Positive polynomials on unbounded equality-constrained domains

Javier Peña* [1], Juan C. Vera† [2], and Luis F. Zuluaga‡ [3]

[1] *Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, USA 15213*

[2] *Tilburg School of Economics and Management, Tilburg University, Tilburg, The Netherlands*

[3] *Faculty of Business Administration, University of New Brunswick, Fredericton, NB, Canada E3B 5A3*

April 14, 2011

Abstract

Certificates of non-negativity are fundamental tools in optimization. A "certificate" is generally understood as an expression that makes the non-negativity of the function in question evident. Some classical certificates of non-negativity are Farkas Lemma and the S-lemma. The *lift-and-project* procedure can be seen as a certificate of non-negativity for affine functions over the union of two polyhedra. These certificates of non-negativity underlie powerful algorithmic techniques for various types of optimization problems. Recently, more elaborate sum-of-squares certificates of non-negativity for higher degree polynomials have been used to obtain powerful numerical techniques for solving polynomial optimization problems, particularly for mixed integer programs and non-convex binary programs. We present a new certificate of non-negativity for polynomials over the intersection of a closed set S and the zero set of a given polynomial $h(x)$. The certificate is written in terms of the set of non-negative polynomials over S and the ideal generated by $h(x)$. Our certificate of non-negativity yields a copositive programming reformulation for a very general class of polynomial optimization problems. This copositive programming formulation generalizes Burer's copositive formulation for binary programming and offers an avenue for the development of new algorithms to solve polynomial optimization problems. In particular, the copositive formulation could be used to obtain new semidefinite programming relaxations for binary programs, as our approach is different and complementary to the conventional approaches to obtain semidefinite programming relaxations via matrix relaxations.

1 Introduction

Certificates of non-negativity are fundamental tools in optimization. A "certificate" is generally understood as an expression that makes the non-negativity of the function in question evident. Some classical certificates of non-negativity are Farkas Lemma and the S-lemma.

*jfp@andrew.cmu.edu (Corresponding author)

†j.c.veralizcano@uvt.nl

‡lzuluaga@unb.ca

The former is a certificate of non-negativity for affine functions over a given polyhedral domain, the latter is a certificates of non-negativity for quadratic functions over the sublevel set of a given quadratic function. The Balas-Ceria-Cornuéjols lift-and-project procedure [1] can be seen as a certificate of non-negativity for affine functions over the union of two polyhedra. More elaborate certificates of non-negativity for higher degree polynomials over a basic semi-algebraic set include the classical Pólya's Theorem [8], and the more modern Schmüdgen's Theorem [15] and Putinar's Theorem [13]. These theorems give certificates of non-negativity for polynomials over a given basic semi-algebraic set. The more recent work of Nie et al [11], Demmel et al [6], and Marshall [10] provide additional certificates of non-negativity via gradient and KKT ideals. These certificates of non-negativity underlie powerful algorithmic techniques for various types of optimization problems, particularly for mixed integer programs and non-convex binary programs.

We present a new certificate of non-negativity for polynomials over the intersection of a closed domain S and the zero set of a given polynomial $h(x)$. It is evident that if $p(x)$ is non-negative on the domain S, then $p(x)+h(x)q(x)$ is non-negative on the domain $S\cap h^{-1}(0)$ for any polynomial $q(x)$. We show that under suitable conditions on $h(x)$ and S, the converse of this statement holds as well, thereby establishing a certificate of non-negativity for polynomials on $S\cap h^{-1}(0)$ in terms of non-negative polynomials on S. We note that for the case when S is not compact, previous results in [6, 10, 11] give certificates of non-negativity for a polynomial $p(x)$ in terms of the gradient of $p(x)$ or the KKT ideal involving $p(x)$ and the polynomials defining S. By contrast, our certificate of non-negativity is written purely in terms of the set of non-negative polynomials over S and the ideal generated by $h(x)$. This property of our certificate of non-negativity yields some interesting consequences. In particular, it leads to a canonical convexification procedure for polynomial optimization problems. Our convexification procedure yields an equivalent formulation of polynomial optimization problems as linear conic programs over the dual of the cone of copositive forms. This formulation is inspired by Burer's dual copositive formulation of binary quadratic programming problems [3, 4]. Indeed, the latter can be recovered as a special case of our convexification procedure (see Section 5). These copositive programming formulations offer an avenue for the development of new algorithms to solve polynomial optimization problems. In particular, the copositive formulation could be used to obtain new semidefinite programming relaxations for binary programs, as our approach is different and complementary to the conventional approaches to obtain semidefinite programming relaxations via matrix relaxations [16, 9].

The suitable conditions that ensure the validity of our certificate of non-negativity are related to the behavior of the "zeros at infinity" of the polynomials $h(x)$ on the set S. The formalization of this condition is stated in terms of the *horizon cone* of the set S and the homogeneous component of the polynomial $h(x)$. Loosely speaking, the conditions presented in [3, 4, 2] for the convexification of binary quadratic programming problems are special cases of the more general zeros at infinity condition presented here.

The main parts of the paper are organized as follows. Section 2 motivates and formally states our main result; namely, the certificate of non-negativity presented in Theorem 3. Section 3 provides insight about the zeros at infinity condition in Theorem 3. Section 4 describes a canonical convexification procedure for equality-constrained polynomial optimization problems over the non-negative orthant, and its natural extension to inequality-constrained polynomial optimization problems. The procedure is a generic reformulation of these classes of problems as a linear conic program over the cone of completely positive forms on $\mathbb{R}^n_+$. Section 5 gives further applications of our main theorem. Section 6 presents the technical proofs of the main theorems in the paper.

2 A new certificate of non-negativity

To motivate our certificate of non-negativity in Theorem 3 below, we begin by recalling a key result in the Balas-Ceria-Cornuejols lift-and-project procedure [1] and cast it as a certificate of non-negativity. Specifically, assume $\{x \in \mathbb{R}^n : Ax \le b\} \subseteq \{x \in \mathbb{R}^n : 0 \le x_j \le 1\}$ for some fixed $j \in \{1, \ldots, n\}$. In [1, Theorem 2.10] Balas, Ceria, and Cornuéjols show the following characterization of the valid inequalities for $\{x \in \mathbb{R}^n : Ax \le b,\, x_j \in \{0,1\}\}$:

$$\begin{array}{c} \alpha^{\mathrm{T}} x + \beta \ge 0 \\ \text{for all } x \in \{x \in \mathbb{R}^n : Ax \le b,\, x_j \in \{0,1\}\} \end{array} \iff \begin{array}{c} \exists\, u, v \ge 0 \text{ and } u_0, v_0 \in \mathbb{R} \text{ such that} \\ \begin{array}{rcl} \alpha &=& A^{\mathrm{T}} u + e_j u_0 \\ \alpha &=& A^{\mathrm{T}} v + e_j v_0 \\ \beta &=& -b^{\mathrm{T}} u + u_0 \\ \beta &=& -b^{\mathrm{T}} v. \end{array} \end{array} \tag{1}$$

Notice that (1) is a certificate of non-negativity for linear polynomials over the set $\{x \in \mathbb{R}^n : Ax \le b,\, x_j \in \{0,1\}\}$. This can be seen more clearly by introducing some notation.

For a given positive integer d, let $\mathbb{R}_d[x] = \mathbb{R}_d[x_1, \ldots, x_n]$ denote the set of real polynomials of degree at most d in n variables. For a given $S \subseteq \mathbb{R}^n$, let $\mathcal{P}_d(S) \subseteq \mathbb{R}_d[x]$ be defined as the set of non-negative polynomials of degree at most d on S, that is,

$$\mathcal{P}_d(S) := \{p \in \mathbb{R}_d[x] : p(x) \ge 0\ \forall x \in S\}.$$

Letting $S = \{x : Ax \le b\}$, $h(x) = x_j(1 - x_j)$, and using Farkas Lemma to characterize $\mathcal{P}_1(S)$, it is not difficult to show that (1) can be rewritten as:

$$\mathcal{P}_1(S \cap h^{-1}(0)) = (x_j \mathcal{P}_1(S) + (1 - x_j)\mathcal{P}_1(S) + h(x)\mathbb{R}) \cap \mathbb{R}_1[x]. \tag{2}$$

In (2), the non-negativity is now evident. The condition $S \subseteq \{x \in \mathbb{R}^n : 0 \le x_j \le 1\}$ implies that any element of $x_j\mathcal{P}_1(S)$, $(1 - x_j)\mathcal{P}_1(S)$ is non-negative in $S \cap h^{-1}(0)$, while any element of $h(x)\mathbb{R}$ vanishes in $S \cap h^{-1}(0)$.

For higher degree polynomials there are more elaborate certificates of non-negativity such as the classical Pólya's Theorem [8], and the more modern Schmüdgen's Theorem [15] and Putinar's Theorem [13]. These theorems give certificates of non-negativity for polynomials over a given basic semi-algebraic set $\{x : g_i(x) \ge 0, i = 1, \ldots, m\}$ in terms of the preordering or the quadratic module generated by $g_1, \ldots, g_m$ respectively.

Compared to these more elaborated certificates, the certificate of non-negativity (2) has a noteworthy characteristic; namely, that like Farkas Lemma, and the S-Lemma, the certificate is constructed using polynomials of bounded degree (i.e., the polynomials in the right-hand of (2) have bounded degree). This is possible thanks to (2) being written in terms of polynomials that are non-negative on the "simple" set S to certify the non-negativity of polynomials in the more "complex" set $S \cap h^{-1}(0)$.

As mentioned earlier, equation (2) is the key result behind the lift-and-project procedure in [1]. A sequential application of (2) yields a convexification procedure for mixed-integer linear programs. Moreover, an algorithmic implementation of the lift-and-project procedure constituted the foundation for the development of the general purpose Branch-Bound-and-Cut algorithms that are so successfully used today to solve general mixed-integer programming problems (a fascinating account of these developments can be found in [5]).

Naturally, the question arises of whether results similar to (2) can be found for more general sets, and whether such results could be used algorithmically to improve solution methods for more general problems. The first question was positively answered when the set S is compact in [12]. Specifically, from Theorem 1 and Corollary 2 in [12] the following result readily follows.

Theorem 1. *Assume $S \subseteq \mathbb{R}^n$ is compact and $h \in \mathcal{P}_d(S)$. Then*

$$\mathcal{P}_d(S \cap h^{-1}(0)) = \text{closure}(\mathcal{P}_d(S) + h(x)\mathbb{R}_{d-\deg(h)}[x]). \quad (3)$$

Preliminary results on the algorithmic use of Theorem 1 to improve solution methods for non-convex quadratic binary programs were recently presented in [7].

Our main contribution is to show that under a suitable additional condition, Theorem 1 also holds in the more general case when the set S is unbounded. To get an idea of what this condition might be, notice that one inclusion in (3) holds for any S; namely $\mathcal{P}_d(S \cap h^{-1}(0)) \supseteq \text{closure}(\mathcal{P}_d(S) + h(x)\mathbb{R}_{d-\deg(h)}[x])$. On the other hand, the other inclusion in Theorem 1 may fail when S is unbounded. To see this, consider the following case and the generic counterexample in Section 7.

Example 2. *Let $d = 2$, $S = \mathbb{R}^2_+$, and $h(x) = x_1x_2 + 1$. Note that because $S \cap h^{-1}(0) = \emptyset$, then $\mathcal{P}_2(S \cap h^{-1}(0)) = \mathbb{R}_2[x]$. On the other hand, for any $t \in \mathbb{R}_+$ consider the point $x^t = (t, -1/t)$, and notice that $h(x^t) = 0$, and $\lim_{t\to\infty} x^t = (\infty, 0) \in S$ (loosely speaking). So although $h(x)$ does not have a zero in S, it has a "zero at infinity" in S. This zero at infinity "limits" the polynomials in $\mathcal{P}_d(S) + h(x)\mathbb{R}_{d-\deg(h)}[x]$, since for any $p \in \mathcal{P}_d(S) + h(x)\mathbb{R}_{d-\deg(h)}[x]$, $\lim_{t\to\infty} p(x^t) \geq 0$. Thus, for example $-x_1^2 \in \mathcal{P}_2(S \cap h^{-1}(0))$, but $-x_1^2 \notin \mathcal{P}_d(S) + h(x)\mathbb{R}_{d-\deg(h)}[x]$ as $\lim_{t\to\infty} -(x_1^t)^2 = \infty$.*

So in order for Theorem 1 to hold when S is unbounded, a condition is needed on the zeros at infinity of $h(x)$ in S. This condition is formally stated below.

Given a polynomial $h \in \mathbb{R}[x]$ let $\tilde{h}(x)$ denote the homogeneous component of h of highest total degree. In other words, $\tilde{h}(x)$ is obtained by dropping from h the terms whose total degree is less than $\deg(h)$. Recall that given $S \subseteq \mathbb{R}^n$ the *horizon cone* S^∞ is defined as (see, e.g., [14]):

$$S^\infty := \{y \in \mathbb{R}^n : \text{ there exist } x^k \in S,\ \lambda^k \in \mathbb{R}_+,\ k = 1, 2, \ldots \text{ such that } \lambda^k \downarrow 0 \text{ and } \lambda^k x^k \to y\}.$$

We are now ready to state our main result, namely a certificate of non-negativity for elements of $\mathcal{P}_d(S \cap h^{-1}(0))$ in terms of $\mathcal{P}_d(S)$ for the general case when S might be unbounded.

Theorem 3. *Assume $K \subseteq \mathbb{R}^n$ is a closed convex pointed cone. Let $S \subseteq K$ be a closed set and $h \in \mathcal{P}_d(S)$ be such that*

$$(S \cap h^{-1}(0))^\infty = S^\infty \cap \tilde{h}^{-1}(0). \quad (4)$$

Then

$$\mathcal{P}_d(S \cap h^{-1}(0)) = \text{closure}(\mathcal{P}_d(S) + h(x)\mathbb{R}_{d-\deg(h)}[x]). \quad (5)$$

Proof. See Section 6. □

As it is shown in Section 6.2, the hypotheses of Theorem 3 allow us to use previous results from [12] in its proof . The special case $K = \mathbb{R}^n_+$ leads to some interesting consequences that we discuss in Section 4.

Condition (4) concerns the behavior of "zeros at infinity" of the polynomial $h(x)$ on the set S. We show in Section 7 that the statement of Theorem 3 generically fails when this assumption is violated. The condition $h \in \mathcal{P}_d(S)$ can be replaced by $\deg(h) \leq d/2$ as shown in Corollary 4. Section 3 further elaborates on these conditions.

Corollary 4. *The statement of Theorem 3 holds if the hypothesis $h \in \mathcal{P}_d(S)$ is changed to* $\deg(h) \leq d/2$.

Proof. Let $h_1(x) = h(x)^2 \in \mathcal{P}_d(S)$. We have

$$\begin{aligned}(S \cap h_1^{-1}(0))^\infty &= (S \cap h^{-1}(0))^\infty \\ &= S^\infty \cap \tilde{h}^{-1}(0) && \text{by assumption} \\ &= S^\infty \cap \tilde{h}_1^{-1}(0) && \text{as } \tilde{h}_1 = \tilde{h}^2.\end{aligned}$$

Applying Theorem 3, we obtain

$$\begin{aligned}\mathcal{P}_d(S \cap h^{-1}(0)) &= \mathcal{P}_d(S \cap h_1^{-1}(0)) \\ &= \text{closure}(\mathcal{P}_d(S) + h_1(x)\mathbb{R}_{d-\deg(h_1)}[x]) \\ &\subseteq \text{closure}(\mathcal{P}_d(S) + h(x)\mathbb{R}_{d-\deg(h)}[x]) \\ &\subseteq \mathcal{P}_d(S \cap h^{-1}(0)).\end{aligned}$$

□

Remark 5. By repeatedly applying Theorem 3 and Corollary 4, we obtain the following more general version of (5). Assume $K \subseteq \mathbb{R}^n$ is a closed convex pointed cone. Let $S \subseteq K$ be a closed set. Let $h_i \in \mathbb{R}_d[x]$ be given and put $S_i = \{x \in S : h_j(x) = 0, j < i\}$. Assume that for each $i = 1, \ldots, m$

(i) $h_i \in \mathcal{P}_d(S_i)$ or $\deg(h_i) \leq d/2$.

(ii) $(S_i \cap h_i^{-1}(0))^\infty = S_i^\infty \cap \tilde{h}_i^{-1}(0)$.

Then

$$\mathcal{P}_d\left(S \cap \bigcap_{i=1}^m h_i^{-1}(0)\right) = \text{closure}\left(\mathcal{P}_d(S) + \sum_{i=1}^m h_i(x)\mathbb{R}_{d-\deg(h_i)}[x]\right).$$

3 About the "zeros at infinity" condition

We next present some basic results that shed light into the condition (4) in Theorem 3. These results will also allow us to illustrate some particular applications of Theorem 3 in Section 5.

The following proposition shows that in order to check if condition (4) holds, it is only necessary to check one inclusion.

Proposition 6. *For any $S \subseteq \mathbb{R}^n$ and $h \in \mathbb{R}[x]$ we have $(S \cap h^{-1}(0))^\infty \subseteq S^\infty \cap \tilde{h}^{-1}(0)$.*

Proof. Let $d = \deg(h)$, and assume $y \in (S \cap h^{-1}(0))^\infty$. Then there are sequences $x^k \in S, \lambda^k \in \mathbb{R}_+,\ k = 1, \ldots$ such that $h(x^k) = 0$, $\lambda^k \downarrow 0$ and $\lambda^k x^k \to y$. Thus, in particular, $y \in S^\infty$. On the other hand, for $\ell < d$ let $f_\ell(x)$ be the homogeneous component of $h(x)$ of degree ℓ. We have that

$$\tilde{h}(y) = h_d(y) = \lim_{k\to\infty}(\lambda^k)^d h_d(x^k) = \lim_{k\to\infty}(\lambda^k)^d\left(h(x^k) - \sum_{\ell<d} f_\ell(x^k)\right) = \lim_{k\to\infty}\sum_{\ell<d}(\lambda^k)^{d-\ell} f_\ell(\lambda^k x^k) = 0.$$

□

The next proposition shows that the horizon cone of a non-empty polyhedron is the *recession cone* of the polyhedron.

Proposition 7. *Let $A \in \mathbb{R}^{n\times m}$, $b \in \mathbb{R}^m$, and $L = \{x \in \mathbb{R}^n : Ax \geq b\}$. If $L \neq \emptyset$, then*

$$L^\infty = \{y \in \mathbb{R}^n : Ay \geq 0\}.$$

Proof. Assume $y \in L^\infty$, then there exist sequences $\lambda^k \geq 0$, $\lambda^k \downarrow 0$, $x^k \in L$, such that $\lambda^k x^k \to y$. It follows that $Ay = \lim_{k\to\infty} A(\lambda^k x^k) = \lim_{k\to\infty} \lambda^k(Ax^k) \geq \lim_{k\to\infty} \lambda^k b = 0$. Now, assume $y \in \mathbb{R}^n$ is such that $Ay \geq 0$. Let $x \in L$. Define $\lambda^k = 1/k$, and $x_k = x + ky$. Notice that $Ax^k = A(x+ky) = Ax \geq b$, and $\lim_{k\to\infty} \lambda^k x_k = \lim_{k\to\infty}(1/k)(x+ky) = y$. Thus $y \in L^\infty$. □

Proposition 9 below provides sufficient conditions for (4) when S is a finite union of polyhedra. This result allows us to apply the convexification procedure presented in Section 4 to linearly constrained mixed-integer programs with a non-linear objective. Its proof relies on the following lemma.

Lemma 8. *Let $S \in \mathbb{R}^n$, and $\mathcal{I} \subseteq \{1,\ldots,n\}$. If there exist $M \in \mathbb{R}$ such that $|x_\mathcal{I}| \leq M$ for all $x \in S$, then $S^\infty \subseteq \{y \in \mathbb{R}^n : y_\mathcal{I} = 0\}$.*

Proof. Assume $y \in S^\infty$, then there exist sequences $\lambda^k \geq 0$, $\lambda^k \downarrow 0$, and $x^k \in S$ such that $\lambda^k x^k \to y$. In particular, $|y_\mathcal{I}| = \lim_{k\to\infty} \lambda^k |x^k_\mathcal{I}| \leq \lim_{k\to\infty} \lambda^k M = 0$. Thus $y_\mathcal{I} = 0$. □

Proposition 9. *Let $L \subseteq \mathbb{R}^n$ be a polyhedron, and for $i = 1,\ldots,m$ let $\mathcal{I}_i \subseteq \{1,\ldots,n\}$, and $c_i \in \mathbb{R}^{\mathcal{I}_i}$. Put $L_i := \{x \in L : x_{\mathcal{I}_i} = c_i\}$ for $i = 1,\ldots,m$. Assume*

$$\hat{L} = \bigcup_{i=1}^m L_i \neq \emptyset,$$

and assume also that there exist $M_i \in \mathbb{R}$, $i = 1,\ldots,m$ such that $|x_{\mathcal{I}_i}| \leq M_i$ for all $x \in L$, and all $i = 1,\ldots,m$. Then

$$\hat{L}^\infty = L^\infty.$$

Proof. $\hat{L} \subseteq L$ and thus $\hat{L}^\infty \subseteq L^\infty$. On the other hand, since $\hat{L} \neq \emptyset$, there exists $i \in \{1,\ldots,m\}$ such that $\hat{L} \supseteq L_i \neq \emptyset$. Therefore,

$$\begin{aligned} \hat{L}^\infty &\supseteq {L_i}^\infty \\ &= L^\infty \cap \{x \in \mathbb{R}^n : x_{\mathcal{I}_i} = 0\} \quad \text{(by Proposition 7)} \\ &= L^\infty \quad \text{(by Lemma 8).} \end{aligned}$$

□

4 A canonical convexification procedure

One of the consequences of Theorem 3 is a canonical convexification procedure for polynomial optimization problems. Observe that $\mathbb{R}_d[x]$ is a vector space of dimension $N(n,d) := \binom{n+d}{d}$ over $\mathbb{R}$. Assume $\{\phi_i : i = 1,\ldots,N(n,d)\}$ is a basis of $\mathbb{R}_d[x]$ and let $M_d : \mathbb{R}^n \to \mathbb{R}^{N(n,d)}$ be the mapping $a \mapsto (\phi_i(a))_{i=1,\ldots,N(n,d)}$. Assume also that $\mathbb{R}^{N(n,d)}$ is endowed with an inner product $\langle\cdot,\cdot\rangle$ and let $C_d : \mathbb{R}_d[x] \to \mathbb{R}^{N(n,d)}$ be the corresponding *coefficient mapping* such that for all $p \in \mathbb{R}_d[x]$ and $a \in \mathbb{R}^n$

$$p(a) = \langle C_d(p), M_d(a)\rangle.$$

Consider the problem

$$\begin{array}{lll} \inf & q(x) & \\ \text{s.t.} & h_i(x) = 0 \quad i = 1,\ldots,m & \\ & x \geq 0, & \end{array} \tag{6}$$

for some given polynomials $q, h_1, \dots, h_m \in \mathbb{R}_d[x]$. This problem can be rewritten as

$$\begin{array}{ll} \inf & \langle C_d(q), M_d(x)\rangle \\ \text{s.t.} & \langle C_d(h_i), M_d(x)\rangle = 0 \quad i = 1, \dots, m \\ & x \geq 0, \end{array}$$

which in turn can be rewritten as

$$\begin{array}{lll} \inf & \langle C_d(q), Y\rangle & \\ \text{s.t.} & \langle C_d(h_i), Y\rangle = 0 & i = 1, \dots, m \\ & Y \in \{M_d(x) : x \geq 0\}. & \end{array} \tag{7}$$

The latter reformulation suggests the following natural convexification procedure. Consider the cone of *copositive forms* of degree d, that is,

$$\text{Copos}(n, d) := \{C_d(p) : p \in \mathcal{P}_d(\mathbb{R}^n_+)\}.$$

Notice that the cone of copositive matrices can be identified with $\text{Copos}(n, 2)$. The cone of *completely positive d-forms* is the dual cone $\text{Copos}(n, d)^*$. It is easy to see that

$$\text{Copos}(n, d)^* = \text{coco}\{M_d(x) : x \geq 0\},$$

where coco stands for the conic convex hull. The following result formalizes an equivalence between (6) and a canonical convexification of (7).

Theorem 10. *Let $q, h_1, \dots, h_m \in \mathbb{R}_d[x]$ be such that for $i = 1, \dots, m$*

(i) $\deg(h_i) = d$ *and* $h_i \in \mathcal{P}_d(S_i)$, *where* $S_i = \{x \in \mathbb{R}^n_+ : h_j(x) = 0, j < i\}$, *and*

(ii) $\left(S_i \cap h_i^{-1}(0)\right)^\infty = S_i^\infty \cap \tilde{h}_i^{-1}(0)$.

Then the problem (6) *is equivalent to the linear conic program*

$$\begin{array}{lll} \inf & \langle C_d(q), Y\rangle & \\ \text{s.t.} & \langle C_d(h_i), Y\rangle = 0, & i = 1, \dots, m \\ & \langle C_d(1), Y\rangle = 1 & \\ & Y \in \text{Copos}(n, d)^*. & \end{array} \tag{8}$$

More precisely:

(a) *The optimal values of* (6) *and* (8) *are the same.*

(b) *One of the two problems* (6) *and* (8) *attains its optimal value if and only if the other one does. When that is the case, the set of optimal solutions to* (8) *is*

$$\text{conv}\{M_d(x) : x \textit{ is an optimal solution to } (6)\}.$$

Proof. See Section 6. ☐

Remark 11. The condition $\deg(h_i) = d$ in Theorem 10 can be relaxed to $\deg(h_i) \leq d$. In this case problem (6) is equivalent to

$$\begin{array}{lll} \inf & \langle C_d(q), Y\rangle & \\ \text{s.t.} & \langle C_d(g_i h_i), Y\rangle = 0, & i = 1, \dots, m \\ & \langle C_d(1), Y\rangle = 1 & \\ & Y \in \text{Copos}(n, d)^* & \end{array}$$

where each $g_i \in \mathbb{R}_{d-\deg(h_i)}[x]$, $i = 1, \dots, m$ is any polynomial satisfying the following two conditions:

(i) $\mathbb{R}^n_+ \cap (g_i h_i)^{-1}(0) = \mathbb{R}^n_+ \cap h_i^{-1}(0)$

(ii) $\mathbb{R}^n_+ \cap (\widetilde{g_i h_i})^{-1}(0) = \mathbb{R}^n_+ \cap \tilde{h}_i^{-1}(0)$.

For instance, each g_i can be chosen to be

$$g_i(x) = (x_1 + \cdots + x_n + 1)^{d - \deg(h_i)}.$$

Furthermore, if $\deg(h_i) \leq d/2$ then the condition $h_i \in \mathcal{P}_d(S_i)$ can be dropped as well, just as in Corollary 4.

Condition (i) in Remark 11 ensures that the feasible set of (6) does not change if each constraint $h_i(x) = 0,\ i = 1, \ldots, m$ is replaced with $g_i(x)h_i(x) = 0$. Condition (ii) ensures that $\left(S_i \cap h_i^{-1}(0)\right)^\infty = S_i^\infty \cap \tilde{h}_i^{-1}(0)$ if and only if $\left(S_i \cap (g_i h_i)^{-1}(0)\right)^\infty = S_i^\infty \cap (\widetilde{g_i h_i})^{-1}(0)$ for $i = 1, \ldots, m$.

The convexification procedure presented in Theorem 10 for the equality constrained problem (6) has a natural extension for inequality constrained problems.

Theorem 12. *Consider the problem*

$$\begin{array}{ll} \inf & q(x) \\ \text{s.t.} & g_i(x) \geq 0 \quad \forall i \in \mathcal{I} \\ & h_j(x) = 0 \quad \forall j \in \mathcal{J} \\ & x \geq 0, \end{array} \tag{9}$$

Suppose $\{1, \ldots, m\} = \mathcal{I} \cup \mathcal{J}$ *is an ordering of the set of indices* $\mathcal{I} \cup \mathcal{J}$ *such that for*

$$U_k = \begin{cases} \{x \geq 0 : g_k(x) \geq 0\} & \text{if } k \in \mathcal{I} \\ \{x \geq 0 : h_k(x) = 0\} & \text{if } k \in \mathcal{J} \end{cases}, \quad \text{and } S_k = \bigcap_{j<k} U_j$$

the following conditions hold. For every $k \in \mathcal{I}$,

(i) $\deg(h_k) = d$

(ii) $\left(S_k \cap h_k^{-1}(0)\right)^\infty = S_k^\infty \cap \tilde{h}_k^{-1}(0)$,

and for every $k \in \mathcal{J}$,

(iii) $\deg(g_k) = d$

(iv) $\left(S_k \cap g_k^{-1}(\mathbb{R}_+)\right)^\infty = S_k^\infty \cap \tilde{g}_k^{-1}(\mathbb{R}_+)$.

Then, the problem (9) *is equivalent to the linear conic program*

$$\begin{array}{lll} \inf & \langle C_{2d}(q(x)), Y \rangle & \\ \text{s.t.} & \langle C_{2d}((g_i(x) - t_i^d)^2), Y \rangle = 0, & i \in \mathcal{I} \\ & \langle C_{2d}(h_j^2(x)), Y \rangle = 0, & j \in \mathcal{J} \\ & \langle C_{2d}(1), Y \rangle = 1 & \\ & Y \in \text{Copos}(n + |\mathcal{I}|, 2d)^*. & \end{array} \tag{10}$$

Proof. We reduce this case to the case of Theorem 10, where we only have equalities. To do so, we use a new slack variable t_i for each inequality. That is, for each $i \in \mathcal{I}$ let $G_i(x,t) = g_i(x) - t_i^d$, where t is a vector of $|\mathcal{I}|$ new variables. We can reformulate (9) as

$$\begin{array}{ll} \inf & q(x) \\ \text{s.t.} & G_i(x,t) = 0 \quad \forall i \in \mathcal{I} \\ & h_j(x) = 0 \qquad \forall j \in \mathcal{J} \\ & x, t \geq 0. \end{array} \tag{11}$$

Notice that for all $i \in \mathcal{I}$ we have $\deg(G_i(x,t)) = d$. Hence by Remark 11 it is enough to show that

$$\{S_i \times \mathbb{R}^m_+ \cap G_i^{-1}(0)\}^\infty = (S_i \times \mathbb{R}^m_+)^\infty \cap \tilde{G}_i^{-1}(0).$$

From Proposition 6, it suffices to show

$$(S_i \times \mathbb{R}^m_+)^\infty \cap \tilde{G}_i^{-1}(0) \subseteq \{S_i \times \mathbb{R}^m_+ \cap G_i^{-1}(0)\}^\infty. \tag{12}$$

To do this, assume $(x,t) \in (S_i \times \mathbb{R}^m_+)^\infty$ is such that $\tilde{G}_i(x,t) = 0$. Then $x \in S_i^\infty$ and $\tilde{g}_i(x) = t_i^d \geq 0$. By condition (iv), $x \in \left(S_i \cap g_i^{-1}(\mathbb{R}_+)\right)^\infty$.

Hence there exist $x^{(k)} \in S$ and $\lambda^{(k)} \downarrow 0$ such that $\lambda^{(k)}x^{(k)} \to x$ and $g_i(x^{(k)}) \geq 0$. Let $t^{(k)}$ be defined by

$$t_j^{(k)} = \begin{cases} g_i(x^{(k)})^{1/d} & \text{if } j = i \\ t_j/\lambda^{(k)} & \text{otherwise.} \end{cases}$$

Then $t^{(k)} \in \mathbb{R}^m_+$ and $G_i(x^{(k)}, t^{(k)}) = g_i(x^{(k)}) - (t_i^{(k)})^d = 0$. To finish the proof of (12) we need to show $\lambda^{(k)}t^{(k)} \to t$. For $j \neq i$, $t_j^{(k)}\lambda^{(k)} = t_j$. Also $\lambda^{(k)}t_i^{(k)} = g_i(x^{(k)})^{1/d}\lambda^{(k)}$. For $\ell \leq d$ let $f_\ell(x)$ be the homogeneous component of $g_i(x)$ of degree ℓ. We have then,

$$(\lambda^{(k)})^d(g_i(x^{(k)}) - \tilde{g}_i(x^{(k)})) = \sum_{\ell<d}(\lambda^{(k)})^d f_\ell(x^{(k)}) = \sum_{\ell<d}(\lambda^{(k)})^{d-\ell} f_\ell(x^{(k)}\lambda^{(k)}) \to 0.$$

Thus,

$$\begin{aligned} \lim_{k\to\infty} t_i^{(k)}\lambda^{(k)} &= \lim_{k\to\infty} g_i(x^{(k)})^{1/d}\lambda^{(k)} = \left(\lim_{k\to\infty} \tilde{g}_i(\lambda^{(k)}x^{(k)}) + (\lambda^{(k)})^d(g_i(x^{(k)}) - \tilde{g}_i(x^{(k)}))\right)^{1/d} \\ &= \tilde{g}_i(x)^{1/d} = t_i. \end{aligned}$$

□

Remark 13. Proceeding as in Remark 11, the condition $\deg(h_i) = d$ in Theorem 10 can be relaxed to $\deg(h_i) \leq d$.

5 Examples

5.1 Mixed binary (non-convex) quadratic programming

Let $q(x) = x^{\mathrm{T}}Qx + 2c^{\mathrm{T}}x$ and consider the problem

$$\begin{array}{ll} \min & q(x) \\ & a_i^{\mathrm{T}}x = b_i,\ i = 1,\ldots,m_\ell \\ & x_j \in \{0,1\},\ j = 1,\ldots,m_b \\ & x \geq 0, \end{array} \tag{13}$$

where $m_b \leq n$, the number of variables.

Proceeding as in [3], we assume the following two conditions hold. First, (13) is feasible and second, for each $j = 1,\ldots,m_b$

$$x \geq 0,\ a_i^{\mathrm{T}}x = b_i,\ i = 1,\ldots,m_\ell \Rightarrow 0 \leq x_j \leq 1. \tag{14}$$

This ensures that for $j = 1,\ldots,m_b$

$$x_j(1 - x_j) \in \mathcal{P}_2(\mathbb{R}^n_+ \cap \{x : a_i^{\mathrm{T}}x = b_i,\ i = 1,\ldots,m_\ell\}). \tag{15}$$

Rewrite (13) as

$$\begin{array}{ll} \min & q(x) \\ & (a_i^{\mathrm{T}}x - b_i)^2 = 0,\ i = 1, \ldots, m_\ell \\ & x_j(1 - x_j) = 0,\ j = 1, \ldots, m_b \\ & x \geq 0. \end{array} \tag{16}$$

Problem (16) satisfies the conditions of Theorem 10. Indeed, condition (i) follows readily from (15) and the fact that $(a_i^{\mathrm{T}}x - b_i)^2 \in \mathcal{P}_2(\mathbb{R}^n_+)$ for $i = 1, \ldots, m_\ell$. Condition (ii) holds for the first m_ℓ constraints by Proposition 7. Condition (ii) also holds for the remaining m_b constraints by Proposition 9 and (14). It thus follows from Theorem 10 that (16) is equivalent to the linear conic program

$$\begin{array}{lll} \inf & \langle C_2(q(x)), Y\rangle & \\ \text{s.t.} & \langle C_2((a_i^{\mathrm{T}}x - b_i)^2), Y\rangle = 0 & i = 1, \ldots, m_\ell \\ & \langle C_2(x_j(1 - x_j)), Y\rangle = 0 & j = 1, \ldots, m_b \\ & \langle C_2(1), Y\rangle = 1 & \\ & Y \in \mathrm{Copos}(n, 2)^*. & \end{array} \tag{17}$$

It is easy to show that (17) is equivalent to the dual copositive programming formulation for (13) derived by Burer [3].

5.2 Mixed binary quadratically constrained quadratic programming

Now we consider the extension of (13) allowing quadratic constraints

$$\begin{array}{ll} \min & q_0(x) \\ & q_k(x) = 0,\ k = 1, \ldots, m_q \\ & a_i^{\mathrm{T}}x = b_i,\ i = 1, \ldots, m_\ell \\ & x_j \in \{0, 1\},\ j = 1, \ldots, m_b \\ & x \geq 0, \end{array} \tag{18}$$

where $m_b \leq n$, and $q_i \in \mathbb{R}_2[x]$ for $i = 0, \ldots, m_q$.

Burer [3] asks what we need to assume about the quadratic constraints in order to have the equivalence between (18) and its copositive reformulation (19)

$$\begin{array}{lll} \inf & \langle C_2(q_0(x)), Y\rangle & \\ \text{s.t.} & \langle C_2(q_k(x)), Y\rangle = 0 & i = 1, \ldots, m_q \\ & \langle C_2((a_i^{\mathrm{T}}x - b_i)^2), Y\rangle = 0 & i = 1, \ldots, m_\ell \\ & \langle C_2(x_j(1 - x_j)), Y\rangle = 0 & j = 1, \ldots, m_b \\ & \langle C_2(1), Y\rangle = 1 & \\ & Y \in \mathrm{Copos}(n, 2)^*. & \end{array} \tag{19}$$

As the generic counterexample in Section 7 shows, condition (4) provides the precise answer to this question.

Let $L := \{x \geq 0 : a_i^{\mathrm{T}}x = b_i,\ i = 1, \ldots, m_\ell\}$ and $S = \{x \in L : x_j \in \{0, 1\},\ j = 1, \ldots, m_b\}$. Burer [3] shows that if for all $k = 1, \ldots, m_q$

$$q_k \in \mathcal{P}_2(L), \text{ and } d \in L^\infty, d_j \neq 0 \Rightarrow \frac{\partial q_k(x)}{\partial x_j} = 0, \tag{20}$$

then (18) is equivalent to (19). Condition (20) implies the assumptions of Theorem 10. The following example shows that (20) is actually more restrictive than (4). Consider

$$\begin{array}{ll} \min & x^2+4x \\ & q(x,y,z) := (x-2)^2 + y(2x-3) = 0 \\ & (y+z-1)^2 = 0 \\ & x,y,z \geq 0. \end{array} \tag{21}$$

Let $L = \{(x,y,z) \geq 0 : y+z=1\}$. Notice $(x,y,z) \in L \Rightarrow 0 \leq y \leq 1$ and then $q(x,y,z) = (1-y)(x-2)^2 + y(x-1)^2 \in P_2(L)$. Also, $L^\infty = \{(x,0,0) : x \geq 0\}$, and thus (20) is not satisfied in this case. However,

$$\begin{aligned} (L \cap q^{-1}(0))^\infty = \{(1,1,0),(2,0,1)\}^\infty = \{(0,0,0)\} = \\ L^\infty \cap \{(x,y,z) : x^2 - 2yx = 0\} = L^\infty \cap \tilde{q}^{-1}(0). \end{aligned}$$

Thus, setting $S_1 = \mathbb{R}^3_+$, $h_1(x,y,z) = (y+z-1)^2$, $S_2 = L$, and $h_2(x,y,z) = q(x,y,z)$, it follows from Theorem 10 that (21) is equivalent to the copositive reformulation given by (19).

5.3 An example with $d = 3$

Next we consider MAX3SAT. Given a 3CNF (i.e. with at most 3 variables per clause) formula Φ with n variables and m clauses, we want to know the maximum number of clauses in Φ that can be satisfied simultaneously by a truth assignment.

We associate to Φ a polynomial $P_\Phi(x,y)$ with $2n$ variables and m monomials. To do this, with a literal X_i we associate the variable x_i and with $\neg X_i$ we associate y_i. With a clause we associate the monomial formed by the product of the variables associated with its literals. With Φ we associated $P_\Phi(x,y)$, the polynomial formed by the sum of the monomials associated with the clauses from Φ.

Given a truth assignment $\sigma : \{x_1,\ldots,x_n\} \to \{0,1\}$, $P_\Phi(1-\sigma,\sigma)$ is the number of clauses not satisfied by σ. Thus, the maximum number of clauses in Φ that any assignment can satisfy is equal to

$$\begin{array}{lll} \max & m - P_\Phi(x,y) & \\ \text{s.t.} & x_i + y_i = 1 & i=1,\ldots,n \\ & x_i y_i = 0 & i = 1,\ldots,n \\ & x,y \geq 0. & \end{array} \tag{22}$$

As suggested by Remark 11, problem (22) can be rewritten as

$$\begin{array}{lll} \max & m - P_\Phi(x,y) & \\ \text{s.t.} & (x_i+y_i+1)(x_i+y_i-1)^2 = 0 & i=1,\ldots,n \\ & x_i^2 y_i = 0 & i = 1,\ldots,n \\ & x,y \geq 0. & \end{array} \tag{23}$$

Notice that for every i

$$x_i^2 y_i, (x_i+y_i+1)(x_i+y_i-1)^2 \in P_3(\mathbb{R}^{2n}_+).$$

By Proposition 7 we have for every i and every polyhedron $S \subseteq \mathbb{R}^{2n}$,

$$\begin{aligned} \{(x,y) \in S : (x_i+y_i+1)(x_i+y_i-1)^2 = 0\}^\infty &= \{(x,y) \in S : x_i + y_i = 1\}^\infty \\ &= \{(c,d) \in S^\infty : c_i + d_i = 0\} \\ &= \{(c,d) \in S^\infty : (c_i+d_i)^3 = 0\}. \end{aligned}$$

Let $S_i = \{(x,y) \in \mathbb{R}^{2n}_+ : x_j + y_j = 1,\ j = 1,\dots,n \text{ and } x_j y_j = 0,\ j = 1,\dots,i-1\}$. By Proposition 9 we have for every i,

$$\begin{aligned}
\{(x,y) \in S_i : x_i^2 y_i = 0\}^\infty &= (\{(x,y) \in S_i : x_i = 0\} \cup \{(x,y) \in S_i : y_i = 0\})^\infty \\
&= \{(x,y) \in S_i : x_i = 0\}^\infty \cup \{(x,y) \in S_i : y_i = 0\}^\infty \\
&= \{(c,d) \in S_i^\infty : c_i = 0\} \cup \{(c,d) \in S_i^\infty : d_i = 0\} \\
&= \{(c,d) \in S_i^\infty : c_i^2 d_i = 0\}.
\end{aligned}$$

Using Theorem 10 the problem (23) is equivalent to the linear conic program

$$\begin{array}{lll}
\inf & \langle C_3(m - P_\Phi(x,y)), Z\rangle & \\
\text{s.t.} & \langle C_3((x_i + y_i + 1)(x_i + y_i - 1)^2), Z\rangle = 0 & i = 1,\dots,n \\
& \langle C_3(x_i^2 y_i, Z\rangle = 0 & i = 1,\dots,n \\
& \langle C_3(1), Z\rangle = 1 & \\
& Z \in \text{Copos}(2n,3)^*. &
\end{array}$$

5.4 On the non-negativity assumption of the restrictions

Now we show that the assumption $h_j \in \mathcal{P}_d(S_j)$ in Theorem 10 can not be dropped. Consider

$$\begin{array}{ll}
\min & 4x - y - 2x^2 - 2xy - y^2 \\
& x^2 - xy = 0 \\
& y^2 - y = 0 \\
& x, y \geq 0,
\end{array} \tag{24}$$

Notice that the nonegativity condition does not hold for any of the two constraints. The related copositive program is

$$\begin{array}{ll}
\inf & \left\langle \begin{bmatrix} 0 & 2 & -1/2 \\ 2 & -2 & -1 \\ -1/2 & -1 & -1 \end{bmatrix}, Y \right\rangle \\
\text{s.t.} & \left\langle \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1/2 \\ 0 & -1/2 & 0 \end{bmatrix}, Y \right\rangle = 0 \\
& \left\langle \begin{bmatrix} 0 & 0 & -1/2 \\ 0 & 0 & 0 \\ -1/2 & 0 & 1 \end{bmatrix}, Y \right\rangle = 0 \\
& \left\langle \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, Y \right\rangle = 1 \\
& Y \in \text{Copos}(2,2)^*.
\end{array} \tag{25}$$

It is easy to check that the feasible set of (24) is $\{(0,0),(0,1),(1,1)\}$, where $(1,1)$ is optimal with optimal value -2. On the other hand it is easy to check that

$$Y = \frac{1}{8}\begin{bmatrix} 8 & 2 & 4+2\sqrt{2} \\ 2 & 2+\sqrt{2} & 2+\sqrt{2} \\ 4+2\sqrt{2} & 2+\sqrt{2} & 4+2\sqrt{2} \end{bmatrix}$$

is feasible for (25) and has objective value $-1-\sqrt{2} < -2$. Thus the two problems are not equivalent. Actually it can be shown that Y is optimal for (25).

To obtain the non-negativity of the constraints it is always possible to increase the value of d. For example, we can rewrite (24) as

$$\begin{array}{ll}
\min & q_0 := 4x - y - 2x^2 - 2xy - y^2 \\
& q_1(x,y) := (x^2 - xy)^2 = 0 \\
& q_2(x,y) := (y^2 - y)^2 = 0 \\
& x, y \geq 0,
\end{array} \tag{26}$$

obtaining that both $q_1, q_2 \in P_4(\mathbb{R}^2_+)$. Notice also that

$$(\mathbb{R}^2_+)^\infty \cap \tilde{q}_2^{-1}(0) = \{(x,0) : x \geq 0\} = (\mathbb{R}^2_+ \cap q_2^{-1}(0))^\infty.$$

Also,

$$\{(x,0) : x \geq 0\}^\infty \cap \tilde{q}_1^{-1}(0) = \{(0,0)\} = (\{(x,0) : x \geq 0, q_1(x,0) = 0\})^\infty.$$

Thus applying Theorem 12, (24) is equivalent to

$$\begin{array}{rl} \inf & \langle C_4(q_0(x)), Y\rangle \\ \text{s.t.} & \langle C_4(q_2(x)), Y\rangle = 0 \quad i = 1,2 \\ & \langle C_4(1), Y\rangle = 1 \\ & Y \in \mathrm{Copos}(2,4)^*. \end{array} \tag{27}$$

Note that Proposition 9 implies that the both the *key condition* introduced in [3, eq. (1)] and the *weak key condition* introduced in [2, eq. (4)] are sufficient for condition **(ii)** in Theorem 10 to hold in the special case of linearly constrained mixed-binary quadratic programs. In [2, after Theorem 2.1] an implicit question is raised about why, unlike the *key condition* [3, eq. (1)], the *weak key condition* [2, eq. (4)] is not sufficient for the completely copositive convexification procedure to produce an equivalent formulation for linearly constrained mixed-binary quadratic programs. From Theorem 10, the reason is clear: while the *key condition* ensures that the non-negativity condition **(i)** in Theorem 10 holds, the *weak key condition* does not. However, as the example in this section shows, this can be overcome at the price of increasing the degree of the polynomials involved in the formulation of the problem.

6 Proofs of Theorem 3 and Theorem 10

The proof of Theorem 3 relies on Theorem 1 via a suitable *compactification* procedure.

6.1 Compactification

Assume $K \subseteq \mathbb{R}^n$ is a closed convex pointed cone. Fix a point $a \in \mathbb{R}^n$ such that $a^{\mathrm{T}}x > 0$ for all $x \in K$. For $x \in \mathbb{R}^n$ let

$$\overline{x} := \frac{1}{1 + a^{\mathrm{T}}x}\begin{bmatrix}1\\ x\end{bmatrix}.$$

On the other hand, for $f \in \mathbb{R}_d[x]$, let $\overline{f} \in \mathbb{R}[y_0, y_1, \ldots, y_n]$ be defined by

$$\overline{f}(y_0, y_1, \ldots, y_n) := f(y_1/y_0, \ldots, y_n/y_0) \cdot y_0^{\deg(f)}.$$

Finally, for $S \subseteq K$ let

$$\overline{S} := \mathrm{closure}\{\overline{x} : x \in S\}.$$

The above compactification procedure establishes a parallel between the sets $\mathcal{P}_d(S)$ and $\mathcal{P}_d(\overline{S})$ as it is formally described in the following lemma.

Lemma 14. *Let $p \in \mathbb{R}[x]$ be a polynomial of degree d and $S \subseteq \mathbb{R}^n$. Then*

(i) $p(x) = \overline{p}(\overline{x})(1 + a^{\mathrm{T}}x)^{\deg p}$.

(ii) $p \in \mathcal{P}_d(S)$ *implies* $\overline{p} \in \mathcal{P}_d(\overline{S})$.

(iii) $p \in \mathrm{int}(\mathcal{P}_d(S))$ *implies* $\overline{p} \in \mathrm{int}(\mathcal{P}_d(\overline{S}))$.

Proof. First, (i) readily follows by the construction of $\overline{p}$ and $\overline{x}$. Next, (ii) follows from (i). Finally, to prove (iii) assume $\overline{p} \notin \operatorname{int}(\mathcal{P}_d(\overline{S}))$. Since $\overline{S}$ is compact, there exists $y \in \overline{S}$ such that $\overline{p}(y) \leq 0$. Write $y = \lim_{k\to\infty} \overline{x^k}$ with $x^k \in S$. Thus

$$0 \geq \overline{p}(y) = \lim_{k\to\infty} \overline{p}(\overline{x^k}) = \lim_{k\to\infty} p(x^k)(1 + a^{\mathrm{T}}x^k)^{-\deg p},$$

and hence for any $\varepsilon > 0$ there is N such that $k > N$ implies $p(x^k) < \varepsilon(1 + a^{\mathrm{T}}x^k)^{\deg p}$. Therefore $p \notin \operatorname{int}(\mathcal{P}_d(S))$. □

Our proof will also rely on the following lemma:

Lemma 15.

$$\left(S \cap h^{-1}(0)\right)^{\infty} = S^{\infty} \cap \tilde{h}^{-1}(0) \Leftrightarrow \overline{S \cap h^{-1}(0)} = \overline{S} \cap \overline{h}^{-1}(0).$$

Proof. This will follow from the following observations:

(A) For every $z \in \mathbb{R}^{n+1}$, $z_0 > 0$ implies $z \in \overline{S \cap h^{-1}(0)}$ if and only if $z \in \overline{S} \cap \overline{h}^{-1}(0)$.

(B) For every $y \in \mathbb{R}^n$, $\overline{h}(0, y) = \tilde{h}(y)$.

(C) $S^{\infty} = \{y \in \mathbb{R}^n_{++} : (0, y/a^{\mathrm{T}}y) \in \overline{S}\} \cup \{0\}$.

From (A) it follows that

$$\begin{aligned}\overline{S \cap h^{-1}(0)} = \overline{S} \cap \overline{h}^{-1}(0) \Leftrightarrow &\{y \in \mathbb{R}^n_+ : a^{\mathrm{T}}y = 1,\ (0, y) \in \overline{S \cap h^{-1}(0)}\} \\ &= \{y \in \mathbb{R}^n_+ : a^{\mathrm{T}}y = 1,\ (0, y) \in \overline{S} \cap \overline{h}^{-1}(0)\}.\end{aligned} \tag{28}$$

On the other hand, from (B) it follows that

$$\left(S \cap h^{-1}(0)\right)^{\infty} = \{y \in \mathbb{R}^n_{++} : (0, y/a^{\mathrm{T}}y) \in \overline{S \cap h^{-1}(0)}\} \cup \{0\}. \tag{29}$$

Furthermore, from (B) and (C) it follows that

$$\begin{aligned}S^{\infty} \cap \tilde{h}^{-1}(0) &= \left(\{y \in \mathbb{R}^n_{++} : (0, y/a^{\mathrm{T}}y) \in \overline{S}\} \cup \{0\}\right) \cap \tilde{h}^{-1}(0) \\ &= \{y \in \mathbb{R}^n_{++} : (0, y/a^{\mathrm{T}}y) \in \overline{S} \cap \overline{h}^{-1}(0)\} \cup \{0\}.\end{aligned} \tag{30}$$

Putting (29) and (30) together we get

$$\begin{aligned}\left(S \cap h^{-1}(0)\right)^{\infty} = S^{\infty} \cap \tilde{h}^{-1}(0) \Leftrightarrow &\{y \in \mathbb{R}^n_+ : a^{\mathrm{T}}y = 1,\ (0, y) \in \overline{S \cap h^{-1}(0)}\} \\ &= \{y \in \mathbb{R}^n_+ : a^{\mathrm{T}}y = 1,\ (0, y) \in \overline{S} \cap \overline{h}^{-1}(0)\}.\end{aligned} \tag{31}$$

The equivalence now follows from (28) and (31). □

6.2 Proof of Theorem 3

It is immediate that

$$\mathcal{P}_d(S) + h(x)\mathbb{R}_{d-\deg(f)}[x] \subseteq \mathcal{P}_d(S \cap h^{-1}(0)).$$

and consequently,

$$\operatorname{closure}(\mathcal{P}_d(S) + h(x)\mathbb{R}_{d-\deg(f)}[x]) \subseteq \mathcal{P}_d(S \cap h^{-1}(0)).$$

Next, observe that $\mathcal{P}_d(S \cap h^{-1}(0))$ is a closed convex cone with non-empty interior. Hence $\mathcal{P}_d(S \cap h^{-1}(0)) = \text{closure}\big(\text{int}\big(\mathcal{P}_d(S \cap h^{-1}(0))\big)\big)$. Thus to prove the inclusion $\text{closure}(\mathcal{P}_d(S) + h(x)\mathbb{R}_{d-\deg(f)}[x]) \supseteq \mathcal{P}_d(S \cap h^{-1}(0))$, it suffices to show that

$$\text{int}(\mathcal{P}_d(S \cap h^{-1}(0))) \subseteq \mathcal{P}_d(S) + h(x)\mathbb{R}_{d-\deg(h)}[x]. \tag{32}$$

To that end, assume $p \in \text{int}(\mathcal{P}_d(S \cap h^{-1}(0))$. Then Lemma 14(iii) yields

$$\overline{p} \in \text{int}(\mathcal{P}_d(\overline{S \cap h^{-1}(0)})) = \text{int}(\mathcal{P}_d(\overline{S} \cap \overline{h}^{-1}(0))).$$

But since $\overline{S}$ is compact, Theorem 1 implies that $\text{int}(\mathcal{P}_d(\overline{S} \cap \overline{h}^{-1}(0))) \subseteq \mathcal{P}_d(\overline{S}) + \overline{h}(y)\mathbb{R}_{d-\deg h}[y]$. Therefore,

$$\overline{p} \in \mathcal{P}_d(\overline{S}) + \overline{h}(y)\mathbb{R}_{d-\deg h}[y].$$

To finish notice that if $\overline{p}(y) = r(y) + \overline{h}(y)q(y)$, where $r(y) \in \mathcal{P}_d(\overline{S})$ and $q(y) \in \mathbb{R}_{d-\deg h}[y]$ then by Lemma 14(i)

$$p(x) = \overline{p}(\overline{x})(1 + a^{\mathrm{T}}x)^d = r(\overline{x})(1 + a^{\mathrm{T}}x)^d + h(x)q(\overline{x})(1 + a^{\mathrm{T}}x)^{d-\deg(h)}.$$

Notice that $r(\overline{x})(1 + a^{\mathrm{T}}x)^d \in \mathcal{P}_d(S)$ and $q(\overline{x})(1 + a^{\mathrm{T}}x)^{d-\deg(h)} \in \mathbb{R}_{d-\deg h}[x]$. Consequently $p \in \mathcal{P}_d(S) + h(x)\mathbb{R}_{d-\deg(h)}[x]$, proving (32). □

6.3 Proof of Theorem 10

We will prove the Theorem in the special case $m = 1$. The general case follows via a straightforward induction argument. For ease of notation, we shall write h for h_1. Observe that for $U \subseteq \mathbb{R}^n$ and $q(x) \in \mathbb{R}_d[x]$ we have

$$\begin{aligned}
\inf\{q(x) : x \in U\} &= \inf\{\langle C_d(q), M_d(x)\rangle : x \in U\} \\
&= \inf\{\langle C_d(q), Y\rangle : Y \in M_d(U)\} \\
&= \inf\{\langle C_d(q), Y\rangle : Y \in \text{conv}(M_d(U))\} \\
&= \inf\{\langle C_d(q), Y\rangle : Y \in \text{coco}(M_d(U)),\ \langle C_d(1), Y\rangle = 1\}.
\end{aligned} \tag{33}$$

We claim that to finish the proof it suffices to show

$$\text{coco}\left(M_d(S \cap h^{-1}(0))\right) = \text{coco}\left(M_d(S)\right) \cap \{Y \in \mathbb{R}^N : \langle C_d(h), Y\rangle = 0\}. \tag{34}$$

Here is a proof of the claim. Putting (33) and (34) together, we readily obtain part (a) of Theorem 10. For part (b) of Theorem 10, we simply need to show that if Y is an optimal solution to (8), then $Y \in \text{conv}\{M_d(x) : x$ is an optimal solution to (6)$\}$. Assume Y is indeed an optimal solution to (8). In particular, $Y \in \text{coco}\left(M_d(S)\right) \cap \{Y \in \mathbb{R}^N : \langle C_d(h), Y\rangle = 0\} \cap \{Y \in \mathbb{R}^N : \langle C_d(1), Y\rangle = 1\}$. Hence from (33) and (34) it follows that $Y \in \text{conv}\{M_d(x) :$ x is a feasible solution to (6)$\}$, i.e., there exist $x^j, j = 1, \ldots, K$ feasible solutions to (6) and $\lambda_j > 0, j = 1, \ldots, K$ such that

$$Y = \sum_{j=1}^{K} \lambda_j x^j,\ \sum_{j=1}^{K} \lambda_j = 1.$$

Let v^* be the optimal value of (6), which is the same as the optimal of (8) by part (a) of Theorem 10. Then it follows that

$$v^* = \langle C_d(q), Y\rangle = \sum_{j=1}^{K} \lambda_j q(x^j) \geq \sum_{j=1}^{K} \lambda_j v^* = v^*.$$

Since each $\lambda_j > 0$, we must necessarily have $q(x^j) = v^*$ for each $j = 1, \dots, K$. Therefore $x^j, j = 1, \dots, K$ are optimal solutions to (6) and part (b) of Theorem 10 follows.

Next we provide a proof of (34). Notice that for $U \subseteq \mathbb{R}^n$,

$$\begin{aligned} C_d(\mathcal{P}_d(U)) &= \{C_d(p) : p(u) \geq 0 \text{ for all } u \in U\} \\ &= \{C_d(p) : \langle C_d(p), M_d(u)\rangle \geq 0 \text{ for all } u \in U\} \\ &= \{C_d(p) : \langle C_d(p), Y\rangle \geq 0 \text{ for all } Y \in M_d(U)\} \\ &= M_d(U)^*, \end{aligned}$$

and consequently $C_d(\mathcal{P}_d(U))^* = M_d(U)^{**} = \operatorname{coco} M_d(U)$. Applying this to $U = S \cap h^{-1}(0)$ and using Theorem 3, we get

$$\begin{aligned} \operatorname{coco}\left(M_d(S \cap h^{-1}(0))\right) &= C_d(\mathcal{P}_d(S \cap h^{-1}(0)))^* \\ &= C_d(\operatorname{closure}(\mathcal{P}_d(S) + h(x)\mathbb{R}))^* \qquad \text{(by Theorem 3)} \\ &= \operatorname{closure}(C_d(\mathcal{P}_d(S) + h(x)\mathbb{R}))^* \\ &= C_d(\mathcal{P}_d(S) + h(x)\mathbb{R})^* \\ &= (C_d(\mathcal{P}_d(S)) + C_d(h(x)\mathbb{R}))^* \\ &= \operatorname{coco}(M_d(S) \cap C_d(h(x)\mathbb{R})^*. \end{aligned}$$

To finish, observe that

$$\begin{aligned} C_d(h(x)\mathbb{R})^* &= \{Y \in \mathbb{R}^{N(n,d)} : \langle C_d(ch), Y\rangle \geq 0 \text{ for all } c \in \mathbb{R}\} \\ &= \{Y \in \mathbb{R}^{N(n,d)} : \langle C_d(ch), Y\rangle = 0 \text{ for all } c \in \mathbb{R}\} \\ &= \{Y \in \mathbb{R}^{N(n,d)} : \langle C_d(h), Y\rangle = 0\}. \end{aligned}$$

□

7 A generic counterexample

We next show that indeed the statement of Theorem 3 generically fails if condition (4) is violated. For simplicity assume $K = \mathbb{R}^n_+$, and let $a = e$, that is, the vector of all-ones. Assume condition (4) in Theorem 3 does not hold. Then by Lemma 15(ii) there exists $t \in \Delta_n$ such that $(0, t) \in \bar{S} \cap \bar{h}^{-1}(0) \setminus \overline{S \cap h^{-1}(0)}$. Since $\overline{S \cap h^{-1}(0)}$ is a closed subset of Δ_{n+1} there exists $\epsilon > 0$ such that

$$y \in \Delta_{n+1}, y \in \overline{S \cap h^{-1}(0)} \Rightarrow \|y - (0, t)\| > \epsilon.$$

Take

$$p(x) := (1 + e^{\mathrm{T}}x)^{d-2}\left(1 + \sum_{i=1}^{n}(x_i - t_i(1 + e^{\mathrm{T}}x))^2 - \epsilon^2(1 + e^{\mathrm{T}}x)^2\right).$$

We claim that $p(x) \in \mathcal{P}_d(S \cap h^{-1}(0))$ but $p(x) \notin \operatorname{closure}(\mathcal{P}_d(S) + h(x)\mathbb{R}_{d-\deg(h)}[x])$. To show that, we compactify: First, note that if $y \in \overline{S \cap h^{-1}(0)} \subseteq \Delta_{n+1}$ we have

$$\bar{p}(y) = \|y - (0, t)\|^2 - \epsilon^2 > 0.$$

Thus by Lemma 14(i), $p(x) \in \mathcal{P}_d(S \cap h^{-1}(0))$. On the other hand, $\bar{p}(0, t) = -\epsilon^2$. Hence by continuity there exists $\delta > 0$ such that

$$\|q - p\| < \delta \Rightarrow \bar{q}(0, t) < -\epsilon^2/2. \tag{35}$$

Now to show $p(x) \notin \text{closure}(\mathcal{P}_d(S) + h(x)\mathbb{R}_{d-\deg(h)}[x])$ we proceed by contradiction. Assume there exist $r(x) \in \mathcal{P}_d(S)$ and $s(x) \in \mathbb{R}_{d-\deg(h)}[x]$ such that $\|r + hs - p\| < \delta$. From (35) we get

$$\bar{r}(0,t) + \bar{h}(0,t)\bar{s}(0,t) < -\epsilon^2/2. \quad (36)$$

But, this is a contradiction because $\bar{h}(0,t) = 0$, and $\bar{r}(0,t) \geq 0$ since $(0,t) \in \bar{S}$ and $\bar{r} \in \mathcal{P}_d(\bar{S})$. □

References

[1] E. Balas, S. Ceria, and G. Cornuéjols. A lift-and-project cutting plane algorithm for mixed 0–1 programs. *Math Program.*, 58:295–324, 1993.

[2] I. Bomze and F. Jarre. A note on Burer's copositivity representation of mixed binary QPs. *Optimization Letters*, 4(3):465–472, 2010.

[3] S. Burer. On the copositive representation of binary and continuous nonconvex quadratic programs. *Math. Program.*, 120(2):479–495, 2009.

[4] S. Burer. Copositive programming. In J. B. Lasserre and M. Anjos, editors, *Handbook of Semidefinite, Cone and Polynomial Optimization: Theory, Algorithms, Software and Applications*. To Appear.

[5] G. Cornuéjols. Revival of the Gomory cuts in the 1990s. *Annals of Operations Research*, 149:63–66, 2007.

[6] J. Demmel, J. Nie, and V. Powers. Representations of positive polynomials on non-compact semi-algebraic sets via KKT ideals. *J. Pure Appl. Algebra*, 209(1):189–200, 2007.

[7] B. Ghaddar, M. Anjos, and J. Vera. An iterative scheme for valid inequality generation in binary quadratic programming. In *15th Conference on integer programming and combinatorial optimization (IPCO XV)*, 2011.

[8] G. Hardy, J. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, New York, second edition, 1988.

[9] L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM J. on Optim.*, 1(2):166–190, 1991.

[10] M. Marshall. Representations of non-negative polynomials, degree bounds, and applications to optimization. *Can. J. Math*, 61(1):205–221, 2009.

[11] J. Nie, J. Demmel, and B. Sturmfels. Minimizing polynomials via sum of squares over the gradient ideal. *Math. Program.*, 106(3, Ser. A):587–606, 2006.

[12] J. Peña, J. Vera, and L. Zuluaga. Exploiting equalities in polynomial programming. *Operations Research Letters*, 36:223–228, 2008.

[13] M. Putinar. Positive polynomials on compact sets. *Indiana Univ. Math. J.*, 42:969–984, 1993.

[14] T. Rockafellar and R. Wets. *Variational Analysis*. Springer–Verlag, Berlin, 1998.

[15] K. Schmüdgen. The K-moment problem for compact semi-algebraic sets. *Math. Ann.*, 289:203–206, 1991.

[16] N. Shor. Class of global minimum bounds of polynomial functions. *Cybernetics*, 23:731–734, 1987.