$$b_{|S|} = b_0 + \sum_{i=0}^{|S|-1} n^i b_i \tag{13}$$

This follows by induction hypothesis and because again $g_\ell(J \cup S)u_{J \cup S} \leq -\beta$ and therefore, $|v_S| \leq \sum_{i=0}^{|S|-1} \left(\sum_{\substack{I \in B \\ |I|=i}} |v_I| \right) + \sqrt{-2\beta} + \sum_{\emptyset \neq I \subset J} |v_I|$.

From (13), for any $S \in B \cap \mathcal{P}_t(N)$, we have that $|v_S|$ is bounded by $b_t = (n^{t-1} + 1)b_{t-1} = n^{O(t^2)}b_0 = \frac{n^{O(t^2)}}{\sqrt{P}}$. □

5 Application in 0/1 polynomial optimization

In this section we use the developed technique to prove an integrality gap result for the unconstrained 0/1 n -variate polynomial optimization problem. We start with the following definition of SoS hierarchy.

Definition 5.1 The SoS hierarchy at level t for the unconstrained 0/1 optimization problem with the objective function $f(x) : \{0, 1\}^n \rightarrow \mathbb{R}$, denoted by $\text{SoS}_t(f(x))$, is given by the feasible points y_I^N for each $I \subseteq N$ of the following semidefinite program

$$\sum_{I \subseteq N} y_I^N = 1, \tag{14}$$

$$\sum_{I \subseteq N} y_I^N Z_I Z_I^\top \geq 0, \quad \text{where } Z_I \in \mathbb{R}^{\mathcal{P}_t(N)} \tag{15}$$

Furthermore, the objective function of $\text{SoS}_t(f(x))$ is $\sum_{I \subseteq N} f(x_I)y_I^N$.

The main result of this section is the following theorem.

Theorem 5.1 *The class of unconstrained k -degree 0/1 n -variate polynomial optimization problems cannot be solved exactly with a $k - 1$ level of SoS hierarchy.*

Proof For every $k \leq n$ we give an unconstrained n -variate polynomial optimization problem with an objective function $f(x)$ of degree k such that $\text{SoS}_{k-1}(f(x))$ has an integrality gap. Consider a maximization problem with the following objective function over $\{0, 1\}^n$: $f(x) = \sum_{\emptyset \neq I \in \mathcal{P}_k(N)} \binom{n-|I|}{k-|I|} (-1)^{|I|+1} \prod_{i \in I} x_i$. We prove that the following solution is super-optimal and feasible for $\text{SoS}_{k-1}(f(x))$

$$y_I^N = \begin{cases} \alpha, & \forall I \in N, \quad |I| \geq n - k + 1 \\ -\epsilon, & I = \emptyset \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

where $\alpha > 0$ is such that $\sum_{\emptyset \neq I \subseteq N} y_I^N = 1 + \epsilon$ and ϵ is small enough.

Using a counting argument it is easy to check that the objective function is equal to

$$f(x) = \sum_{\substack{K \subseteq N \\ |K|=k}} \sum_{\substack{J \subseteq K \\ J \neq \emptyset}} (-1)^{|J|+1} \prod_{j \in J} x_j$$

Now, consider any 0/1 solution. For every $K \subseteq N$ of size $|K| = k$, a partial summation $\sum_{\emptyset \neq J \subseteq K} (-1)^{|J|+1} \prod_{j \in J} x_j$ takes value one, if for at least one $j \in K$, $x_j = 1$, and zero otherwise. Therefore, whenever $n - k + 1$ coordinates of x are set to 1, any set K of size k has at least one $j \in K$ such that $x_j = 1$. Thus the integral optimum is $\binom{n}{k}$ for any solution $x \in \{0, 1\}^n$ such that at least $n - k + 1$ coordinates are set to 1.

On the other hand the objective value for the SoS solution (16) is given by the formula

$$\sum_{I \subseteq N} f(x_I) y_I^N = \sum_{\substack{I \subseteq N \\ |I| \geq n-k+1}} f(x_I) y_I^N = \binom{n}{k} \sum_{\substack{I \subseteq N \\ |I| \geq n-k+1}} y_I^N = (1 + \epsilon) \binom{n}{k}$$

where the first equality comes from the fact that $f(x_\emptyset) = 0$ and the second from the fact that $f(x_I)$ evaluates to $\binom{n}{k}$ for any $I \subseteq N$, $|I| \geq n - k + 1$.

Finally, we prove that the solution (16) is feasible for $\text{SOs}_{k-1}(f(x))$. The constraint (14) is satisfied by definition. In order to prove that the constraint (15) is satisfied, we apply Lemma 3.1 with the collection $\mathcal{C} = \{I \oplus S \mid I \in \mathcal{P}_{k-1}(N)\}$ of subsets of N , for $S = N$, and get that

$$D + \sum_{I \in \mathcal{P}_n(N) \setminus \mathcal{C}} y_I^N Z_{t,S}^{-1} Z_I (Z_{t,S}^{-1} Z_I)^\top = D - \epsilon Z_{t,S}^{-1} Z_\emptyset (Z_{t,S}^{-1} Z_\emptyset)^\top \tag{17}$$

where D is a diagonal matrix with diagonal entries equal to $\alpha \geq 1/2^n$. Since the nonzero eigenvalue of the rank one matrix $Z_{t,S}^{-1} Z_\emptyset (Z_{t,S}^{-1} Z_\emptyset)^\top$ is equal to $(Z_{t,S}^{-1} Z_\emptyset)^\top Z_{t,S}^{-1} Z_\emptyset \leq |\mathcal{P}_t(N)| t^{2t} = n^{O(t)}$, one can choose $\epsilon = 1/n^{Ct}$, for sufficiently large C , such that by the Weyl's inequality we have that the matrix in (17) is PSD. □

6 Discussion

In this paper we introduced a technique for analysing the positive semidefiniteness of the matrices in the Sum-of-Squares hierarchy. The technique works by partially diagonalizing such a matrix in order to allow a diagonal dominance argument. The key reason the partial diagonalization can be done is that the moment matrices of the hierarchy admit a well-understood decomposition into a sum of rank-one matrices $Z_I Z_I^\top$ defined by zeta vectors Z_I for every $I \subseteq N$. The partial diagonalization is possible whenever we choose a collection of rank-one matrices such that the corresponding zeta vectors form a basis, and the resulting diagonal matrix consists of the coefficients of the chosen rank-one matrices in the decomposition. We introduced a natural way of choosing such a basis of zeta vectors, but it is not clear if our choice is the only interesting way of doing this.

A question for further study is thus if it is possible to find partial diagonalizations in other systematic ways of choosing the basis of zeta vectors. This is particularly interesting, since one drawback of our technique is that of diagonal dominance: applying the technique requires that we are in a situation where the negative terms in the

decomposition of the moment matrix are small compared to the positive terms that are selected to be the diagonal entries. Hence in order to prove positive semidefiniteness, we would like to be able to select only “large” positive values on the diagonal. How the large values are distributed among the coefficients of the decomposition is entirely dependent on the constraints of the problem and the solution to the hierarchy.

Finally, another natural point of interest is to find further integrality gap constructions where the technique developed in this paper can be applied.

Acknowledgements The authors would like to express their gratitude to anonymous reviewers for several comments that greatly improved the presentation of this paper.

7 Appendix: Derivation of the SoS hierarchy

In this section we derive the formulation of the SoS hierarchy used in Sect. 2 and give the missing proofs. In our notation we follow the survey by Rothvoß [24] and we use several known derivations [17]. Let $y \in \mathbb{R}^{\mathcal{P}_{2t+2}(N)}$ be a vector indexed by the subsets of $\{1, \dots, n\}$ of size at most $2t + 2$, and $M_{t+1}(y)$ the *moment matrix of the variables* y defined by $[M_{t+1}(y)]_{I,J} = y_{I \cup J}$, for I, J subsets of N such that $|I|, |J| \leq t + 1$. Similarly, for every constraint ℓ define the *moment matrix of the constraint* ℓ as $[M_t^\ell(y)]_{I,J} = \sum_{i=1}^n A_{\ell i} y_{I \cup J \cup \{i\}} - b_\ell y_{I \cup J}$, where $|I|, |J| \leq t$.

Definition 7.1 The SoS hierarchy at level t for the set K , denoted by $\text{SoS}_t(K)$, is given by the following semidefinite program

$$y_\emptyset = 1, \quad (18)$$

$$M_{t+1}(y) \succeq 0, \quad (19)$$

$$M_t^\ell(y) \succeq 0 \quad \text{for every constraint } \ell \quad (20)$$

Change of variables. A point in the SoS hierarchy is given by a vector $y \in \mathbb{R}^{\mathcal{P}_{2t+2}(N)}$, as seen in Definition 7.1. We now change this variable to a vector that is indexed by *all* the subsets of N in order to obtain a useful decomposition of the moment matrix as a sum of rank-one matrices. Here it is not necessary to distinguish between the moment matrix of the variables and constraints, hence in what follows we denote a generic vector by $w \in \mathbb{R}^{\mathcal{P}_{2d}(N)}$, where d is either t or $t + 1$.

Definition 7.2 Let $w \in \mathbb{R}^{\mathcal{P}_{2d}(N)}$. For every $I \in \mathcal{P}_{2d}(N)$, define a vector $w^N \in \mathbb{R}^{\mathcal{P}_n(N)}$ such that

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

To simplify the notation, we note that the moment matrix of the variables is structurally similar to the moment matrix of the constraints: if $z \in \mathbb{R}^{\mathcal{P}_{2t}(N)}$ is a vector such that $z_I = \sum_{i=1}^n A_{\ell i} y_{I \cup \{i\}} - b_\ell y_I$ for some ℓ , then $[M_t^\ell(y)]_{I,J} = z_{I \cup J}$. Hence, the following lemma holds for the moment matrix of variables and constraints.

Lemma 7.1 *Let $w \in \mathbb{R}^{\mathcal{P}_{2d}(N)}$, and $M \in \mathbb{R}^{\mathcal{P}_d(N) \times \mathcal{P}_d(N)}$ such that $M_{I,J} = w_{I \cup J}$. Then*

$$M = \sum_{H \subseteq N} w_H^N Z_H Z_H^\top$$

Proof Since $M_{I,J} = w_{I \cup J}$, we have by the change of variables 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 $\chi_{I \cup J}(H)$ is the 0-1 indicator function such that $\chi_I(H) = 1$ if and only if $I \cup J \subseteq H$. On the other hand, $[Z_H Z_H^\top]_{I,J} = [Z_H]_I [Z_H]_J = 1$ if $I \cup J \subseteq H$, and 0 otherwise. Therefore $[Z_H Z_H^\top]_{I,J} = \chi_{I \cup J}(H)$. \square

Lemma 7.2 *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*

$$z_I^N = g_\ell(x_I) y_I^N \tag{21}$$

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 $z_I = \sum_{I \subseteq H \subseteq N} z_H^N$. Substituting (21) yields

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

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. \square

The above discussion justifies Definition 2.1.

References

1. Bansal, N., Pruhs, K.: The geometry of scheduling. *SIAM J. Comput.* **43**(5), 1684–1698 (2014)
2. Barak, B., Chan, S. O., Kothar P. K.: Sum of squares lower bounds from pairwise independence. In: Proceedings of the 47th Annual ACM Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, pp. 97–106 (2015)
3. Bhaskara, A., Charikar, M., Vijayaraghavan, A., Guruswami, V., Zhou, Y.: Polynomial integrality gaps for strong SDP relaxations of densest k -subgraph. In: Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, pp. 388–405 (2012)
4. Carr, R.D., Fleischer, L., Leung, V. J., Phillips, C. A.: Strengthening integrality gaps for capacitated network design and covering problems. In: Proceedings of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms, January 9–11, 2000, San Francisco, CA, USA, pp. 106–115 (2000)
5. Cheung, K.K.H.: Computation of the Lasserre ranks of some polytopes. *Math. Oper. Res.* **32**(1), 88–94 (2007)
6. Cheung, M., Mestre, J., Shmoys, D. B., Verschae, J.: A primal-dual approximation algorithm for min-sum single-machine scheduling problems. In: Proceedings of the 14th International Workshop on Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX 2011, and 15th International Workshop, RANDOM 2011, Princeton, NJ, USA, pp. 135–146 (2011)
7. Chlantac, E., Tulsiani, M.: Convex relaxations and integrality gaps. In: Anjos, M.F., Lasserre, J.B. (eds.) *Handbook on Semidefinite, Conic and Polynomial Optimization*, pp. 139–169. Springer, Boston (2012)
8. Fawzi, H., Saunderson, J., Parrilo, P. A.: Sparse sum-of-squares certificates on finite abelian groups. In: 54th IEEE Conference on Decision and Control, CDC 2015, Osaka, Japan, 2015, pp. 5909–5914 (2015)
9. Grigoriev, D.: Complexity of positivstellensatz proofs for the knapsack. *Comput. Complex.* **10**(2), 139–154 (2001)
10. Grigoriev, D.: Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theor. Comput. Sci.* **259**(1–2), 613–622 (2001)
11. Karlin, A. R., Mathieu, C., Nguyen, C. T.: Integrality gaps of linear and semi-definite programming relaxations for knapsack. In: Proceedings of the 15th International Conference on Integer Programming and Combinatorial Optimization, IPCO 2011, New York, NY, USA, pp. 301–314 (2011)
12. Karp, R.M.: Reducibility among combinatorial problems. In: Proceedings of a Symposium on the Complexity of Computer Computations, Yorktown Heights, New York, pp. 85–103 (1972)
13. Kurpisz, A., Leppänen, S., Mastrolilli, M.: Tight Sum-of-Squares lower bounds for binary polynomial optimization problems. In: Proceedings of the 43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, Rome, Italy, pp. 78:1–14 (2016)
14. Kurpisz, A., Leppänen, S., Mastrolilli, M.: On the hardest problem formulations for the 0/1 Lasserre hierarchy. In: Proceedings of the 42nd International Colloquium on Automata, Languages, and Programming, ICALP 2015, Kyoto, Japan, pp. 872–885 (2015)
15. Kurpisz, A., Leppänen, S., Mastrolilli, M.: Sum-of-squares hierarchy lower bounds for symmetric formulations. In: Proceedings of the 18th International Conference on Integer Programming and Combinatorial Optimization, IPCO 2016, Liège, Belgium, pp. 362–374 (2016)
16. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. *SIAM J. Optim.* **11**(3), 796–817 (2001)
17. 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)
18. Laurent, M.: Lower bound for the number of iterations in semidefinite hierarchies for the cut polytope. *Math. Oper. Res.* **28**(4), 871–883 (2003)
19. Meka, R., Potechin, A., Wigderson, A.: Sum-of-squares lower bounds for planted clique. In: Proceedings of the 47th Annual ACM Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, pp. 87–96 (2015)
20. Cheung, M., Shmoys, D. B.: A primal-dual approximation algorithm for min-sum single-machine scheduling problems. [arXiv:1612.03339v1](https://arxiv.org/abs/1612.03339v1) [cs.DS]
21. Moore, M.J.: An n job, one machine sequencing algorithm for minimizing the number of late jobs. *Manag. Sci.* **15**, 102–109 (1968)

22. O'Donnell, R., Zhou, Y.: Approximability and proof complexity. In: Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013, New Orleans, Louisiana, USA, pp. 1537–1556. SIAM (2013)
23. Parrilo, P.: Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. Ph.D. thesis, California Institute of Technology (2000)
24. Rothvoß, T.: The Lasserre hierarchy in approximation algorithms. Lecture Notes for the MAPSP 2013 - Tutorial (June 2013)
25. Sakaue, S., Takeda, A., Kim, S., Ito, N.: Exact sdp relaxations with truncated moment matrix for binary polynomial optimization problems. In: Proceedings of the 5th International Conference on Continuous Optimization ICCOPT 2016, Tokyo, Japan (2016)
26. Schoenebeck, G.: Linear level Lasserre lower bounds for certain k-csps. In: Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, Philadelphia, PA, USA, pp. 593–602 (2008)
27. Shor, N.: Class of global minimum bounds of polynomial functions. *Cybernetics* **23**(6), 731–734 (1987)
28. Tulsiani, M.: CSP gaps and reductions in the Lasserre hierarchy. In: Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, pp. 303–312 (2009)
29. Wolsey, L.A.: Facets for a linear inequality in 0–1 variables. *Math. Program.* **8**, 168–175 (1975)