

# A new approximation hierarchy for polynomial conic optimization\*

Peter J.C. Dickinson<sup>†</sup>      Janez Povh<sup>‡</sup>

February 6, 2019

## Abstract

In this paper we consider polynomial conic optimization problems, where the feasible set is defined by constraints in the form of given polynomial vectors belonging to given nonempty closed convex cones, and we assume that all the feasible solutions are non-negative. This family of problems captures in particular polynomial optimization problems (POPs), polynomial semi-definite polynomial optimization problems (PSDPs) and polynomial second-order cone-optimization problems (PSOCPs).

We propose a new general hierarchy of linear conic optimization relaxations inspired by an extension of Pólya's Positivstellensatz for homogeneous polynomials being positive over a basic semi-algebraic cone contained in the non-negative orthant, introduced in [9]. We prove that based on some classic assumptions, these relaxations converge monotonically to the optimal value of the original problem.

Adding a redundant polynomial positive semi-definite constraint to the original problem drastically improves the bounds produced by our method. We provide an extensive list of numerical examples that clearly indicate the advantages and disadvantages of our hierarchy. In particular, in comparison to the classic approach of sum-of-squares, our new method provides reasonable bounds on the optimal value for POPs, and strong bounds for PSDPs and PSOCPs, even outperforming the sum-of-squares approach in these latter two cases.

**Keywords:** *polynomial conic optimization; polynomial semi-definite programming; polynomial second-order cone programming; approximation hierarchy; linear programming; semi-definite programming*

---

\*Published as: Dickinson, P.J.C. & Povh, J. Comput Optim Appl (2019). <https://doi.org/10.1007/s10589-019-00066-0>

<sup>†</sup>University of Groningen, University of Vienna, and University of Twente, Dep. Appl. Mathematics, P.O. Box 217, 7500 AE Enschede, The Netherlands. Email: p.j.c.dickinson@utwente.nl

<sup>‡</sup>University of Ljubljana, Faculty of mechanical engineering, Aškerčeva ulica 6, 1000 Ljubljana, Slovenia. Email: janez.povh@fs.uni-lj.si

## 1 Introduction

*Polynomial optimization* problems (POPs) are those for which we want to optimize a polynomial objective function over a feasible set defined by a set of polynomial inequalities (we call such a feasible set a *basic semi-algebraic set*). Several NP-hard problems can be formulated in this way. We provide a short list of some examples of such problems: (i) optimizing a homogeneous quadratic polynomial over the non-negative orthant, which is equivalent to testing the matrix copositivity [8, 32]; (ii) 0-1 linear optimization problems [48, 49, 55]; (iii) quadratic optimization problems including, for example, the Quadratic assignment problem, the Graph partitioning problem and the MAX-CUT problem [10, 28, 40, 41, 42, 46, 45]. Polynomial optimization is also a strong tool in control, see [11]. A good overview of the possible applications of polynomial optimization can also be found in [3].

Polynomial optimization problems attract a lot of attention in theoretical and applied mathematics. Real algebraic geometry and semi-algebraic geometry are sub-fields in algebra that are strongly related to polynomial optimization problems. Since these problems are, in general, very difficult, it is a natural choice to look for tractable relaxations. These relaxations are often based on some variant of a “*Positivstellensatz*” for given semi-algebraic sets [43, 44, 47]. Many researchers have proposed hierarchies of such relaxations that are based on moment and sums-of-squares (SOS) approximations of the original problem, and give semi-definite optimization/programming (SDP) problems. Lasserre [22] proposed a hierarchy of semi-definite optimization problems that, under certain conditions, converge with their optimal values to the optimal value of the original polynomial optimization problem, see also [38, 50]. We refer the reader to a 2010 book by Lasserre [23] and a 2008 chapter by Laurent [27] for beautiful and comprehensive overviews of the results from this area.

In polynomial optimization we have inequality constraints, and a natural extension to this is to consider other conic constraints, giving us, for example, polynomial semi-definite optimization (PSDP) and polynomial second-order cone optimization (PSOCP). The feasible set in the first case is primarily defined by the constraint that a given matrix polynomial (i.e., a matrix whose entries are polynomials) is positive semi-definite. In the second case the feasible set is primarily determined by the constraint that a given polynomial vector must belong to the second-order cone.

In the papers [14, 15, 16, 20] the authors considered how to extend the SOS approach to obtain good relaxations for PSDP problems, see also [11, 19]. The hierarchy of relaxations, which contain SOS polynomials and matrix SOS polynomials, is recaptured in

## Subsection 4.2.

The PSOCs can be simply rewritten as classic POPs and later approached using tools from this area. The authors of the paper [21] proposed a general lift-and-project and reformulation-linearization technique that can also be efficiently applied to this family of problems.

If the feasible set for a POP is contained in the non-negative orthant then the existing approaches still work and the sign constraints can be modelled explicitly, i.e., by adding constraints  $x_i \geq 0$  to the set of polynomial constraints. The same idea works for PSDP and PSOCP. In [9, Theorem 2.4] a new Positivstellensatz was presented that is tailored for the non-negativity case and can be used to develop new, linear, conic optimization approximation hierarchies for POPs over the non-negative semi-algebraic sets.

The main contributions of this paper are:

- Based on [9, Theorem 2.4], we propose a new hierarchy of linear conic optimization relaxations that can be uniformly applied to general polynomial conic optimization problems (over the non-negative orthant), including the special cases listed above.
- We prove that under conditions closely related to the compactness of the feasible set and non-redundancy in its description, the optimal values of the proposed hierarchy are monotonic and asymptotically convergent to the optimal value of the original problem.
- Numerical evaluations are considered for POPs, and it is found that in these examples, our method performs poorly in comparison with the classic SOS method. However, by adding an additional (redundant) positive semi-definite constraint to the original problem, our new method then performs reasonably well in comparison to the classic SOS approach.
- We also provide numerical evaluations for PSDPs and PSOCs, which demonstrate that in these examples our new method performs very well in comparison to the classic SOS approach. This performance is further improved by adding a (redundant) positive semi-definite constraint to the original problem, which implies SOS-like semi-definite programming constraints in our hierarchy. The enhanced hierarchy often outperforms the classic SOS approach also on POP problems.

Our work is also inline with recent results from Ahmadi and Majumdar [2] and Lasserre et al. [25], who developed new approximation hierarchies that would overcome the complexity drawbacks of the existing SOS hierarchies. New linear optimization/programming (LP) and second-order cone (SOC) hierarchies are presented in [2]. In [25] the authors proposed an LP-SDP hierarchy, but with only one SDP constraint with which they can control the dimension. They also provided proof of convergence for the hierarchy and numerical results, which together with the numerical results from [30] show that these

approaches have a strong potential. A sparse version of this hierarchy is provided in [53] and can solve much larger problem instances. Another provably convergent hierarchy, called the mismatch hierarchy, is proposed in [17], which solves complex polynomial optimization problems with several thousand variables and constraints arising in electrical engineering.

Throughout the paper we will use  $\mathbb{N}$  to denote the set of non-negative integers  $\{0, 1, 2, \dots\}$  and use  $\mathbb{R}_+$  to denote the set of non-negative real numbers.

## 2 Polynomial Conic optimization

We define a polynomial conic optimization problem to be one with the following form

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{g}_i(\mathbf{x}) \in \mathcal{K}_i^* \quad \text{for all } i = 1, \dots, p, \\ & \mathbf{x} \in \mathbb{R}_+^n, \end{aligned} \tag{1}$$

where

- i.*  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is a polynomial of degree  $d_0 \geq 1$ ,
- ii.* For  $i = 1, \dots, p$ , we have that  $\mathbf{g}_i : \mathbb{R}^n \rightarrow \mathbb{R}^{m_i}$  is a polynomial vector of degree  $d_i \geq 1$ , i.e.,  $d_i$  is the maximum degree of the polynomials  $(\mathbf{g}_i(\mathbf{x}))_j$  for  $j \in \{1, \dots, m_i\}$ ,
- iii.* For  $i = 1, \dots, p$ , we have that  $\mathcal{K}_i^* \subseteq \mathbb{R}^{m_i}$  is a nonempty closed convex cone.

For  $i = 1, \dots, p$  we will in fact consider the cone  $\mathcal{K}_i^*$  to be the dual cone to a nonempty convex cone  $\mathcal{K}_i$ , i.e.,  $\mathcal{K}_i^* = \{\mathbf{u} \in \mathbb{R}^{m_i} \mid \langle \mathbf{u}, \mathbf{w} \rangle \geq 0 \text{ for all } \mathbf{w} \in \mathcal{K}_i\}$  where  $\langle \mathbf{u}, \mathbf{w} \rangle := \mathbf{u}^\top \mathbf{w} = \sum_{j=1}^{m_i} u_j w_j$ . For a nonempty closed convex cone  $\mathcal{K}_i^*$  in Euclidean space, there always exists such a cone  $\mathcal{K}_i$ , for example, the dual cone to  $\mathcal{K}_i^*$ . For the theory in this paper, the cones  $\mathcal{K}_i$  need not be closed, whereas dual cones in Euclidean space are always closed, and it is for this reason that we have denoted our cones in this unusual manner (i.e., the dual cone being in the problem). Note that the sets  $\mathcal{K}_i$  and  $\mathcal{K}_i^*$  are convex cones in Euclidean space, i.e.,

$$\mathbf{y}, \mathbf{z} \in \mathcal{K}_i, \quad \varphi, \theta \in \mathbb{R}_+ \quad \Rightarrow \quad \varphi \mathbf{y} + \theta \mathbf{z} \in \mathcal{K}_i,$$

and likewise for  $\mathcal{K}_i^*$ . They should not be confused with polynomial cones, which are often discussed in relation to polynomial optimization.

Problems such as (1) include as special cases (i) POPs over the non-negative orthant (for  $m_i = 1$  and  $\mathcal{K}_i = \mathcal{K}_i^* = \mathbb{R}_+$  for all  $i$ ), (ii) PSDPs over the non-negative orthant (for  $\mathcal{K}_i^*$  being linearly isomorphic to a positive semi-definite cone for all  $i$ ) and (iii) PSOCs

over the non-negative orthant (here we take  $\mathcal{K}_i^*$  to be linearly isomorphic to a second-order cone for all  $i$ ). This will be discussed further in Section 4, where examples of such problems will also be looked at.

Note that polynomial equality constraints can be included through the cones  $\mathcal{K}_i = \mathbb{R}$ ,  $\mathcal{K}_i^* = \{0\}$ . Also note that if we have a polynomial conic optimization problem, similar to the form (1), with a compact feasible set but without the sign constraint  $\mathbf{x} \in \mathbb{R}_+^n$ , then through a change of coordinates we can move the feasible set into the non-negative orthant and can then add the (redundant) sign constraint (see Example 4.3). For the main results in this paper we need a compactness-like assumption (Assumption 2.2), therefore having  $\mathbf{x} \in \mathbb{R}_+^n$  explicitly in the problem is not very restrictive.

For ease of notation, throughout the paper we will let  $\mathcal{F}$  be the feasible set of (1), i.e.

$$\mathcal{F} = \{\mathbf{x} \in \mathbb{R}_+^n \mid \mathbf{g}_i(\mathbf{x}) \in \mathcal{K}_i^* \text{ for } i = 1, \dots, p\}. \quad (2)$$

We will make the following two assumptions about the problem (1):

**Assumption 2.1.** *For all  $i$  we assume that  $\mathcal{K}_i$  has a nonempty interior.*

**Assumption 2.2.** *We assume that one of the constraints in the problem (1) is of the form  $R - \mathbf{a}^\top \mathbf{x} \in \mathbb{R}_+$ , where  $\mathbf{a} \in \mathbb{R}^n$  is a vector with all entries being strictly positive.*

These assumptions guarantee asymptotic convergence of the hierarchy of lower bounds introduced in the following section. Without these assumptions we would still have valid lower bounds, but we would not be able to guarantee their convergence.

The first assumption is equivalent to having  $\mathcal{K}_i^* \cap (-\mathcal{K}_i^*) = \{\mathbf{0}\}$  for all  $i$ , see, e.g., [4, Theorem 2.3]. Within the conic optimization literature, this property is referred to as  $\mathcal{K}_i^*$  being pointed; however, we shall avoid this term due to the confusion with alternative meanings of the term pointed. If we have a cone  $\mathcal{K}_i^*$  such that  $\mathcal{K}_i^* \cap (-\mathcal{K}_i^*) \neq \{\mathbf{0}\}$  then it is possible to project this constraint into a smaller subspace with the cone having the required property in this subspace<sup>1</sup>. We thus see that for simple cones Assumption 2.1 is relatively mild. However, it should be noted that for cones with a complicated definition it might be more difficult to ensure that this holds.

Also note that if the feasible set of problem (1) is bounded, then we can always add a redundant constraint to the problem of the form required in Assumption 2.2. This assumption holding means that provided  $\mathcal{F} \neq \emptyset$  then  $\mathcal{F}$  is compact and the optimal solution to problem (1) is obtained. Further discussion of these assumptions will be provided in Subsection 3.3.

---

<sup>1</sup>Considering the linear space  $\mathcal{L}_i = \mathcal{K}_i^* \cap (-\mathcal{K}_i^*) = \mathcal{K}_i^\perp$ , we have that  $\mathcal{L}_i^\perp \cap \mathcal{K}_i^* = (\mathcal{L}_i + \mathcal{K}_i)^*$  is a closed convex cone with  $(\mathcal{L}_i^\perp \cap \mathcal{K}_i^*) \cap (-(\mathcal{L}_i^\perp \cap \mathcal{K}_i^*)) = \{\mathbf{0}\}$  and  $\mathbf{g} \in \mathcal{K}_i^* \Leftrightarrow \text{Proj}_{\mathcal{L}_i^\perp}(\mathbf{g}) \in \mathcal{L}_i^\perp \cap \mathcal{K}_i^*$ .

### 3 Approximation

#### 3.1 New hierarchy

For an optimization problem  $(P)$ , we let  $\text{val}(P)$  denote its optimal value. From the definition of the infimum, we have that  $\text{val}(1)$  is equal to the optimal value of the following problem:

$$\begin{aligned} \max_{\lambda} \quad & \lambda \\ \text{s.t.} \quad & f(\mathbf{x}) - \lambda \geq 0 \quad \text{for all } \mathbf{x} \in \mathcal{F} \end{aligned} \tag{3}$$

Letting  $D = \max_i \{d_i \mid i = 0, \dots, p\}$  and  $\mathbf{e}$  be the all ones vector, we now introduce the following problems for  $r \in \mathbb{N} := \{0, 1, 2, \dots\}$ :

$$\begin{aligned} \max_{\lambda, \mathbf{y}} \quad & \lambda \\ \text{s.t.} \quad & (1 + \mathbf{e}^\top \mathbf{x})^{D-d_0+r} (f(\mathbf{x}) - \lambda) \geq_c \sum_{i=1}^p \sum_{\substack{\boldsymbol{\alpha} \in \mathbb{N}^n: \\ \mathbf{e}^\top \boldsymbol{\alpha} \leq D-d_i+r}} \mathbf{x}^\alpha \langle \mathbf{y}_{i,\boldsymbol{\alpha}}, \mathbf{g}_i(\mathbf{x}) \rangle \\ & \mathbf{y}_{i,\boldsymbol{\alpha}} \in \mathcal{K}_i \quad \text{for all } i = 1, \dots, p, \quad \boldsymbol{\alpha} \in \mathbb{N}^n : \mathbf{e}^\top \boldsymbol{\alpha} \leq D - d_i + r. \end{aligned} \tag{4_r}$$

The “ $\geq_c$ ” means that we are comparing the coefficients of the polynomials, for example,  $ax_1^3 + bx_1x_2 \geq_c cx_1^3 + dx_2^2$  iff  $a \geq c$  and  $b \geq 0$  and  $0 \geq d$  (see also Example 3.1 below). This should not be confused with the NP-hard problem of checking whether one polynomial is greater than or equal to another for all  $\mathbf{x}$ . It is, however, a simple sufficient condition for one polynomial being greater than or equal to another for all  $\mathbf{x} \in \mathbb{R}_+^n$ .

In Subsection 3.2 we will see that

$$\text{val}(4_0) \leq \text{val}(4_1) \leq \text{val}(4_2) \leq \dots \leq \lim_{r \rightarrow \infty} \text{val}(4_r) \leq \text{val}(1), \tag{5}$$

and in Subsection 3.4 we will see that if Assumptions 2.1 and 2.2 hold then we have

$$\lim_{r \rightarrow \infty} \text{val}(4_r) = \text{val}(1). \tag{6}$$

Written out explicitly, the problem  $(4_r)$  is a linear conic optimization problem over the cones  $\mathcal{K}_i$ . Therefore, if we can optimize over combinations of these cones then we can solve  $(4_r)$ . The problem  $(4_r)$  is in fact a linear conic optimization problem with

- $\binom{n+D-d_i+r}{n}$  cone constraints of being in  $\mathcal{K}_i$ , for  $i \in \{1, \dots, p\}$ ,
- $\binom{n+D+r}{n}$  inequality constraints,
- $1 + \sum_{i=1}^p m_i \binom{n+D-d_i+r}{n}$  variables.

This compares favourably with other standard approximation hierarchies for polynomial optimization. One of the advantages of our new approach is that although as  $r$  increases, the number of cones we have to optimize over increases dramatically, the cones themselves remain the same. This is in contrast to the sum-of-squares approach for polynomial matrix problems, as will be discussed in Section 4.

Another advantage of our new approach is that we are not dependent on the form of  $\mathcal{K}_i$ , we only require that it is a convex cone that we can optimize over. This increases the range of possible future applications.

Before discussing the theory behind this new approach, we will first consider an illustrative example. This example is purposely picked to be a simple example for which the convergence of our method is slow. In Section 4 we will consider more representative examples, which fully demonstrate the strength of our new method.

*Example 3.1.* Let us consider the following polynomial optimization problem:

$$\begin{aligned} \min_x \quad & x \\ \text{s.t.} \quad & x^2 = x \\ & 2x \geq 1. \end{aligned}$$

Trivially,  $x = 1$  is the only feasible point for this problem and thus the optimal value is equal to 1. This problem is in the form required with

$$\begin{aligned} f(x) = x, & & g_1(x) = x^2 - x, & & \mathcal{K}_1^* = \{0\}, & & \mathcal{K}_1 = \mathbb{R}, \\ & & g_2(x) = 2x - 1, & & \mathcal{K}_2^* = \mathbb{R}_+, & & \mathcal{K}_2 = \mathbb{R}_+. \end{aligned}$$

The  $r$ th level of our new hierarchy is given by

$$\begin{aligned} \max_{\lambda, \mathbf{y}, \mathbf{z}} \quad & \lambda \\ \text{s.t.} \quad & (1+x)^{r+1}(x-\lambda) \geq_c \sum_{i=0}^r y_i x^i (x^2 - x) + \sum_{i=0}^{r+1} z_i x^i (2x - 1) \\ & \lambda, y_0, \dots, y_r \in \mathbb{R}, \quad z_0, \dots, z_{r+1} \in \mathbb{R}_+. \end{aligned}$$

We have

$$\begin{aligned}
(1+x)^{r+1}(x-\lambda) &= x^{r+2} + \sum_{i=1}^{r+1} x^i \left( \binom{r+1}{i-1} - \lambda \binom{r+1}{i} \right) - \lambda, \\
\sum_{i=0}^r y_i x^i (x^2 - x) &= x^{r+2} y_r + \sum_{i=2}^{r+1} x^i (y_{i-2} - y_{i-1}) - x y_0, \\
\sum_{i=0}^{r+1} z_i x^i (2x - 1) &= 2x^{r+2} z_{r+1} + \sum_{i=1}^{r+1} x^i (2z_{i-1} - z_i) - z_0
\end{aligned}$$

Therefore, the  $r$ th level of the hierarchy is equivalent to the following, where after the vertical line we show which monomial term this constraint corresponds to:

$$\begin{array}{ll}
\max_{\lambda, \mathbf{y}, \mathbf{z}} & \lambda \\
\text{s.t.} & 1 \geq y_r + 2z_{r+1} \\
& \binom{r+1}{i-1} - \lambda \binom{r+1}{i} \geq y_{i-2} - y_{i-1} + 2z_{i-1} - z_i \quad \text{for all } i = 2, \dots, r+1 \\
& \binom{r+1}{0} - \lambda \binom{r+1}{1} \geq -y_0 + 2z_0 - z_1 \\
& -\lambda \geq -z_0 \\
& \lambda, y_0, \dots, y_r \in \mathbb{R}, \quad z_0, \dots, z_{r+1} \in \mathbb{R}_+.
\end{array}
\left| \begin{array}{l} x^{r+2} \\ x^i \\ x^1 \\ x^0 \end{array} \right.$$

By first eliminating the  $\mathbf{y}$  variable, this problem can be solved analytically to give an optimal value of  $1 - (1 + 2^{r+1})^{-1}$ , which tends towards 1, the optimal value of the original problem, as  $r$  tends to infinity.

### 3.2 Monotonically increasing lower bounds

In this subsection we will prove the inequality relations from (5).

**Proposition 3.2.** *We have  $\text{val}(4_r) \leq \text{val}(3) = \text{val}(1)$  for all  $r \in \mathbb{N}$ .*

*Proof.* From the definition of the infimum it is trivial to see that  $\text{val}(3) = \text{val}(1)$ . It is also trivial to see that if  $(\lambda, \mathbf{y})$  is feasible for  $(4_r)$  then  $\lambda$  is feasible for  $(3)$ .  $\square$

**Proposition 3.3.** *We have  $\text{val}(4_r) \leq \text{val}(4_{r+1})$  for all  $r \in \mathbb{N}$ .*

*Proof.* We will show that if  $(\lambda, \mathbf{y}) \in \mathbb{R} \times \left( \mathcal{K}_i^{\left| \{ \alpha \in \mathbb{N}^n \mid \mathbf{e}^\top \alpha \leq D - d_i + r \} \right|} \right)_{i \in \{1, \dots, p\}}$  is feasible for  $(4_r)$  then there exists  $\tilde{\mathbf{y}} \in \left( \mathcal{K}_i^{\left| \{ \alpha \in \mathbb{N}^n \mid \mathbf{e}^\top \alpha \leq D - d_i + r + 1 \} \right|} \right)_{i \in \{1, \dots, p\}}$  such that  $(\lambda, \tilde{\mathbf{y}})$  is feasible for  $(4_{r+1})$ .

Such a  $\tilde{\mathbf{y}}$  is constructed by letting

$$\tilde{\mathbf{y}}_{i, \alpha} = \mathbf{y}_{i, \alpha} + \sum_{\substack{j=1, \dots, n: \\ \alpha_j \geq 1}} \mathbf{y}_{i, \alpha - \mathbf{e}_j} \in \mathcal{K}_i \quad \text{for all } \begin{array}{l} i = 1, \dots, p, \\ \alpha \in \mathbb{N}^n : \mathbf{e}^\top \alpha < D - d_i + (r+1), \end{array}$$

$$\tilde{\mathbf{y}}_{i, \alpha} = \sum_{\substack{j=1, \dots, n: \\ \alpha_j \geq 1}} \mathbf{y}_{i, \alpha - \mathbf{e}_j} \in \mathcal{K}_i \quad \text{for all } \begin{array}{l} i = 1, \dots, p, \\ \alpha \in \mathbb{N}^n : \mathbf{e}^\top \alpha = D - d_i + (r+1), \end{array}$$

where  $\mathbf{e}_j \in \mathbb{R}^n$  is the unit vector with the  $j$ th entry equal to one and all the other entries equal to zero, with membership of  $\mathcal{K}_i$  following from this being a convex cone.

This construction implies that

$$\sum_{i=1}^p \sum_{\substack{\boldsymbol{\alpha} \in \mathbb{N}^n \\ \mathbf{e}^\top \boldsymbol{\alpha} \leq D-d_i+(r+1)}} \mathbf{x}^\alpha \langle \tilde{\mathbf{y}}_{i,\boldsymbol{\alpha}}, \mathbf{g}_i(\mathbf{x}) \rangle =_c (1 + \mathbf{e}^\top \mathbf{x}) \sum_{i=1}^p \sum_{\substack{\boldsymbol{\alpha} \in \mathbb{N}^n \\ \mathbf{e}^\top \boldsymbol{\alpha} \leq D-d_i+r}} \mathbf{x}^\alpha \langle \mathbf{y}_{i,\boldsymbol{\alpha}}, \mathbf{g}_i(\mathbf{x}) \rangle.$$

Having  $(\lambda, \mathbf{y})$  feasible for (4<sub>r</sub>), and  $(1 + \mathbf{e}^\top \mathbf{x})$  with all the coefficients non-negative, this implies that

$$(1 + \mathbf{e}^\top \mathbf{x}) \sum_{i=1}^p \sum_{\substack{\boldsymbol{\alpha} \in \mathbb{N}^n \\ \mathbf{e}^\top \boldsymbol{\alpha} \leq D-d_i+r}} \mathbf{x}^\alpha \langle \mathbf{y}_{i,\boldsymbol{\alpha}}, \mathbf{g}_i(\mathbf{x}) \rangle \leq_c (1 + \mathbf{e}^\top \mathbf{x})^{D-d_0+(r+1)} (f(\mathbf{x}) - \lambda). \quad \square$$

### 3.3 Analysis of the Assumptions

Before we look at the convergence of our new approximation hierarchy, we will first need to analyse the assumptions we made about problem (1). We will show that Assumptions 2.1 and 2.2 imply two properties, introduced below. The presence of these properties simplifies the workings in Subsection 3.4. We could have in fact assumed that these properties are held in place of Assumptions 2.1 and 2.2. This would have had the advantage of encapsulating a larger class of problems. However, we chose not to do this as these properties are somewhat more complicated than Assumptions 2.1 and 2.2.

Before defining these properties, we first need to introduce some new notation. For  $i \in \{1, \dots, p\}$  and  $\mathbf{w} \in \mathbb{R}^{m_i}$  we define the polynomial  $g_{i,\mathbf{w}}(\mathbf{x}) :=_c \langle \mathbf{w}, \mathbf{g}_i(\mathbf{x}) \rangle$ , and note that  $\deg(g_{i,\mathbf{w}}) \leq d_i$ . The first property is then as follows:

**Property 3.4.** *For all  $i \in \{1, \dots, p\}$  there exists  $\boldsymbol{\gamma} \in \mathcal{K}_i$  such that  $\deg(g_{i,\boldsymbol{\gamma}}) = d_i$ .*

An intuitive justification for Property 3.4 is that it ensures that none of the highest-degree terms are trivially redundant. If Property 3.4 does not hold for some  $i$  then we can remove all the degree  $d_i$  terms from  $\mathbf{g}_i$  without affecting the feasible set  $\mathcal{F}$ .

For  $i \in \{1, \dots, p\}$ , we now let  $\tilde{\mathbf{g}}_i(\mathbf{x})$  be the highest-order homogeneous part of  $\mathbf{g}_i(\mathbf{x})$ , i.e., the homogeneous polynomial vector  $\mathbf{g}_i(\mathbf{x})$  with all the terms of degree strictly less than  $d_i$  removed (and similarly for  $\tilde{f}$ ). For example, if

$$\mathbf{g}_1(\mathbf{x}) = \begin{pmatrix} x_1 x_2^2 + x_1 \\ x_2 - 1 \\ 16 - 4x_1^2 - 2x_2^3 \end{pmatrix}, \quad \text{then} \quad \tilde{\mathbf{g}}_1(\mathbf{x}) = \begin{pmatrix} x_1 x_2^2 \\ 0 \\ -2x_2^3 \end{pmatrix}.$$

We also let  $\tilde{\mathcal{F}} = \{\mathbf{x} \in \mathbb{R}_+^n \mid \tilde{\mathbf{g}}_i(\mathbf{x}) \in \mathcal{K}_i^* \text{ for } i = 1, \dots, p\}$ . The second property is then as follows:

**Property 3.5.** We have  $\tilde{f}(\mathbf{x}) > 0$  for all  $\mathbf{x} \in \tilde{\mathcal{F}} \setminus \{\mathbf{0}\}$ .

An intuitive justification for Property 3.5 is that it ensures that the problem is well behaved at infinity.

From the following two results it immediately follows that Assumptions 2.1 and 2.2 imply Properties 3.4 and 3.5, respectively.

**Proposition 3.6.** For  $i \in \{1, \dots, p\}$  let  $\widehat{\mathcal{K}}_i = \{\mathbf{w} \in \mathcal{K}_i \mid \deg(g_{i,\mathbf{w}}) = d_i\}$ , where  $\mathcal{K}_i \subseteq \mathbb{R}^{m_i}$ ,  $g_{i,\mathbf{w}} \in \mathbb{R}[\mathbf{x}]$  and  $d_i \in \mathbb{N}$  are as given earlier in the paper. Now, consider the following statements:

- i.  $\text{int } \mathcal{K}_i \neq \emptyset$ ;
- ii.  $\widehat{\mathcal{K}}_i \neq \emptyset$ ;
- iii.  $\widehat{\mathcal{K}}_i$  is a dense subset of  $\mathcal{K}_i$ ;
- iv.  $\widehat{\mathcal{K}}_i^* = \mathcal{K}_i^*$ .

We then have

$$(i) \Rightarrow (ii) \Leftrightarrow (iii) \Rightarrow (iv).$$

*Proof.* We begin by noting that  $\deg(g_{i,\mathbf{w}}) \leq d_i$  for all  $\mathbf{w} \in \mathbb{R}^{m_i}$  and  $\widehat{\mathcal{K}}_i \subseteq \mathcal{K}_i$ . We split the remainder of the proof into the following parts:

$(i) \Rightarrow (ii)$ : Consider an arbitrary  $\mathbf{w} \in \text{int } \mathcal{K}_i$ . If  $\mathbf{w} \in \widehat{\mathcal{K}}_i$  then we are done, and from now on in this part of the proof we will assume that  $\mathbf{w} \notin \widehat{\mathcal{K}}_i$ , i.e.  $\deg(g_{i,\mathbf{w}}) < d_i$ .

There exists  $j \in \{1, \dots, m_i\}$  such that  $d_i = \deg((\mathbf{g}_i(\mathbf{x}))_j) = \deg(g_{i,\mathbf{e}_j})$ , where  $\mathbf{e}_j \in \mathbb{R}^{m_i}$  is the unit vector with the  $j$ th entry equal to one and all the other entries equal to zero. For all  $\varepsilon \neq 0$  we then have  $\deg(g_{i,\mathbf{w}+\varepsilon\mathbf{e}_j}) = \deg(g_{i,\mathbf{w}} + \varepsilon g_{i,\mathbf{e}_j}) = d_i$ . The proof of this part is then completed by noting that as  $\mathbf{w} \in \text{int } \mathcal{K}_i$  there exists  $\varepsilon \neq 0$  such that  $\mathbf{w} + \varepsilon\mathbf{e}_j \in \mathcal{K}_i$ .

$(ii) \Rightarrow (iii)$ : Suppose that  $\exists \boldsymbol{\gamma} \in \widehat{\mathcal{K}}_i$  and consider an arbitrary  $\mathbf{w} \in \mathcal{K}_i \setminus \widehat{\mathcal{K}}_i$ . Then  $\deg(g_{i,\boldsymbol{\gamma}}) = d_i > \deg(g_{i,\mathbf{w}})$ , and thus for all  $\varepsilon \neq 0$  we have

$$\deg(g_{i,\mathbf{w}+\varepsilon\boldsymbol{\gamma}}) = \deg(g_{i,\mathbf{w}} + \varepsilon g_{i,\boldsymbol{\gamma}}) = d_i.$$

The proof of this part is then completed by noting that as  $\mathcal{K}_i$  is a convex cone we have  $\mathbf{w} + \varepsilon\boldsymbol{\gamma} \in \mathcal{K}_i$  for all  $\varepsilon > 0$ .

$(iii) \Leftarrow (ii)$ : This trivially follows from the fact that  $\mathcal{K}_i \neq \emptyset$ .

(iii)  $\Rightarrow$  (iv): If  $\widehat{\mathcal{K}}_i$  is a dense subset of  $\mathcal{K}_i$  then  $\text{cl } \widehat{\mathcal{K}}_i = \text{cl } \mathcal{K}_i$  and from a well-known result for sets in Euclidean space we have  $\mathcal{K}_i^* = (\text{cl } \mathcal{K}_i)^* = (\text{cl } \widehat{\mathcal{K}}_i)^* = \widehat{\mathcal{K}}_i^*$ , e.g., [4, Theorem 2.1].  $\square$

**Proposition 3.7.** *If Assumption 2.2 holds for the problem (1) then  $\widetilde{\mathcal{F}} = \{\mathbf{0}\}$ .*

*Proof.* First, note that  $\mathbf{0} \in \widetilde{\mathcal{F}}$  follows from  $\widetilde{\mathbf{g}}_i(\mathbf{x})$  being homogeneous and  $\mathcal{K}_i^*$  being a nonempty closed cone for all  $i$ . We then complete the proof by noting that if Assumption 2.2 holds for the problem (1) then

$$\widetilde{\mathcal{F}} = \{\mathbf{x} \in \mathbb{R}_+^n \mid \widetilde{\mathbf{g}}_i(\mathbf{x}) \in \mathcal{K}_i^* \text{ for } i = 1, \dots, p\} \subseteq \{\mathbf{x} \in \mathbb{R}_+^n \mid -\mathbf{a}^\top \mathbf{x} \in \mathbb{R}_+\} = \{\mathbf{0}\}. \quad \square$$

### 3.4 Asymptotic Convergence

In this subsection we look at the theory behind how our new hierarchy converges. In Section 4 we will then look at some numerical examples to see how it performs in practice.

In order to prove the convergence we need to begin by considering the following Positivstellensatz:

**Theorem 3.8.** [9, Theorems 2.4 and 4.1] *Consider the homogeneous polynomials  $f_i : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  for  $i \in \{0\} \cup \mathcal{I}$  such that  $f_0(\mathbf{x}) > 0$  for all  $\mathbf{x} \in \mathbb{R}_+^{n+1} \setminus \{\mathbf{0}\} : f_i(\mathbf{x}) \geq 0 \forall i \in \mathcal{I}$ . Then for some  $R \in \mathbb{N}$  there exists a subset  $\mathcal{J} \subseteq \mathcal{I}$  of finite cardinality and homogeneous polynomials  $q_j : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  for  $j \in \mathcal{J}$ , with all of their coefficients non-negative and  $\deg(q_j) + \deg(f_j) = R + \deg(f_0)$ , such that  $(\mathbf{e}^\top \mathbf{x})^R f_0(\mathbf{x}) \geq_c \sum_{j \in \mathcal{J}} f_j(\mathbf{x}) q_j(\mathbf{x})$ .*

The restriction on the degrees was not in the original theorems, but can be added as all of the polynomials are homogeneous. There are no limitations on the set  $\mathcal{I}$ , it could in fact index uncountably many polynomials  $f_i$ .

Before we can use this Positivstellensatz, we first need to adapt it to apply for non-homogeneous polynomials. In order to do this we use the notation from Property 3.5, i.e.,  $\widetilde{g}(\mathbf{x})$  is the polynomial  $g(\mathbf{x})$  with all terms of degree strictly less than  $\deg(g)$  removed. We then obtain the following Positivstellensatz:

**Theorem 3.9.** *Consider polynomials  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$  for  $i \in \{0\} \cup \mathcal{I}$  such that*

- i.  $f_0(\mathbf{x}) > 0$  for all  $\mathbf{x} \in \mathbb{R}_+^n : f_i(\mathbf{x}) \geq 0 \forall i \in \mathcal{I}$ , and
- ii.  $\widetilde{f}_0(\mathbf{x}) > 0$  for all  $\mathbf{x} \in \mathbb{R}_+^n \setminus \{\mathbf{0}\} : \widetilde{f}_i(\mathbf{x}) \geq 0 \forall i \in \mathcal{I}$ .

*Then for some  $R \in \mathbb{N}$  there exists a subset  $\mathcal{J} \subseteq \mathcal{I}$  of finite cardinality and polynomials  $q_j : \mathbb{R}^n \rightarrow \mathbb{R}$  for  $j \in \mathcal{J}$ , with all of their coefficients non-negative and  $\deg(q_j) + \deg(f_j) \leq R + \deg(f_0)$ , such that  $(1 + \mathbf{e}^\top \mathbf{x})^R f_0(\mathbf{x}) \geq_c \sum_{j \in \mathcal{J}} f_j(\mathbf{x}) q_j(\mathbf{x})$ .*

*Proof.* For a polynomial  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  of degree  $d$ , we define  $g^H(\mathbf{x}, x_{n+1}) : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  to be the unique homogeneous polynomial of degree  $d$  such that  $g(\mathbf{x}) = g^H(\mathbf{x}, 1)$ . We then have  $\tilde{g}(\mathbf{x}) = g^H(\mathbf{x}, 0)$ , e.g., for  $g(x_1, x_2) = x_1x_2^2 + x_1$  we have

$$\begin{aligned} g^H(x_1, x_2, x_3) &= x_1x_2^2 + x_1x_3^2, \\ g^H(x_1, x_2, 1) &= x_1x_2^2 + x_1 = g(x_1, x_2), \\ g^H(x_1, x_2, 0) &= x_1x_2^2 = \tilde{g}(x_1, x_2). \end{aligned}$$

Using this notation, for  $f_0$  given in the theorem we have

- i.  $f_0^H(\mathbf{x}, 1) > 0$  for all  $\mathbf{x} \in \mathbb{R}_+^n : f_i^H(\mathbf{x}, 1) \geq 0 \forall i \in \mathcal{I}$ , and
- ii.  $f_0^H(\mathbf{x}, 0) > 0$  for all  $\mathbf{x} \in \mathbb{R}_+^n \setminus \{\mathbf{0}\} : f_i^H(\mathbf{x}, 0) \geq 0 \forall i \in \mathcal{I}$ .

As all the polynomials are homogeneous, this is equivalent to

$$f_0^H(\mathbf{x}, x_{n+1}) > 0 \text{ for all } (\mathbf{x}, x_{n+1}) \in \mathbb{R}_+^{n+1} \setminus \{(\mathbf{0}, 0)\} : f_i^H(\mathbf{x}, x_{n+1}) \geq 0 \forall i \in \mathcal{I}.$$

From Theorem 3.8 we find that for some  $R \in \mathbb{N}$  there exists a subset  $\mathcal{J} \subseteq \mathcal{I}$  of finite cardinality and homogeneous polynomials  $q_j : (\mathbb{R}^n \times \mathbb{R}) \rightarrow \mathbb{R}$  for  $j \in \mathcal{J}$ , with all of their coefficients non-negative and  $\deg(q_j) + \deg(f_j) = R + \deg(f_0)$ , such that  $(\mathbf{e}^\top \mathbf{x} + x_{n+1})^R f_0^H(\mathbf{x}, x_{n+1}) \geq_c \sum_{j \in \mathcal{J}} f_j^H(\mathbf{x}, x_{n+1}) q_j(\mathbf{x}, x_{n+1})$ . Letting  $x_{n+1} = 1$ , we then obtain the required result.  $\square$

We are now ready to apply this result for our case:

**Theorem 3.10.** *Let  $\mathcal{F}$  be as given in (2) with Properties 3.4 and 3.5 holding, and consider a polynomial  $f$  of degree  $d_0$  and  $\lambda \in \mathbb{R}$  such that  $f(\mathbf{x}) - \lambda > 0$  for all  $\mathbf{x} \in \mathcal{F}$ . Then there exists  $r \in \mathbb{N}$  and  $\mathbf{y}_{i,\alpha} \in \mathcal{K}_i$  for  $i = 1, \dots, p$ ,  $\alpha \in \mathbb{N}^n : \mathbf{e}^\top \alpha \leq D - d_i + r$  such that*

$$(1 + \mathbf{e}^\top \mathbf{x})^{D-d_0+r} (f(\mathbf{x}) - \lambda) \geq_c \sum_{i=1}^p \sum_{\substack{\alpha \in \mathbb{N}^n \\ \mathbf{e}^\top \alpha \leq D-d_i+r}} \mathbf{x}^\alpha \langle \mathbf{y}_{i,\alpha}, \mathbf{g}_i(\mathbf{x}) \rangle$$

*Proof.* In order to apply Theorem 3.9 we let  $f_0(\mathbf{x}) = f(\mathbf{x}) - \lambda$  and  $f_i \in \mathbb{R}[\mathbf{x}]$  for  $i$  in some set  $\mathcal{I}$  be such that

$$\{f_i \mid i \in \mathcal{I}\} = \{g_{i,\mathbf{w}} : i = 1, \dots, p, \quad \mathbf{w} \in \widehat{\mathcal{K}}_i\},$$

where  $\widehat{\mathcal{K}}_i$  is as given in Proposition 3.6 and recall from this proposition that Property 3.4

implies that  $\widehat{\mathcal{K}}_i^* = \mathcal{K}_i^*$ . We then have

$$\begin{aligned}\widetilde{f}_0(\mathbf{x}) &= \widetilde{f}(\mathbf{x}), \\ \{\mathbf{x} \in \mathbb{R}_+^n \mid f_i(\mathbf{x}) \geq 0 \forall i \in \mathcal{I}\} &= \{\mathbf{x} \in \mathbb{R}_+^n \mid \langle \mathbf{w}, \mathbf{g}_i(\mathbf{x}) \rangle \geq 0 \forall i = 1, \dots, p, \mathbf{w} \in \widehat{\mathcal{K}}_i\} \\ &= \{\mathbf{x} \in \mathbb{R}_+^n \mid \mathbf{g}_i(\mathbf{x}) \in \widehat{\mathcal{K}}_i^* \forall i = 1, \dots, p\} \\ &= \{\mathbf{x} \in \mathbb{R}_+^n \mid \mathbf{g}_i(\mathbf{x}) \in \mathcal{K}_i^* \forall i = 1, \dots, p\} = \mathcal{F}, \\ \{\mathbf{x} \in \mathbb{R}_+^n \mid \widetilde{f}_i(\mathbf{x}) \geq 0 \forall i \in \mathcal{I}\} &= \{\mathbf{x} \in \mathbb{R}_+^n \mid \langle \mathbf{w}, \widetilde{\mathbf{g}}_i(\mathbf{x}) \rangle \geq 0 \forall i = 1, \dots, p, \mathbf{w} \in \widehat{\mathcal{K}}_i\} \\ &= \{\mathbf{x} \in \mathbb{R}_+^n \mid \widetilde{\mathbf{g}}_i(\mathbf{x}) \in \widehat{\mathcal{K}}_i^* \forall i = 1, \dots, p\} \\ &= \{\mathbf{x} \in \mathbb{R}_+^n \mid \widetilde{\mathbf{g}}_i(\mathbf{x}) \in \mathcal{K}_i^* \forall i = 1, \dots, p\} = \widetilde{\mathcal{F}}.\end{aligned}$$

This, in addition to Property 3.5 holding, implies that the requirements of  $f_0$  for Theorem 3.9 hold, and thus for some  $R \in \mathbb{N}$  there exists a finite set

$$\mathcal{J} \subseteq \{(i, \mathbf{w}) \mid i \in \{1, \dots, p\}, \mathbf{w} \in \widehat{\mathcal{K}}_i\}$$

and polynomials  $q_{i, \mathbf{w}} : \mathbb{R}^n \rightarrow \mathbb{R}$  for  $(i, \mathbf{w}) \in \mathcal{J}$ , with non-negative coefficients and  $\deg(q_{i, \mathbf{w}}) + d_i \leq R + d_0$ , such that

$$(1 + \mathbf{e}^\top \mathbf{x})^R (f(\mathbf{x}) - \lambda) \geq_c \sum_{(i, \mathbf{w}) \in \mathcal{J}} \langle \mathbf{w}, \mathbf{g}_i(\mathbf{x}) \rangle q_{i, \mathbf{w}}(\mathbf{x}).$$

We can write  $q_{i, \mathbf{w}}$  in the form

$$q_{i, \mathbf{w}} = \sum_{\substack{\boldsymbol{\alpha} \in \mathbb{N}^n: \\ \mathbf{e}^\top \boldsymbol{\alpha} \leq R + d_0 - d_i}} q_{i, \mathbf{w}, \boldsymbol{\alpha}} \mathbf{x}^\boldsymbol{\alpha},$$

where  $q_{i, \mathbf{w}, \boldsymbol{\alpha}}$  are non-negative scalars for all  $i, \mathbf{w}, \boldsymbol{\alpha}$ . Using this notation we have

$$\begin{aligned}(1 + \mathbf{e}^\top \mathbf{x})^R (f(\mathbf{x}) - \lambda) &\geq_c \sum_{i=1}^p \sum_{\substack{\boldsymbol{\alpha} \in \mathbb{N}^n: \\ \mathbf{e}^\top \boldsymbol{\alpha} \leq R + d_0 - d_i}} \mathbf{x}^\boldsymbol{\alpha} \langle \mathbf{y}_{i, \boldsymbol{\alpha}}, \mathbf{g}_i(\mathbf{x}) \rangle, \\ \text{where } \mathbf{y}_{i, \boldsymbol{\alpha}} &= \sum_{\substack{\mathbf{w} \in \mathcal{K}_i: \\ (i, \mathbf{w}) \in \mathcal{J}}} q_{i, \mathbf{w}, \boldsymbol{\alpha}} \mathbf{w} \in \mathcal{K}_i.\end{aligned}$$

Now letting  $r = \max\{0, R + d_0 - D\}$  and considering Proposition 3.3 we obtain the required result.  $\square$

*Remark 3.11.* Suppose that the polynomial  $f$  and the polynomial vectors  $\mathbf{g}_i$  are in fact homogeneous, i.e.,  $(\mathbf{g}_i(\mathbf{x}))_j$  is a homogeneous polynomial of degree  $d_i$  for all  $j = 1, \dots, m_i$ . Then from Theorem 3.8 (using the techniques from Theorem 3.10), we find that if

$f(\mathbf{x}) > 0$  for all  $\mathbf{x} \in \mathcal{F} \setminus \{\mathbf{0}\}$ , then there exist  $r \in \mathbb{N}$  and  $\mathbf{y}_{i,\alpha} \in \mathcal{K}_i$  for  $i = 1, \dots, p$ ,  $\alpha \in \mathbb{N} : \mathbf{e}^\top \alpha = D - d_i + r$  such that

$$(\mathbf{e}^\top \mathbf{x})^{D-d_0+r} f(\mathbf{x}) \geq_c \sum_{i=1}^p \sum_{\substack{\alpha \in \mathbb{N}^n: \\ \mathbf{e}^\top \alpha = D-d_i+r}} \mathbf{x}^\alpha \langle \mathbf{y}_{i,\alpha}, \mathbf{g}_i(\mathbf{x}) \rangle \quad (7)$$

### 3.5 Enhancing the convergence

The main characteristic of our hierarchy, the conic linearity, has an affect on the tightness of the bounds it provides. The bounds improve if we increase  $r$ , but sometimes we would like to improve these bounds already for the beginning members of the hierarchy. One way to make such improvements and stay tractable is to add to the original problem the redundant constraint

$$\begin{pmatrix} 1 & \mathbf{x}^\top \\ \mathbf{x} & \mathbf{x}\mathbf{x}^\top \end{pmatrix} \in \mathcal{PSD}^{n+1}, \quad (8)$$

and then apply our hierarchy, where  $\mathcal{PSD}^{n+1}$  denotes the set of positive semi-definite matrices of the order  $n + 1$ . This gives us (extra) positive semi-definite variables in the hierarchy, but these positive semi-definite variables would be limited to having the order  $n + 1$ , unlike in the classic SOS approach, where the order of the positive semi-definite constraints quickly grows to an intractable level. This approach is closely related to that taken by Peña, Vera and Zuluaga [39], which has been demonstrated to be very effective, and it was its effectiveness in this paper that provided the inspiration for us to consider it. It is also aligned with a recent hierarchy of Lasserre et al. [25], where each member of the hierarchy is an SDP with a bounded size and some linear constraints. Considered on the dual side, this approach is closely related to considering truncated moment matrices, but with subtle differences caused by the multiplication with  $\mathbf{x}^\alpha$  (see problem (12)).

We tested this approach on several examples in Section 4, and the numerical results show that it improves the bounds significantly.

Another approach to improve our hierarchy would be to add optimality conditions to the original problem, similar to was previously done for Lasserre's hierarchy [7, 34, 37]. If the cones  $\mathcal{K}_i^*$  from (1) are semi-algebraic, then we can write out the polynomial inequalities for the cones explicitly, and then apply the SOS or linear optimization approach from this subsection. However, this could result in a lot of terms with a very high degree, and thus large semi-definite or linear optimization instances. We have not tested this approach.

### 3.6 Further discussion on convergence

One advantage of the moment and sums-of-squares-based hierarchies is the occurrence of a finite convergence [12, 13, 26, 35]. Recently Nie has provided in [33, 36] an extensive and clear study of the convergence of Lasserre's hierarchy. He showed that under some

generic assumptions, the so-called flat truncation is a necessary and sufficient condition for finite convergence. A similar property also holds for non-commutative polynomials [5, 18].

We do not yet have results of this type for our hierarchy - neither for the original nor for the enhanced version. Actually, we hope that this paper will motivate other researchers to pursue this direction.

We know that for our new hierarchy (without enhancement) we would not generally expect to get finite convergence, and in Example 3.1 we saw an example with an asymptotic but not finite convergence<sup>2</sup>. If  $\mathcal{K}_1, \dots, \mathcal{K}_m$  are all polyhedral cones, then the optimization problems in our hierarchy would all be linear ones and thus we cannot expect them to encapsulate the nonlinearity in the problem. However, after adding (8) the resulting hierarchy converges in practice much better and we believe a variant of Nie’s results can be obtained.

While the convergence results for Lasserre’s hierarchy are beautiful results, they are limited by the fact that in practice the finite convergence can be at a level of the hierarchy that is intractable. The main advantage of our new hierarchy is that much higher levels of the hierarchy remain tractable, as will be seen in Section 4.

## 4 Applications of the new hierarchy

In this section we will consider generalized classes of polynomial conic optimization problems from the literature, and consider how the application of our new approach to these problems compares with previous approaches. In particular we consider:

- POP over the non-negative orthant, where we found that our hierarchy in its standard form performs poorly compared to state-of-the-art methods, but performs comparatively well when we added the redundant SDP constraint (8) to the original problem before applying our hierarchy.
- PSDP over the non-negative orthant, where we found that our hierarchy (in its standard form) performs comparatively well in comparison to state-of-the-art methods, and performs even better when adding the redundant SDP constraint (8) to the original problem.
- PSOCP over the non-negative orthant, where we found that our hierarchy outperforms state-of-the-art methods in its standard form.

---

<sup>2</sup>This example does not in fact satisfy Assumption 2.2, but we can add the extra redundant constraint “ $3 - 2x \in \mathbb{R}_+$ ” to the original problem to satisfy this assumption without changing the optimal values in our hierarchy.

For all the examples in this section, for ease of implementation we used YALMIP [29] within MATLAB [1], which can handle polynomials directly. This converted the optimization problems from our hierarchy (and state-of-the-art hierarchies) into LP, SOC and SDP optimization problems, which were then in almost all cases solved by SeDuMi [52]. We also used MOSEK [31] to compute the values for (11) in Table 2 and the SOS bounds in Table 9.

Although using YALMIP made the implementations easier, it meant that most of the computation times were spent in constructing the problems, rather than in solving them. These total computation times can be reduced through better, more direct implementations, and thus are not representative measures of the performance. For this reason no computation times have been included. Instead, we consider whether the optimization problems can be solved, or if there are memory problems, along with comparing the size of these problems.

All the computations were carried out on a laptop with two 2.30-GHz CPU cores and 4.0 GB of RAM. The code for these examples is available from the authors on request.

## 4.1 Constrained polynomial optimization over the non-negative orthant

In this subsection we will apply our method to POPs over the non-negative orthant. We will see that our method performs poorly in comparison to the classic SOS approach. If, however, we add the constraint (8) to the original problem before applying our method then the performance is drastically improved, even occasionally outperforming the classic SOS approach. We are, however, of the opinion that the classic SOS approach is currently still superior to our new method for these problems and instead this subsection should be thought of as an illustration of how our new method works, in preparation for the subsequent subsections where we apply our (enhanced) method to the more complicated cases of PSDPs and PSOCPs.

A polynomial  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is defined to be a sum-of-squares (SOS) polynomial if there exists a polynomial vector  $\mathbf{h} : \mathbb{R}^n \rightarrow \mathbb{R}^q$  such that  $f(\mathbf{x}) =_c \mathbf{h}(\mathbf{x})^\top \mathbf{h}(\mathbf{x}) =_c \sum_{i=1}^q h_i(\mathbf{x})^2$ , and we then trivially have that  $f(\mathbf{x}) \geq 0$  for all  $\mathbf{x} \in \mathbb{R}^n$ .

Letting  $\mathbf{v}_d(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^{\binom{n+d}{n}}$  be such that  $\mathbf{v}_d(\mathbf{x}) =_c (\mathbf{x}^\alpha)_{\alpha \in \mathbb{N}, \mathbf{e}^\top \alpha \leq d}$ , it is well known that a polynomial  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  of degree at most  $2d$  is SOS if and only if there exists an order  $\binom{n+d}{n}$  positive semi-definite matrix  $A$ , denoted  $A \in \mathcal{PSD}^{\binom{n+d}{n}}$ , such that  $f(\mathbf{x}) =_c \mathbf{v}_d(\mathbf{x})^\top A \mathbf{v}_d(\mathbf{x})$ , and there is a linear relationship between the elements of  $A$  and the coefficients of  $f$ . This allows us to optimize over the set of SOS polynomials using positive semi-definite optimization [23, 27].

Let us consider a constrained polynomial optimization problem over the non-negative

orthant:

$$\begin{aligned}
& \min_{\mathbf{x}} && f(\mathbf{x}) \\
& \text{s.t.} && g_i(\mathbf{x}) \geq 0 \quad \text{for all } i = 1, \dots, p, \\
& && x_i \geq 0 \quad \text{for all } i = 1, \dots, n.
\end{aligned} \tag{9}$$

Note that this is a special instance of (1) with  $\mathcal{K}_i^* = \mathbb{R}_+$  for all  $i$ .

If the sign constraints  $x_i \geq 0$  are considered as the usual polynomial inequalities then (9) is a special instance of a classic constrained polynomial optimization problem, formulated in [23, problem (5.2)].

Due to the hardness of (9), several approximation hierarchies are proposed in the literature. Probably one of the most well-known and tight is the SOS hierarchy, which consists of the following approximation problems (for each  $r \in \mathbb{N}$ ):

$$\begin{aligned}
& \max_{\lambda, \{a_i\}} && \lambda \\
& \text{s.t.} && f(\mathbf{x}) - \lambda =_c a_0(\mathbf{x}) + \sum_{i=1}^n x_i a_i(\mathbf{x}) + \sum_{i=1}^p g_i(\mathbf{x}) a_{n+i}(\mathbf{x}) \\
& && a_0(\mathbf{x}) \text{ is SOS of degree at most } 2\lfloor \frac{1}{2}(D+r) \rfloor \\
& && a_i(\mathbf{x}) \text{ is SOS of degree at most } 2\lfloor \frac{1}{2}(D+r-1) \rfloor \text{ for } i = 1, \dots, n \\
& && a_{n+i}(\mathbf{x}) \text{ is SOS of degree at most } 2\lfloor \frac{1}{2}(D+r-d_i) \rfloor \text{ for all } i = 1, \dots, p.
\end{aligned} \tag{10}$$

In practice this method provides very good approximations. If we know some constant  $N$  such that the feasible set for (9) is contained in the Euclidean ball centred at the origin with the radius  $N$  (note that this is essentially our Assumption 2.2) then we can add this ball constraint to the problem (9) and then the solutions of the dual problems (10) converge to the optimal value of (9), when  $r \rightarrow \infty$  (see [22, Theorem 4.2], [23, Theorem 5.6]). However, as  $r$  (and/or  $D$ ) increase, the hierarchy members contain increasingly larger positive semi-definite constraints that are more and more difficult to handle. The problem from the hierarchy corresponding to fixed  $r$  has:

- One positive semi-definite constraint of the order  $\binom{n+\lfloor \frac{1}{2}(D+r) \rfloor}{n}$ ,
- $n$  positive semi-definite constraints of the order  $\binom{n+\lfloor \frac{1}{2}(D+r-1) \rfloor}{n}$ ,
- a positive semi-definite constraint of the order  $\binom{n+\lfloor \frac{1}{2}(D+r-d_i) \rfloor}{n}$  for each  $i \in \{1, \dots, p\}$ ,
- $\binom{n+D+r}{n}$  equality constraints,
- $1 + \frac{1}{2} \binom{n+\lfloor \frac{1}{2}(D+r) \rfloor}{n} \left[ \binom{n+\lfloor \frac{1}{2}(D+r) \rfloor}{n} + 1 \right] + \frac{1}{2} n \binom{n+\lfloor \frac{1}{2}(D+r-1) \rfloor}{n} \left[ \binom{n+\lfloor \frac{1}{2}(D+r-1) \rfloor}{n} + 1 \right] + \sum_{i=1}^p \frac{1}{2} \binom{n+\lfloor \frac{1}{2}(D+r-d_i) \rfloor}{n} \left[ \binom{n+\lfloor \frac{1}{2}(D+r-d_i) \rfloor}{n} + 1 \right]$  variables.

Our hierarchy from Section 3 applied to (9) can be simplified to

$$\begin{aligned}
& \max_{\lambda, \{y_{i,\alpha}\}} \lambda \\
& \text{s.t.} \quad (1 + \mathbf{e}^\top \mathbf{x})^{D-d_0+r} (f(\mathbf{x}) - \lambda) \geq_c \sum_{i=1}^p g_i(\mathbf{x}) \sum_{\substack{\alpha \in \mathbb{N}^n: \\ \mathbf{e}^\top \alpha \leq D-d_i+r}} y_{i,\alpha} \mathbf{x}^\alpha \\
& \quad y_{i,\alpha} \geq 0 \quad \text{for all } i, \alpha
\end{aligned} \tag{11}$$

This hierarchy is usually weaker, but on the other hand contains problems that are easier to solve. It contains

- $\binom{n+D+r}{n} + \sum_{i=1}^p \binom{n+D-d_i+r}{n}$  inequality constraints,
- $1 + \sum_{i=1}^p \binom{n+D-d_i+r}{n}$  variables.

By adding the redundant PSD constraint (8) to the original problem and then applying our hierarchy to give the following problem adds additional  $\binom{n+D+r-2}{n}$  PSD constraints of the order  $n+1$  and an additional  $\frac{1}{2} \left[ \binom{n+D+r-2}{n} \right] \left[ \binom{n+D+r-2}{n} + 1 \right]$  variables:

$$\begin{aligned}
& \max_{\lambda, \{y_{i,\alpha}\}, \{Z_\beta\}} \lambda \\
& \text{s.t.} \quad (1 + \mathbf{e}^\top \mathbf{x})^{D-d_0+r} (f(\mathbf{x}) - \lambda) \geq_c \sum_{i=1}^p g_i(\mathbf{x}) \sum_{\substack{\alpha \in \mathbb{N}^n: \\ \mathbf{e}^\top \alpha \leq D-d_i+r}} y_{i,\alpha} \mathbf{x}^\alpha + \dots \\
& \quad \dots + \sum_{\substack{\beta \in \mathbb{N}^n: \\ \mathbf{e}^\top \beta \leq D+r-2}} \mathbf{x}^\beta \mathbf{v}_1(\mathbf{x})^\top Z_\beta \mathbf{v}_1(\mathbf{x}) \\
& \quad y_{i,\alpha} \geq 0 \quad \text{for all } i, \alpha \\
& \quad Z_\beta \in \mathcal{PSD}^{n+1} \quad \text{for all } \beta
\end{aligned} \tag{12}$$

*Remark 4.1.* Suppose we have  $d_0 = d_1 = \dots = d_p = 2$ . For  $r = 0$  the SOS relaxation (10) is equivalent to the enhanced relaxation (12), since every feasible solution for the first problem yields a feasible solution with the same objective value for the second problem and vice versa.

In Table 1 we consider the explicit problem sizes that this would give for an instance with  $r = 4$ ,  $n = d_0 = 8$ ,  $p = d_1 = d_2 = 2$ . As we see, the SOS problem, (10), is on the boundary or even beyond the reach of state-of-the-art SDP solvers that are based on the interior point methods, but are still solvable with some other SDP solvers, like SDPNAL [57], while our new LP approximation (11) is solvable with most LP solvers. The new SDP hierarchy (12) contains a lot of SDP variables of small size and is solvable with SDP solvers that enable parallelization, see ,e.g., [51, 56].

Hierarchy	SDP constraints		Linear constraints		Number of variables
	Order	Number	Type	Number	
SOS hierarchy, (10)	3003	1	=	125,970	12,798,787
	1,287	10			
New LP hierarchy, (11)	-	-	$\geq$	213,486	87,517
New SDP hierarchy, (12)	9	43,758	$\geq$	213,486	2,056,627

**Table 1:** Considering a polynomial optimization problem with  $n = d_0 = 8$ ,  $p = d_1 = d_2 = 2$  the table above gives the size of the approximation problems (10), (11) and (12) for  $r = 4$ .

*Example 4.2.* In the paper [25], in particular instances C4\_2, C4\_4, C4\_6, C6\_2, C6\_4, C6\_6, C8\_2, C8\_4, C10\_4, C20\_2, the authors considered some polynomial optimization problems of the form (9) such that for all feasible points of this problem and all  $i \in \{1, \dots, m\}$  we have  $g_i(\mathbf{x}) \leq 1$ .<sup>3</sup> For these problems the authors compared the standard SOS approach, a bounded SOS approach [24] and an LP approach based on Krivine-Stengle’s Positivstellensatz to compute the lower bounds on these problems.

For these instances we computed the lower bounds provided by our new LP hierarchy, (11), and SDP hierarchy, (12). The new LP hierarchy was computed for all  $r$  for which we did not get “out of memory” warnings, while the new SDP hierarchy was computed for  $r = 1$ . The results for this are in Table 2.

As we can see, our new LP approximation, (11), does not perform very well in comparison to the other methods, but our new SDP approximation, (12), does perform extremely well, always giving the tightest lower bound of all the approximations in the instances considered. It is even strictly better than all of the other bounds for instances C4\_6 and C6\_6.

In order to compare the sizes of these problems, in Table 3 we consider the size of the hierarchies for the case C6\_6. For  $r = 1$  we see that not only does our new SDP hierarchy, (12), provide a better lower bound than the (bounded) SOS approach, but the size of the problem, in terms of the order of the largest PSD constraints, is also significantly smaller.

## 4.2 Polynomial semi-definite optimization over the non-negative orthant

In this subsection we will apply our method to PSDPs over the non-negative orthant. We will see that our method performs very well on these problems, especially after adding the redundant constraint (8) to the original problem, but the method does not suffer as severely from memory problems as the classic SOS approach.

<sup>3</sup>Note that for compact optimization problems we can always normalize the constraints so that this inequality holds; however, doing this involves computing a good upper bound of  $g_i$  on the feasibility set, which is itself a demanding task.

Instance	$n$	Problem (11)		val (12), $r = 1$	KS-bound	(B)SOS bound	Optimum
		$\underline{r}$ and $\bar{r}$	val (11)				
C4_2	4	$\underline{r} = 3$ $\bar{r} = 11$	-0.900 -0.299	<b>-0.250</b>	-0.310	<b>-0.250</b>	-0.250
C4_4	4	$\underline{r} = 8$ $\bar{r} = 9$	-3.90 -2.12	<b>-0.696</b>	-0.88542	<b>-0.696</b>	-0.696
C4_6	4	$\underline{r} = 4$ $\bar{r} = 8$	-2.00 -1.30	<b>-0.956</b>	Infeasible	-1.193	-
C6_2	6	$\underline{r} = 5$ $\bar{r} = 8$	-0.900 -0.545	<b>-0.250</b>	-0.350	<b>-0.250</b>	-0.250
C6_4	6	$\underline{r} > 11$	-	<b>-0.696</b>	-1.11	<b>-0.696</b>	-0.696
C6_6	6	$\underline{r} = 6$ $\bar{r} = 11$	-2.54 -1.64	<b>-0.889</b>	Infeasible/ Out of Memory	-1.193	-
C8_2	8	$\underline{r} > 7$	-	<b>-0.250</b>	-0.425	<b>-0.250</b>	-0.250
C8_4	8	$\underline{r} > 15$	-	<b>-0.696</b>	-1.11	<b>-0.696</b>	-0.696
C10_2	10	$\underline{r} > 9$	-	<b>-0.250</b>	-0.425	<b>-0.250</b>	-0.250
C20_2	20	$\underline{r} > 19$	-	<b>-0.250</b>	-0.589	<b>-0.250</b>	-0.250

**Table 2:** The results of Example 4.2 are presented, considering instances from the paper [25]. Problems (11) and (12) are our new LP and SDP hierarchies, respectively, whose optimal values give lower bounds on the true optimal value (given in the final column). The last three columns give the results reported from [25], with KS-bound and (B)SOS-bound being the alternative lower bounds on the true optimal value using state-of-the-art methods. The (B)SOS-bound is the optimal value from both the standard SOS method and the bounded SOS method.  $\underline{r}$  is the lowest value of  $r$  such that the problem (11) was feasible, while  $\bar{r}$  is the highest value of  $r$  that we could compute this problem for without getting “out of memory” warnings. For each instance the best bound is in bold.

In [14, 15, 16, 20] the authors considered PSDPs, and how the SOS approach could be extended for positive semi-definite constraints using the concept of matrix SOS.

Let us consider a (scalar) polynomial  $f$  and the matrix polynomials  $G_i : \mathbb{R}^n \rightarrow \mathcal{S}^{m_i}$  (i.e.,  $((G_i)_{k\ell} = (G_i)_{\ell k}$  is a polynomial for all  $k, \ell = 1, \dots, m_i$ ). We define

$$d_0 = \deg(f), \quad d_i = \deg(G_i) = \max_{k,\ell} \deg(G_i)_{k\ell}.$$

Following [14, 15] we define a PSDP over the non-negative orthant as

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & G_i(\mathbf{x}) \in \mathcal{PSD}^{m_i} \quad \text{for all } i = 1, \dots, p, \\ & x_i \geq 0 \quad \text{for all } i = 1, \dots, n. \end{aligned} \tag{13}$$

This is again a very hard problem, so tractable relaxations are very desirable. The standard SOS approach to (13) is based on using matrix SOS constraints.

A polynomial matrix  $G : \mathbb{R}^n \rightarrow \mathcal{S}^m$  is matrix SOS if there exists a matrix polynomial  $H : \mathbb{R}^n \rightarrow \mathbb{R}^{q \times m}$  such that  $G(\mathbf{x}) =_c H(\mathbf{x})^\top H(\mathbf{x})$ , and we then trivially have that  $G(\mathbf{x}) \in \mathcal{PSD}^m$  for all  $\mathbf{x} \in \mathbb{R}^n$ . Similar to the standard SOS case (see, e.g., [15, 20]), we

Hierarchy	Optimal Value	PSD constraints		Number of...		
		Order	Number	"=" const.	" $\geq$ " const.	Variables
KS, $r = 5$	Infeasible	-	-	1,947,792	80,730	80,730
KS, $r = 6$	OoM	-	-	5,245,786	376,740	376,740
(11), $r = 11$	-1.64	-	-	0	224,707	123,761
(12), $r = 1$	-0.889	7	462	0	1,786	13,007
SOS, $r = 1$	-1.193	84	7	1,716	10	25,001
BSOS, $d = 1, k = 3$	-1.193	84	1	924	23	3,593

**Table 3:** Comparing the size of the approximation hierarchies for the case C6.6 in Example 4.2.

have that  $G(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathcal{S}^m$  is a matrix SOS of degree at most  $2d$  if and only if there exists  $A \in \mathcal{PSD}^{m \binom{n+d}{n}}$  such that

$$G(\mathbf{x}) =_c (I_m \otimes \mathbf{v}_d(\mathbf{x}))^\top A (I_m \otimes \mathbf{v}_d(\mathbf{x})),$$

where " $\otimes$ " denotes the Kronecker product and  $I_m$  denotes the identity matrix of the order  $m$ . Also, similar to before, there is a linear relationship between the elements of  $A$  and the coefficients of the elements of  $G(\mathbf{x})$ , and thus we can optimize over the set of matrix SOS polynomials using positive semi-definite optimization. We can also construct the SOS relaxation hierarchy for (13) for each  $r \in \mathbb{N}$  [14, 15]:

$$\begin{aligned}
& \max_{\lambda, \{a_i, A_i\}} \lambda \\
\text{s.t. } & f(\mathbf{x}) - \lambda =_c a_0(\mathbf{x}) + \sum_{i=1}^n x_i a_i(\mathbf{x}) + \sum_{i=1}^p \langle G_i(\mathbf{x}), A_i(\mathbf{x}) \rangle \\
& a_0(\mathbf{x}) \text{ is SOS of degree at most } 2 \lfloor \frac{1}{2}(D+r) \rfloor \\
& a_i(\mathbf{x}) \text{ is SOS of degree at most } 2 \lfloor \frac{1}{2}(D+r-1) \rfloor \text{ for } i = 1, \dots, n \\
& A_i(\mathbf{x}) \text{ is matrix SOS, } \deg(A_i) \leq 2 \lfloor \frac{1}{2}(D+r-d_i) \rfloor \text{ for all } i = 1, \dots, p.
\end{aligned} \tag{14}$$

where  $D = \max\{d_i : i = 0, 1, \dots, p\}$ . Under some conditions related to the compactness of the feasibility set for (13), the optimal values of this hierarchy converge to the optimal value of (13) [14, Theorem 2.2].

Similar to before, the order of the PSD constraints in this problem grows very quickly with  $r$ . In fact for each  $i \in \{1, \dots, p\}$ , the SOS constraint for  $A_i(\mathbf{x})$  is equivalent to a PSD constraint of the order  $m_i \binom{n + \lfloor (D+r-d_i)/2 \rfloor}{n}$ .

Alternatively, by taking  $\mathcal{K}_i = \mathcal{PSD}^{m_i} = \mathcal{K}_i^*$  for all  $i$ , we can apply our hierarchy from Section 3 to problem (13), both with and without the additional PSD constraint (8) in the original problem, to give the following problems, respectively:

$$\begin{aligned}
& \max_{\lambda, \{Y_{i,\alpha}\}} \lambda \\
& \text{s.t.} \quad (1 + \mathbf{e}^\top \mathbf{x})^{D-d_0+r} (f(\mathbf{x}) - \lambda) \geq_c \sum_{i=1}^p \sum_{\substack{\alpha \in \mathbb{N}^n: \\ \mathbf{e}^\top \alpha \leq D-d_i+r}} \mathbf{x}^\alpha \langle Y_{i,\alpha}, G_i(\mathbf{x}) \rangle \\
& \quad Y_{i,\alpha} \in \mathcal{PSD}^{m_i} \quad \text{for all } i, \alpha,
\end{aligned} \tag{15}$$

$$\begin{aligned}
& \max_{\lambda, \{Y_{i,\alpha}\}, \{Z_\beta\}} \lambda \\
& \text{s.t.} \quad (1 + \mathbf{e}^\top \mathbf{x})^{D-d_0+r} (f(\mathbf{x}) - \lambda) \geq_c \sum_{i=1}^p \sum_{\substack{\alpha \in \mathbb{N}^n: \\ \mathbf{e}^\top \alpha \leq D-d_i+r}} \mathbf{x}^\alpha \langle Y_{i,\alpha}, G_i(\mathbf{x}) \rangle + \dots \\
& \quad \dots + \sum_{\substack{\beta \in \mathbb{N}^n: \\ \mathbf{e}^\top \beta \leq D+r-2}} \mathbf{x}^\beta \mathbf{v}_1(\mathbf{x})^\top Z_\beta \mathbf{v}_1(\mathbf{x}) \\
& \quad Y_{i,\alpha} \in \mathcal{PSD}^{m_i} \quad \text{for all } i, \alpha, \\
& \quad Z_\beta \in \mathcal{PSD}^{n+1} \quad \text{for all } \beta.
\end{aligned} \tag{16}$$

Here, in comparison to the matrix SOS approach, we have many more SDP constraints, but the order of the positive semi-definite constraints is fixed.

In Table 4 we consider the sizes of these approximation problems for  $r = 4$  with  $n = 8$  variables and  $p = 5$  matrix SDP constraints of the order 10 with degrees  $d_0 = 4$  and  $d_i = 2$ ,  $i = 1, \dots, 5$ .

Hierarchy	SDP constraints		Linear constraints		Number of variables
	Order	Number	Type	Number	
Matrix SOS hierarchy, (10)	1650	5	=	12,870	7,042,696
	495	1			
	165	8			
New hierarchy, (15)	10	15,015	$\geq$	12,870	825,826
New hierarchy, (16)	10	15,015	$\geq$	12,870	960,961
	9	3003			

**Table 4:** Complexity of the  $r = 4$  member of matrix SOS hierarchy (14) and our new hierarchies (15) and (16), for  $n = 8$ ,  $p = 5$ ,  $d_0 = 4$  and  $d_i = 2$ ,  $m_i = 10$ ,  $i = 1, \dots, 5$ .

The difference is significant. While the Matrix SOS problem is already out of reach for most of the SDP solvers, our new hierarchies are manageable by at least those SDP solvers that can explore parallelization, like [56, 51].

*Example 4.3.* Let us consider the example from [14, Section II E]:

$$\min_{\mathbf{x}} f(\mathbf{x}) = -x_1^2 - x_2^2 \quad \text{s.t.} \quad G(\mathbf{x}) = \begin{bmatrix} 1 - 4x_1x_2 & x_1 \\ x_1 & 4 - x_1^2 - x_2^2 \end{bmatrix} \in \mathcal{PSD}^2.$$

This has the optimal solutions  $(0, 2)$  and  $(0, -2)$  with the optimal value  $-4$ .

As reported in [14, Table II], the SOS approximation hierarchy (14) applied to this problem reduces for  $r = 0$  to semi-definite optimization problems with 2 PSD constraints of the orders 3 and 2, giving an optimal value of  $-4$ , which is already the optimal value of the original problem.

We can transform the semi-definite constraint from (17) into two (scalar) polynomial constraints and apply the approach from Subsection 4.1. However, in [14, Table I] it is reported that only the 7<sup>th</sup> member of the hierarchy (10) yields the optimal value of  $-4$ . This member of the hierarchy has 3 PSD constraints of the orders 36, 28 and 21.

To apply our hierarchy to this problem, we first need to translate its feasible set into the non-negative orthant. To do this we substitute  $x_1$  by  $x_1 - 2$  and  $x_2$  by  $x_2 - 2$  to give the following problem.

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) = -(x_1 - 2)^2 - (x_2 - 2)^2 \\ \text{s.t.} \quad & G(\mathbf{x}) = \begin{bmatrix} 1 - 4(x_1 - 2)(x_2 - 2) & x_1 - 2 \\ x_1 - 2 & 4 - (x_1 - 2)^2 - (x_2 - 2)^2 \end{bmatrix} \in \mathcal{PSD}^2. \end{aligned}$$

Note that the semi-definite constraint implies that  $4 - (x_1 - 2)^2 - (x_2 - 2)^2 \geq 0$ ; therefore,  $x_1, x_2 \geq 0$ , i.e., after the substitution the sign constraint is redundant and we can add it without affecting the feasible region.

Our hierarchy (15) for this problem reduces for  $r = 1$  to

$$\begin{aligned} \max_{\lambda, \{Y\}} \quad & \lambda \\ \text{s.t.} \quad & (1 + x_1 + x_2)(-(x_1 - 2)^2 - (x_2 - 2)^2 - \lambda) \\ & \geq_c \langle Y_{(0,0)} + x_1 Y_{(1,0)} + x_2 Y_{(0,1)}, G(\mathbf{x}) \rangle \\ & Y_{(0,0)}, Y_{(1,0)}, Y_{(0,1)} \in \mathcal{PSD}^2 \end{aligned} \tag{17}$$

The optimal value for this is also equal to  $-4$ , i.e., the member of the hierarchy corresponding to  $r = 1$  is already giving the optimal value of the original problem. Note that the problem from our hierarchy has only 3 PSD constraints, all being of the order 2.

*Example 4.4.* Let us consider the following family of PSDPs:

$$\begin{aligned}
 \min_{\mathbf{x}} \quad & \text{trace } G(\mathbf{x}) \\
 \text{s.t.} \quad & G(\mathbf{x}) \in \mathcal{PSD}^n \\
 & \mathbf{x} \in \mathbb{R}_+^n,
 \end{aligned} \tag{18}$$

where

$$G(\mathbf{x}) = \begin{bmatrix} (1 + x_1/2)^2 & 0 & \dots & 0 \\ 0 & (1 + x_2/2)^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (1 + x_n/2)^2 \end{bmatrix} - \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}^\top.$$

It is shown in Appendix A that for the problem (18) we have:

- The optimal value is equal to  $n - 1$ ;
- The matrix SOS hierarchy, (14), for  $r \in \{0, 1\}$  has an optimal value of zero;
- Our new hierarchy, (15), for  $r \in \{0, 1\}$  has an optimal value of zero;
- Our new hierarchy, (16), for  $r = 0$  has an optimal value of zero.

In Table 5 we report on the numerical results for higher levels of these hierarchies, and in Table 6 we compare how large these problems are for  $n = 21$ . From these results we see that, in terms of optimal values, our new hierarchies perform very competitively in comparison with the matrix SOS approach. Moreover, in Table 6, we see that the sizes of our hierarchies, in terms of the order of the largest PSD constraint, are much smaller than the matrix SOS approach. We also see that adding the redundant constraint (8) to the original problem before applying our hierarchy to give the problem (16), significantly improves the lower bound in comparison to the problem (15).

$n$	optimum	val (15)	val (16)		Matrix SOS, (14)
		$r = 2$	$r = 1$	$r = 2$	$r = 2$
3	2	1.600	1.948	2.000	2.000
6	5	4.444	4.865	5.000	5.000
9	8	7.385	7.790	8.000	8.000
12	11	1.035	1.072	11.000	11.000
15	14	13.333	13.651	14.000	14.000
18	17	16.319	16.587	17.000	17.000
21	20	19.310	19.525	20.000	Out of Memory

**Table 5:** Numerical results for polynomial SDP problem (18), computed for different approximation hierarchies.

Type	r	Optimal value	PSD const.	
			Order	#
New hierarchy, (15)	2	19.310	21	253
New hierarchy, (16)	1	19.525	22	22
			21	22
	2	20.000	22	253
Matrix SOS, (14)	2	20.000	462	1
			253	1
			22	21

**Table 6:** Comparing the size of some instances of approximation hierarchies, in terms of order and number of PSD constraints, of the problem in Example 4.4 for  $n = 21$ .

### 4.3 Polynomial second-order cone optimization over the non-negative orthant

In this subsection we will apply our method to PSOCs over the non-negative orthant. We will see that our method performs very well on these problems, especially after adding the redundant constraint (8) to the original problem, convincingly outperforming the classic SOS approach.

Suppose that the cones  $\mathcal{K}_i^*$  in (1) are the second-order cones (SOC):

$$\mathcal{SOC}^{m_i} := \{(z_1, \mathbf{z}) \in \mathbb{R} \times \mathbb{R}^{m_i-1} : \|\mathbf{z}\|_2 \leq z_1\}.$$

Note that  $(z_1, \mathbf{z}) \in \mathbb{R} \times \mathbb{R}^{m_i-1}$  belongs to  $\mathcal{SOC}^{m_i}$  iff  $z_1 \geq 0$  and  $z_1^2 - \sum_{i=2}^{m_i} z_i^2 \geq 0$ .

For  $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^{m_i}$  being a polynomial vector of degree  $d$ , we call  $\mathbf{g}(\mathbf{x}) \in \mathcal{SOC}^{m_i}$  a polynomial second-order cone constraint (PSOC). It can be transformed into two polynomial inequalities, one of the order  $d$  and one of the order  $2d$ . Therefore, we can handle PSOCs as classic polynomial optimization problems and use approaches from Subsection 4.1.

Since SOCs are tractable, i.e., we can solve linear conic optimization problems over these cones numerically efficiently, it is a natural idea to approximate PSOCs also using the new hierarchy from Section 3 directly. We demonstrate this with the following two examples.

*Example 4.5.* Let us consider the following simple PSOC optimization problem:

$$\begin{aligned} \min_{x_1, x_2} \quad & x_1^2 + x_2^2 \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{x}) = \begin{bmatrix} x_1^2 - x_2^2 \\ x_1 x_2 \\ x_1 + x_2 + 1 \end{bmatrix} \in \mathcal{SOC}^3 \\ & x_1, x_2 \in \mathbb{R}_+. \end{aligned} \tag{19}$$

Its minimum, obtained by Mathematica [54], is attained at  $\mathbf{x}^* = (1.6180, 0)^\top$  and is equal to 2.6180. Our hierarchy reduces for fixed  $r$  to

$$\begin{aligned} & \max_{\lambda, \{\mathbf{y}_\alpha\}} \lambda \\ & \text{s.t. } (1 + x_1 + x_2)^r (x_1^2 + x_2^2 - \lambda) \geq_c \sum_{\mathbf{e}^\top \alpha \leq r} x^\alpha \langle \mathbf{y}_\alpha, \mathbf{g}(\mathbf{x}) \rangle \\ & \mathbf{y}_\alpha \in \mathcal{SOC}^3 \quad \forall \mathbf{e}^\top \alpha \leq r. \end{aligned} \quad (20)$$

The optimal values of (20) for  $r \in \{0, \dots, 5\}$  are reported in the second row of Table 7.

We can replace the PSOC constraint  $\mathbf{g}(\mathbf{x}) \in \mathcal{SOC}^3$  by two equivalent polynomial constraints:

$$x_1^2 - x_2^2 \geq 0, \quad (x_1^2 - x_2^2)^2 - (x_1 x_2)^2 - (x_1 + x_2 + 1)^2 \geq 0$$

and apply the SOS hierarchy (10). The lower bounds provided by the SOS hierarchy are shown in the fourth row of Table 7. We can see that in this example our hierarchy outperforms the SOS hierarchy. This is especially apparent when we consider the sizes of the problems in Table 8. We also consider adding the redundant PSD constraint (8) to the original problem and then applying our hierarchy, and when doing this the  $r = 1$  member of the hierarchy gives the exact optimal value (within numerical errors) using a very small optimisation problem, as is also shown in Tables 7 and 8.

Hierarchy	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$
New, without (8)	1.0000	2.0000	2.5000	2.6000	2.6154	2.6176
New, with (8)	1.0000	2.6180	2.6180	2.6180	2.6180	2.6180
SOS	0.0000	0.0000	0.0110	0.0850	0.9895	2.2275

**Table 7:** The optimal values of our hierarchy (20) (without (8)) compared to SOS hierarchy (10), obtained by replacing the PSOC constraint with two polynomial constraints. We also consider adding the additional redundant constraint (8) to the problem (19) before applying our hierarchy.

Type	r	Optimal value	PSD constraints		# $\mathcal{SOC}^3$ const.
			order	number	
New, without (8)	5	2.6176	-	-	21
New, with (8)	1	2.6180	3	3	3
SOS	5	2.2275	15	3	0
			10	1	
			6	1	

**Table 8:** We compare the sizes of the optimization problems, in terms of the size and the number of PSD and  $\mathcal{SOC}^3$  constraints, for some approximation hierarchies for the problem in Example 4.5.

*Example 4.6.* We have also evaluated our hierarchy on seven randomly generated polynomial second-order cone-optimization problems in three variables. Each such problem has two polynomial SOC constraints of degree two and dimension four. More precisely, for  $i = 1, 2$  we generated polynomials  $g_1^i(\mathbf{x}), \dots, g_4^i(\mathbf{x})$ , which have all possible monomials in three variables of the degree two with coefficients generated randomly from  $[-0.3, 0.7]$  following the uniform distribution. We then replaced  $g_1^i$  with  $\tilde{g}_1^i = \sum_{k=1}^4 g_k^i$  and set the PSOC  $\mathbf{g}^i = (\tilde{g}_1^i, g_2^i, g_3^i, g_4^i)^\top \in \mathcal{SOC}^4$ . When we have generated both PSOCs, we set the objective function  $f(\mathbf{x}) = \tilde{g}_1^1(\mathbf{x}) + \tilde{g}_1^2(\mathbf{x}) + h(\mathbf{x})$ , where  $h$  is polynomials of the degree two with random uniform coefficients from  $[-0.5, 0.5]$ .

For each random polynomial second-order cone optimization problem we computed the optimum value using Mathematica [54] and the optimal values of the first three members of our hierarchy (with and without adding the additional SDP constraint (8) to the original problem) and of the standard SOS hierarchy. Recall that the SOS hierarchy is applied after reformulating each polynomial SOC constraint into two polynomial inequalities, as described at the beginning of this subsection. The results are in Table 9.

As we can see, for these examples our new hierarchy (without (8)) provides better bounds through solving smaller optimization problems in comparison to the standard SOS method. If we add (8) to the original problem, then the hierarchy is even better.

problem	optimum	New hierarchy, without (8)			New hierarchy, with (8)			SOS hierarchy		
		$r = 0$	$r = 1$	$r = 2$	$r = 0$	$r = 1$	$r = 2$	$r = 0$	$r = 1$	$r = 2$
RpolySOC_3_2_4a	2.362	2.103	<b>2.362</b>	<b>2.362</b>	2.103	<b>2.362</b>	<b>2.362</b>	1.339	2.319	<b>2.362</b>
RpolySOC_3_2_4b	unknown	1.463	1.858	1.920	1.463	1.915	<b>1.927</b>	1.039	1.354	1.502
RpolySOC_3_2_4c	1.503	1.453	1.500	<b>1.503</b>	<b>1.503</b>	<b>1.503</b>	<b>1.503</b>	1.496	1.496	1.496
RpolySOC_3_2_4d	3.204	<b>3.204</b>	<b>3.204</b>	<b>3.204</b>	<b>3.204</b>	<b>3.204</b>	<b>3.204</b>	<b>3.204</b>	<b>3.204</b>	<b>3.204</b>
RpolySOC_3_2_4e	1.408	<b>1.408</b>	<b>1.408</b>	<b>1.408</b>	<b>1.408</b>	<b>1.408</b>	<b>1.408</b>	1.405	1.405	1.405
RpolySOC_3_2_4f	0.797	0.730	0.795	<b>0.797</b>	0.730	<b>0.797</b>	<b>0.797</b>	0.569	0.569	0.570
RpolySOC_3_2_4g	1.812	1.770	1.810	<b>1.812</b>	<b>1.812</b>	<b>1.812</b>	<b>1.812</b>	1.358	1.368	1.369
RpolySOC_3_2_4h	0.691	0.558	0.658	0.681	0.678	<b>0.691</b>	<b>0.691</b>	0.275	0.396	0.397
PSD constraints, (Order $\times$ Num.)	-	-	-	-	$4 \times 1$	$4 \times 4$	$4 \times 10$	$10 \times 1$ $4 \times 5$	$10 \times 4$ $4 \times 2$	$20 \times 1$ $10 \times 5$ $4 \times 2$
Number of $\mathcal{SOC}^4$	2	8	20	20	2	8	20	-	-	-
Number of " $\geq$ " const.	10	20	35	35	10	20	35	2	2	0
Number of "=" const.	0	0	0	0	0	0	0	35	56	84
Number of variables	9	33	81	81	19	73	181	108	243	506

**Table 9:** This table contains the results for eight randomly generated polynomial second-order cone-optimization problems. In the top half of the table, the first column contains the internal name of the problem, the second column contains the optimum value computed by Mathematica [54] (note that for RpolySOC\_3\_2\_4b Mathematica did not return the optimum value, so we report it as unknown), the columns 3–5 contain optimal values of our hierarchy corresponding to  $r = 0, 1, 2$ , the columns 6–8 contain the optimal value of the SOS hierarchy corresponding to  $r = 0, 1, 2$  and the columns 9–11 contain the optimal values of our hierarchy after adding the redundant constraint (8) to the original problem. The bottom five rows of the table give the size of the approximation problems, noting that each PSOC optimization problem considered has the same size and thus gives the same sized approximations. For each instance the best bound is in bold.

## 5 Conclusion

In this paper we considered polynomial conic optimization over the non-negative orthant. We proposed a hierarchy of relaxations that are in the form of conic linear optimization problems. We demonstrated that when the feasible set is contained in the non-negative orthant we can successfully apply the proposed hierarchy to constrained polynomial optimization problems (POPs), polynomial semi-definite optimization problems (PSDPs) and polynomial second-order cone-optimization problems (PSOCPs). We proved that the hierarchy is monotonic and gives bounds that converge asymptotically to the optimum value of the original problem, if some classic assumptions are satisfied. Through the addition of a redundant PSD constraint in the original problem, the proposed hierarchy performs reasonably well for POPs, in comparison to the classic SOS approach, and outperforms the classic SOS approach for PSDPs and PSOCPs, while also being cheaper to compute. The preliminary numerical evaluation confirms that this hierarchy deserves a deeper theoretical/numerical study and efficient code development. In particular, it would be of interest to consider the dual problems of our hierarchy, investigating whether similar results to that for the SOS/moment hierarchies hold with regards to the flat extension conditions [6], flat truncations [33] and extracting optimal solutions [13, Section 2].

## Acknowledgements

The research for this paper was started while P.J.C. Dickinson was at the Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, The Netherlands. It was then continued while this author was at the Department of Statistics and Operations Research, University of Vienna, Austria, and then again after he joined the Department of Applied Mathematics, University of Twente, The Netherlands. This author would like to gratefully acknowledge support from the Netherlands Organisation for Scientific Research (NWO) through grant no. 613.009.021.

The second author started this research when he was affiliated to the Faculty of Information Studies in Novo Mesto, Slovenia, and continued the work after moving to the University of Ljubljana, Slovenia. He wishes to thank the Slovenian Research Agency for support via the program P1-0383 and projects J1-8132, N1-0057, N1-0071.

Both authors would also like to thank the anonymous referees for their useful comments with regards to this paper.

## References

- [1] Matlab version (r2012a), 2012. The MathWorks, Natick, MA, USA.
- [2] A. A. Ahmadi and A. Majumdar. DSOS and SDSOS optimization: LP and SOCP-based alternatives to sum of squares optimization. In *Information Sciences and Systems (CISS), 2014 48th Annual Conference on*, pages 1–5. IEEE, 2014.
- [3] A. A. Ahmadi and A. Majumdar. Some applications of polynomial optimization in operations research and real-time decision making. *Optimization Letters*, 10(4):709–729, 2016.
- [4] A. Berman. *Cones, Matrices and Mathematical Programming*, volume 79 of *Lecture Notes in Economics and Mathematical Systems*. Springer Verlag, 1973.
- [5] S. Burgdorf, I. Klep, and J. Povh. *Optimization of polynomials in non-commuting variables*. Springer, 2016.
- [6] R. E. Curto and L. A. Fialkow. Truncated k-moment problems in several variables. *Journal of Operator Theory*, 54:189–226, 2005.
- [7] J. Demmel, J. Nie, and V. Powers. Representations of positive polynomials on non-compact semialgebraic sets via KKT ideals. *Journal of Pure and Applied Algebra*, 209(1):189–200, 2007.
- [8] P. J. C. Dickinson. *The Copositive Cone, the Completely Positive Cone and their Generalisations*. PhD thesis, University of Groningen, Groningen, The Netherlands, 2013.
- [9] P. J. C. Dickinson and J. Povh. On an extension of Pólya’s Positivstellensatz. *Journal of Global Optimization*, 61(4):615–625, 2015.
- [10] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the Association for Computing Machinery*, 42(6):1115–1145, 1995.
- [11] D. Henrion and A. Garulli. *Positive polynomials in control*, volume 312. Springer Science & Business Media, 2005.
- [12] D. Henrion and J.-B. Lasserre. Gloptipoly: Global optimization over polynomials with matlab and sedumi. *ACM Transactions on Mathematical Software*, 29(2):165–194, 2003.
- [13] D. Henrion and J.-B. Lasserre. *Detecting Global Optimality and Extracting Solutions in GloptiPoly*, volume 312 of *Lecture Notes in Control and Information Science*, pages 293–310. Springer, Berlin, Heidelberg, 2005.

- [14] D. Henrion and J.-B. Lasserre. Convergent relaxations of polynomial matrix inequalities and static output feedback. *IEEE Transactions on Automatic Control*, 51(2):192–202, 2006.
- [15] C. W. Hol and C. W. Scherer. Sum of squares relaxations for polynomial semidefinite programming. In *Proc. Symp. on Mathematical Theory of Networks and Systems (MTNS), Leuven, Belgium, 2004*.
- [16] C. W. J. Hol and C. W. Scherer. A sum-of-squares approach to fixed-order  $H_\infty$ -synthesis. In *Positive polynomials in control*, volume 312 of *Lecture Notes in Control and Inform. Sci.*, pages 45–71. Springer, Berlin, 2005.
- [17] C. Jozs and D. K. Molzahn. Moment/sum-of-squares hierarchy for complex polynomial optimization. *arXiv preprint arXiv:1508.02068*, 2015.
- [18] I. Klep and J. Povh. Constrained trace-optimization of polynomials in freely non-commuting variables. *Journal of Global Optimization*, 64(2):325–348, 2016.
- [19] I. Klep and M. Schweighofer. Pure states, positive matrix polynomials and sums of hermitian squares. *Indiana University Mathematics Journal*, 59(3):857–874, 2010.
- [20] M. Kojima. Sums of squares relaxations of polynomial semidefinite programs. *Department of Mathematical and computing Sciences, Tokyo Institute of Technology, Meguro, Tokyo, 2003*.
- [21] M. Kojima, S. Kim, and H. Waki. A general framework for convex relaxation of polynomial optimization problems over cones. *Journal of the Operations Research Society of Japan*, 46(2):125–144, 2003.
- [22] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.
- [23] J. B. Lasserre. *Moments, positive polynomials and their applications*, volume 1 of *Imperial College Press Optimization Series*. Imperial College Press, London, 2010.
- [24] J. B. Lasserre. A lagrangian relaxation view of linear and semidefinite hierarchies. *SIAM Journal on Optimization*, 23(3):1742–1756, 2013.
- [25] J. B. Lasserre, K.-C. Toh, and S. Yang. A bounded degree SOS hierarchy for polynomial optimization. *EURO Journal on Computational Optimization*, pages 1–31, 2015.
- [26] M. Laurent. Semidefinite representations for finite varieties. *Mathematical Programming*, 109(1):1–26, 2007.

- [27] M. Laurent. *Sums of squares, moment matrices and optimization over polynomials*, volume 149 of *The IMA Volumes in Mathematics and its Applications*, pages 157–270. Springer, 2009.
- [28] M. Laurent and F. Rendl. Semidefinite programming and integer programming. In G. N. K. Aardal and R. Weismantel, editors, *Discrete Optimization*, volume 12 of *Handbooks in Operations Research and Management Science*, pages 393 – 514. Elsevier, 2005.
- [29] J. Löfberg. YALMIP : A toolbox for modeling and optimization in matlab. In *In Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.
- [30] A. Marandi, J. Dahl, and E. de Klerk. A numerical evaluation of the bounded degree sum-of-squares hierarchy of Lasserre, Toh, and Yang on the pooling problem. *Annals of Operations Research*, 265(1):67–92, 2018.
- [31] MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 7.1 (Revision 28)*, 2015.
- [32] K. G. Murty and S. N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39(2):117–129, 1987.
- [33] J. Nie. Certifying convergence of Lasserre’s hierarchy via flat truncation. *Mathematical Programming*, 142(1–2):485–510, 2013.
- [34] J. Nie. An exact jacobian SDP relaxation for polynomial optimization. *Mathematical Programming*, 137(1–2):225–255, 2013.
- [35] J. Nie. Polynomial optimization with real varieties. *SIAM Journal on Optimization*, 23(3):1634–1646, 2013.
- [36] J. Nie. Optimality conditions and finite convergence of Lasserre’s hierarchy. *Mathematical Programming*, 146(1–2):97–121, 2014.
- [37] J. Nie, J. Demmel, and B. Sturmfels. Minimizing polynomials via sum of squares over the gradient ideal. *Mathematical Programming*, 106(3):587–606, 2006.
- [38] P. A. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, Pasadena, California, 2000.
- [39] J. Peña, J. Vera, and L. F. Zuluaga. Computing the stability number of a graph via linear and semidefinite programming. *SIAM Journal on Optimization*, 18(1):87–105, 2008.

- [40] J. Povh. Contribution of copositive formulations to the graph partitioning problem. *Optimization: A Journal of Mathematical Programming and Operations Research*, pages 1–13, 2011.
- [41] J. Povh and F. Rendl. A copositive programming approach to graph partitioning. *SIAM Journal on Optimization*, 18(1):223–241, 2007.
- [42] J. Povh and F. Rendl. Copositive and semidefinite relaxations of the quadratic assignment problem. *Discrete Optimization*, 6(3):231–241, 2009.
- [43] M. Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42(3):969–984, 1993.
- [44] M. Putinar and F.-H. Vasilescu. Positive polynomials on semi-algebraic sets. *Comptes Rendus de l’Académie des Sciences - Series I - Mathematics*, 328(7):585–589, 1999.
- [45] F. Rendl, G. Rinaldi, and A. Wiegele. Solving max-cut to optimality by intersecting semidefinite and polyhedral relaxations. *Math. Program.*, 121(2, Ser. A):307–335, 2010.
- [46] S. Sahni and T. Gonzalez.  $P$ -complete approximation problems. *Journal of the Association for Computing Machinery*, 23(3):555–565, 1976.
- [47] K. Schmüdgen. The  $K$ -moment problem for compact semi-algebraic sets. *Mathematische Annalen*, 289(2):203–206, 1991.
- [48] A. Schrijver. *Theory of linear and integer programming*. Wiley, 1999.
- [49] A. Schrijver. *Combinatorial optimization*. Springer, 2003.
- [50] M. Schweighofer. Optimization of polynomials on compact semialgebraic sets. *SIAM Journal on Optimization*, 15(3):805–825, 2005.
- [51] M. Simionica and J. Povh. A parallel implementation of the boundary point method. In *SOR ’17 proceedings, Bled, Slovenia*, pages 114–119, 2017.
- [52] J. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11–12:625–653, 1999. Version 1.05 available from <http://fewcal.kub.nl/sturm>.
- [53] T. Weisser, J.-B. Lasserre, and K.-C. Toh. Sparse-BSOS: a bounded degree SOS hierarchy for large scale polynomial optimization with sparsity. 2017.
- [54] Wolfram Research, Inc. Mathematica 11.1, 2017.
- [55] L. Wolsey. *Integer Programming*. A Wiley-Interscience publication. Wiley, 1998.

[56] M. Yamashita, K. Fujisawa, and M. Kojima. SDPARA: Semidefinite programming algorithm parallel version. *Parallel Computing*, 29(8):1053–1067, 2003.

[57] X.-Y. Zhao, D. Sun, and K.-C. Toh. A newton-cg augmented lagrangian method for semidefinite programming. *SIAM Journal on Optimization*, 20(4):1737–1765, 2010.

## A Proofs for Example 4.4

**Lemma A.1.** *We have  $\text{val} (18) = n - 1$ .*

*Proof.* If we consider  $\mathbf{x} = 2\mathbf{e}_1$  then we have  $G(2\mathbf{e}_1) = \text{Diag}(0, 1, \dots, 1) \in \mathcal{PSD}^n$  and thus  $\text{val} (18) \leq \text{trace } G(2\mathbf{e}_1) = n - 1$ . Now assume for the sake of contradiction that  $\text{val} (18) < n - 1$ . This is possible only if there exists  $\mathbf{x} \in \mathbb{R}_+^n$  such that at least two of the on-diagonal entries of  $G(\mathbf{x})$  are strictly less than one (otherwise, we have  $n - 1$  diagonal entries that are at least 1, and the remaining is nonnegative, hence  $\text{trace } G(\mathbf{x}) \geq n - 1$ ). Without loss of generality the first two on-diagonal entries of  $G(\mathbf{x})$  are strictly less than one, implying that  $x_1, x_2 > \frac{4}{3}$ . The positive semi-definiteness of  $G$  implies that the determinant of the submatrix of  $G$  corresponding of the first two rows and columns is non-negative, but

$$G_{1,1}(\mathbf{x}) \cdot G_{2,2}(\mathbf{x}) - G_{1,2}(\mathbf{x})^2 < 1 - \frac{16}{9} < 0,$$

which is a contradiction. Hence,  $\text{val} (18) = n - 1$ . □

Considering the approximation hierarchies discussed in this paper for problem (18), we have the following results:

**Lemma A.2.** *The optimal value of our new hierarchy, (15) (without the additional constraint (8) in the original problem), for  $r \in \{0, 1\}$  is equal to zero.*

*Proof.* Our hierarchy, (15), for  $r = 0$  and  $r = 1$  reduces to the following respectively:

$$\begin{aligned} \max_{\lambda, Y} \quad & \lambda \\ \text{s.t.} \quad & \text{trace } G(\mathbf{x}) - \lambda \geq_c \langle Y, G(\mathbf{x}) \rangle \\ & Y \in \mathcal{PSD}^n, \end{aligned} \tag{21}$$

$$\begin{aligned} \max_{\lambda, \{Y\}} \quad & \lambda \\ \text{s.t.} \quad & (1 + \mathbf{e}^\top \mathbf{x})(\text{trace } G(\mathbf{x}) - \lambda) \geq_c \langle Y_0, G(\mathbf{x}) \rangle + \sum_{i=1}^n x_i \langle Y_i, G(\mathbf{x}) \rangle \\ & Y_i \in \mathcal{PSD}^n \quad \text{for all } i = 0, \dots, n. \end{aligned} \tag{22}$$

Considering  $\lambda = 0$  and  $Y$  equal to the identity matrix gives us a feasible point of (21), and thus  $\text{val}(21) \geq 0$ .

From Proposition 3.3 we have  $\text{val}(21) \leq \text{val}(22)$ . We will now complete the proof by showing that  $\text{val}(22) \leq 0$ .

For an arbitrary feasible point of (22), we consider the following coefficient inequalities explicitly:

$$\begin{aligned} 1 : \quad & n - \lambda \geq \text{trace } Y_0, \\ x_i^2 : \quad & \frac{1}{4} \geq -\frac{3}{4}(Y_0)_{ii} + (Y_i)_{ii}, \\ x_i^3 : \quad & -\frac{3}{4} \geq -\frac{3}{4}(Y_i)_{ii}. \end{aligned}$$

This implies that for all  $i$  we have  $(Y_0)_{ii} \geq \frac{4}{3}(Y_i)_{ii} - \frac{1}{3} \geq \frac{4}{3} - \frac{1}{3} = 1$ , which in turn implies that  $\lambda \leq n - \text{trace } Y_0 \leq n - n = 0$ , and thus  $\text{val}(22) \leq 0$ .  $\square$

**Lemma A.3.** *The optimal value of both of our new hierarchy, (16) (with the additional constraint (8) in the original problem), and the Matrix SOS hierarchy, (14), are equal to zero for  $r = 0$ .*

*Proof.* These problems both reduce to the following problem:

$$\begin{aligned} \max_{\lambda, Y, v, \mathbf{w}, Z} \quad & \lambda \\ \text{s.t.} \quad & \text{trace } G(\mathbf{x}) - \lambda \geq_c \langle Y, G(\mathbf{x}) \rangle + \left\langle \begin{pmatrix} v & \mathbf{w}^\top \\ \mathbf{w} & Z \end{pmatrix}, \begin{pmatrix} 1 & \mathbf{x}^\top \\ \mathbf{x} & \mathbf{x}\mathbf{x}^\top \end{pmatrix} \right\rangle \\ & Y \in \mathcal{PSD}^n, \quad \begin{pmatrix} v & \mathbf{w}^\top \\ \mathbf{w} & Z \end{pmatrix} \in \mathcal{PSD}^{n+1}. \end{aligned} \quad (23)$$

Considering  $\lambda = v = 0$ ,  $\mathbf{w} = \mathbf{0}$ ,  $Z = 0$  and  $Y$  equal to the identity matrix gives us a feasible point of (23), and thus  $\text{val}(23) \geq 0$ .

For an arbitrary feasible point of (23), we consider the following coefficient inequalities explicitly:

$$\begin{aligned} 1 : \quad & n - \lambda \geq \text{trace } Y + v, \\ x_i^2 : \quad & -\frac{3}{4} \geq -\frac{3}{4}y_{ii} + z_{ii}. \end{aligned}$$

Considering the positive semi-definiteness constraints, we also have  $0 \leq v$  and  $0 \leq z_{ii}$  for all  $i$ . Therefore,  $y_{ii} \geq 1 + \frac{4}{3}z_{ii} \geq 1$  for all  $i$  and  $\lambda \leq n - \text{trace } Y - v \leq n - n - 0 = 0$ , implying that  $\text{val}(23) \leq 0$ .  $\square$

**Lemma A.4.** *The optimal value of the Matrix SOS hierarchy, (14), is equal to zero for  $r = 1$ .*

*Proof.* This problem reduces to the following problem:

$$\begin{aligned}
& \max_{\lambda, Y, \{v_i\}, \{\mathbf{w}_i\}, \{Z_i\}} \lambda \\
& \text{s.t.} \quad \text{trace } G(\mathbf{x}) - \lambda =_c \langle Y, G(\mathbf{x}) \rangle + \left\langle \begin{pmatrix} v_0 & \mathbf{w}_0^\top \\ \mathbf{w}_0 & Z_0 \end{pmatrix}, \begin{pmatrix} 1 & \mathbf{x}^\top \\ \mathbf{x} & \mathbf{x}\mathbf{x}^\top \end{pmatrix} \right\rangle + \dots \\
& \quad \quad \quad \dots + \sum_{i=1}^n x_i \left\langle \begin{pmatrix} v_i & \mathbf{w}_i^\top \\ \mathbf{w}_i & Z_i \end{pmatrix}, \begin{pmatrix} 1 & \mathbf{x}^\top \\ \mathbf{x} & \mathbf{x}\mathbf{x}^\top \end{pmatrix} \right\rangle \\
& \quad \quad \quad Y \in \mathcal{PSD}^n, \quad \begin{pmatrix} v_i & \mathbf{w}_i^\top \\ \mathbf{w}_i & Z_i \end{pmatrix} \in \mathcal{PSD}^{n+1} \quad \text{for } i = 0, \dots, n.
\end{aligned}$$

For  $i, j \in \{1, \dots, n\}$  such that  $i = j$ , considering the coefficient equalities for  $x_i^3$  and  $x_i^2 x_j$  explicitly, for any feasible point to this problem we have  $0 = (Z_i)_{ii}$  and  $0 = 2(Z_i)_{ij} + (Z_i)_{jj}$ . As  $Z_i \in \mathcal{PSD}^n$ , this implies that  $Z_i$  is equal to the zero matrix for all  $i$ . Therefore, this problem reduces to the problem (23), which we have already shown to have an optimal value equal to zero.  $\square$