Journal of the Operations Research
Society of Japan
2003, Vol. 46, No. 2, 125-144

A GENERAL FRAMEWORK FOR CONVEX RELAXATION OF POLYNOMIAL OPTIMIZATION PROBLEMS OVER CONES

Masakazu Kojima
Tokyo Institute of Technology

Sunyoung Kim*
Ewha Women's University

Hayato Waki
Tokyo Institute of Technology

(Received Received June 14, 2002)

Abstract The class of POPs (Polynomial Optimization Problems) over cones covers a wide range of optimization problems such as 0-1 integer linear and quadratic programs, nonconvex quadratic programs and bilinear matrix inequalities. This paper presents a new framework for convex relaxation of POPs over cones in terms of linear optimization problems over cones. It provides a unified treatment of many existing convex relaxation methods based on the lift-and-project linear programming procedure, the reformulation-linearization technique and the semidefinite programming relaxation for a variety of problems. It also extends the theory of convex relaxation methods, and thereby brings flexibility and richness in practical use of the theory.

Keywords: Optimization, convex relaxation, noncovex program, quadratic program, semidefinite program, second order cone program

1. Introduction

Various convex relaxation methods have been studied intensively and extensively in recent years. For 0-1 integer LPs (Linear Programs), a lift-and-project LP procedure by Balas-Ceria-Cornuéjols [1], the RLT (Reformulation-Linearization Technique) by Sherali-Adams [17] and an SDP (Semidefinite Programming) relaxation method by Lovász-Schrijver [12] were regarded as their pioneering works. They had been modified, generalized and extended to various problems and methods; the RLT [18, 19] for 0-1 mixed integer polynomial programs, the SCRM (Successive Convex Relaxation Method) [7, 8] for QOPs (Quadratic Optimization Problems), the SDP relaxation [9, 10][1] of polynomial programs, and SOCP (Second Order Cone Programming) relaxations [5, 6] for QOPs. These methods share the following basic idea.

(i) Add (redundant) valid inequality constraints to a target optimization problem in the n-dimensional Euclidean space $\mathbb{R}^n$.
(ii) Lift the problem with the additional inequality constraints in $\mathbb{R}^n$ to an equivalent optimization problem in a symmetric matrix space; the resulting problem is an LP with additional rank-1 and positive semidefinite constraints on its matrix variables.
(iii) Relax the rank-1 constraint (and positive semidefinite constraint in cases of the RLT and the lift-and-project LP procedure) so that the resulting feasible region is convex.
(iv) Project the relaxed lifted problem in the matrix space back to the original Euclidean space $\mathbb{R}^n$.

*Research supported by Kosef R004-2001-000-00200 and the intramural research grant of Ewha women's university 2002.

[1]Lasserre called his relaxation presented in these papers the LMI (Linear Matrix Inequality) relaxation, but we call it the SDP relaxation because the latter name is more popular than the former.

In some special cases such as the max cut problem [3] or some classes of 0-1 QOPs [6, 15, 23, 24, 26], (i) is not included, but it is necessary in general. In fact, (i) is a key issue in the papers [1, 5, 7–10, 12, 17–19] mentioned above; they employ various techniques of constructing effective valid inequality constraints to strengthen their convex relaxations. Lasserre's SDP relaxation method [9] is very powerful as a theoretical result in the sense that optimal values of fairly general polynomial programs having a compact feasible region can be approximated as closely as desired by solving a finite sequence of SDP relaxations. However, as we require higher accuracy for approximate optimal values, the size of SDP relaxations to be solved increases rapidly. This creates a major difficulty in applying the SDP relaxation method even to medium scale problems. Practically, computing better bounds for optimal values efficiently (inexpensively) is a critical and important issue, specifically when convex relaxation is utilized in the branch and bound method. Related to this issue, Kim-Kojima [5] recently showed that their SOCP relaxation is a reasonable compromise between the effectiveness of the SDP relaxation and the low computational cost of the lift-and-project LP relaxation. Some comparison among the techniques employed in [9], [12] and [17] was reported in [11].

The purpose of this article is to present a general and flexible framework for convex relaxation methods for POPs (polynomial optimization problems) over cones. This new framework is a slight extension of the existing framework for polynomial programs [9, 10, 18, 19]. However, it not only provides a unified treatment of various existing convex relaxation methods mentioned above, but also leads to their further extensions using LOPs (Linear Optimization Problems) over cones [16]; in particular, it highlights SOCP relaxations [5, 6, 14] of POPs over cones, and bring richness in the theory and practice of convex relaxations of nonconvex programs.

In Section 2, we begin with a POP over cones in the Euclidean space; the problem can be either a target problem itself to be solved or the one that we have derived from a 0-1 LP, a QOP or a polynomial program by adding some valid polynomial constraints over cones as in (i). (See the problem (1) below for an example of a POP over cones.) Then, instead of (ii) and (iii), we just relax the problem by an LOP over cones in a higher dimensional Euclidean space. This is described in the latter part of Section 2. The linearization technique used there is the same as the one in the lift-and-project LP procedure [1] and the RLT [17–19].

We should emphasize that adding valid polynomial constraints over cones to an optimization problem to be solved is essential to have tighter convex relaxations. An important feature of the new framework is flexibility and variety in constructing valid constraints, which can be polynomial SDP constraints, polynomial SOCP constraints or even more general polynomial constraints over cones. In Sections 3 and 4, we discuss how we construct such constraints in detail.

In Section 5, we show how our convex relaxation by LOPs over cones works on the problem (1) below.

$$\left.\begin{array}{ll} \text{maximize} & -2x_1 + x_2 \\ \text{subject to} & x_1 \geq 0,\ x_2 \geq 0,\ x_1^2 + (x_2-1)^2 - 1 \geq 0,\ \left\| \begin{pmatrix} x_1+1 \\ x_2 \end{pmatrix} \right\| \leq 2. \end{array}\right\} \quad (1)$$

Here x_1 and x_2 denote real variables. See Figure 1 in Section 5. This problem is used as an illustrative example of POPs over cones throughout the paper.

Section 6 is devoted to the theory of the convex relaxation of POPs over cones. Based on the duality theorem of LOPs over cones, we present a sufficient condition for the convex relaxation to attain a given bound for the objective values of a POP over cones. This part is regarded as a generalization of Theorem 4.2 (b) of [9] where Lasserre presented a characterization of bounds attained by his SDP relaxation for the objective values of a polynomial program. We also generalize a characterization of the SDP relaxation given by Fujie-Kojima [2] (see also [7]) for QOPs to POPs over cones.

2. Polynomial Optimization Problems over Cones and Their Linearization

Let $\mathcal{K}$ be a closed convex cone in the m-dimensional Euclidean space $\mathbb{R}^m$, $f_0(\boldsymbol{x}), f_1(\boldsymbol{x}), \dots, f_m(\boldsymbol{x})$ polynomials in real variables $x_1, x_2, \dots, x_n$, $\boldsymbol{x} = (x_1, x_2, \dots, x_n)^T$ a variable vector in $\mathbb{R}^n$, and $\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \dots, f_m(\boldsymbol{x}))^T \in \mathbb{R}^m$ a system of polynomials. We are concerned with the POP (Polynomial Optimization Problem) over the cone $\mathcal{K}$:

$$\text{maximize } f_0(\boldsymbol{x}) \text{ subject to } \boldsymbol{x} \in \mathcal{F} \equiv \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{f}(\boldsymbol{x}) \in \mathcal{K}\}. \tag{2}$$

In our succeeding discussions, we often deal with the cases where $\mathcal{K}$ is represented as the Cartesian product of closed convex cones $\mathcal{K}_i \subset \mathbb{R}^{m_i}$ $(i = 1, 2, \dots, k)$ such that $\mathcal{K} = \mathcal{K}_1 \times \mathcal{K}_2 \times \cdots \times \mathcal{K}_k$, where we assume that $\sum_{i=1}^k m_i = m$. The POP in such cases is written as

$$\text{maximize } f_0(\boldsymbol{x}) \text{ subject to } \boldsymbol{x} \in \mathcal{F} \equiv \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{f}_i(\boldsymbol{x}) \in \mathcal{K}_i \ (i = 1, 2, \dots, k)\}. \tag{3}$$

Here each $\boldsymbol{f}_i(\boldsymbol{x}) \in \mathbb{R}^{m_i}$ denotes a system of polynomials in real variables $x_1, x_2, \dots, x_n$.

The nonnegative orthant $\mathbb{R}^\ell_+$ and the cone $\mathcal{S}^\ell_+$ of $\ell \times \ell$ symmetric positive semidefinite matrices are popular closed convex cones which have been used in LPs (Linear Programs) and SDPs (Semidefinite Programs), respectively. We identify the space $\mathcal{S}^\ell$ of $\ell \times \ell$ symmetric matrices with the $\ell(\ell+1)/2$-dimensional Euclidean space in our succeeding discussion; hence we regard the positive semidefinite matrix cone $\mathcal{S}^\ell_+$ in $\mathcal{S}^\ell$ as a closed convex cone in $\mathbb{R}^{\ell(\ell+1)/2}$. We also consider the p-order cone

$$\mathcal{N}^{1+\ell}_p \equiv \{\boldsymbol{v} = (v_0, v_1, v_2, \dots, v_\ell)^T \in \mathbb{R}^{1+\ell} : \sum_{i=1}^{\ell} |v_i|^p \le v_0^p\}.$$

Here $p \ge 1$. Following the convention, we use the notation

$$\mathcal{N}^{1+\ell}_\infty \equiv \{\boldsymbol{v} = (v_0, v_1, v_2, \dots, v_\ell)^T \in \mathbb{R}^{1+\ell} : |v_i| \le v_0 \ (i = 1, 2, \dots, \ell)\}.$$

Among the $(1+\ell)$-dimensional p-order cones $\mathcal{N}^{1+\ell}_p$ $(p \ge 1)$, we are particularly interested in the case where $p = 2$ and the case where $p = +\infty$. The former is corresponding to the second order cone, and LOPs over second order cones are known as SOCPs (Second Order Cone Programs). In the latter, $\boldsymbol{v} \in \mathcal{N}^{1+\ell}_\infty$ is characterized as a system of linear inequalities $-v_0 \le v_i \le v_0$ $(i = 1, 2, \dots, \ell)$. For example, we can rewrite lower and upper bound constraints $\underline{b}_i \le v_i \le \bar{b}_i$ $(i = 1, 2, \dots, \ell)$ as

$$\begin{pmatrix} 1 \\ \left(2v_1 - (\bar{b}_1 + \underline{b}_1)\right) / \left(\bar{b}_1 - \underline{b}_1\right) \\ \left(2v_2 - (\bar{b}_2 + \underline{b}_2)\right) / \left(\bar{b}_2 - \underline{b}_2\right) \\ \vdots \\ \left(2v_\ell - (\bar{b}_\ell + \underline{b}_\ell)\right) / \left(\bar{b}_\ell - \underline{b}_\ell\right) \end{pmatrix} \in \mathcal{N}^{1+\ell}_\infty.$$

Here $-\infty < \underline{b}_i < \bar{b}_i < +\infty$ $(i = 1, 2, \dots, \ell)$. More general p-order cones were used in [22] where Xue and Ye proposed interior-point methods for minimization of a sum of p-norms. It should be noted that $\mathcal{N}^{1+\ell}_p \subset \mathcal{N}^{1+\ell}_q$ if $1 \le p \le q$.

A good example which shows a difference between polynomial programs [9, 10, 19, 21] and POPs over cones is a BMI (Bilinear Matrix Inequality): Find $\boldsymbol{x} = (x_0, x_1, x_2, \ldots, x_p)^T \in \mathbb{R}^{1+p}$ and $\boldsymbol{y} = (y_0, y_1, y_2, \ldots, y_q)^T \in \mathbb{R}^{1+q}$ satisfying $\sum_{i=0}^{p} \sum_{j=0}^{q} \boldsymbol{A}_{ij} x_i y_j \in \mathcal{S}^{\ell}_{+}, x_0 = 1$ and $y_0 = 1$, where $\boldsymbol{A}_{ij} \in \mathcal{S}^{\ell}$ $(i = 0, 1, 2, \ldots, p,\ j = 0, 1, 2, \ldots, q)$. We can rewrite the BMI as a POP over cones:

$$\text{maximize } \lambda \ \text{ subject to } \sum_{i=0}^{p} \sum_{j=0}^{q} \boldsymbol{A}_{ij} x_i y_j - \lambda \boldsymbol{I} \in \mathcal{S}^{\ell}_{+} \ \text{ and } 1 - \lambda \geq 0,$$

where $\boldsymbol{I}$ denotes the $\ell \times \ell$ identity matrix; the BMI has a solution if and only if the maximal objective value of the POP over cones above is nonnegative. In their paper [7], Kojima-Tunçel dealt with this problem as a special case of the conic quadratic inequality representation. If we take $n = 2$, $k = 4$,

$$\begin{aligned} g_0(\boldsymbol{x}) &= -2x_1 + x_2,\ g_1(\boldsymbol{x}) = x_1,\ g_2(\boldsymbol{x}) = x_2, \\ g_3(\boldsymbol{x}) &= x_1^2 + x_2^2 - 2x_2,\ \boldsymbol{g}_4(\boldsymbol{x}) = \begin{pmatrix} 2 \\ x_1 + 1 \\ x_2 \end{pmatrix}, \\ \mathcal{K}_1 &= \mathcal{K}_2 = \mathcal{K}_3 = \mathbb{R}_+ \ \text{ (the cone of nonnegative numbers)}, \\ \mathcal{K}_4 &= \mathcal{N}^3_2 \ \text{ (the 3-dimensional second order cone)}, \end{aligned}$$

we can rewrite the problem (1) as a POP

$$\text{maximize} \quad g_0(\boldsymbol{x}) \quad \text{subject to} \quad g_j(\boldsymbol{x}) \in \mathbb{R}_+ \ (j = 1, 2, 3),\ \boldsymbol{g}_4(\boldsymbol{x}) \in \mathcal{N}^3_2. \tag{4}$$

Let us illustrate how we create valid polynomial constraints over cones in the example (4) above. First we observe that

$$g_5(\boldsymbol{x}) \equiv x_1 x_1 \in \mathbb{R}_+,\ g_6(\boldsymbol{x}) \equiv x_1 x_2 \in \mathbb{R}_+,\ g_7(\boldsymbol{x}) \equiv x_2 x_2 \in \mathbb{R}_+ \tag{5}$$

are valid polynomial constraints over the cone $\mathbb{R}_+$ of nonnegative numbers for the problem (4). We consider the pair of constraints $g_1(\boldsymbol{x}) \in \mathbb{R}_+$ and $\boldsymbol{g}_4(\boldsymbol{x}) \in \mathcal{N}^3_2$ over cones in the problem (4). Another constraint is obtained by "multiplying" them as

$$\boldsymbol{g}_8(\boldsymbol{x}) \equiv g_1(\boldsymbol{x}) \boldsymbol{g}_4(\boldsymbol{x}) \in \mathcal{N}^3_2. \tag{6}$$

Similarly we can derive valid constraint

$$\boldsymbol{g}_9(\boldsymbol{x}) \equiv g_2(\boldsymbol{x}) \boldsymbol{g}_4(\boldsymbol{x}) \in \mathcal{N}^3_2. \tag{7}$$

By adding the valid constraints in (5), (6) and (7) to the problem (4), we obtain a POP over cones

$$\left. \begin{array}{ll} \text{maximize} & g_0(\boldsymbol{x}) \\ \text{subject to} & g_j(\boldsymbol{x}) \in \mathbb{R}_+ \ (j = 1, 2, 3, 5, 6, 7),\ \boldsymbol{g}_k(\boldsymbol{x}) \in \mathcal{N}^3_2 \ (k = 4, 8, 9), \end{array} \right\} \tag{8}$$

which is equivalent to the original problem (1).

Now we return to the general POP (2), and show how we linearize it. Let $\mathbb{Z}_+$ denote the set of nonnegative integers. For every variable vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$ and every $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{Z}^n_+ \backslash \{\boldsymbol{0}\}$, we use the notation $\boldsymbol{x}^{\boldsymbol{a}}$ for the term $x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$.

Then we can write any polynomial $f(\boldsymbol{x})$ in the variables $x_1, x_2, \ldots, x_n$ in the form $f(\boldsymbol{x}) = \gamma + \sum_{\boldsymbol{a}\in\mathcal{A}} c(\boldsymbol{a})\boldsymbol{x}^{\boldsymbol{a}}$ for some finite subset $\mathcal{A}$ of $\mathbb{Z}^n_+ \backslash\{\boldsymbol{0}\}$ and some $c(\boldsymbol{a}) \in \mathbb{R}$ $(\boldsymbol{a} \in \mathcal{A})$. We call $\mathcal{A}$ the support of the polynomial $f(\boldsymbol{x})$. Replacing each $\boldsymbol{x}^{\boldsymbol{a}}$ of the polynomial $f(\boldsymbol{x})$ by a single variable $y_{\boldsymbol{a}} \in \mathbb{R}$, we define a linearization $F((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}))$ of $f(\boldsymbol{x})$ as $F((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) = \gamma + \sum_{\boldsymbol{a}\in\mathcal{A}} c(\boldsymbol{a})y_{\boldsymbol{a}}$. Here $(y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})$ denotes a vector consisting of real variables $y_{\boldsymbol{a}}$ $(\boldsymbol{a} \in \mathcal{A})$. We can naturally extend the definition of the linearization $F((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}))$ of a polynomial $f(\boldsymbol{x})$ to a system of polynomials $\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x}))$.

Let $F_0((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}))$ and $\boldsymbol{F}((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}))$ denote the linearizations of the polynomial $f_0(\boldsymbol{x})$ and the polynomial system $\boldsymbol{f}(\boldsymbol{x})$, respectively. Here we assume for simplicity of notation that all polynomials $f_0(\boldsymbol{x}), f_1(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})$ share a common support $\mathcal{A}$; if the term $\boldsymbol{x}^{\boldsymbol{a}}$ does not appear in a polynomial for some $\boldsymbol{a} \in \mathcal{A}$, then we let the corresponding coefficient $c(\boldsymbol{a})$ of the term $\boldsymbol{x}^{\boldsymbol{a}}$ zero. A linearization of the POP (2) is

$$\text{maximize } F_0((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \ \text{ subject to } \boldsymbol{F}((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \in \mathcal{K}. \tag{9}$$

Since F_0 and $\boldsymbol{F}$ are linear functions in variables $y_{\boldsymbol{a}}$ $(\boldsymbol{a} \in \mathcal{A})$, the problem (9) forms an LOP over the cone $\mathcal{K}$. By construction, we know that the LOP (9) serves as a convex relaxation of the POP (2) over cones. In fact, for any feasible solution $\boldsymbol{x}$ of the POP (2) over $\mathcal{K}$, $(y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}) = (\boldsymbol{x}^{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})$ gives a feasible solution of the LOP (9) over $\mathcal{K}$ at which the objective value of the function $F_0((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}))$ takes the same objective value $f_0(\boldsymbol{x})$ as the POP (2) over $\mathcal{K}$.

Similarly we can derive a linearization of the POP (3)

$$\text{maximize } F_0((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \ \text{ subject to } \boldsymbol{F}_i((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \in \mathcal{K}_i \ (i = 1, 2, \ldots, k). \tag{10}$$

Here $F_0((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}))$ and $\boldsymbol{F}_i((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}))$ mean the linearizations of the polynomial $f_0(\boldsymbol{x})$ and the polynomial system $\boldsymbol{f}_i(\boldsymbol{x})$ $(i = 1, 2, \ldots, k)$, respectively. When each cone $\mathcal{K}_i$ is either the nonnegative orthant of the Euclidean space, a positive semidefinite matrix cone, or a second order cone, interior-point methods can be applied to solve the LOP (10) effectively.

The linearization of the POP (8) leads to an SOCP

$$\left.\begin{array}{ll} \text{maximize} & -2y_{10} + y_{01} \\ \text{subject to} & y_{10} \in \mathbb{R}_+,\ y_{01} \in \mathbb{R}_+,\ y_{20} + y_{02} - 2y_{01} \in \mathbb{R}_+, \\ & y_{20} \in \mathbb{R}_+,\ y_{11} \in \mathbb{R}_+,\ y_{02} \in \mathbb{R}_+, \\ & \begin{pmatrix} 2 \\ y_{10} + 1 \\ y_{01} \end{pmatrix} \in \mathcal{N}^3_2,\ \begin{pmatrix} 2y_{10} \\ y_{20} + y_{10} \\ y_{11} \end{pmatrix} \in \mathcal{N}^3_2,\ \begin{pmatrix} 2y_{01} \\ y_{11} + y_{01} \\ y_{02} \end{pmatrix} \in \mathcal{N}^3_2, \end{array}\right\} \tag{11}$$

which serves as a convex relaxation of the POP (1).

3. Universally Valid Polynomial Constraints over Cones

Given a POP of the form (3) over cones, there are two types of valid polynomial constraints over cones which we can add to the original constraints for strengthening a resulting convex relaxation. The one is a uvp-constraint (universally valid polynomial constraint) over a cone that can be added to any POPs. We say that $\boldsymbol{f}(\boldsymbol{x}) \in \mathcal{K} \subset \mathbb{R}^\ell$ is a uvp-constraint if it holds for every $\boldsymbol{x} \in \mathbb{R}^n$. The other is a polynomial constraint over a cone that is a consequence of some of original polynomial constraints over cones and some uvp-constraints. The former holds for every $\boldsymbol{x} \in \mathbb{R}^n$, while the latter holds for all feasible solutions of the original problem but not necessarily for all $\boldsymbol{x} \in \mathbb{R}^n$. This section contains some representative and useful examples of valid polynomial constraints of the first type, and the next section the second type.

3.1. Positive semidefinite matrix cones

Let $\boldsymbol{u}$ be a mapping from $\mathbb{R}^n$ into $\mathbb{R}^\ell$ whose jth component u_j is a polynomial in $x_1, x_2, \ldots, x_n$. Then the $\ell \times \ell$ symmetric matrix $\boldsymbol{u}(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x})^T$ is positive semidefinite for any $\boldsymbol{x} \in \mathbb{R}^n$; $\boldsymbol{u}(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x})^T \in \mathcal{S}^\ell_+$. Thus we can add the polynomial constraint $\boldsymbol{u}(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x})^T \in \mathcal{S}^\ell_+$ over cones to any POP over cones.

The popular SDP relaxation of nonconvex QOPs in the variable vector $\boldsymbol{x} \in \mathbb{R}^n$ is derived by taking $\boldsymbol{u}(\boldsymbol{x}) = (1, x_1, x_2, \ldots, x_n)^T$. In this case, the constraint $\boldsymbol{u}(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x})^T \in \mathcal{S}^{1+n}_+$ is written as

$$\begin{pmatrix} 1 & x_1 & x_2 & \cdots & x_n \\ x_1 & x_1x_1 & x_1x_2 & \cdots & x_1x_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n & x_nx_1 & x_nx_2 & \cdots & x_nx_n \end{pmatrix} \in \mathcal{S}^{1+n}_+, \tag{12}$$

and the corresponding linearization yields the standard positive semidefinite constraint for nonconvex QOPs. See, for example, [2, 7].

Specifically we have

$$\begin{pmatrix} 1 & x_1 & x_2 \\ x_1 & x_1x_1 & x_1x_2 \\ x_2 & x_2x_1 & x_2x_2 \end{pmatrix} \in \mathcal{S}^3_+,$$

when $n = 2$. We can add this positive semidefinite constraint to the POP (8). The corresponding linearization becomes

$$\begin{pmatrix} 1 & y_{10} & y_{01} \\ y_{10} & y_{20} & y_{11} \\ y_{01} & y_{11} & y_{02} \end{pmatrix} \in \mathcal{S}^3_+, \tag{13}$$

which we can add to the SOCP relaxation (11) of the POP (8).

In his paper [9, 10], Lasserre (implicitly) used a family of stronger uvp-constraints over positive semidefinite matrix cones than the standard positive semidefinite constraint of the form (12). He took $\boldsymbol{u}(\boldsymbol{x}) = \boldsymbol{u}^r(\boldsymbol{x})$ to be the column vector consisting of a basis for real-valued polynomials of degree r

$$1, x_1, x_2, \ldots, x_n, x_1^2, x_1x_2, \ldots, x_1x_n, x_2^2, x_2x_3, \ldots, x_n^2, \ldots, x_1^r, \ldots, x_n^r. \tag{14}$$

The linearization of the matrix $\boldsymbol{u}^r(\boldsymbol{x})(\boldsymbol{u}^r(\boldsymbol{x}))^T$ is often called the moment matrix. See (15) below for the case that $n = 2$ and $r = 2$. Although the class of uvp-constraints $\boldsymbol{u}^r(\boldsymbol{x})(\boldsymbol{u}^r(\boldsymbol{x}))^T \in \mathcal{S}^\ell_+$, where ℓ denotes the dimension of the vector $\boldsymbol{u}^r(\boldsymbol{x})$, may be theoretically powerful enough to attain better bounds for optimal values and/or the exact optimal values of a wide class of polynomial programs as shown in Theorem 4.2 of [9], other types of uvp-constraints are of importance in practice mainly because SDPs with large scale positive semidefinite matrix variables are difficult to solve.

When $n = 2$ and $r = 2$, the constraint $\boldsymbol{u}^2(\boldsymbol{x})(\boldsymbol{u}^2(\boldsymbol{x}))^T \in \mathcal{S}^6_+$ becomes

$$\begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1x_2 \\ x_2^2 \end{pmatrix} \begin{pmatrix} 1 & x_1 & x_2 & x_1^2 & x_1x_2 & x_2^2 \end{pmatrix}$$

$$\equiv \begin{pmatrix} 1 & x_1 & x_2 & x_1^2 & x_1x_2 & x_2^2 \\ x_1 & x_1^2 & x_1x_2 & x_1^3 & x_1^2x_2 & x_1x_2^2 \\ x_2 & x_1x_2 & x_2^2 & x_1^2x_2 & x_1x_2^2 & x_2^3 \\ x_1^2 & x_1^3 & x_1^2x_2 & x_1^4 & x_1^3x_2 & x_1^2x_2^2 \\ x_1x_2 & x_1^2x_2 & x_1x_2^2 & x_1^3x_2 & x_1^2x_2^2 & x_1x_2^3 \\ x_2^2 & x_1x_2^2 & x_2^3 & x_1^2x_2^2 & x_1x_2^3 & x_2^4 \end{pmatrix} \in \mathcal{S}^6_+,$$

which we can add to the problem (8), and its linearization is

$$\begin{pmatrix} 1 & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{pmatrix} \in \mathcal{S}^6_+, \tag{15}$$

which can be used for the SOCP relaxation (11) of the problem (8).

3.2. Second order cones

This section presents another class of uvp-constraints over cones. Let $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ be mappings from $\mathbb{R}^n$ into $\mathbb{R}^\ell$ whose components are polynomials in $x_1, x_2, \ldots, x_n$. By the Cauchy-Schwartz inequality, we see that $\left(\boldsymbol{f}_1(\boldsymbol{x})^T\boldsymbol{f}_2(\boldsymbol{x})\right)^2 \leq \left(\boldsymbol{f}_1(\boldsymbol{x})^T\boldsymbol{f}_1(\boldsymbol{x})\right)\left(\boldsymbol{f}_2(\boldsymbol{x})^T\boldsymbol{f}_2(\boldsymbol{x})\right)$ holds. We can rewrite this inequality as a constraint over $\mathcal{N}^3_2$:

$$\begin{pmatrix} \boldsymbol{f}_1(\boldsymbol{x})^T\boldsymbol{f}_1(\boldsymbol{x}) + \boldsymbol{f}_2(\boldsymbol{x})^T\boldsymbol{f}_2(\boldsymbol{x}) \\ \boldsymbol{f}_1(\boldsymbol{x})^T\boldsymbol{f}_1(\boldsymbol{x}) - \boldsymbol{f}_2(\boldsymbol{x})^T\boldsymbol{f}_2(\boldsymbol{x}) \\ 2\boldsymbol{f}_1(\boldsymbol{x})^T\boldsymbol{f}_2(\boldsymbol{x}) \end{pmatrix} \in \mathcal{N}^3_2.$$

To derive another useful uvp-constraints, we consider a trivial relation

$$\boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})^T - \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})^T \in \mathcal{S}^\ell_+ \tag{16}$$

for a mapping $\boldsymbol{f}$ from $\mathbb{R}^n$ into $\mathbb{R}^\ell$ whose components are polynomials in $x_1, x_2, \ldots, x_n$. Let $\boldsymbol{C}$ be an arbitrary $\ell \times \ell$ symmetric positive semidefinite matrix, $\boldsymbol{L}$ a $p \times \ell$ matrix satisfying $\boldsymbol{C} = \boldsymbol{L}^T\boldsymbol{L}$, where $1 \leq p \leq \ell$. Then the inequality $(\boldsymbol{L}\boldsymbol{f}(\boldsymbol{x}))^T(\boldsymbol{L}\boldsymbol{f}(\boldsymbol{x})) \leq \left(\boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{C}\boldsymbol{f}(\boldsymbol{x})\right)$ is true for any $\boldsymbol{x} \in \mathbb{R}^\ell$. We can convert this inequality to

$$\begin{pmatrix} 1 + \boldsymbol{C} \bullet \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})^T \\ 1 - \boldsymbol{C} \bullet \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})^T \\ 2\boldsymbol{L}\boldsymbol{f}(\boldsymbol{x}) \end{pmatrix} \in \mathcal{N}^{2+p}_2, \tag{17}$$

which forms a uvp-constraint over $\mathcal{N}^{2+p}_2$, and we obtain

$$\begin{pmatrix} 1 + \boldsymbol{C} \bullet \boldsymbol{H}((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \\ 1 - \boldsymbol{C} \bullet \boldsymbol{H}((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \\ 2\boldsymbol{L}\boldsymbol{F}((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \end{pmatrix} \in \mathcal{N}^{2+p}_2 \tag{18}$$

as its linearization. Here $\boldsymbol{H}((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}))$ and $\boldsymbol{F}((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}))$ denote the linearizations of $\boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})^T$ and $\boldsymbol{f}(\boldsymbol{x})$, respectively, and $\boldsymbol{A} \bullet \boldsymbol{B}$ stands for the inner product of two matrices $\boldsymbol{A},\ \boldsymbol{B} \in \mathcal{S}^p$; $\boldsymbol{A} \bullet \boldsymbol{B} = \sum_{i=1}^p \sum_{j=1}^p A_{ij}B_{ij}$.

Notice that any polynomial constraint over a second order cone

$$\begin{pmatrix} h_0(\boldsymbol{x}) \\ \boldsymbol{h}(\boldsymbol{x}) \end{pmatrix} \in \mathcal{N}_2^{1+\ell},$$

can be rewritten as

$$\begin{pmatrix} h_0(\boldsymbol{x}) & \boldsymbol{h}(\boldsymbol{x})^T \\ \boldsymbol{h}(\boldsymbol{x}) & h_0(\boldsymbol{x})\boldsymbol{I} \end{pmatrix} \in \mathcal{S}_+^{1+\ell},$$

where $h_i(\boldsymbol{x})$ denotes a polynomial in $x_1, x_2, \ldots, x_n$ $(i = 0, 1, 2, \ldots, \ell)$, $\boldsymbol{h}(\boldsymbol{x})$ the polynomial system $(h_1(\boldsymbol{x}), h_2(\boldsymbol{x}), \ldots, h_\ell(\boldsymbol{x}))^T \in \mathbb{R}^\ell$ and $\boldsymbol{I}$ the $\ell \times \ell$ identity matrix. Specifically we can rewrite (16) as a valid polynomial inequality over $\mathcal{S}_+^{1+\ell}$

$$\begin{pmatrix} 1 & \boldsymbol{f}(\boldsymbol{x})^T \\ \boldsymbol{f}(\boldsymbol{x}) & \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})^T \end{pmatrix} \in \mathcal{S}_+^{1+\ell}. \tag{19}$$

As for the linearization of the constraint above, we obtain

$$\begin{pmatrix} 1 & \boldsymbol{F}((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}))^T \\ \boldsymbol{F}((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) & \boldsymbol{H}((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \end{pmatrix} \in \mathcal{S}_+^{1+\ell}. \tag{20}$$

Then we may regard (18) as a further relaxation of (20). In view of the effectiveness of convex relaxations, adding the valid constraint (17) over $\mathcal{N}_2^{2+p}$ does not result in a stronger convex relaxation than adding the valid constraint (19). Concerning their computational costs, however, (18) is much cheaper than (20). This fact was presented in the paper [5] for a special case in which $\boldsymbol{f}(\boldsymbol{x}) = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$ was taken.

4. Deriving Valid Polynomial Constraints over Cones

4.1. Kronecker products of positive semidefinite matrix cones

For every pair of an $\ell \times \ell$ matrix $\boldsymbol{A}$ and an $m \times m$ matrix $\boldsymbol{B}$, we use the notation $\boldsymbol{A} \otimes \boldsymbol{B}$ for their Kronecker product. It is well-known that ν is an eigenvalue of the $\ell m \times \ell m$ matrix $\boldsymbol{A} \otimes \boldsymbol{B}$ if and only if ν is represented as the product of an eigenvalue λ of $\boldsymbol{A}$ and an eigenvalue μ of $\boldsymbol{B}$ (see, for example, [4]). Hence, if $\boldsymbol{A}$ and $\boldsymbol{B}$ are positive semidefinite then so is their Kronecker product $\boldsymbol{A} \otimes \boldsymbol{B}$.

Suppose that the POP (3) has two constraints over positive semidefinite matrix cones

$$\boldsymbol{f}_i(\boldsymbol{x}) \in \mathcal{K}_i \equiv \mathcal{S}_+^{m_i} \text{ and } \boldsymbol{f}_j(\boldsymbol{x}) \in \mathcal{K}_j \equiv \mathcal{S}_+^{m_j}. \tag{21}$$

Here we assume that every component of mappings $\boldsymbol{f}_i : \mathbb{R}^n \to \mathcal{S}^{m_i}$ and $\boldsymbol{f}_j : \mathbb{R}^n \to \mathcal{S}^{m_j}$ is a polynomial in $x_1, x_2, \ldots, x_n$. In view of the discussion above, if $\boldsymbol{x} \in \mathbb{R}^n$ satisfies the constraint (21) then it satisfies

$$\boldsymbol{f}_i(\boldsymbol{x}) \otimes \boldsymbol{f}_j(\boldsymbol{x}) \in \mathcal{S}_+^{m_i m_j}. \tag{22}$$

As a result, we can add (22) to the POP (3) as a valid polynomial constraint over $\mathcal{S}_+^{m_i m_j}$.

We describe two special cases of interests as follows. The first one is the simplest case of the constraints (21) for $m_i = m_j = 1$. In this case, both $\mathcal{S}_+^{m_i}$ and $\mathcal{S}_+^{m_j}$ become the set $\mathbb{R}_+$ of nonnegative numbers. It follows that (21) and (22) can be rewritten as

$$f_i(\boldsymbol{x}) \geq 0 \text{ and } f_j(\boldsymbol{x}) \geq 0$$

and

$$f_i(\boldsymbol{x})f_j(\boldsymbol{x}) \geq 0,$$

respectively. A higher order polynomial inequality can always be derived in this way from two polynomial inequalities in the problem (3). This is an essential technique used in the RLT [17–20].

Let us show the second case, which played an essential role in the papers [9, 10]. We let $\boldsymbol{u}^r(\boldsymbol{x})$ denote the column vector consisting of a basis given in (14) for real-valued polynomials of degree r, for some positive integer r. Take $m_i = 1$, $\boldsymbol{f}_j(\boldsymbol{x}) = \boldsymbol{u}^r(\boldsymbol{x})(\boldsymbol{u}^r(\boldsymbol{x}))^T$ and m_j to be the dimension of the vector $\boldsymbol{u}^r(\boldsymbol{x})$. Then (21) and (22) become

$$f_i(\boldsymbol{x}) \geq 0 \ \text{ and } \boldsymbol{u}^r(\boldsymbol{x})(\boldsymbol{u}^r(\boldsymbol{x}))^T \in \mathcal{S}_+^{m_j}$$

and

$$f_i(\boldsymbol{x})\boldsymbol{u}^r(\boldsymbol{x})(\boldsymbol{u}^r(\boldsymbol{x}))^T \in \mathcal{S}_+^{m_j},$$

respectively.

4.2. Hadamard products of p-order cones ($p \geq 1$)

We use the symbol $\circ$ to denote the Hadamard product of two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ in $\mathbb{R}^{1+\ell}$, $\boldsymbol{v} \circ \boldsymbol{w} = (v_0w_0, v_1w_1, v_2w_2, \ldots, v_\ell w_\ell)^T$, and to indicate the Hadamard product of two closed convex cones $\mathcal{V}$ and $\mathcal{W}$ in $\mathbb{R}^{1+\ell}$, $\mathcal{V} \circ \mathcal{W} = \{\boldsymbol{v} \circ \boldsymbol{w} : \boldsymbol{v} \in \mathcal{V},\ \boldsymbol{w} \in \mathcal{W}\}$. If $1 \leq p \leq q \leq +\infty$, then $\mathcal{N}_p^{1+\ell} \circ \mathcal{N}_q^{1+\ell} \subset \mathcal{N}_p^{1+\ell}$. In fact, if $\boldsymbol{z} = \boldsymbol{v} \circ \boldsymbol{w} \in \mathcal{N}_p^{1+\ell} \circ \mathcal{N}_q^{1+\ell}$ then

$$\begin{aligned} \sum_{j=1}^{\ell} |z_j|^p &= \sum_{j=1}^{\ell} |v_j w_j|^p \\ &\leq w_0^p \sum_{j=1}^{\ell} |v_j|^p \ (\text{since } |w_j| \leq w_0 \ (j = 1, 2, \ldots, \ell)) \\ &\leq (w_0 v_0)^p \ (\text{since } \boldsymbol{v} \in \mathcal{N}_p^{1+\ell}) \\ &= z_0^p. \end{aligned}$$

Hence $\mathcal{N}_p^{1+\ell} \circ \mathcal{N}_q^{1+\ell} \subset \mathcal{N}_p^{1+\ell}$. Recall that $\mathcal{N}_p^{1+\ell} \subset \mathcal{N}_q^{1+\ell}$ when $1 \leq p \leq q \leq +\infty$.

Suppose that $\mathcal{K}_i = \mathcal{N}_p^{1+\ell}$ and $\mathcal{K}_j = \mathcal{N}_q^{1+\ell}$ hold for some p and q such that $1 \leq p \leq q \leq +\infty$ in problem (3); either or both of p and q can be $+\infty$. Then $\boldsymbol{f}_i(\boldsymbol{x}) \circ \boldsymbol{f}_j(\boldsymbol{x}) \in \mathcal{N}_p^{1+\ell}$ provides a valid polynomial constraint for the problem (3).

It should be also noted that the cone $\mathcal{N}_q^{1+\ell}$ is symmetric with respect to the coordinates $1, 2, \ldots, \ell$. Thus, for any permutation $(s_1, s_2, \ldots, s_\ell)$ of $(1, 2, \ldots, \ell)$,

$$\boldsymbol{f}_i(\boldsymbol{x}) \circ \left((\boldsymbol{f}_j(\boldsymbol{x}))_0, (\boldsymbol{f}_j(\boldsymbol{x}))_{s_1}, (\boldsymbol{f}_j(\boldsymbol{x}))_{s_2}, \ldots, (\boldsymbol{f}_j(\boldsymbol{x}))_{s_\ell}\right)^T \in \mathcal{N}_p^{1+\ell}$$

remains a valid polynomial constraint for the POP (3).

If $h(\boldsymbol{x}) \in \mathbb{R}_+$ and $\boldsymbol{f}_j(\boldsymbol{x}) \in \mathcal{N}_p^{1+\ell}$, then we have $h(\boldsymbol{x})\boldsymbol{f}_j(\boldsymbol{x}) \in \mathcal{N}_p^{1+\ell}$. This relation is straightforward, and can be also derived from the above discussion if we take

$$\boldsymbol{f}_i(\boldsymbol{x}) = (h(\boldsymbol{x}), h(\boldsymbol{x}), \ldots, h(\boldsymbol{x}))^T \in \mathbb{R}_+^{1+\ell}.$$

4.3. Linear transformation of cones

Any feasible solution $\boldsymbol{x}$ of the POP (3) satisfies

$$\boldsymbol{T}(\boldsymbol{f}_1(\boldsymbol{x}), \boldsymbol{f}_2(\boldsymbol{x}), \ldots, \boldsymbol{f}_j(\boldsymbol{x})) \in \mathcal{K}_0 \tag{23}$$

if $\boldsymbol{T}$ is a linear mapping from $\mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \times \cdots \times \mathbb{R}^{m_j}$ into $\mathbb{R}^{m_0}$ satisfying $\boldsymbol{T}(\mathcal{K}_1 \times \mathcal{K}_2 \times \cdots \times \mathcal{K}_j) \subset \mathcal{K}_0$. Here $j \geq 1$ and $\mathcal{K}_0$ stands for a closed convex cone in $\mathbb{R}^{m_0}$. Thus (23) serves as a valid polynomial constraint over $\mathcal{K}_0$ for the POP (3). As a result, we can add (23) to the POP (3) or replace some of the constraints $\boldsymbol{f}_i(\boldsymbol{x}) \in \mathcal{K}_i$ $(i = 1, 2, \ldots, j)$ by (23) before producing other candidates of valid polynomial constraints over cones for the POP (3). If we compare the linearization of the constraint (23) with the set of the linearizations of the original constraints $\boldsymbol{f}_i(\boldsymbol{x}) \in \mathcal{K}_i$ $(i = 1, 2, \ldots, j)$, however, the former is weaker than the latter; therefore we gain nothing for the effectiveness of convex relaxation. The purpose of applying a linear transformation to the constraints of the POP (3) is to create a smaller size or more tractable valid polynomial constraint over a cone combining some constraints of the POP (3) to save the computational cost. Recall that we have derived the uvp-constraint (17) and its linearization (18) by applying a linear transformation $\boldsymbol{Y} \in \mathcal{S}^\ell \to \boldsymbol{C} \bullet \boldsymbol{Y} \in \mathbb{R}$ to the uvp-constraint (16) over $\mathcal{S}^\ell_+$.

4.4. Quadratic convexity

Let $\boldsymbol{Q} \in \mathcal{S}^n_+$, $\boldsymbol{q} \in \mathbb{R}^n$ and $\gamma \in \mathbb{R}$. We can transform a convex quadratic inequality $\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{q}^T\boldsymbol{x} + \gamma \leq 0$ to a linear constraint over the second order cone $\mathcal{N}^{2+p}_2$

$$\begin{pmatrix} 1 - \boldsymbol{q}^T\boldsymbol{x} - \gamma \\ 1 + \boldsymbol{q}^T\boldsymbol{x} + \gamma \\ \boldsymbol{2Lx} \end{pmatrix} \in \mathcal{N}^{2+p}_2,$$

where $\boldsymbol{Q} = \boldsymbol{L}^T\boldsymbol{L}$ indicates a factorization of $\boldsymbol{Q}$ for some $p \times n$ matrix $\boldsymbol{L}$. Conversions like this are often used when we transform a quadratically constrained convex quadratic program into an SOCP to which interior-point methods can be applied effectively. In this subsection, we emphasize another advantage of this conversion. That is, after obtaining valid constraints over second order cones, we can apply the technique using the Hadamard product presented in Section 4.2 to create valid higher order polynomial constraints over second order cones.

We now consider the inequality constraint

$$\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - h_1(\boldsymbol{x})h_2(\boldsymbol{x}) \leq 0,\ h_1(\boldsymbol{x}) \geq 0 \text{ and } h_2(\boldsymbol{x}) \geq 0. \tag{24}$$

Here we assume that the $n \times n$ symmetric matrix $\boldsymbol{Q}$ is positive semidefinite, and that $h_1(\boldsymbol{x})$ and $h_2(\boldsymbol{x})$ are polynomials in $x_1, x_2, \ldots, x_n$. Since $\boldsymbol{Q}$ is positive semidefinite, we can find a $p \times n$ matrix $\boldsymbol{L}$ such that $\boldsymbol{Q} = \boldsymbol{L}^T\boldsymbol{L}$. Then (24) turns to be an equivalent polynomial second order cone constraint

$$\begin{pmatrix} h_1(\boldsymbol{x}) + h_2(\boldsymbol{x}) \\ h_1(\boldsymbol{x}) - h_2(\boldsymbol{x}) \\ 2\boldsymbol{L}\boldsymbol{x} \end{pmatrix} \in \mathcal{N}^{2+p}_2.$$

4.5. Adding valid polynomial inequality constraints from numerical computation

Let $e(\boldsymbol{x})$ be a polynomial in real variables $x_1, x_2, \ldots, x_n$ and $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$. We consider a valid polynomial inequality constraint of the form

$$\lambda - e(\boldsymbol{x}) \geq 0 \tag{25}$$

for the POP (3). Here $\lambda \in \mathbb{R}$ serves as a parameter to be chosen such that the constraint (25) is valid inequality constraint for the POP (3). This leads to another POP

$$\text{maximize } e(\boldsymbol{x}) \text{ subject to } \mathcal{F} \equiv \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{f}_i(\boldsymbol{x}) \in \mathcal{K}_i \ (i = 1, 2, \dots, k)\}. \tag{26}$$

If λ is not less than the optimal value λ^* of this problem, then (25) provides a valid inequality constraint for the POP (3). Ideally we want to choose the optimal value λ^* itself for λ. The computation of λ^* is usually as difficult as the original POP (3). Therefore, we apply convex relaxation to the POP (3), and formulate a LOP over cones

$$\text{maximize } E((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \ \text{ subject to } \boldsymbol{F}_i((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \in \mathcal{K}_i \ (i = 1, 2, \dots, k). \tag{27}$$

Here $E((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}))$ means the linearization of the polynomial $e(\boldsymbol{x})$. Let $\hat{\lambda}$ be the optimal value of the relaxed LOP (27). Obviously $\lambda^* \leq \hat{\lambda}$, and it follows that (25) with $\lambda = \hat{\lambda}$ is surely a valid inequality constraint for the POP (3).

Adding a single valid inequality constraint (25) may not be enough to strengthen the convex relaxation of the POP (3). In general, it is necessary to generate multiple valid inequality constraints of the type $\hat{\lambda}_i - e_i(\boldsymbol{x}) \geq 0$ with different polynomials $e_i(\boldsymbol{x})$ and their upper bounds $\hat{\lambda}_i$ over $\mathcal{F}$, and then apply the techniques given in the previous subsections to them and some of the original polynomial constraints of the POP (3) to create new valid polynomial constraints over cones.

The discussion above leads to the idea of a successive convex relaxation of the feasible region $\mathcal{F}$ of the POP (3) [1, 7, 8, 12]. We assume below that the objective function $f_0(\boldsymbol{x})$ of the POP (3) is a linear function. Then, the maximal objective value of the POP (3) coincides with the maximal value of $f_0(\boldsymbol{x})$ over the convex hull c.hull($\mathcal{F}$) of $\mathcal{F}$. As a result, computing (an approximation of) the maximal objective value of the POP (3) is reduced to a tractable description (of an approximation) of the convex hull c.hull($\mathcal{F}$) of $\mathcal{F}$. Suppose that we have a description $\mathcal{P}_q$ of the feasible region of the POP (3) at the qth iteration. Each member of $\mathcal{P}_q$ is a pair of a polynomial system $\boldsymbol{h}(\boldsymbol{x})$ and a closed convex cone $\mathcal{H}$; specifically $\mathcal{P}_0 = \{(\boldsymbol{f}_i, \mathcal{K}_i) : i = 1, 2, \dots, k\}$, and we can rewrite the POP (3) as

$$\text{maximize } f_0(\boldsymbol{x}) \text{ subject to } \boldsymbol{h}(\boldsymbol{x}) \in \mathcal{H} \ ((\boldsymbol{h}(\boldsymbol{x}), \mathcal{H}) \in \mathcal{P}_0).$$

In each iteration, we compute valid inequality constraints for the feasible region $\mathcal{F}$ of the POP (3) with the description $\mathcal{P}_q$ by solving relaxed problems as presented above. Then we refine the description $\mathcal{P}_q$ of $\mathcal{F}$ by incorporating the newly computed valid inequality constraints into the description $\mathcal{P}_q$ to generate a better description $\mathcal{P}_{q+1}$ of $\mathcal{F}$. Here "better" means that the convex relaxation

$$\mathcal{C}_q \equiv \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{H}\left((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})\right) \in \mathcal{H} \ \ ((\boldsymbol{h}(\boldsymbol{x}), \mathcal{H}) \in \mathcal{P}_q)\}$$

shrinks; ideally to the convex hull c.hull($\mathcal{F}$) of $\mathcal{F}$ such that $\mathcal{C}_q \supset \mathcal{C}_{q+1}$ and $\cap_{q=0}^{\infty} \mathcal{C}_q =$c.hull($\mathcal{F}$).

Some existing successive convex relaxation methods use linear functions for $e_i(\boldsymbol{x})$'s, and quadratic valid inequalities of the form $\left(\hat{\lambda}_i - e_i(\boldsymbol{x})\right)\left(\hat{\lambda}_j - e_j(\boldsymbol{x})\right) \geq 0$ are generated and added to $\mathcal{P}_q$. See [1, 7, 8, 12] for more details.

5. A Numerical Example

We investigate and compare three different convex relaxations of the problem (1), an SDP relaxation, an SOCP relaxation and a combination of them. See Figure 1.

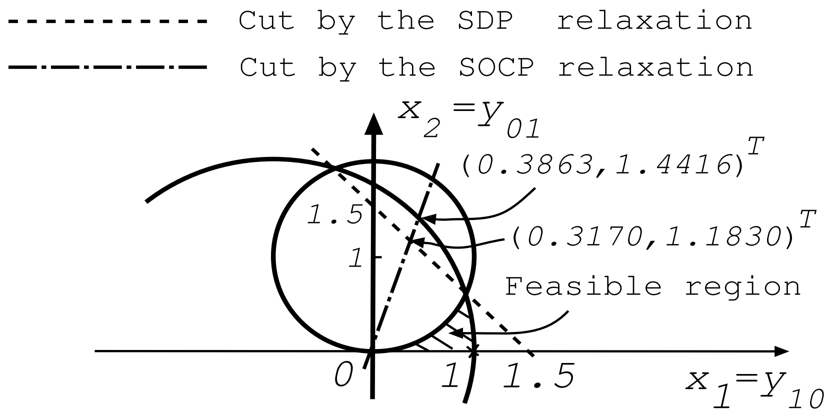

Figure 1: Problem (1): $n = 2$, $m = 4$ and the optimal value 0.0000.

We transform the problem (1) to a QOP

$$\begin{array}{ll} \text{Maximize} & -2x_1 + x_2 \\ \text{subject to} & x_1 \geq 0,\ x_2 \geq 0,\ x_1x_2 \geq 0, \\ & x_1^2 + x_2^2 - 2x_2 \geq 0,\ x_1^2 + x_2^2 + 2x_1 - 3 \leq 0. \end{array}$$

Here we have added a valid quadratic constraint $x_1x_2 \geq 0$. The standard application of the SDP relaxation to this QOP leads us to an SDP

$$\left.\begin{array}{ll} \text{Maximize} & -2y_{10} + y_{01} \\ \text{subject to} & y_{10} \geq 0,\ y_{01} \geq 0,\ y_{11} \geq 0, \\ & y_{20} + y_{02} - 2y_{01} \geq 0,\ y_{20} + y_{02} + 2y_{10} - 3 \leq 0, \\ & \boldsymbol{y} \equiv \begin{pmatrix} 1 & y_{10} & y_{01} \\ y_{10} & y_{20} & y_{11} \\ y_{01} & y_{11} & y_{02} \end{pmatrix} \in \mathcal{S}^3_+, \end{array}\right\} \tag{28}$$

which has an approximate optimal solution

$$\boldsymbol{y}^{\text{sdp}} \equiv \begin{pmatrix} 1 & y^{\text{sdp}}_{10} & y^{\text{sdp}}_{01} \\ y^{\text{sdp}}_{10} & y^{\text{sdp}}_{20} & y^{\text{sdp}}_{11} \\ y^{\text{sdp}}_{01} & y^{\text{sdp}}_{11} & y^{\text{sdp}}_{02} \end{pmatrix} = \begin{pmatrix} 1.0000 & 0.0000 & 1.5000 \\ 0.0000 & 0.3830 & 0.0000 \\ 1.5000 & 0.0000 & 2.6170 \end{pmatrix},$$

and an approximate optimal value $-2y^{\text{sdp}}_{10} + y^{\text{sdp}}_{01} = 1.5000$.

Now we consider an SOCP relaxation of the problem (1). Recall that we have converted the problem (1) into a POP (8) over cones $\mathbb{R}_+$ and $\mathcal{N}^3_2$ by adding some valid constraints given in (5), (6) and (7) to (1), and that we have derived an SOCP relaxation (11) of the POP (8). we obtain an approximate optimal solution by solving the SOCP (11)

$$\boldsymbol{y}^{\text{socp}} \equiv \begin{pmatrix} 1 & y^{\text{socp}}_{10} & y^{\text{socp}}_{01} \\ y^{\text{socp}}_{10} & y^{\text{socp}}_{20} & y^{\text{socp}}_{11} \\ y^{\text{socp}}_{01} & y^{\text{socp}}_{11} & y^{\text{socp}}_{02} \end{pmatrix} = \begin{pmatrix} 1.0000 & 0.3863 & 1.4416 \\ 0.3863 & 0.3863 & 0.0000 \\ 1.4416 & 0.0000 & 2.4969 \end{pmatrix},$$

and an approximate optimal value $-2y^{\text{socp}}_{10} + y^{\text{socp}}_{01} = 0.6691$. Comparing the optimal value 1.5000 of the SDP relaxation (28) with the optimal value 0.6691 of the SOCP relaxation (11), we know that the SOCP relaxation provides a better bound for the optimal value 0.0000 of the original problem (1). It is easily verified that:

- The optimal solution $\boldsymbol{y}^{\rm sdp}$ of the SDP (28) does not satisfy the constraint

$$\begin{pmatrix} 2y_{10} \\ y_{20}+y_{10} \\ y_{11} \end{pmatrix} \in \mathcal{N}_2^3$$

  of the SOCP (11),
- The optimal solution $\boldsymbol{y}^{\rm socp}$ of the SOCP (11) does not satisfy the positive semidefinite constraint

$$\boldsymbol{y} \equiv \begin{pmatrix} 1 & y_{10} & y_{01} \\ y_{10} & y_{20} & y_{11} \\ y_{01} & y_{11} & y_{02} \end{pmatrix} \in \mathcal{S}_+^3$$

  of the SDP (28).

This suggests to combine the SDP and SOCP relaxations to get a tighter bound for the optimal objective value. Indeed if we add the positive semidefinite constraint above to the SOCP (11), the resulting SOCP-SDP problem attains a better bound 0.5490 at $y_{10} = 0.3170$ and $y_{01} = 1.1830$. See Figure 1. We conclude that both SDP and SOCP relaxations are important in this example.

6. Characterization of the Linearization of POPs over Cones

Throughout this section, we deal with the POP of the form (2) over the closed convex cone $\mathcal{K} \subset \mathbb{R}^m$, and we present characterization of the upper bound that is attained by its linearization (9) for its optimal objective value.

6.1. Basic theory

We denote each $f_j(\boldsymbol{x})$ in (2) as $f_j(\boldsymbol{x}) = \gamma_j + \sum_{\boldsymbol{a}\in\mathcal{A}} c_j(\boldsymbol{a})\boldsymbol{x}^{\boldsymbol{a}}$, and its linearization as $F_j((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) = \gamma_j + \sum_{\boldsymbol{a}\in\mathcal{A}} c_j(\boldsymbol{a}) y_{\boldsymbol{a}}$. Here $\gamma_j \in \mathbb{R}$, and we assume that $\gamma_0 = 0$. Let

$$\boldsymbol{F}((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \;=\; (F_1((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})), \ldots, F_m((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})))^T .$$

Then, the linearization of the POP (2) is defined as an LOP over the cone $\mathcal{K}$:

$$\text{maximize} \quad \sum_{\boldsymbol{a}\in\mathcal{A}} c_0(\boldsymbol{a}) y_{\boldsymbol{a}} \quad \text{subject to} \quad \boldsymbol{F}((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \in \mathcal{K}, \tag{29}$$

which is a convex relaxation of the POP (2). Moreover, let $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)^T \in \mathbb{R}^m$, $\boldsymbol{c}(\boldsymbol{a}) = (c_1(\boldsymbol{a}), \ldots, c_m(\boldsymbol{a}))^T \in \mathbb{R}^m$ $(\boldsymbol{a} \in \mathcal{A})$, and let $\langle \cdot, \cdot \rangle$ be any inner product in $\mathbb{R}^m$. For the succeeding discussions, we also present the dual of the LOP (29)

$$\text{minimize} \quad \langle \boldsymbol{\gamma}, \boldsymbol{v} \rangle \quad \text{subject to} \quad \boldsymbol{v} \in \mathcal{G}, \tag{30}$$

where $\mathcal{G} = \{\boldsymbol{v} \in \mathcal{K}^* : c_0(\boldsymbol{a}) + \langle \boldsymbol{c}(\boldsymbol{a}), \boldsymbol{v} \rangle = 0 \ (\boldsymbol{a} \in \mathcal{A})\}$, where $\mathcal{K}^*$ denote the dual cone of $\mathcal{K}$, *i.e.*, $\mathcal{K}^* = \{\boldsymbol{v} \in \mathbb{R}^m : \langle \boldsymbol{z}, \boldsymbol{v} \rangle \geq 0 \ \forall \boldsymbol{z} \in \mathcal{K}\}$. Then we know that

$$\left.\begin{array}{l} \text{the supremum objective value of the POP (2)} \\ \;\leq\; \text{the supremum objective value of the primal LOP (29)} \\ \;\leq\; \langle \boldsymbol{\gamma}, \boldsymbol{v} \rangle \text{ at any } \boldsymbol{v} \in \mathcal{G}. \end{array}\right\} \tag{31}$$

We define the Lagrangian function

$$\begin{array}{rcl} L(\boldsymbol{x}, \boldsymbol{v}) & = & f_0(\boldsymbol{x}) + \langle \boldsymbol{v}, \boldsymbol{f}(\boldsymbol{x}) \rangle \\ & = & \langle \boldsymbol{\gamma}, \boldsymbol{v} \rangle + \displaystyle\sum_{\boldsymbol{a}\in\mathcal{A}} \left(c_0(\boldsymbol{a}) + \langle \boldsymbol{v}, \boldsymbol{c}(\boldsymbol{a}) \rangle \right) \boldsymbol{x}^{\boldsymbol{a}} \\ & & \text{for every } \boldsymbol{x} \in \mathbb{R}^n \text{ and every } \boldsymbol{v} \in \mathcal{K}^*. \end{array}$$

The lemma below characterizes the bound $\zeta = \langle \boldsymbol{\gamma}, \boldsymbol{v} \rangle$ in the last line of (31).

Lemma 6.1.

(i) A pair of $\boldsymbol{v} \in \mathcal{K}^*$ *and* $\zeta \in \mathbb{R}$ *satisfies*

$$L(\boldsymbol{x}, \boldsymbol{v}) = \zeta \text{ for every } \boldsymbol{x} \in \mathbb{R}^n \tag{32}$$

if and only if $\boldsymbol{v} \in \mathcal{G}$ *and* $\zeta = \langle \boldsymbol{\gamma}, \boldsymbol{v} \rangle$.

(ii) Suppose that there exist $\boldsymbol{v} \in \mathcal{K}^*$ *and* $\zeta \in \mathbb{R}$ *satisfying the identity (32). Then the objective value of the primal LOP (29) is bounded by* ζ *from above.*

Proof: The assertion (i) follows from the observation that (32) holds if and only if

$$\langle \boldsymbol{\gamma}, \boldsymbol{v} \rangle = \zeta \text{ and } c_0(\boldsymbol{a}) + \langle \boldsymbol{c}(\boldsymbol{a}), \boldsymbol{v} \rangle = 0 \ (\boldsymbol{a} \in \mathcal{A})$$

hold. The assertion (ii) follows from (31). ∎

From the lemma, we see that

$$\inf_{\boldsymbol{v} \in \mathcal{K}^*} \sup_{\boldsymbol{x} \in \mathbb{R}^n} L(\boldsymbol{x}, \boldsymbol{v}) \leq \inf_{\boldsymbol{v} \in \mathcal{G}} \sup_{\boldsymbol{x} \in \mathbb{R}^n} L(\boldsymbol{x}, \boldsymbol{v}) = \inf_{\boldsymbol{v} \in \mathcal{G}} \langle \boldsymbol{\gamma}, \boldsymbol{v} \rangle.$$

This shows that the upper bound given by the Lagrangian dual of the POP (2) for the optimal value of the POP (2) is at least as effective as the one by the dual LOP (30). In general, there is a gap between these two upper bounds. For example, consider a problem

$$\text{maximize } x^4 - x^2 \text{ subject to } -x^4 + 1 \geq 0.$$

The Lagrangian dual of this problem

$$\inf_{v \in \mathbb{R}_+} \sup_{x \in \mathbb{R}} x^4 - x^2 + v(-x^4 + 1)$$

attains the upper bound +1 for the optimal value 0 of the problem, while the linear relaxation

$$\text{maximize } y_4 - y_2 \text{ subject to } -y_4 + 1 \geq 0$$

forms an unbounded linear program; hence its dual is infeasible.

The theorem below provides a sufficient condition for the primal LOP (29) to attain the supremum objective value of the POP (2).

Theorem 6.2. *Let* $\zeta^* < \infty$ *be the supremum objective value of the POP (2). Suppose that the relation (32) holds for* $\zeta = \zeta^*$ *and some* $\boldsymbol{v} = \boldsymbol{v}^* \in \mathcal{K}^*$. *Then* $\boldsymbol{v}^*$ *is an optimal solution of the dual LOP (30) and the identities*

$$\begin{aligned} \zeta^* &= \text{the supremum objective values of the primal LOP (29)} \\ &= \text{the optimal value of the dual LOP (30) at the optimal solution } \boldsymbol{v}^* \end{aligned}$$

hold.

Proof: In general,

$$\begin{aligned} \zeta^* &= \sup \left\{ \sum_{\boldsymbol{a} \in \mathcal{A}} c_0(\boldsymbol{a}) \boldsymbol{x}^{\boldsymbol{a}} : \boldsymbol{f}(\boldsymbol{x}) \in \mathcal{K} \right\} \\ &\leq \sup \left\{ \sum_{\boldsymbol{a} \in \mathcal{A}} c_0(\boldsymbol{a}) y_{\boldsymbol{a}} : \boldsymbol{F}((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \in \mathcal{K} \right\} \\ &\leq \inf \left\{ \langle \boldsymbol{\gamma}, \boldsymbol{v} \rangle : \boldsymbol{v} \in \mathcal{G} \right\}. \end{aligned}$$

On the other hand, we know in view of Lemma 6.1 that $\boldsymbol{v}^*$ is a feasible solution of the dual LOP (30) with the objective value ζ^*. Therefore the conclusion follows. ∎

The theorem below characterizes the supremum objective value of the primal LOP (29) under the existence of an interior feasible solution of the POP (2).

Theorem 6.3. *Assume that there is an $\bar{\boldsymbol{x}} \in \mathbb{R}^n$ such that $\boldsymbol{f}(\bar{\boldsymbol{x}})$ lies in the interior of $\mathcal{K}$ and that the objective values of the primal LOP (29) is bounded from above. Let ζ denote the supremum objective values of (29). Then there exists $\boldsymbol{v} \in \mathcal{K}^*$ satisfying (32).*

Proof: Let $(\bar{y}_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}) = (\bar{\boldsymbol{x}}^{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})$. Then $\boldsymbol{F}((\bar{y}_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) = \boldsymbol{f}(\bar{\boldsymbol{x}})$, which lies in the interior of the cone $\mathcal{K}$. By the duality theorem (see, for example, [16]), there exists an optimal solution $\boldsymbol{v}$ of the dual LOP (30) with the objective value $\zeta = \langle \boldsymbol{\gamma}, \boldsymbol{v} \rangle$. Thus the desired result follows from Lemma 6.1. ∎

6.2. Convex relaxations of the feasible region of the POP (2)

Let $\mathcal{A}^L = \{\boldsymbol{a} \in \mathcal{A} : \sum_{i=1}^n a_i = 1\}$ and $\mathcal{A}^N = \{\boldsymbol{a} \in \mathcal{A} : \sum_{i=1}^n a_i \geq 2\}$. $\mathcal{A}^L$ and $\mathcal{A}^N$ stand for the subsets of the support $\mathcal{A}$ corresponding to the linear terms and the nonlinear terms, respectively and $\mathcal{A} = \mathcal{A}^L \cup \mathcal{A}^N$. We consider the projection $\widehat{\mathcal{F}}$ of the feasible region of the primal LOP (29) on the $\boldsymbol{x}$-space

$$\widehat{\mathcal{F}} \;=\; \{\boldsymbol{x} = (y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}^L) \in \mathbb{R}^n : \boldsymbol{F}((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \in \mathcal{K} \text{ for some } (y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}^N)\}.$$

Then we may regard $\widehat{\mathcal{F}}$ as a convex relaxation of the feasible region $\mathcal{F}$ of the original POP (2). Characterization of $\widehat{\mathcal{F}}$ is given in this subsection. Such characterization is meaningful when the objective function $f_0(\boldsymbol{x})$ of the POP (2) is a linear function $\sum_{\boldsymbol{a}\in\mathcal{A}^L} c_0(\boldsymbol{a})\boldsymbol{x}^{\boldsymbol{a}}$ because its convex relaxation of (29) can be rewritten as

$$\text{maximize} \sum_{\boldsymbol{a}\in\mathcal{A}^L} c_0(\boldsymbol{a})\boldsymbol{x}^{\boldsymbol{a}} \text{ subject to } \boldsymbol{x} \in \widehat{\mathcal{F}}.$$

Theoretically, we may assume without loss of generality that the objective function of the POP (2) is linear; otherwise replace the nonlinear objective function $f_0(\boldsymbol{x})$ by x_{n+1} and $f_0(\boldsymbol{x}) - x_{n+1} \geq 0$ to the constraint.

Let

$$\begin{aligned} \mathcal{L} &= \{\boldsymbol{v} \in \mathcal{K}^* : \langle \boldsymbol{v}, \boldsymbol{f}(\boldsymbol{x}) \rangle \text{ is linear in } \boldsymbol{x}\}^* \\ &= \{\boldsymbol{v} \in \mathcal{K}^* : \langle \boldsymbol{v}, \boldsymbol{c}(\boldsymbol{a}) \rangle = 0 \ (\boldsymbol{a} \in \mathcal{A}^N)\}^* \end{aligned}$$

or equivalently $\mathcal{L}^* = \{\boldsymbol{v} \in \mathcal{K}^* : \langle \boldsymbol{v}, \boldsymbol{c}(\boldsymbol{a}) \rangle = 0 \ (\boldsymbol{a} \in \mathcal{A}^N)\}$. Then $\mathcal{G} \subset \mathcal{L}^*$; hence if $\boldsymbol{v} \in \mathcal{K}^*$ satisfies (32) then $\boldsymbol{v} \in \mathcal{L}^*$. We now introduce another convex relaxation of the feasible region of the POP (2):

$$\widetilde{\mathcal{F}} \equiv \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{f}(\boldsymbol{x}) \in \mathcal{L}\}. \tag{33}$$

Note that each $\boldsymbol{x} \in \widetilde{\mathcal{F}}$ is characterized as

$$\langle \boldsymbol{v}, \boldsymbol{f}(\boldsymbol{x}) \rangle \equiv \langle \gamma, \boldsymbol{v} \rangle + \sum_{\boldsymbol{a}\in\mathcal{A}^L} \langle \boldsymbol{v}, \boldsymbol{c}(\boldsymbol{a}) \rangle \boldsymbol{x}^{\boldsymbol{a}} \geq 0 \text{ for every } \boldsymbol{v} \in \mathcal{L}^* \subset \mathcal{K}^*.$$

Hence $\widetilde{\mathcal{F}}$ is represented in terms of an infinite number of linear inequalities.

Theorem 6.4.

(i) $\widehat{\mathcal{F}} \subset \widetilde{\mathcal{F}}$.

(ii) Assume that there is an $\bar{\boldsymbol{x}} \in \mathbb{R}^n$ such that $\boldsymbol{f}(\bar{\boldsymbol{x}})$ lies in the interior of $\mathcal{K}$. Then the closure of $\widehat{\mathcal{F}}$ coincides with $\widetilde{\mathcal{F}}$

Proof: (i) Suppose that $\boldsymbol{x} = (y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}^L) \in \widehat{\mathcal{F}}$. Then there exists a $(y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}^N)$ such that $\boldsymbol{F}((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \in \mathcal{K}$. Then, for every $\boldsymbol{v} \in \mathcal{L}^* \subset \mathcal{K}^*$, we see that $\langle \boldsymbol{v}, \boldsymbol{f}(\boldsymbol{x}) \rangle = \langle \boldsymbol{v}, \boldsymbol{F}((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \rangle \geq 0$. Hence $\boldsymbol{x} \in \widetilde{\mathcal{F}}$.

(ii) We have shown above that $\widehat{\mathcal{F}} \subset \widetilde{\mathcal{F}}$. Since $\widetilde{\mathcal{F}}$ is closed, we know that the closure of $\widehat{\mathcal{F}}$ is contained in $\widetilde{\mathcal{F}}$. Hence it suffices to show that the closure of $\widehat{\mathcal{F}}$ contains $\widetilde{\mathcal{F}}$. Assume on the contrary that there is an $\tilde{\boldsymbol{x}} \in \widetilde{\mathcal{F}}$ which does not lie in the closure of $\widehat{\mathcal{F}}$. Then, by the separation theorem (see, for example [13]), there exist $\tilde{c}(\boldsymbol{a}) \in \mathbb{R}$ $(\boldsymbol{a} \in \mathcal{A}^L)$ and $\tilde{\zeta} \in \mathbb{R}$ such that

$$\sum_{\boldsymbol{a} \in \mathcal{A}^L} \tilde{c}(\boldsymbol{a}) \tilde{\boldsymbol{x}}^{\boldsymbol{a}} > \tilde{\zeta} \geq \sum_{\boldsymbol{a} \in \mathcal{A}^L} \tilde{c}(\boldsymbol{a}) \boldsymbol{x}^{\boldsymbol{a}} \text{ for every } \boldsymbol{x} \in \widehat{\mathcal{F}}. \tag{34}$$

Now, replace the objective function $f_0(\boldsymbol{x})$ of the POP (2) and the primal LOP (29) by the linear function $\sum_{\boldsymbol{a} \in \mathcal{A}^L} \tilde{c}(\boldsymbol{a}) \boldsymbol{x}^{\boldsymbol{a}}$ and apply Theorem 6.3. Then there exists a $\boldsymbol{v} \in \mathcal{L}^*$ such that $\sum_{\boldsymbol{a} \in \mathcal{A}^L} \tilde{c}(\boldsymbol{a}) \boldsymbol{x}^{\boldsymbol{a}} + \langle \boldsymbol{v}, \boldsymbol{f}(\boldsymbol{x}) \rangle = \zeta$ for every $\boldsymbol{x} \in \mathbb{R}^n$. Here $\zeta = \sup_{\boldsymbol{x} \in \widehat{\mathcal{F}}} \sum_{\boldsymbol{a} \in \mathcal{A}^L} \tilde{c}(\boldsymbol{a}) \boldsymbol{x}^{\boldsymbol{a}}$. Therefore $\sum_{\boldsymbol{a} \in \mathcal{A}^L} \tilde{c}(\boldsymbol{a}) \tilde{\boldsymbol{x}}^{\boldsymbol{a}} \leq \sum_{\boldsymbol{a} \in \mathcal{A}^L} \tilde{c}(\boldsymbol{a}) \tilde{\boldsymbol{x}}^{\boldsymbol{a}} + \langle \boldsymbol{v}, \boldsymbol{f}(\tilde{\boldsymbol{x}}) \rangle = \zeta \leq \tilde{\zeta}$. This contradicts to the first strict inequality in (34). ∎

The result above is an extension of Theorem 2.1 of Fujie-Kojima [2] for QOPs to POPs over cones. Also the relation $\widehat{\mathcal{F}} \subset \widetilde{\mathcal{F}}$ played an essential role in the convergence analysis of the Kojima-Tunçel [7, 8] successive convex relaxations of nonconvex sets.

6.3. Application to Lasserre's SDP relaxation

We derive Theorem 4.2 (b) of [9], one of the main results shown by Lasserre as a special case of Theorem 6.2. Consider a polynomial program of the form

$$\text{maximize } f_0(\boldsymbol{x}) \text{ subject to } \tilde{f}_j(\boldsymbol{x}) \geq 0 \ (j = 1, 2, \ldots, m). \tag{35}$$

Here we assume that $f_0(\boldsymbol{x})$ and $\tilde{f}_j(\boldsymbol{x})$ are polynomials in $x_1, x_2, \ldots, x_n$ of degree at most ω_0 and ω_j $(j = 1, 2, \ldots, m)$. Let $\tilde{\omega}_j = \lceil \omega_j / 2 \rceil$ and choose a nonnegative integer N not less than $\omega_0/2$ and $\tilde{\omega}_j$ $(j = 1, 2, \ldots, m)$. Define

$$\begin{aligned}
\boldsymbol{u}^r(\boldsymbol{x}) &= \text{the column vector consisting of a basis for real-valued} \\
&\quad\ \text{polynomials of degree } r \text{ given in (14)}, \\
\ell_j &= \text{the dimension of the column vector } \boldsymbol{u}^{N-\tilde{\omega}_j}(\boldsymbol{x}) \ (j = 1, 2, \ldots, m), \\
\ell_{m+1} &= \text{the dimension of the column vector } \boldsymbol{u}^N(\boldsymbol{x}), \\
\mathcal{K}_j &= \mathcal{S}_+^{\ell_j} \ (j = 1, 2, \ldots, m+1), \\
\boldsymbol{f}_j(\boldsymbol{x}) &= \tilde{f}_j(\boldsymbol{x}) \boldsymbol{u}^{N-\tilde{\omega}_j}(\boldsymbol{x}) \left(\boldsymbol{u}^{N-\tilde{\omega}_j}(\boldsymbol{x}) \right)^T \in \mathcal{S}^{\ell_j} \ (j = 1, 2, \ldots, m), \\
\boldsymbol{f}_{m+1}(\boldsymbol{x}) &= \boldsymbol{u}^N(\boldsymbol{x}) \left(\boldsymbol{u}^N(\boldsymbol{x}) \right)^T \in \mathcal{S}_+^{\ell_{m+1}}.
\end{aligned}$$

Then we can derive

$$\text{maximize } f_0(\boldsymbol{x}) \text{ subject to } \boldsymbol{f}_j(\boldsymbol{x}) \in \mathcal{K}_j \ (j = 1, 2, \ldots, m+1) \tag{36}$$

as a POP over $\mathcal{K} = \mathcal{K}_1 \times \mathcal{K}_2 \times \cdots \times \mathcal{K}_{m+1}$, which is equivalent to the polynomial program (35), and its linearization

$$\left.\begin{array}{ll} \text{maximize} & F_0((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \\ \text{subject to} & \boldsymbol{F}_j((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})) \in \mathcal{K}_j \ (j = 1, 2, \ldots, m+1), \end{array}\right\} \tag{37}$$

where $F_0((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}))$ and $\boldsymbol{F}_j((y_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}))$ denote the linearization of $f_0(\boldsymbol{x})$ and $\boldsymbol{f}_j(\boldsymbol{x})$ $(j = 1, 2, \ldots, m+1)$, respectively. We note that the LOP (37) is corresponding to "the LMI problem" (4.5) of Lasserre [9], but any problem corresponding to (36) was not described in [9] explicitly.

Since the dual $\mathcal{K}_j^*$ of each $\mathcal{K}_j = \mathcal{S}_+^{\ell_j}$ is itself, the Lagrangian function is

$$L(\boldsymbol{x}, \boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_{m+1}) = f_0(\boldsymbol{x}) + \sum_{j=1}^{m+1} \boldsymbol{V}_j \bullet \boldsymbol{f}_j(\boldsymbol{x})$$
$$\text{for every } \boldsymbol{x} \in \mathbb{R}^n \ \text{ and } \boldsymbol{V}_j \in \mathcal{S}_+^{\ell_j} \ (j = 1, 2, \ldots, m+1).$$

Let $\zeta \in \mathbb{R}$. Suppose that the identity

$$L(\boldsymbol{x}, \boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_{m+1}) = \zeta \ \text{ for all } \boldsymbol{x} \in \mathbb{R}^n. \tag{38}$$

holds for some $\boldsymbol{V}_j \in \mathcal{S}_+^{\ell_j}$ $(j = 1, 2, \ldots, m)$. Then we know by Lemma 6.1 that $(\boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_{m+1})$ is a feasible solution of the dual LOP of (37). We represent each $\boldsymbol{V}_j \in \mathcal{S}_+^{\ell_j}$ as $\boldsymbol{V}_j = \sum_{k=1}^{\ell_j} \lambda_{jk} \boldsymbol{w}_{jk} \boldsymbol{w}_{jk}^T$, where $\boldsymbol{w}_{jk}$ denotes an eigenvector of $\boldsymbol{V}_j$ and λ_{jk} the eigenvalue of $\boldsymbol{V}_j$ associated with $\boldsymbol{w}_{jk}$. Then

$$\begin{aligned} & L(\boldsymbol{x}, \boldsymbol{V}_1, \ldots, \boldsymbol{V}_{m+1}) \\ = \ & f_0(\boldsymbol{x}) + \sum_{j=1}^{m} \left(\sum_{k=1}^{\ell_j} \lambda_{jk} \boldsymbol{w}_{jk} \boldsymbol{w}_{jk}^T \right) \bullet \left(\tilde{f}_j(\boldsymbol{x}) \boldsymbol{u}^{N-\tilde{\omega}_j}(\boldsymbol{x}) \left(\boldsymbol{u}^{N-\tilde{\omega}_j}(\boldsymbol{x}) \right)^T \right) \\ & + \left(\sum_{k=1}^{\ell_{m+1}} \lambda_{m+1,k} \boldsymbol{w}_{m+1,k} \boldsymbol{w}_{m+1,k}^T \right) \bullet \boldsymbol{u}^N(\boldsymbol{x}) \left(\boldsymbol{u}^N(\boldsymbol{x}) \right)^T \\ = \ & f_0(\boldsymbol{x}) + \sum_{j=1}^{m} \tilde{f}_j(\boldsymbol{x}) \left(\sum_{k=1}^{\ell_j} \lambda_{jk} (\boldsymbol{w}_{jk}^T \boldsymbol{u}^{N-\tilde{\omega}_j}(\boldsymbol{x}))^2 \right) + \sum_{k=1}^{\ell_{m+1}} \lambda_{m+1,k} (\boldsymbol{w}_{m+1,k}^T \boldsymbol{u}^N(\boldsymbol{x}))^2. \end{aligned}$$

Letting

$$\begin{aligned} t_j(\boldsymbol{x}) &= \sum_{k=1}^{\ell_j} (\sqrt{\lambda_{jk}} \boldsymbol{w}_{jk}^T \boldsymbol{u}^{N-\tilde{\omega}_j}(\boldsymbol{x}))^2 \ (j = 1, 2, \ldots, m), \\ q(\boldsymbol{x}) &= \sum_{k=1}^{\ell_{m+1}} (\sqrt{\lambda_{m+1,k}} \boldsymbol{w}_{m+1,k}^T \boldsymbol{u}^N(\boldsymbol{x}))^2, \end{aligned}$$

we obtain $L(\boldsymbol{x}, \boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_{m+1}) = f_0(\boldsymbol{x}) + \sum_{j=1}^{m} \tilde{f}_j(\boldsymbol{x}) t_j(\boldsymbol{x}) + q(\boldsymbol{x})$, for some polynomial $q(\boldsymbol{x})$ of degree at most $2N$ and some polynomials $t_j(\boldsymbol{x})$ of degree at most $2N - \omega_j$ $(j = 1, 2, \ldots, m)$, all sums of squares.

Conversely, suppose that the identity

$$f_0(\boldsymbol{x}) + \sum_{j=1}^{m} \tilde{f}_j(\boldsymbol{x}) t_j(\boldsymbol{x}) + q(\boldsymbol{x}) = \zeta \text{ for all } \boldsymbol{x} \in \mathbb{R}^n \tag{39}$$

holds for some polynomial $q(\boldsymbol{x})$ of degree at most $2N$ and some polynomials $t_j(\boldsymbol{x})$ of degree at most $2N - \omega_j$ $(j = 1, 2, \ldots, m)$, all sums of squares such that

$$q(\boldsymbol{x}) = \sum_{k=1}^{\ell'_{m+1}} (\boldsymbol{w}_{m+1,k}^T \boldsymbol{u}^N(\boldsymbol{x}))^2 \text{ and } t_j(\boldsymbol{x}) = \sum_{k=1}^{\ell'_j} (\boldsymbol{w}_{jk}^T \boldsymbol{u}^{N-\tilde{\omega}_j}(\boldsymbol{x}))^2.$$

Let $\boldsymbol{V}_j = \sum_{k=1}^{\ell'_j} \boldsymbol{w}_{jk} \boldsymbol{w}_{jk}^T$ $(j = 1, 2, \ldots, m+1)$. Then $\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_{m+1}) = \zeta$.

Let $\zeta \in \mathbb{R}$. We have shown that there exist $\boldsymbol{V}_j \in \mathcal{S}_+^{\ell_j}$ $(j = 1, 2, \ldots, m+1)$ satisfying (38) if and only if there exist a polynomial $q(\boldsymbol{x})$ of degree at most $2N$ and polynomials $t_j(\boldsymbol{x})$ of degree at most $2N - \omega_j$ $(j = 1, 2, \ldots, m)$, all sums of squares, satisfying (39). Consequently we obtain the corollary below by Lemma 6.1 and Theorem 6.2.

Corollary 6.5. *Let $\zeta^* < \infty$ be the supremum objective values of the POP (36) over the cone $\mathcal{K} = \mathcal{S}_+^{\ell_1} \times \mathcal{S}_+^{\ell_2} \times \cdots \times \mathcal{S}_+^{\ell_{m+1}}$. Suppose that the identity (39) holds for $\zeta = \zeta^*$ and some polynomial $q(\boldsymbol{x})$ of degree at most $2N$ and some polynomials $t_j(\boldsymbol{x})$ of degree at most $2N - \omega_j$ $(j = 1, 2, \ldots, m)$, all sums of squares. Then*

$$\begin{aligned} \zeta^* &= \text{the supremum of the objective values of the primal LOP (37)} \\ &= \text{the optimal value of the dual LOP of (37),} \end{aligned}$$

and if $\boldsymbol{x}^*$ is an optimal solution of the POP (36), then $(y^*_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A}) = ((\boldsymbol{x}^*)^{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{A})$ is an optimal solution of (37). In addition, if $(\boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_{m+1})$ is an optimal solution of the dual LOP of (37), then

$$\begin{aligned} & f_0(\boldsymbol{x}) + \sum_{j=1}^{m+1} \boldsymbol{V}_j \bullet \boldsymbol{f}_j(\boldsymbol{x}) \\ &= f_0(\boldsymbol{x}) + \sum_{j=1}^{m} \tilde{f}_j(\boldsymbol{x}) \left(\sum_{k=1}^{\ell_j} \lambda_{jk} (\boldsymbol{w}_{jk}^T \boldsymbol{u}^{N-\tilde{\omega}_j}(\boldsymbol{x}))^2 \right) + \sum_{k=1}^{\ell_{m+1}} \lambda_{m+1,k} (\boldsymbol{w}_{m+1,k}^T \boldsymbol{u}^N(\boldsymbol{x}))^2 \\ &= \zeta^* \text{ for every } \boldsymbol{x} \in \mathbb{R}^n \end{aligned}$$

holds. Here $\boldsymbol{w}_{jk}$ denotes an eigenvector of $\boldsymbol{V}_j$ and λ_{jk} the eigenvalue of $\boldsymbol{V}_j$ associated with $\boldsymbol{w}_{jk}$.

The corollary above corresponds to Theorem 4.2 (b) of [9].

7. Concluding Remarks

We have presented a new framework for convex relaxation of POPs over cones in terms of LOPs over cones. Although this framework is quite general and flexible, various important theoretical and practical issues remain.

From a practical point of view, the new framework certainly provides various ways of convex relaxation using LOPs over cones; in particular, the SOCP relaxation is likely to play a more important role. Further theoretical and numerical investigation is necessary to resolve the issue of how we create inexpensive valid polynomial constraints over cones that yield effective convex relaxation. Combining the branch-and-bound method into the new framework is also an interesting approach to utilize the framework.

Exact optimal values of a few special cases [6, 25, 26] of QOPs can be attained by constructing a single SDP relaxation. For some other cases [3, 15, 23] of QOPs, approximate optimal values generated through their SDP relaxations attain some guaranteed percentages of their exact optimal values. However, a single application of SDP relaxation to a general POP (or even a general QOP) neither attain the optimal value nor any guaranteed percentage of the optimal value. To get better bounds for the optimal value, it is necessary to use convex relaxation techniques repeatedly or to incorporate convex relaxation into some other optimization methods such as the branch-and-bound method. The former methodology includes the successive convex relaxation technique [1, 7, 8, 12] presented in Section 4.5, and the RLT (Reformulation-Linearization Technique) [17, 20] which constructs a sequence or a hierarchy of convex relaxation problems and then solves them sequentially (see also [9, 10]). The new framework proposed in this paper covers those techniques.

References

[1] E. Balas, S. Ceria and G. Cornuéjols: A lift-and-project cutting plane algorithm for mixed 0-1 programs. *Mathematical Programming*, **58** (1993) 295–323.

[2] T. Fujie and M. Kojima: Semidefinite programming relaxation for nonconvex quadratic programs. *Journal of Global Optimization*, **10** (1997) 367–380.

[3] M. X. Goemans and D. P. Williamson: Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of Association for Computing Machinery*, **42** (1995) 1115–1145.

[4] A. Graham: *Kronecker Products and Matrix Calculus: with Applications* (John Wiley & Sons, New York, 1981).

[5] S. Kim and M. Kojima: Second order cone programming relaxation of nonconvex quadratic optimization problems. *Optimization Methods and Software*, **15** (2001) 201–224.

[6] S. Kim and M. Kojima: Exact solutions of some nonconvex quadratic optimization problems via SDP and SOCP relaxations. B-375, Dept. of Math. and Comp. Sciences, Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8552, January 2002.

[7] M. Kojima and L. Tunçel: Cones of matrices and successive convex relaxations of nonconvex sets. *SIAM Journal on Optimization* **10** (2000) 750–778.

[8] M. Kojima and L. Tunçel: Discretization and localization in successive convex relaxation methods for nonconvex quadratic optimization problems. *Mathematical Programming*, **89** (2000) 79–111.

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

[10] J. B. Lasserre: An explicit equivalent convex positive semidefinite program for nonlinear 0-1 programs. *SIAM Journal on Optimization*, **12** (2002) 756–769.

[11] M. Laurent: A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming. Report PNA-R0108, CWI, Amsterdam, The Netherlands, 2001.

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

[13] O. L. Mangasarian: *Nonlinear Programming* (McGraw-Hill, New York, 1969).

[14] M. Muramatsu and T. Suzuki: A new second-order cone programming relaxation for max-cut problems. *Journal of the Operations Research Society of Japan*, **46** (2003) 164–177.

[15] Y. Nesterov: Semidefinite relaxation and nonconvex quadratic optimization. *Optimization Methods and Software*, **9** (1998) 141–160.

[16] Y. Nesterov and A. Nemirovskii: *Interior-Point Polynomial algorithms in Convex Programming* (SIAM, Philadelphia, 1994).

[17] H. D. Sherali and W. P. Adams: A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, **3** (1990) 411–430.

[18] H. D. Sherali and W. P. Adams: A of relaxations and convex hull characterizations for mixed-integer zero-one programming problems. *Discrete Applied Mathematics*, **52** (1994) 83–106.

[19] H. D. Sherali and C. H. Tuncbilek: A global optimization algorithm for polynomial programming problems using a reformulation-linearization technique. *Journal of Global Optimization*, **2** (1992) 101–112.

[20] H. D. Sherali and C. H. Tuncbilek: A reformulation-convexification approach for solving nonconvex quadratic programming problems. *Journal of Global Optimization*, **7** (1995) 1–31.

[21] N. Z. Shor: Dual quadratic estimates in polynomial and boolean programming. *Annals of Operations Research*, **25** (1990) 163-168.

[22] G. Xue and Y. Ye: An efficient algorithm for minimizing a sum of p-norms. *SIAM Journal on Optimization*, **10** (2000) 551–579.

[23] Y. Ye: Approximating quadratic programming with bound and quadratic constraints. *Mathematical Programming*, **84** (1999) 219–226.

[24] Y. Ye: Approximating global quadratic optimization with convex quadratic constraints. *Journal of Global Optimization*, **15** (1999) 1–17.

[25] Y. Ye and S. Zhang: New results on quadratic minimization. Technical Report SEEM2001-03, Dept. of System Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, Hong Kong, 2001.

[26] S. Zhang: Quadratic optimization and semidefinite relaxation. *Mathematical Programming*, **87** (2000) 453-465.

Masakazu Kojima
Department of Mathematical and Computing Sciences,
Tokyo Institute of Technology,
2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152-8552 Japan.
E-mail:kojima@is.titech.ac.jp