On the Lasserre/Sum-of-Squares Hierarchy with Knapsack Covering Inequalities *

Adam Kurpisz † Samuli Leppänen ‡ Monaldo Mastrolilli §

July 8, 2014

Abstract

The Lasserre/Sum-of-Squares hierarchy is a systematic procedure to strengthen LP relaxations by constructing a sequence of increasingly tight formulations. For a wide variety of optimization problems, this approach captures the convex relaxations used in the best available approximation algorithms.

The capacitated covering IP is an integer program of the form $\min\{cx : Ux \geq d, 0 \leq x \leq b, x \in \mathbb{Z}^+\}$, where all the entries of $c, U$ and $d$ are nonnegative. One difficulty in approximating the capacitated covering problems lies in the fact that the ratio between the optimal IP solution to the optimal LP solution can be as large as $||d||_\infty$, even when $U$ consists of a single row (i.e. the Min Knapsack problem). Currently the strongest linear program relaxation is obtained by adding (exponentially many) valid knapsack cover ($KC$) inequalities introduced by Wolsey [44], which yields a very powerful way to cope with these problems.

For the Min Knapsack problem we prove that even after non-constant number of levels the Lasserre/Sum-of-Squares hierarchy does not improve the integrality gap of 2 implied by the starting ($KC$) inequalities. Furthermore, we show that the integrality gap of the relaxation stays $M$ for the special case of $\sum_i x_i \geq 1/M$ when starting with the standard Min Knapsack polytope. We note that Min Knapsack admits an FPTAS and our results quantify a fundamental weakness of the Lasserre/Sum-of-Squares hierarchy for this basic problem.

1 Introduction

The Lasserre/Sum-of-Squares (SOS) hierarchy [27, 30, 38, 42] 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 SDP constraints). This approach captures the convex relaxations used in the best available approximation algorithms for a wide variety of optimization problems. For example, the first round of the Lasserre hierarchy for the Independent Set problem implies the Lovász $\theta$-function [34] and for the Max Cut problem it gives the Goemans-Williamson relaxation [23]. The ARV relaxation of the Sparsest Cut [2] problem is no stronger than the relaxation given in the third round of the Lasserre hierarchy, and most recently the subexponential time algorithm for Unique Games [1] is implied by a sublinear number of rounds of the Lasserre hierarchy [9, 28]. Other approximation

*Supported by the Swiss National Science Foundation project 200020-144491/1 “Approximation Algorithms for Machine Scheduling Through Theory and Experiments” and by Scix Project 12.311.
†adam@idsia.ch. IDSIA, Lugano, Switzerland and Wroclaw University of Technology, Poland.
‡samiuli@idsia.ch. IDSIA, Lugano, Switzerland.
§monaldo@idsia.ch. IDSIA, Lugano, Switzerland.

1The “Sum-of-Squares” (SOS) proof system is the dual of the Lasserre hierarchy meaning that the “$d/2$-round Lasserre value” of an optimization problem is equal to the best bound provable using a degree-$d$ SOS proof. For brevity, we will interchange Lasserre hierarchy with SOS hierarchy since they are essentially the same in our context.
guarantees that arise from the first $O(1)$ levels of Lasserre (or weaker) hierarchy can be found in \cite{9, 10, 17, 18, 20, 21, 28, 35, 39}. For a more detailed overview on the use of hierarchies in approximation algorithms, see the recent survey of Chlamtác and Tulsiani \cite{19}, the reading group web-page in \cite{3} and the references therein. Additional results and references can be found in the monograph by Laurent \cite{32}.

The \textit{capacitated covering integer program} (see e.g. \cite{13}) is a general integer program of the form $\min \{ c^T x \mid U x \geq d, 0 \leq x \leq b, x \in \mathbb{Z}^+ \}$, where all the entries of $c, U$ and $d$ are nonnegative. One difficulty in approximating the capacitated covering problems lies in the fact that the ratio between the optimal IP solution value to the optimal LP solution value (\textit{integrality gap}) can be as large as $|d|_\infty$, even when $U$ consists of a single row (i.e. the \textsc{Min Knapsack} problem).

The currently most powerful way to cope with these problems is to strengthen the LP by adding (exponentially many) \textsc{Knapsack Cover} (KC) inequalities introduced by Wolsey \cite{44} (see also \cite{13}), that have proved to be a useful tool to address capacitated covering problems \cite{3, 5, 6, 12, 14, 16, 33}. For the \textsc{Min Knapsack} problem, the improved IP/LP ratio with these inequalities is 2 (and it is tight \cite{13}).

The natural question addressed in this paper revolves around understanding whether the Lasserre hierarchy can help to strengthen the (KC) inequalities used in several capacitated covering problems.

For example, for the \textsc{Min-Sum Single-Machine Scheduling} problem (denoted $1||\sum f_j$), the (KC) inequalities are the essential ingredient for obtaining the $(4 + \varepsilon)$-approximation algorithm in \cite{37} (see also \cite{6} and \cite{16}). Since the \textsc{Min Knapsack} problem is a special case of $1||\sum f_j$, no better than 2-approximation can be obtained (due to the (KC) inequalities integrality gap of 2 \cite{13}). However, no hardness of approximation result is known for $1||\sum f_j$ and, as remarked in \cite{16}, “it is still conceivable (and perhaps likely) that there exists a polynomial time approximation scheme”. With this aim, stronger relaxations are sought since (KC) inequalities are not sufficient.

### 1.1 Related Results

Most of the known lower bounds for the SOS hierarchy originated in the works of Grigoriev \cite{24, 25} (also independently rediscovered later by Schoenebeck \cite{41}). These works show that random 3XOR or 3SAT instances cannot be solved by even $\Omega(n)$ rounds of SOS hierarchy. Subsequent lower bounds, such as those of \cite{11, 43} rely on \cite{25, 41} combined with gadget reductions.

In \cite{24} Grigoriev showed that if $x_1, ..., x_n$ are \pm 1 with $n$ odd, then degree-$(n - 1)$ SOS cannot disprove $x_1 + \ldots + x_n = 0$, even though of course $|x_1 + \ldots + x_n| \geq 1$. (A simplified proof can be found in \cite{26}).

In \cite{13} Cheung considered the case of $x_1, ..., x_n$ constrained to be 0/1 along with the constraint $x_1 + \ldots + x_n \geq \delta(n)$, for some $\delta(n) < 1$ that depends on $n$. He proved that the Lasserre hierarchy requires $n$ levels to converge to the integral polytope. This is shown by providing a feasible solution at level $n - 1$ of value $\frac{n}{n+1}$ (i.e. a solution where the sum of the values assigned to the variables is equal to $\frac{n}{n+1}$), whereas the smallest integral solution has value one. Note that the integrality gap given by Cheung’s solution is $1+1/n$, i.e. arbitrarily close to one for large $n$. Therefore the result in \cite{13} \textit{does not} rule out the possibility that the integrality gap might decrease quickly with the number of levels (recall that the integrality gap of the starting LP can be arbitrarily large). Indeed the main interest in the work of Cheung \cite{13} revolves around understanding how fast the Lasserre hierarchy converges to the integral polytope. The main interest in this paper is understanding how fast the integrality gap of the starting LP reduces. This is conceptually an important difference. For example, for the \textsc{Max-Knapsack} problem the Lasserre hierarchy requires $\Omega(n)$ levels to converge to the integral polytope, but only $O(1/\varepsilon)$ levels to obtain an integrality gap of $O(1 - \varepsilon)$, for any arbitrarily small constant $\varepsilon > 0$ \cite{29}.

In a recent paper \cite{36} by one of the authors, the \textsc{Min Knapsack} problem is studied when the constraint $x_1 + \ldots + x_n \geq 1/M$ (see \cite{11}) is strengthened with some levels of the Lasserre
hierarchy (without using \((KC)\) inequalities). When \(M\) is exponential in the number of levels, it is shown that the integrality gap of \((1)\) remains arbitrarily large, even at level \((n - 1)\) of Lasserre hierarchy (note that this is a tight characterization, since at level \(n\) the solution is integral). More precisely, it is shown that for \textit{any} characterization \(t < n\), a solution with \(\sum^{n}_{i=1}x_{i} \leq \frac{2^{2t+1}}{M}\) is feasible for the Lasserre hierarchy starting with \((1)\) (showing therefore an integrality gap of \(k = \frac{M}{2^{2t+1}}\) for \(M \geq 2^{2t+1}\), i.e. \(M\) exponential in \(t\)). In terms of SOS proof system, the result in [26] implies that degree-\((n - 1)\) SOS cannot refute a solution of value \(1/k\), where \(k\) can be arbitrarily small, exhibiting therefore a much weaker behavior of the SOS hierarchy than the one showed by Cheung [15] (in the latter it was proved that degree-\((n - 1)\) SOS cannot refute a solution of value \(n/(n + 1)\)).

However, the integrality gap in [26] holds when \(M\) is assumed to be exponential in the number of levels (i.e. when the input instance allows numbers exponentially large). In this paper, we complement the result in [26] by considering \textit{any} \(M > 1\) and showing that the Lasserre hierarchy does not help to improve the initial gap of \(1/M\) for the starting \((LP)\) \((1)\), even at level \(t = \sqrt{\log n}\).

With the aim to analyzing the ability of the Lasserre hierarchy to reduce the initial integrality gap, it “is perhaps not too apt” to start with an LP relaxation with an unbounded integrality gap. We show that starting with the the currently best known LP relaxation strengthened with \((KC)\) inequalities (with gap bounded by 2), does not help either to reduce the initial gap.

### 1.2 Our Results

The main contribution of this paper is in analyzing the effectiveness of the Lasserre hierarchy to reduce the integrality gap of the \((KC)\) inequalities. With this aim, we consider the very basic capacitated covering integer program with one single constraint, namely the Min Knapsack problem, that is ubiquitous in many applications. We show that the integrality gap of the Min Knapsack relaxation with \((KC)\) inequalities remains 2, even after \(t = \sqrt{\log n}\) levels. This roughly shows that here the Lasserre hierarchy does not help to strengthen the \((KC)\) inequalities for this basic version of the capacitated covering problem.

As a direct consequence of our analysis we also show another result that is interesting on its own. Consider the following basic LP relaxation:

\[
(LP)\ \ \ \ \min\{\sum^{n}_{i=1}x_{i} : \sum^{n}_{i=1}x_{i} - 1/M \geq 0, x_{i} \in [0, 1] \text{ for } i \in N\}
\]

We prove that for any \(\varepsilon > 0\) and \textit{any} \(M > 1\), a solution with \(\sum^{n}_{i=1}x_{i} \leq \frac{1}{M(1-\varepsilon)}\) is feasible for the Lasserre hierarchy starting with \((LP)\) at level \(t = \sqrt{\log n}\) (showing therefore an integrality gap of \(M(1 - \varepsilon)\)). Note that this result complements those in [15] [36] (see the discussion in Section 1.1 for a comparison).

The proofs are obtained by providing feasible solutions for the Lasserre hierarchy with the claimed bounds. Satisfying the positive semi-definite requirement is one of the major hurdles in constructing the Lasserre hierarchy gap examples. In this paper, we first formulate the Lasserre hierarchy at level \(t\) as a semi-infinite LP after a change of variables. Each linear constraint can be seen as the requirement that the (pseudo)-expected value (see e.g. [4] [8]) of the original constraint multiplied by the square of a polynomial of degree \(t\) being nonnegative. This must hold for \textit{any} polynomial of degree at most \(t\). Proving that a solution satisfies these infinite number of constraints is among the main difficulties. We restrict to “well-structured” solutions.

\[\text{Note that in order to claim that one can optimize over the Lasserre hierarchy in polynomial time, it could be necessary that the number of constraints of the starting LP being polynomial in the input size (see the discussion in [33]). Even though it is not clear whether the Lasserre hierarchy with \((KC)\) inequalities can be solved in polynomial time, our result shows that it does not help to strengthen \((KC)\) inequalities.}\]

\[\text{For ease of presentation we omit a more involved analysis which shows that the claimed results still hold for } t \text{ arbitrarily close to } \log n.\]
and show that this simplifies the positive-semidefinite requirements. Moreover this reduces the number of “interesting” cases by obtaining a set of sufficient conditions that are easier to analyze: for example we show that we can restrict the analysis to univariate polynomials with degree exactly $t$ and real roots. This more structured set of requirements allows us to prove that a solution, with the claimed integrality bounds, is feasible for the Lasserre hierarchy.

We note that most of prior results exhibiting gap instances for the Lasserre hierarchy relaxations do so for problems that are already known to be hard to approximate, under some suitable assumption. Based on this hardness result, one would expect that the Lasserre hierarchy relaxations to have an integrality gap that matches the inapproximability factor. Some exceptions are also known where the known integrality gaps are substantially stronger than the (very weak) hardness bounds known for the problem (see [11] and the references therein), but here it is still conceivable that the apparent “weakness” of the Lasserre hierarchy is due to the inherent complexity of the problem, that has still to be fully understood, and are perhaps indicative of the hardness of approximating. In this paper, our gap construction is a rare exception to this trend, by showing integrality gaps for an “easy” problem that admits an FPTAS.

In the following, due to space limitations, omitted proofs can be found in appendix.

2 The Lasserre Hierarchy

In this section, we provide a definition of the Lasserre hierarchy [30]. In our notation, we mainly follow the survey of Laurent [31]. For an introduction we refer to [40].

Variables and Moment Matrix. Throughout this paper, vectors are written as columns. 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 known as the moment matrix of $y$.

The Lasserre Hierarchy Definition. Let $K$ be defined by the following

$$K := \{x \in [0,1]^n : g_\ell(x) \geq 0 \text{ for } \ell = 1, \ldots, m\}$$ (2)

where $g_\ell(x) = \sum_{i=1}^n g_{i \ell} x_i$ + $g_{\emptyset}$ is a linear function of $x$. Let $g \ast y$ be a vector, often called shift operator, where the $I$-th entry is $(g \ast y)_I = \sum_{i=1}^n g_i y_{I \cup \{i\}}$ + $g_{\emptyset} y_I$.

Definition 1. The Lasserre hierarchy at the $t$-th level, denoted as $\text{LAs}_t(K)$, is the set of vectors $y \in \mathbb{R}^{P_{2t} + 2(N)}$ that satisfy the following

$$y_{\emptyset} = 1$$ (3)

$$M_{t+1}(y) \succeq 0$$ (4)

$$M_t(g \ast y) \succeq 0 \quad \ell = 1, \ldots, m$$ (5)

Equivalent formulation. For any $I \in \mathcal{P}(N)$, consider the following variables $\{y_{I,N \setminus I} : I \subseteq N\}$ defined as follows:

$$y_{I,N \setminus I} := \sum_{H \subseteq N \setminus I} (-1)^{|H|} y_{H \cup I}$$ (6)

Variables $\{y_{I,N \setminus I} : I \subseteq N\}$ have a nice interpretation as (pseudo)probability (see e.g. [7, 8]): $y_{I,N \setminus I}$ can be seen as the (pseudo)probability of the integral solution $\{x_i = 1 : i \in I, x_j = 0 : j \in N \setminus I\}$. At level $n$, variables $\{y_{I,N \setminus I} : I \subseteq N\}$ are actual probabilities.

Given a generic multilinear polynomial $P_d(x) = \sum_{I \in \mathcal{P}(N)} (v_I \prod_{i \in I} x_i)$ of degree $d$ with coefficients $v_I$, let $P_d(x_Q)$ be the evaluation of $P_d(x)$ by setting $x_i = 1$ if $i \in Q$, and $x_i = 0$ if $i \in N \setminus Q$. 

In appendix.
The following gives an equivalent formulation of the Lasserre hierarchy as a semi-infinite linear program that is more suitable for analyzing the integrality gap of the considered problem. The formulation is easily obtained after the change of variables (6) (as explained in the appendix).

Definition 2. The Lasserre hierarchy at the $t$-th level, denoted as $\text{Las}_t(K)$, is the set of solutions $y^N$ that satisfy the following semi-infinite linear program:

$$\sum_{Q \subseteq N} y_{Q,N \setminus Q} = 1$$  \hfill (7)

$$\sum_{Q \subseteq N} y_{Q,N \setminus Q} \cdot p_d^2(x_Q) \geq 0 \quad \forall \text{ polynomial } p_d(x) \text{ of degree } d \leq t + 1$$  \hfill (8)

$$\sum_{Q \subseteq N} y_{Q,N \setminus Q} \cdot g_\ell(x_Q) \cdot p_d^2(x_Q) \geq 0 \quad \forall \text{ polynomial } p_d(x) \text{ of degree } d \leq t \land \forall \ell \in [m]$$  \hfill (9)

3 Min-Knapsack with knapsack covering 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) $\min \left\{ \sum_{i=1}^n c_i x_i : \sum_{i=1}^n p_i x_i \geq P, x \in \{0, 1\}^n \right\}$. It is not hard to show that the natural linear program, obtained by relaxing $x \in \{0, 1\}^n$ to $x \in [0, 1]^n$ in (IP), has an unbounded integrality gap.

3.1 Adding knapsack covering inequalities

By adding the Knapsack Cover (KC) inequalities introduced by Wolsey [44] (see also [13]), the arbitrarily large integrality gap of the natural LP can be reduced to 2 (and it is tight [13]). The (KC) constraints are as follows: $\sum_{j \notin A} p^A_j x_j \geq P - p(A)$ for all $A \subseteq N$, where $p(A) = \sum_{i \in A} p_i$ and $p^A_j = \min \{p_j, P - p(A)\}$. (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).

The following instance [13] shows that the integrality gap implied by (KC) inequalities is 2: we have $n$ items of unit costs and profits. We are asked to select a set of items in order to obtain a profit of at least $1 + 1/(n - 1)$. The resulting linear program formulation with (KC) inequalities is as follows.

$$\begin{align*}
\text{(LP$^+$)} \quad \min & \quad \sum_{j=1}^n x_j \\
\text{s.t.} & \quad \sum_{j=1}^n x_j \geq 1 + 1/(n - 1) \\
& \quad \sum_{j \in N'} x_j \geq 1 \quad \forall N' \subseteq N : |N'| = n - 1 \\
& \quad x_i \in [0, 1] \quad i = 1, \ldots, n
\end{align*}$$

Note that the solution $x_i = 1/(n - 1)$ is a valid fractional solution of value $1 + 1/(n - 1)$ whereas the optimal integral solution has value 2.
4 Lasserre hierarchy integrality gap with \((KC)\) inequalities

In this section we show that the Lasserre hierarchy, starting with \((LP^+)\) (i.e. \(\text{Las}_t(LP^+)\)), does not reduce the integrality gap of 2 even at level \(t = \sqrt{\log n}\). We show this by providing a feasible solution having the desired gap in the following lemma. Throughout this section we assume that \(n\) is a large enough number.

Lemma 1. For any \(Q \subseteq N\) such that \(|Q| = k\), the solution \(y_{Q,N\setminus Q} = \binom{n}{k}^{-1} \Pr(\sum_{i=1}^{n} x_i = k)\) is a feasible solution to \(\text{Las}_t(LP^+)\), where \(t = \sqrt{\log n}\), \(\varepsilon = o(t^{-1})\) and

\[
\Pr\left(\sum_{i=1}^{n} x_i = \log n\right) = \frac{(1 + \varepsilon)n}{(n - 1) \log n} \\
\Pr\left(\sum_{i=1}^{n} x_i = j \frac{n}{t}\right) = \frac{\varepsilon t}{jn} \quad j = 1, \ldots, t, \\
\Pr\left(\sum_{i=1}^{n} x_i = k\right) = 0 \quad \text{for any other} \ k \geq 1, \\
\Pr\left(\sum_{i=1}^{n} x_i = 0\right) = 1 - \sum_{k=1}^{n} \Pr\left(\sum_{i=1}^{n} x_i = k\right)
\]

(10)

Corollary 2. The integrality gap of \(\text{Las}_t(LP^+)\) is 2 for \(t = \sqrt{\log n}\).

Proof. The solution defined in Lemma 1 has an objective value that is arbitrarily close to 1 whereas the optimal integral value is 2. Indeed, with the given solution the objective value is obtained by projecting the Lasserre solution onto the \((LP^+)\) variables (i.e. setting \(x_i = y_{\{i\}}\)) and by (17) it is therefore equal to

\[
\sum_{i=1}^{n} y_{\{i\}} = \sum_{Q \subseteq N} y_{Q,N\setminus Q} |Q| = \frac{n}{n-1}(1 + \varepsilon) + \varepsilon t \rightarrow 1
\]

as \(n \rightarrow \infty\). \(\square\)

4.1 Proof of Lemma 1

Lemma 1 follows by showing that the suggested solution (10) satisfies (7), (8) and (9).

First note that (7) and (8) are satisfied by (10) since it forms a probability distribution. Indeed all the given values are nonnegative and they sum up to one.

We show that solution (10) also satisfies (9). We start by observing that the structure of the solution given in Lemma 1 allows us to consider only univariate polynomials in order to satisfy (8). This is shown in Lemma 3 by using the following preliminary observation.

Lemma 3. Let \(u : N_0 \rightarrow \mathbb{R}\) be a function such that \(u(k) \geq 0\) for any \(k > 0\). If

\[
\sum_{k=0}^{n} u(k)p_d^2(k) \geq 0 \quad \forall \text{ univariate polynomials } p_d \text{ of degree } d \leq t
\]

then

\[
\sum_{Q \subseteq N} u(|Q|) \binom{n}{|Q|}^{-1} P_d^2(x_Q) \geq 0 \quad \forall \text{ multivariate polynomials } P_d \text{ of degree } d \leq t
\]
Lemma 4. If the solution of Lemma 2 satisfies the following then (9) holds.

\[ \sum_{k=1}^{n} \Pr \left( \sum_{i=1}^{n} x_i = k \right) \left( k - 2 \right) p_d^2(k) \geq \left( 1 + \frac{1}{n-1} \right) p_d^2(0) \quad \forall \text{ univariate } p_d \text{ of degree } d \leq t \tag{11} \]

Proof. We first show that (11) implies that (9) holds for the constraint \( \sum_{j=1}^{n} x_j \geq 1 + 1/(n-1) \).

Note that for large \( n \) we assign the probabilities such that \( y_{Q,N \setminus Q} = 0 \) for all \( Q \) when \( |Q| = 1 \).

Let for \( k \geq 0 \)

\[ u(k) = \Pr \left( \sum_{i=1}^{n} x_i = k \right) \left( k - 1 - \frac{1}{n-1} \right) \]

In the solution given in Lemma 4 we have \( y_{Q,N \setminus Q} = y_{W,N \setminus W} \) whenever \( |Q| = |W| \). This allows us to write \( y_{Q,N \setminus Q} = \binom{n}{k}^{-1} \Pr(\sum_{i=1}^{n} x_i = k) \). By the latter and by Lemma 3, the inequality (9) is fulfilled by satisfying the following for all univariate polynomials \( p_d \) of degree \( d \leq t \)

\[ \sum_{k=1}^{n} \Pr \left( \sum_{i=1}^{n} x_i = k \right) \left( k - 1 - \frac{1}{n-1} \right) p_d^2(k) \geq \Pr \left( \sum_{i=1}^{n} x_i = 0 \right) \left( 1 + \frac{1}{n-1} \right) p_d^2(0) \]

Note that the latter is clearly implied if (11) is satisfied.

Next, we will show that (11) also implies that (9) is satisfied for all the knapsack covering constraints, w.l.o.g. for \( \sum_{j=1}^{n-1} x_j \geq 1 \). This constraint is not symmetric and we need to deal with the asymmetry first. According to Definition 2, (9) holds for the considered constraint \( g_d(x) = \sum_{j=1}^{n-1} x_j - 1 \) if

\[ \sum_{Q \subseteq N} y_{Q,N \setminus Q} g_d(x_Q) P_d^2(x_Q) = \sum_{n \in \mathbb{N}} y_{Q,N \setminus Q} (|Q| - 1) P_d^2(x_Q) + \sum_{n \in \mathbb{N}} y_{Q,N \setminus Q} (|Q| - 2) P_d^2(x_Q) \geq 0 \]

Again, for large \( n \) we assign the probabilities such that \( y_{Q,N \setminus Q} = 0 \) for all \( Q \) when \( |Q| = 1 \), so the only negative term corresponds to the empty set. Moving it to the right-hand side yields

\[ \sum_{n \in \mathbb{N}} y_{Q,N \setminus Q} (|Q| - 1) P_d^2(x_Q) + \sum_{n \in \mathbb{N}} y_{Q,N \setminus Q} (|Q| - 2) P_d^2(x_Q) \geq y_{\emptyset,N} P_d^2(x_{\emptyset}) \]

This relation is implied by the following stronger requirement

\[ \sum_{Q \neq \emptyset} y_{Q,N \setminus Q} (|Q| - 2) P_d^2(x_Q) \geq y_{\emptyset,N} P_d^2(x_{\emptyset}) \quad \forall \text{ multivariate } P_d \text{ of degree } d \leq t \tag{12} \]

By setting

\[ u(k) = \begin{cases} -\Pr(\sum_{i=1}^{n} x_i = 0) & k = 0 \\ \Pr(\sum_{i=1}^{n} x_i = k) (k - 2) & \forall k > 0 \end{cases} \]

Lemma 3 implies (12) if the following is satisfied for all univariate polynomials \( p_d \) of degree \( d \leq t \)

\[ \sum_{k=1}^{n} \Pr \left( \sum_{i=1}^{n} x_i = k \right) (k - 2) p_d^2(k) \geq \Pr \left( \sum_{i=1}^{n} x_i = 0 \right) p_d^2(0) \]

This in turn is implied by (11). \( \square \)
Satisfying the condition of Lemma 4. We now show that the solution given in the statement of Lemma 1 satisfies (11) and that there exists an \( \varepsilon = o(t^{-1}) \) as claimed. We start by giving a helpful lemma to simplify the analysis (the proof can be found in appendix).

**Lemma 5.** In order to prove that given probabilities \( \Pr(\sum_{i=1}^{n} x_i = k) \) satisfy (11) it is enough to prove that they satisfy (14) for polynomials with the following properties:

(a) all the roots \( r_1, \ldots, r_t \) are real,

(b) all the roots are in the range, \( 1 \leq r_j \leq n \) for all \( j = 1, \ldots, t \),

(c) the degree is exactly \( t \).

The fundamental theorem of algebra states that a polynomial of one variable of degree \( t \) has exactly \( t \) roots. We prove that the probabilities satisfy (11) by expressing the generic univariate polynomial \( p_t(x) \) using its roots \( r_1, \ldots, r_t \), so that (11) becomes

\[
\sum_{k=1}^{n} \Pr\left(\sum_{i=1}^{n} x_i = k\right) (k-2) (r_1-k)^2 \cdots (r_t-k)^2 \geq \left(1 + \frac{1}{n-1}\right) r_1^2 \cdots r_t^2 \quad (13)
\]

To show that (13) is satisfied we separate two cases: when all of the roots of the polynomial are greater than 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 \( \beta = \log n \).

1. \( r_j \geq \alpha \) for all \( j \). It is sufficient to show that the left–hand side term in (13) corresponding to \( k = \beta \) satisfies

\[
\Pr\left(\sum_{i=1}^{n} x_i = \beta\right) (\beta-2) (r_1-\beta)^2 \cdots (r_t-\beta)^2 \geq \left(1 + \frac{1}{n-1}\right) r_1^2 \cdots r_t^2
\]

Replacing the probability we get

\[
\frac{n}{n-1} \frac{1 + \varepsilon}{\beta} (\beta-2) (r_1-\beta)^2 \cdots (r_t-\beta)^2 \geq \frac{n}{n-1} r_1^2 \cdots r_t^2 \iff 1 + \varepsilon \geq \prod_{i=1}^{t} \left(\frac{r_i}{r_i - \beta}\right)^2 \frac{1}{\frac{1}{1 - 2\beta^{-1}}}
\]

Since by Lemma 5 (b) and assumption, all roots \( r_j \) satisfy \( \alpha \leq r_j \leq n \). Since \( \frac{r_j}{r_j - \beta} \leq \frac{\alpha}{\alpha - \beta} \) it is sufficient that the following 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 \( u = T_{t+1} \), \( l = 1, \ldots, t \) such that \( \Pr(\sum_{i=1}^{n} x_i = u) > 0 \) and \( |u - r_i| \geq \frac{t}{n} \) for all \( i = 1, \ldots, t \). Let \( u \) be such a point. It is sufficient to show that we can satisfy

\[
\Pr\left(\sum_{i=1}^{n} x_i = u\right) (u-2) (r_1-u)^2 \cdots (r_t-u)^2 \geq \frac{n}{n-1} r_1^2 \cdots r_t^2
\]

We have \( \Pr(\sum_{i=1}^{n} x_i = u) = \frac{\varepsilon}{u} \) 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
\[
\varepsilon \left( \frac{n}{2t} \right)^{2t} \geq \frac{n^2}{n-1} n^{2t-2} \alpha^2
\]
which gives us the requirement that
\[
\varepsilon \geq \frac{2\alpha^2}{n^2} (2t)^{2t} \frac{n}{n-1}
\]
These two cases suggest that we fix \( \varepsilon \) as
\[
\varepsilon = \max \left\{ \frac{1}{1 - 2\beta^{-1}} \left( 1 - \frac{\beta}{\alpha} \right)^{-2t} - 1, \frac{n - 1}{n-1} \frac{2\alpha^2}{n^2} (2t)^{2t} \right\}
\]
The proof has now been reduced to showing that with this choice of \( \varepsilon \) we have \( \varepsilon t \to 0 \), i.e., \( \varepsilon = o(t^{-1}) \). Assume \( \varepsilon = \frac{1}{1 - 2\beta^{-1}} \left( 1 - \frac{\beta}{\alpha} \right)^{-2t} - 1 \). Then
\[
\varepsilon t = t \left( \frac{1}{1 - 2\beta^{-1}} \left( 1 - \frac{\beta}{\alpha} \right)^{-2t} - 1 \right) \leq t \left( \frac{1}{1 - 2\beta^{-1}} e^{4t \beta^{-1}} - 1 \right)
\]
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
\[
\varepsilon t \leq t \left( \frac{1}{1 - 2\beta^{-1}} \cdot 8t \frac{\beta}{\alpha} + t \left( \frac{1}{1 - 2\beta^{-1}} - 1 \right) \right) = \frac{8}{1 - 2\beta^{-1}} \cdot t^2 \frac{\beta}{\alpha} + \frac{2t^2 \beta^{-1}}{1 - 2\beta^{-1}}
\]
The right-hand side goes to 0 if \( \varepsilon \to 0 \) as \( t \to 0 \) and \( t \to 0 \) as \( n \to \infty \). This is clearly the case when we have \( \alpha = \log^3 n, \beta = \log n \) and \( t = \sqrt{\log n} \).

Next, assume \( \varepsilon = \frac{n - 1}{n-1} \frac{2\alpha^2}{n^2} (2t)^{2t} \). Then
\[
\varepsilon t = t \frac{n - 1}{n-1} \frac{2\alpha^2}{n^2} (2t)^{2t},
\]
which immediately yields the condition on \( \alpha \) and \( t \) that we need \( \frac{t^2 \alpha^2}{n^2} (2t)^{2t} \to 0 \) as \( n \to \infty \).

Substituting \( \alpha = \log^3 n \) and \( t = \sqrt{\log n} \) allows us to write this as
\[
\frac{t^2 \alpha^2}{n^2} (2t)^{2t} = \sqrt{\log n} \frac{\log^6 n}{n^2} (2\sqrt{\log n})^{2\sqrt{\log n}}
\]
By a change of variables of the form \( w = \sqrt{\log n} \) we get
\[
\frac{w^{2w+13} 2w^2}{e^{2w^2}} \leq \frac{w^{4w+13} \log w}{e^{2w^2}} = \frac{e^{(4w+13) \log w}}{e^{2w^2}} = e^{(4w+13) \log w - 2w^2}
\]
which tends to 0 as \( n \to \infty \).

4.2 Min-Knapsack

We remark that our analysis implies the following fact that is interesting by its own. Consider the linear program relaxation in (1). The following lemma provides a feasible solution for \( \text{Las}^L(LP) \) with integrality gap arbitrarily close to \( 1/M \). The proof is omitted since it is similar to the proof of Lemma [1].
Lemma 6. For any $Q \subseteq N$ such that $|Q| = k$, the solution $y_{Q,N\setminus Q} = \binom{n}{k}^{-1} \Pr(\sum_{i=1}^{n} x_i = k)$ is a feasible solution to $\text{Las}_t(LP)$, where $t = \sqrt{\log n}$, $\varepsilon = o(t^{-1})$ and

\[
\Pr \left( \sum_{i=1}^{n} x_i = \log n \right) = \frac{(1 + \varepsilon)}{M \log n}
\]

\[
\Pr \left( \sum_{i=1}^{n} x_i = j \frac{n}{t} \right) = \frac{\varepsilon t}{jn} \quad j = 1, \ldots, t,
\]

\[
\Pr \left( \sum_{i=1}^{n} x_i = k \right) = 0 \quad \text{for any other } k \geq 1,
\]

\[
\Pr \left( \sum_{i=1}^{n} x_i = 0 \right) = 1 - \sum_{k=1}^{n} \Pr \left( \sum_{i=1}^{n} x_i = k \right)
\]

(14)

Corollary 7. The integrality gap of $\text{Las}_t(LP)$ is arbitrarily close to $M$ for $t = \sqrt{\log n}$.

Proof. The solution defined in Lemma 6 has an objective value that is arbitrarily close to $1/M$ whereas the optimal integral value is $1:

\[
\sum_{i=1}^{n} y_{i} = \sum_{Q \subseteq N} y_{Q,N\setminus Q} |Q| = \frac{1}{M} (1 + \varepsilon) \varepsilon t \rightarrow 1/M
\]

as $n \rightarrow \infty$. \qed

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

References

[1] S. Arora, B. Barak, and D. Steurer. Subexponential algorithms for unique games and related problems. In FOCS, pages 563–572, 2010.

[2] S. Arora, S. Rao, and U. V. Vazirani. Expander flows, geometric embeddings and graph partitioning. J. ACM, 56(2), 2009.

[3] N. Bansal. \url{http://www.win.tue.nl/~nikhil/hierarchies/index.html}.

[4] N. Bansal, N. Buchbinder, and J. Naor. Randomized competitive algorithms for generalized caching. In STOC, pages 235–244, 2008.

[5] N. Bansal, A. Gupta, and R. Krishnaswamy. A constant factor approximation algorithm for generalized min-sum set cover. In SODA, pages 1539–1545, 2010.

[6] N. Bansal and K. Pruhs. The geometry of scheduling. In FOCS, pages 407–414, 2010.

[7] B. Barak, F. G. S. L. Brandão, A. W. Harrow, J. A. Kelner, D. Steurer, and Y. Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In STOC, pages 307–326, 2012.

[8] B. Barak, J. A. Kelner, and D. Steurer. Rounding sum-of-squares relaxations. CoRR \{(to appear in STOC 2014), abs/1312.6652, 2013.

[9] B. Barak, P. Raghwendra, and D. Steurer. Rounding semidefinite programming hierarchies via global correlation. In FOCS, pages 472–481, 2011.

[10] M. Bateni, M. Charikar, and V. Guruswami. Maxmin allocation via degree lower-bounded arborescences. In STOC, pages 543–552, 2009.

[11] A. Bhaskara, M. Charikar, A. Vijayaraghavan, V. Guruswami, and Y. Zhou. Polynomial integrality gaps for strong sdp relaxations of densest $k$-subgraph. In SODA, pages 388–405, 2012.
[12] T. Carnes and D. B. Shmoys. Primal-dual schema for capacitated covering problems. In IPCO, pages 288–302, 2008.

[13] R. D. Carr, L. Fleischer, V. J. Leung, and C. A. Phillips. Strengthening integrality gaps for capacitated network design and covering problems. In SODA, pages 106–115, 2000.

[14] D. Chakrabarty, E. Grant, and J. Könemann. On column-restricted and priority covering integer programs. In IPCO, pages 355–368, 2010.

[15] K. K. H. Cheung. Computation of the Lasserre ranks of some polytopes. Math. Oper. Res., 32(1):88–94, 2007.

[16] M. Cheung and D. B. Shmoys. A primal-dual approximation algorithm for min-sum single-machine scheduling problems. In APPROX-RANDOM, pages 135–146, 2011.

[17] E. Chlamtac. Approximation algorithms using hierarchies of semidefinite programming relaxations. In FOCS, pages 691–701, 2007.

[18] E. Chlamtac and G. Singh. Improved approximation guarantees through higher levels of SDP hierarchies. In APPROX-RANDOM, pages 49–62, 2008.

[19] E. Chlamtac and M. Tulsiani. Convex relaxations and integrality gaps. In to appear in Handbook on semidefinite, conic and polynomial optimization. Springer.

[20] M. Cygan, F. Grandoni, and M. Mastroiilli. How to sell hyperedges: The hypermatching assignment problem. In S. Khanna, editor, SODA, pages 342–351. SIAM, 2013.

[21] W. F. de la Vega and C. Kenyon-Mathieu. Linear programming relaxations of maxcut. In SODA, pages 53–61, 2007.

[22] M. A. Goberna and M. A. López. Linear semi-infinite programming theory: An updated survey. European Journal of Operational Research, 143(2):390–405, 2002.

[23] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. J. ACM, 42(6):1115–1145, 1995.

[24] D. Grigoriev. Complexity of positivstellensatz proofs for the knapsack. Computational Complexity, 10(2):139–154, 2001.

[25] D. Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. Theor. Comput. Sci., 259(1-2):613–622, 2001.

[26] D. Grigoriev, E. A. Hirsch, and D. V. Pasechnik. Complexity of semi-algebraic proofs. In STACS, pages 419–430, 2002.

[27] D. Grigoriev and N. Vorobjov. Complexity of null-and positivstellensatz proofs. Ann. Pure Appl. Logic, 113(1-3):153–160, 2001.

[28] V. Guruswami and A. K. Sinop. Lasserre hierarchy, higher eigenvalues, and approximation schemes for graph partitioning and quadratic integer programming with psd objectives. In FOCS, pages 482–491, 2011.

[29] A. R. Karlin, C. Mathieu, and C. T. Nguyen. Integrality gaps of linear and semi-definite programming relaxations for knapsack. In IPCO, pages 301–314, 2011.

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

[31] M. Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0-1 programming. Math. Oper. Res., 28(3):470–496, 2003.

[32] M. Laurent. Sums of squares, moment matrices and optimization over polynomials. Emerging Applications of Algebraic Geometry, (149):157–270, 2009.

[33] R. Levi, A. Lodi, and M. Sviridenko. Approximation algorithms for the capacitated multi-item lot-sizing problem via flow-cover inequalities. Math. Oper. Res., 33(2):461–474, 2008.

[34] L. Lovász. On the shannon capacity of a graph. IEEE Transactions on Information Theory, 25:1–7, 1979.
A. Magen and M. Moharrami. Robust algorithms for on minor-free graphs based on the Sherali-Adams hierarchy. In APPROX-RANDOM, pages 258–271, 2009.

M. Mastrolilli. The Lasserre hierarchy in almost diagonal form. CoRR, abs/1312.6493, 2014.

J. Mestre and J. Verschae. A 4-approximation for scheduling on a single machine with general cost function. CoRR, abs/1403.0298, 2014.

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

P. Raghavendra and N. Tan. Approximating csps with global cardinality constraints using sdphierarchies. In SODA, pages 373–387, 2012.

T. Rothvoß. Directed Steiner tree and the Lasserre hierarchy. CoRR, abs/1111.5473, 2011.

G. Schoenebeck. Linear level Lasserre lower bounds for certain k-csps. In STOC, pages 593–602, 2008.

N. Shor. Class of global minimum bounds of polynomial functions. Cybernetics, 23(6):731–734, 1987.

M. Tulsiani. Csp gaps and reductions in the Lasserre hierarchy. In STOC, pages 303–312, 2009.

L. A. Wolsey. Facets for a linear inequality in 0–1 variables. Mathematical Programming, 8:168–175, 1975.

Appendix

A An equivalent formulation of the Lasserre Hierarchy

In the following we provide an equivalent formulation of the Lasserre hierarchy as a semi-infinite linear program that is more suitable for analyzing the integrality gap of the considered problem.

In optimization theory, semi-infinite programming (SIP) is an optimization problem with a finite number of variables and an infinite number of constraints, or an infinite number of variables and a finite number of constraints (see e.g. [22]). It is well-known (and easy to see) that any SDP program can be written as a semi-infinite linear program. The suggested equivalent formulation of the Lasserre hierarchy as a semi-infinite linear program is obtained after a change of variables, as explained in the following.

Change of variables. We use the generic vector $w \in \mathbb{R}^{P(N)}$ to denote either the vector $y \in \mathbb{R}^{P(N)}$ of variables, or the shifted vector $g * y$, for any $g \in \mathbb{R}^{P(N)}$. We will use the following change of variables.

Definition 3. Let $w \in \mathbb{R}^{P(N)}$. For any $I, J \in P(N)$, we define

$$w_{I,J} := \sum_{H \subseteq J} (-1)^{|H|} w_{H \cup I}$$

(15)

Let $w^N \in \mathbb{R}^{P(N)}$ be such that the $I$-th entry, with $I \subseteq N$, is equal to

$$w^N_I := w_{I,N \setminus I}$$

(16)

Variables $\{y^N_I : I \subseteq N\}$ have a nice interpretation as (pseudo)probability (see e.g. [7,8]): $y^N_I$ can be seen as the (pseudo)probability of the integral solution $\{x_i = 1 : i \in I, x_j = 0 : j \in N \setminus I\}$. At level $n$, variables $\{y^N_I : I \subseteq N\}$ are actual probabilities. (Note that $y^N_I$ can be seen as the linearization of $\prod_{i \in I} x_i \prod_{j \in N \setminus I} (1 - x_j)$, which gives equation (14) by replacing each monomial $\prod_{i \in I} x_i$ with $y_I$.)

At any level, any solution $y$ can be written as a linear combination of these (pseudo)probabilities, as the following lemma shows.
Lemma 8. \([31]\) For any \(J \subseteq N\) and \(w \in \mathbb{R}^{P(N)}\) we have

\[
w_{J} = \sum_{I \subseteq N} w_{I \cup J, N \setminus (I \cup J)} \tag{17}\]

The next lemma shows that \(z_{I, N \setminus I}\) is the value of \(g(x)\) when \(x\) is equal to \(\{x_i = 1 : i \in I, x_j = 0 : j \in N \setminus I\}\) multiplied by the corresponding (pseudo)probability \(y_{I, N \setminus I}\). We use \(g(x)\) to denote the evaluation of \(g(x)\) by setting \(x_i = 1\) if \(i \in I\), and \(x_i = 0\) if \(i \in N \setminus I\).

Lemma 9. \([31]\) For any \(g, y \in \mathbb{R}^{P(N)}\) and \(z = g \ast y\) we have

\[
z_{I, N \setminus I} = g(x) \cdot y_{I, N \setminus I} \tag{18}\]

The Lasserre Hierarchy as a SIP. For any nonnegative integer \(t\), let \(z = g \ast y\), with \(y, g \in \mathbb{R}^{P_{2t+2}(N)}\), and consider \(M_t(z)\). In the following lemma, we interpret the requirement that \(M_t(z)\) be positive semidefinite as infinitely many linear constraints of certain form.

Lemma 10. For \(t \in \mathbb{N}_0\), \(y, g \in \mathbb{R}^{P_{2t+2}(N)}\) and \(z = g \ast y\), by \((17)\) we have

\[
M_t(z) \geq 0 \iff \sum_{Q \subseteq N} y_{Q, N \setminus Q} g(x_Q) \left( \sum_{I \in P_t(N) \atop I \subseteq Q} v_I \right)^2 \geq 0 \quad \forall v \in \mathbb{R}^{P_t(N)}
\]

Proof. By definition, \(M_t(z) \geq 0 \iff v^T M_t(z) v = \sum_{I,J} v_I v_J z_{I \cup J} \geq 0 \quad \forall v \in \mathbb{R}^{P_t(N)}\). Collecting the terms with \(z_{I \cup J} = z_K\), we can write the sum as

\[
\sum_{I,J} v_I v_J z_{I \cup J} = \sum_{K \in P_t(N)} z_K \sum_{I,J \in P_t(N) \atop I \cup J = K} v_I v_J,
\]

and (by using \((17)\) substituting \(z_K = \sum_{H \subseteq N \setminus K} z_{H \cup K, N \setminus (H \cup K)}\) we get

\[
\sum_{K \in P_t(N)} \left( \sum_{H \subseteq N \setminus K} z_{H \cup K, N \setminus (H \cup K)} \right) \sum_{I,J \in P_t(N) \atop I \cup J = K} v_I v_J.
\]

Replacing \(Q = H \cup K\) this sum becomes

\[
\sum_{Q \subseteq N} z_{Q, N \setminus Q} \sum_{I,J \in P_t(N) \atop I \cup J \subseteq Q} v_I v_J = \sum_{Q \subseteq N} y_{Q, N \setminus Q} g(x_Q) \left( \sum_{I \in P_t(N) \atop I \subseteq Q} v_I \right)^2
\]

Remark 11. Given a generic multilinear polynomial of degree \(d\), \(P_d(x) = \sum_{I \in P_d(N)} (v_I \prod_{i \in I} x_i)\), let \(P_d(x_Q)\) be the evaluation of \(P_d(x)\) by setting \(x_i = 1\) if \(i \in Q\), and \(x_i = 0\) if \(i \in N \setminus Q\). Then the previous lemma establishes that Definition 2 is an equivalent formulation of the lasserre hierarchy, as \(P_d(x_Q) = \sum_{I \in P_d(N) \atop I \subseteq Q} v_I\).
Then, consider a vector $v$ the symmetry in the coefficients of $P_d(x_Q) = \sum_{I \in \mathcal{P}_d(N)} v_I$, where $v_I$ is the coefficient of the monomial $\prod_{i \in I} x_i$. Let

$$F_Q(v) = \left( \sum_{I \in \mathcal{P}_d(N), I \subseteq Q} v_I \right)^2.$$  

We observe that the function $F_Q(v)$ is convex in $v$. Now, also the following function $G(v)$ is convex, since it is composed of a nonnegative sum of convex functions (apart from the constant term corresponding to the term $Q = \emptyset$):

$$G(v) = \sum_{Q \subseteq N} u(|Q|) \left( \frac{n}{|Q|} \right)^{-1} P^2_d(x_Q) = \sum_{Q \subseteq N} u(|Q|) \left( \frac{n}{|Q|} \right)^{-1} F_Q(v)^2$$

Then, consider a vector $v$ such that $v_I \neq v_J$ for some $I, J$ such that $|I| = |J|$. By switching the entries $v_I$ and $v_J$ in the vector $v$, we obtain another vector $v'$ such that $G(v) = G(v')$ due to the symmetry in the coefficients of $G(v)$. By convexity of $G(v)$, it then holds

$$G \left( \frac{v + v'}{2} \right) \leq \frac{1}{2} G(v) + \frac{1}{2} G(v') = G(v)$$

and the entries $w_I, w_J$ in the vector $w = (v + v')/2$ are equal. Therefore, $G(v)$ attains its minimum when $v_I = v_J$ for all $I, J$ such that $|I| = |J|$.

Next we show that this means we can replace $P_d(x_Q)$ with a univariate polynomial of degree $d$ when minimizing $G(v)$. Indeed, due to the symmetry shown above, the evaluation of $P_d$ reduces to

$$P_d(x_Q) = \sum_{I \in \mathcal{P}_d(N)} v_I \prod_{i \in I} x_i = \sum_{j=0}^d v'_j \left( \sum_{i=1}^n x_i \right)^j$$

for some coefficients $v'_j$, $j = 1, ..., n$. We only briefly sketch the argument: Consider any term in $(\sum_{i=1}^n x_i)^j$ after expanding the exponentiation, say $x_{i_1}^{j_1} \cdots x_{i_q}^{j_q}$. Since each $x_i \in \{0, 1\}$, the term reduces to the monomial $x_{i_1} \cdots x_{i_q}$. We group together all the monomials with the same number of variables and note that due to symmetry, every grouped monomial with the same number of variables has the same coefficient. The resulting polynomial is a sum of monomials similar to $P_d(x_Q)$. Setting appropriately the coefficients $v'_j$ we get exactly $P_d(x_Q)$.

We note that hence $P_d(x_Q)$ depends only on the value of the sum $\sum_{i=1}^n x_i$, which can take values between 0 and $n$. This allows us to write instead of $P_d(x_Q)$ a generic univariate polynomial $p_d(k)$ of degree $d$, where $k = \sum_{i=1}^n x_i$. Therefore $G(v)$ is nonnegative if

$$\sum_{k=0}^n u(k)p^2_d(k) \geq 0$$

which proves the lemma.

**Proof of Lemma 5**

First notice that (11) (and so (13)) is equivalent to

$$\sum_{k=1}^n \Pr \left( \sum_{i=1}^n x_i = k \right) \left( k - 2 \prod_{j=1}^t \left( \frac{r_j - k}{r_j} \right) \right)^2 \geq \left( 1 + \frac{1}{n-1} \right)$$

where for the fixed $n$ the right-hand side is constant.
(a) Let \( p(t) \) be the univariate polynomial with \( 2q \) complex roots (complex roots appear in conjugate pairs) i.e. \( r_{2j-1} = a_j + b_j i \), \( r_{2j} = a_j - b_j i \) for \( j = 1, \ldots, q \) and the rest real roots. Let \( p'(t) \) be the polynomial with all real roots such that \( r'_{2j-1} = r'_{2j} = \sqrt{a_{2j}^2 + b_{2j}^2} \) 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,

\[
\sum_{k=1}^n \Pr (\sum_{i=1}^n x_i = k) (k-2) \prod_{j=1}^t \left( \frac{r_i - k}{r_j} \right)^2 \geq \sum_{k=1}^n \Pr (\sum_{i=1}^n x_i = k) (k-2) \prod_{j=1}^t \left( \frac{r'_i - k}{r'_j} \right)^2
\]

(b) Let \( p(t) \) be the univariate polynomial with all positive roots but one i.e. \( r_1 = -a \), for \( a > 0 \). Let \( p'(t) \) be the univariate polynomial with all positive roots such that \( r'_1 = a \) and \( r'_j = r_j, j > 1 \). Since for any \( k \in [n] \)

\[
\left( \frac{-a - k}{-a} \right)^2 \geq \left( \frac{a - k}{a} \right)^2
\]

and follows that,

\[
\sum_{k=1}^n \Pr (\sum_{i=1}^n x_i = k) (k-2) \prod_{j=1}^t \left( \frac{r_i - k}{r_j} \right)^2 \geq \sum_{k=1}^n \Pr (\sum_{i=1}^n x_i = k) (k-2) \prod_{j=1}^t \left( \frac{r'_i - k}{r'_j} \right)^2
\]

Now, let \( p(t) \) be the univariate polynomial with \( r_1 \in (0,1) \) and \( r_j \geq 1, \text{ for } j > 1 \). Let \( p'(t) \) be the univariate polynomial with \( r_1 = 1 \) and \( r'_j = r_j, j > 1 \).

Since for any \( k \in [n] \)

\[
\left( \frac{r_1 - k}{r_1} \right)^2 \geq \left( \frac{1 - k}{1} \right)^2
\]

and follows that,

\[
\sum_{k=1}^n \Pr (\sum_{i=1}^n x_i = k) (k-2) \prod_{j=1}^t \left( \frac{r_i - k}{r_j} \right)^2 \geq \sum_{k=1}^n \Pr (\sum_{i=1}^n x_i = k) (k-2) \prod_{j=1}^t \left( \frac{r'_i - k}{r'_j} \right)^2
\]

Next, let \( p(t) \) be the univariate polynomial with \( r_t = an \) for \( a > 1 \) and \( r_j \in [1,n], \text{ for } j \neq t \). Let \( p'(t) \) be the univariate polynomial with \( r_t = n \) and \( r'_j = r_j, j \neq t \).

Since for any \( k \in [n] \)

\[
\left( \frac{an - k}{an} \right)^2 \geq \left( \frac{n - k}{n} \right)^2
\]

and follows that,

\[
\sum_{k=1}^n \Pr (\sum_{i=1}^n x_i = k) (k-2) \prod_{j=1}^t \left( \frac{r_i - k}{r_j} \right)^2 \geq \sum_{k=1}^n \Pr (\sum_{i=1}^n x_i = k) (k-2) \prod_{j=1}^t \left( \frac{r'_i - k}{r'_j} \right)^2
\]
(c) Let $p(t)$ be the univariate polynomial with degree $s < t$ with all real roots. Let $p'(t)$ 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

$$\sum_{k=1}^{n} \Pr(\sum_{i=1}^{n} x_i = k) (k-2) \prod_{j=1}^{t} \left( \frac{r_j - k}{r'_j} \right)^2 \geq \sum_{k=1}^{n} \Pr(\sum_{i=1}^{n} x_i = k) (k-2) \prod_{j=1}^{t} \left( \frac{r'_j - k}{r'_j} \right)^2$$