Research Reports on Mathematical and Computing Sciences
Series B : Operations Research

Department of Mathematical and Computing Sciences
Tokyo Institute of Technology
2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152-8552 Japan

Generalized Lagrangian Duals and Sums of Squares Relaxations of Sparse Polynomial Optimization Problems

Sunyoung Kim†, Masakazu Kojima‡ and Hayato Waki⋆

Research Report B-395, September 2003. Revised July 2004.

Abstract.
Sequences of generalized Lagrangian duals and their SOS (sums of squares of polynomials) relaxations for a POP (polynomial optimization problem) are introduced. The sparsity of polynomials in the POP is used to reduce the sizes of the Lagrangian duals and their SOS relaxations. It is proved that the optimal values of the Lagrangian duals in the sequence converge to the optimal value of the POP using a method from the penalty function approach. The sequence of SOS relaxations is transformed into a sequence of SDP (semidefinite program) relaxations of the POP, which correspond to duals of modification and generalization of SDP relaxations given by Lasserre for the POP.

Key words.
Polynomial Optimization Problem, Sparsity, Global Optimization, Lagrangian Relaxation, Lagrangian Dual, Sums of Squares Optimization, Semidefinite Program, Semidefinite Program Relaxation

† Department of Mathematics, Ewha Women's University, 11-1 Dahyun-dong, Sudaemoon-gu, Seoul 120-750 Korea. A considerable part of this work was conducted while this author was visiting Tokyo Institute of Technology. Research supported by Kosef R004-000-2001-00200. *skim@ewha.ac.kr*

‡ Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152-8552 Japan. Research supported by Grant-in-Aid for Scientific Research on Priority Areas 16016234. *kojima@is.titech.ac.jp*

⋆ Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152-8552 Japan. *Hayato.Waki@is.titech.ac.jp*

1 Introduction

POPs (Polynomial optimization problems or optimization problems with polynomial objective and constraints) represent a broad range of applications in science and engineering. Convex relaxation methods that relax nonconvex constraints and objective function of POPs with convex ones have been widely used for POPs. Various convex relaxation methods exist for POPs, which include quadratic, linear and 0-1 integer programming as special cases. After a lift-and-project LP procedure for 0-1 integer linear programs by Balas-Ceria-Cornuéjols [1], the RLT (Reformulation-Linearization Technique) by Sherali-Adams [18] and an SDP (Semidefinite Programming) relaxation method by Lovász-Schrijver [11] were introduced. Many convex relaxation methods such as the RLT [20, 19] for 0-1 mixed integer polynomial programs, the SCRM (Successive Convex Relaxation Method) [8, 9] for QOPs (Quadratic Optimization Problems), SOCP (Second Order Cone Programming) relaxations [4, 5] for QOPs, and SOS (Sums of Squares) relaxations for POPs [13] have been proposed.

Recently, an important theoretical development has been made by Lasserre [10] toward achieving optimal values of POPs. His method to obtain a sequence of SDP relaxations can be considered as a primal approach which is divided into two phases, according to the paper [6]. The first phase is to transform a given POP into an equivalent optimization problem by adding valid polynomial constraints over semidefinite cones to the POP. The resulting optimization problem forms a PSDP (polynomial semidefinite programming problem). In the second phase, the PSDP is linearized to an SDP. In theory, we can add any size and/or number of valid polynomial constraints over positive semidefinite cones to the POP in the first phase. Lasserre [10] proposed a systematic way of adding valid polynomial constraints over semidefinite cones to the original POP in the first phase to generate a sequence of SDP relaxations in the second phase. It was proved that when the feasible region of the POP is compact, its optimal value can be approximated within any accuracy by the sequence of SDP relaxations. However, the size of an SDP relaxation to be solved in the sequence increases very rapidly as higher accuracy for an approximation to the optimal value of the POP is required.

It is known that practical solvability of SDPs depends on their sizes. A great deal of efforts continue to be made to resolve the difficulty of solving large scale SDPs. See the survey paper [12] and references therein. From a computational point of view, a close approximation to the optimal value of a POP is hard to achieve without efficient methods to handle large scale SDPs. Rapid increase of the size of SDP relaxations in the sequence in [10] as higher accuracy desired makes it difficult to achieve their optimal values in practice. As a result, only small-size POPs can be solved by the method in [10].

The purpose of this paper is to present generalized Lagrangian duals and their SOS (sums of squares) relaxations [13] for sparse POPs. This approach may be regarded as dual of Lasserre's SDP relaxations [10] mentioned above. For the standard Lagrangian function of a POP, it is common that nonnegative numbers are assigned to Lagrangian multipliers. Instead of selecting nonnegative numbers for Lagrangian multipliers, we choose Lagrangian multipliers to be SOS polynomials satisfying similar sparsity to associated constraint polynomials. Then, we define a generalized Lagrangian dual for a POP over such SOS polynomial multipliers, and provide a theoretical foundation for the generalized Lagrangian dual. After a sequence of sets of SOS polynomials is constructed, *e.g.* SOS polynomials of increasing degree, for Lagrangian multipliers, a sequence of Lagrangian duals is obtained. We derive a sufficient condition for the sequence of Lagrangian duals to attain the optimal value of the POP, based on the idea of the penalty function method. For practical purposes, each Lagrangian dual in the sequence is relaxed to an SOS optimization problem, which is further converted into an equivalent SDP. Thus we have sequences of SOS relaxations and SDP relaxations of the POP. The resulting sequence of SDP relaxations corresponds to dual of the sequence of SDP relaxations obtained from the primal approach.

An advantage of the dual approach in this paper is that the sparsity of objective and constraint polynomials in a POP can be exploited to reduce the size of the SDP relaxations. The size of the SDP relaxations depends on the supports of the polynomials in the dual approach, whereas the size of the SDP relaxations from the primal approach depends on the degree of the polynomials.

This paper is organized as follows. In Section 2, we introduce a POP with a compact feasible region and describe the representation of the sparsity of the POP. Two types of POPs are shown for different characterizations of the compact feasible region. Section 3 includes the derivation of generalized Lagrangian duals of the two POPs over Lagrangian multipliers from a set of SOS polynomials satisfying similar sparsity to the associated constraint polynomials. We show a sufficient condition for the Lagrangian duals to attain the optimal value of the original POP. In Section 4, we present numerical methods to approximate optimal values of the generalized Lagrangian duals numerically using SOS relaxations. The numerical methods take advantage of the sparsity of the original POP to produce SDP relaxations with smaller size. The relationship between the resulting SDP relaxations and the SDP relaxations in [10] is also discussed. Section 5 contains the proof of a Lemma which plays an important role in Section 3. In Section 6, we report preliminary numerical results for a sparse and structured POP, and show an computational advantage of the proposed SOS and SDP relaxations. Section 7 is devoted to concluding discussions.

Throughout the paper, we use the following notation: Let $\mathbb{R}^n$ and $\mathbb{Z}^n_+ \subset \mathbb{R}^n$ denote the n-dimensional Euclidean space and the set of n-dimensional nonnegative integer vectors, respectively. Let $f_j : \mathbb{R}^n \to \mathbb{R}$ be a real valued polynomial in $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ $(j = 0, 1, 2, \ldots, m)$. We denote each polynomial $f_j(\boldsymbol{x})$ as $f_j(\boldsymbol{x}) = \sum_{\boldsymbol{a} \in \mathcal{F}_j} c_j(\boldsymbol{a})\boldsymbol{x}^{\boldsymbol{a}}$, where a nonempty finite subset $\mathcal{F}_j$ of $\mathbb{Z}^n_+$ denotes a support of the polynomial $f_j(\boldsymbol{x})$, $c_j(\boldsymbol{a}) \in \mathbb{R}$ and $\boldsymbol{x}^{\boldsymbol{a}} = x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$ for every $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in \mathcal{F}_j$ $(j = 0, 1, 2, \ldots, m)$ and $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. Let r_j denote the degree of each polynomial $f_j(\boldsymbol{x})$ $(j = 0, 1, 2, \ldots, m)$; $r_j = \max\{\sum_{i=1}^n a_j : \boldsymbol{a} \in \mathcal{F}_j\}$.

2 Polynomial optimization problems and sparsity

We consider the POP (polynomial optimization problem):

$$\text{minimize } f_0(\boldsymbol{x}) \;\; \text{subject to } f_j(\boldsymbol{x}) \geq 0 \; (j = 1, 2, \ldots, m). \tag{1}$$

Let us focus on the support $\mathcal{F}_j$ of the polynomial $f_j(\boldsymbol{x})$ $(j = 0, 1, 2, \ldots, m)$ to describe the sparsity of the POP (1). A polynomial $f(\boldsymbol{x})$ of degree r or its support $\mathcal{F}$ is called sparse if the number of elements in the support $\mathcal{F}$ is much smaller than the number of elements in the support $\mathcal{G}(\xi) \equiv \{\boldsymbol{a} \in \mathbb{Z}^n_+ : \sum_{i=1}^m a_i \leq \xi\}$ of a general fully dense polynomial of degree ξ. In particular, if the number of indices in $I_+(\mathcal{F}) \equiv \{i : a_i > 0 \;$ for some $\boldsymbol{a} \in \mathcal{F}\}$ is much smaller than n, then $f(\boldsymbol{x})$ is sparse. We present two examples below.

Example 2.1. A box constraint POP. Let $m = n$ and $f_j(\boldsymbol{x}) = 1 - x_j^2$ $(j = 1, 2, \ldots, n)$. In this case, we have $\mathcal{F}_j = \{\boldsymbol{0},\ 2\boldsymbol{e}^j\}$ $(j = 1, 2, \ldots, n)$. Each $\mathcal{F}_j$ has two elements and $I_+(\mathcal{F}_j) = \{j\}$. Here $\boldsymbol{e}^j$ denotes the jth unit coordinate vector of $\mathbb{R}^n$ with 1 in the jth component and 0 elsewhere.

Example 2.2. Let $m = n$, $f_1(\boldsymbol{x}) = \alpha_1 - x_1^2$ and $f_j(\boldsymbol{x}) = \alpha_j - (\beta_{j-1}x_{j-1} - x_j)^2$ $(j = 2, 3, \ldots, n)$, where α_j $(j = 1, 2, \ldots, n)$ and β_j $(j = 1, 2, \ldots, n-1)$ denote positive numbers. Then, $\mathcal{F}_1 = \{\boldsymbol{0}, 2\boldsymbol{e}^1\}$, $\mathcal{F}_j = \{\boldsymbol{0},\ 2\boldsymbol{e}^{j-1}, \boldsymbol{e}^{j-1} + \boldsymbol{e}^j, 2\boldsymbol{e}^j\}$ $(j = 2, 3, \ldots, n)$. Each $\mathcal{F}_j$ has at most four elements and $I_+(\mathcal{F}_j)$ contains at most two indices.

Another example is given in Section 6 with preliminary numerical results.

Let F denote the feasible region of the POP (1);

$$F \;=\; \{\boldsymbol{x} \in \mathbb{R}^n : f_j(\boldsymbol{x}) \geq 0 \ (j = 1, 2, \ldots, m)\}.$$

Throughout the paper, we assume that F is nonempty and bounded. Then, the POP (1) has a finite optimal value ζ^* at an optimal solution $\boldsymbol{x}^* \in F$;

$$\zeta^* \;=\; \min\{f_0(\boldsymbol{x}) : \boldsymbol{x} \in F\} = f_0(\boldsymbol{x}^*) \ \text{ and } \boldsymbol{x}^* \in F.$$

In what it follows, we further need a bound $\rho > 0$ to be known explicitly for the feasible region. We are concerned with the following cases:

- $F \subset C_\rho \equiv \{\boldsymbol{x} \in \mathbb{R}^n : \rho^2 - x_i^2 \geq 0 \ (i = 1, 2, \ldots, n)\}$ or
- $F \subset B_\rho \equiv \{\boldsymbol{x} \in \mathbb{R}^n : \rho^2 - \boldsymbol{x}^T\boldsymbol{x} \geq 0\}$.

If $F \subset C_\rho$ holds then the POP (1) is equivalent to

$$\text{minimize } f_0(\boldsymbol{x}) \ \text{ subject to } f_j(\boldsymbol{x}) \geq 0 \ (j = 1, 2, \ldots, m) \ \text{ and } \boldsymbol{x} \in C_\rho. \tag{2}$$

Example 2.1 is a special case of the POP (2) where we take $m = 0$ and $\rho = 1$. If $F \subset B_\rho$ is satisfied, then we have $F \subset C_\rho$; hence the two POPs (1) and (2) are equivalent to

$$\text{minimize } f_0(\boldsymbol{x}) \ \text{ subject to } f_j(\boldsymbol{x}) \geq 0 \ (j = 1, 2, \ldots, m) \ \text{ and } \boldsymbol{x} \in B_\rho. \tag{3}$$

In the next section, we present a (generalized) Lagrangian function, a Lagrangian relaxation, a Lagrangian dual and its SOS relaxation for each of the POPs (2) and (3). For both POPs, the Lagrangian dual converges to the optimal value ζ^* of the original POP (1) under a moderate assumption. But, only the SOS relaxation derived from the POP (3) is guaranteed to converge to ζ^*, while the SOS relaxation from the POP (2) inherits more sparsity of the original POP (1) than the one from the POP (3).

3 Generalized Lagrangian duals

3.1 Lagrangian functions

Let $\overline{\Sigma}$ denote the set of sums of squares of polynomials in $\boldsymbol{x} \in \mathbb{R}^n$;

$$\overline{\Sigma} \;=\; \left\{ \sum_{i=1}^{k} \chi_i(\boldsymbol{x})^2 : \begin{array}{l} \chi_i \ \text{ is a polynomial in } \boldsymbol{x} \in \mathbb{R}^n \ (i = 1, 2, \ldots, k) \\ \text{and } k \text{ is any finite positive integer} \end{array} \right\}.$$

We define two types of (generalized) Lagrangian functions $L_B : B_\rho \times \overline{\Sigma}^m \to \mathbb{R}$ for POP (3) and $L_C : \mathbb{R}^n \times \overline{\Sigma}^{m+n} \to \mathbb{R}$ for POP (2):

$$\begin{aligned} L_B(\boldsymbol{x}, \boldsymbol{\varphi}) &= f_0(\boldsymbol{x}) - \sum_{j=1}^{m} \varphi_j(\boldsymbol{x}) f_j(\boldsymbol{x}) \\ L_C(\boldsymbol{x}, \boldsymbol{\varphi}, \boldsymbol{\psi}) &= L_B(\boldsymbol{x}, \boldsymbol{\varphi}) - \sum_{i=1}^{n} \psi_i(\boldsymbol{x})(\rho^2 - x_i^2) \\ &= f_0(\boldsymbol{x}) - \sum_{j=1}^{m} \varphi_j(\boldsymbol{x}) f_j(\boldsymbol{x}) - \sum_{i=1}^{n} \psi_i(\boldsymbol{x})(\rho^2 - x_i^2). \end{aligned}$$

Here $\overline{\Sigma}^\ell$ denotes the Cartesian product of ℓ-tuples of $\overline{\Sigma}$;

$$\overline{\Sigma}^\ell = \left\{ (\varphi_1, \varphi_2, \ldots, \varphi_\ell) : \varphi_j \in \overline{\Sigma} \ (j = 1, 2, \ldots, \ell) \right\} \ (\ell = m \ \text{ or } m+n).$$

Let $(\boldsymbol{\varphi}, \boldsymbol{\psi}) \in \overline{\Sigma}^{m+n}$. Then

$$\left. \begin{array}{l} L_C(\boldsymbol{x}, \boldsymbol{\varphi}, \boldsymbol{\psi}) \leq f_0(\boldsymbol{x}) \ \text{ if } \boldsymbol{x} \in F \bigcap C_\rho, \\ L_B(\boldsymbol{x}, \boldsymbol{\varphi}) \leq f_0(\boldsymbol{x}) \ \text{ if } \boldsymbol{x} \in F \bigcap B_\rho, \\ L_C(\boldsymbol{x}, \boldsymbol{\varphi}, \boldsymbol{\psi}) \leq L_B(\boldsymbol{x}, \boldsymbol{\varphi}) \ \text{ if } \boldsymbol{x} \in B_\rho. \end{array} \right\} \tag{4}$$

3.2 Lagrangian relaxations and duals

We introduce a (generalized) Lagrangian relaxation of the POP (2):

$$L_C^*(\boldsymbol{\varphi}, \boldsymbol{\psi}) = \inf \{ L_C(\boldsymbol{x}, \boldsymbol{\varphi}, \boldsymbol{\psi}) : \boldsymbol{x} \in \mathbb{R}^n \}$$

for each fixed $(\boldsymbol{\varphi}, \boldsymbol{\psi}) \in \overline{\Sigma}^{m+n}$, and a (generalized) Lagrangian relaxation of the POP (3):

$$L_B^*(\boldsymbol{\varphi}) = \inf \{ L_B(\boldsymbol{x}, \boldsymbol{\varphi}) : \boldsymbol{x} \in B_\rho \}$$

for each fixed $\boldsymbol{\varphi} \in \overline{\Sigma}^m$. Let $(\boldsymbol{\varphi}, \boldsymbol{\psi}) \in \overline{\Sigma}^{m+n}$. By (4), we see that

$$\left. \begin{array}{l} L_C^*(\boldsymbol{\varphi}, \boldsymbol{\psi}) \leq \zeta^* = \min\{ f_0(\boldsymbol{x}) : \boldsymbol{x} \in F \} \text{ if } F \subset C_\rho, \\ L_C^*(\boldsymbol{\varphi}, \boldsymbol{\psi}) \leq L_B^*(\boldsymbol{\varphi}) \leq \zeta^* \text{ if } F \subset B_\rho. \end{array} \right\} \tag{5}$$

For every $(\boldsymbol{\Sigma}, \boldsymbol{\Xi}) \subset \overline{\Sigma}^{m+n}$, we define a (generalized) Lagrangian dual of the POP (2):

$$\text{maximize} \quad L_C^*(\boldsymbol{\varphi}, \boldsymbol{\psi}) \quad \text{subject to} \quad (\boldsymbol{\varphi}, \boldsymbol{\psi}) \in \boldsymbol{\Sigma} \times \boldsymbol{\Xi}. \tag{6}$$

For every $\boldsymbol{\Sigma} \subset \overline{\Sigma}^m$, we define a (generalized) Lagrangian dual of the POP (3):

$$\text{maximize} \quad L_B^*(\boldsymbol{\varphi}) \quad \text{subject to} \quad \boldsymbol{\varphi} \in \boldsymbol{\Sigma}. \tag{7}$$

Let $L_C^*(\boldsymbol{\Sigma} \times \boldsymbol{\Xi})$ and $L_B^*(\boldsymbol{\Sigma})$ denote the optimal values of (6) and (7), respectively;

$$L_C^*(\boldsymbol{\Sigma} \times \boldsymbol{\Xi}) = \sup_{(\boldsymbol{\varphi}, \boldsymbol{\psi}) \in \boldsymbol{\Sigma} \times \boldsymbol{\Xi}} L_C^*(\boldsymbol{\varphi}, \boldsymbol{\psi}) \ \text{ and } L_B^*(\boldsymbol{\Sigma}) = \sup_{\boldsymbol{\varphi} \in \boldsymbol{\Sigma}} L_B^*(\boldsymbol{\varphi}).$$

It follows from (5) that

$$\left.\begin{array}{l} L_C^*(\mathbf{\Sigma}\times\mathbf{\Xi}) \leq \zeta^* \text{ if } F\subset C_\rho, \\ L_C^*(\mathbf{\Sigma}\times\mathbf{\Xi}) \leq L_B^*(\mathbf{\Sigma}) \leq \zeta^* \text{ if } F\subset B_\rho \end{array}\right\} \tag{8}$$

holds for every $(\mathbf{\Sigma},\mathbf{\Xi})\subset\overline{\Sigma}^{m+n}$.

Assume that $F\subset C_\rho$. Then, the two POPs (1) and (2) are equivalent. If we restrict $(\boldsymbol{\varphi},\boldsymbol{\psi})$ to the nonnegative orthant $\mathbb{R}^{m+n}_+$ of $\mathbb{R}^{m+n}$, $L_C(\boldsymbol{x},\boldsymbol{\varphi},\boldsymbol{\psi})$ becomes the standard Lagrangian function for the POP (2). It is well-known that a positive duality gap exists in general between the standard Lagrangian dual $L_C^*(\mathbb{R}^{m+n}_+)$ and the optimal value ζ^* of the POP (2). In fact, consider a simple example of $f_0(x)=x^3$ $(\forall x\in\mathbb{R})$, where $n=1,\ m=0$. Then

$$\begin{array}{rcl} \zeta^* &=& \min\left\{x^3 : x^2\leq 1\right\} = -1, \\ L_C(x,\psi) &=& x^3-\psi(1-x^2)\ (\forall x\in\mathbb{R},\ \forall\psi\in\mathbb{R}_+), \\ L_C^*(\psi) &=& -\infty\ (\forall\psi\in\mathbb{R}_+);\quad \text{hence } L_C^*(\mathbb{R}_+)=-\infty. \end{array}$$

As we take a larger set $\mathbf{\Sigma}\times\mathbf{\Xi}\subset\overline{\Sigma}^{m+n}$, the duality gap between $L_C^*(\mathbf{\Sigma}\times\mathbf{\Xi})$ and ζ^* is expected to decrease. In the next section, we derive a sufficient condition on $\mathbf{\Sigma}\times\mathbf{\Xi}\subset\overline{\Sigma}^{m+n}$ for $L_C^*(\mathbf{\Sigma}\times\mathbf{\Xi})$ to attain the optimal value ζ^* of the POP (2). We regard $(\boldsymbol{\varphi},\boldsymbol{\psi})\in\overline{\Sigma}^{m+n}$ as "a penalty parameter" in the derivation, and

$$\Phi_C(\boldsymbol{x},\boldsymbol{\varphi},\boldsymbol{\psi}) = -\sum_{j=1}^m \varphi_j(\boldsymbol{x})f_j(\boldsymbol{x}) - \sum_{i=1}^n \psi_i(\boldsymbol{x})(\rho^2-x_i^2) \tag{9}$$

(the terms added to the objective function $f_0(\boldsymbol{x})$
in the construction of the Lagrangian function $L_C(\boldsymbol{x},\boldsymbol{\varphi},\boldsymbol{\psi})$)

as "a penalty function" with a choice of penalty parameters $(\boldsymbol{\varphi},\boldsymbol{\psi})=(\boldsymbol{\varphi}^p,\boldsymbol{\psi}^p)\in\overline{\Sigma}^{m+n}$ $(p\in\mathbb{Z}_+)$ such that

$$\begin{array}{l} \text{if } \boldsymbol{x}\in F \text{ then } \Phi(\boldsymbol{x},\boldsymbol{\varphi}^p,\boldsymbol{\psi}^p)\to 0 \ \text{ as } p\to\infty, \\ \text{if } \boldsymbol{x}\notin F \text{ then } \Phi(\boldsymbol{x},\boldsymbol{\varphi}^p,\boldsymbol{\psi}^p)\to \infty \ \text{ as } p\to\infty. \end{array}$$

Additional properties of the penalty function $\Phi_C(\boldsymbol{x},\boldsymbol{\varphi}^p,\boldsymbol{\psi}^p)$ are described in Lemma 3.2. It is shown that the Lagrangian function $L_C(\boldsymbol{x},\boldsymbol{\varphi},\boldsymbol{\psi})=f_0(\boldsymbol{x})+\Phi_C(\boldsymbol{x},\boldsymbol{\varphi},\boldsymbol{\psi})$ with the penalty parameter $(\boldsymbol{\varphi},\boldsymbol{\psi})=(\boldsymbol{\varphi}^p,\boldsymbol{\psi}^p)\in\overline{\Sigma}^{m+n}$ has a global minimizer $\boldsymbol{x}^p$ over $\mathbb{R}^n$ with the objective value $f_0(\boldsymbol{x}^p)\to\zeta^*$ as $p\to\infty$ and that $\{\boldsymbol{x}^p\}$ has an accumulation point in the optimal solution set of the POP (2).

3.3 Main theorem

We need some additional notation and symbols. Take a real number $\gamma\geq 1$ such that

$$\left.\begin{array}{rcl} |f_j(\boldsymbol{x})/\gamma| &\leq& 1 \ \text{ if } \|\boldsymbol{x}\|_\infty\leq\sqrt{2}\rho, \\ |f_j(\boldsymbol{x})/\gamma| &\leq& \|\boldsymbol{x}/\rho\|_\infty^r \ \text{ if } \|\boldsymbol{x}\|_\infty\geq\sqrt{2}\rho \end{array}\right\} \tag{10}$$

$(j = 0, 1, 2, \ldots, m)$. Define

$$\left.\begin{array}{rcl} \varphi_j^p(\boldsymbol{x}) &=& (1 - f_j(\boldsymbol{x})/\gamma)^{2p} \ \ (j = 1, 2, \ldots, m, \ p \in \mathbb{Z}_+), \\ \boldsymbol{\varphi}^p(\boldsymbol{x}) &=& (\varphi_1^p(\boldsymbol{x}), \varphi_2^p(\boldsymbol{x}), \ldots, \varphi_m^p(\boldsymbol{x})) \ \ (p \in \mathbb{Z}_+), \\ \psi_i^p(\boldsymbol{x}) &=& ((m+2)\gamma/\rho^2)\,(x_i/\rho)^{2r(p+1)} \ \ (i = 1, 2, \ldots, n, \ p \in \mathbb{Z}_+), \\ \boldsymbol{\psi}^p(\boldsymbol{x}) &=& (\psi_1^p(\boldsymbol{x}), \psi_2^p(\boldsymbol{x}), \ldots, \psi_n^p(\boldsymbol{x})) \ \ (p \in \mathbb{Z}_+). \end{array}\right\} \quad (11)$$

Here $r = \max\{r_j : j = 0, 1, 2, \ldots, m\}$ denotes the maximum of the degree r_j of the polynomial $f_j(\boldsymbol{x})$ $(j = 0, 1, 2, \ldots, m)$. Obviously $(\boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) \in \overline{\Sigma}^{m+n}$ $(p \in \mathbb{Z}_+)$.

Theorem 3.1. *Assume that* $\boldsymbol{\Xi} \times \boldsymbol{\Sigma} \subset \overline{\Sigma}^{m+n}$ *contains an infinite subsequence of* $\{(\boldsymbol{\varphi}^p, \boldsymbol{\psi}^p)\ (p \in \mathbb{Z}_+)\}$. *Then* $L^*_C(\boldsymbol{\Sigma} \times \boldsymbol{\Xi}) = \zeta^*$ *if* $F \subset C_\rho$, *and* $L^*_C(\boldsymbol{\Sigma} \times \boldsymbol{\Xi}) = L^*_B(\boldsymbol{\Sigma}) = \zeta^*$ *if* $F \subset B_\rho$.

To prove the theorem, we use the following lemma whose proof is given in Section 5.

Lemma 3.2. *Suppose that* $F \subset C_\rho$. *Let* p^* *be the smallest nonnegative integer such that* $(m+2)n - 2^{r(p^*+1)} \leq 0$.

(a) *If* $\sqrt{2}\rho \leq \|\boldsymbol{x}\|_\infty$, *then*

$$L_C(\boldsymbol{x}, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) \equiv f_0(\boldsymbol{x}) + \Phi_C(\boldsymbol{x}, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) \geq \|\boldsymbol{x}/\rho\|_\infty^{2rp} \geq \|\boldsymbol{x}/\rho\|_\infty^2$$

for every $p \geq p^*$.

(b) *If* $\tilde{\boldsymbol{x}} \notin F$ *and* $\kappa \in \mathbb{R}$, *then there exist* $\delta > 0$ *and* $\tilde{p} \geq p^*$ *such that*

$$L_C(\boldsymbol{x}, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) \equiv f_0(\boldsymbol{x}) + \Phi_C(\boldsymbol{x}, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) \geq \kappa$$

for every $p \geq \tilde{p}$ *and* $\boldsymbol{x} \in U_\delta(\tilde{\boldsymbol{x}})$. *Here* $U_\delta(\tilde{\boldsymbol{x}}) = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \tilde{\boldsymbol{x}}\| < \delta\}$.

(c) *If* $\hat{\boldsymbol{x}} \in F$ *and* $\epsilon > 0$, *then there exist* $\delta > 0$ *and* $\tilde{p} \geq p^*$ *such that*

$$-\epsilon \leq \Phi_C(\boldsymbol{x}, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p)$$

for every $\boldsymbol{x} \in U_\delta(\hat{\boldsymbol{x}})$ *and* $p \geq \tilde{p}$.

(d) $\lim_{p\to\infty} L^*_C(\boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) = \zeta^*$.

Proof of Theorem 3.1: Note that $B_\rho \subset C_\rho$. Hence, by (8), we only need to show that if $F \subset C_\rho$ then $L^*_C(\boldsymbol{\Sigma} \times \boldsymbol{\Xi}) = \zeta^*$. Suppose that $F \subset C_\rho$. Since the sequence $\{(\boldsymbol{\varphi}^p, \boldsymbol{\psi}^p)\ (p \in \mathbb{Z}_+)\}$ lies in the set $\boldsymbol{\Sigma} \times \boldsymbol{\Xi}$ by the assumption, we see that

$$\limsup_{p\to\infty} L^*_C(\boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) \leq L^*_C(\boldsymbol{\Sigma} \times \boldsymbol{\Xi}) \leq \zeta^*.$$

By (d) of Lemma 3.2,

$$\limsup_{p\to\infty} L^*_C(\boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) \geq \lim_{p\to\infty} L^*_C(\boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) = \zeta^*.$$

Therefore $L^*_C(\boldsymbol{\Sigma} \times \boldsymbol{\Xi}) = \zeta^*$. ∎

3.4 Construction of $\Sigma \times \Xi$ satisfying the assumption of Theorem 3.1

For every nonempty subset $\mathcal{A}$ of $\mathbb{Z}^n_+$, we define

$$\Sigma(\mathcal{A}) \;=\; \left\{ \sum_{i=1}^{k} \chi_i(\boldsymbol{x})^2 : \begin{array}{l} \chi_i \text{ is a polynomial in } \boldsymbol{x} \in \mathbb{R}^n \text{ with a support in } \mathcal{A} \\ (i = 1,2,\ldots,k) \text{ and } k \text{ is any finite positive integer} \end{array} \right\}.$$

Suppose that $\mathcal{A}^q_j$ $(j = 1,2,\ldots,m,\ q \in \mathbb{Z}_+)$ and $\mathcal{B}^q_i$ $(i = 1,2,\ldots,n,\ q \in \mathbb{Z}_+)$ are nonempty finite subsets of $\mathbb{Z}^n_+$ such that

$$\varphi^{\lambda(q)}_j \in \Sigma(\mathcal{A}^q_j) \ \text{ and } \psi^{\lambda(q)}_i(\boldsymbol{x}) \in \Sigma(\mathcal{B}^q_i) \text{ if } q \geq q^* \tag{12}$$

for some $q^* \in \mathbb{Z}_+$ and some mapping λ from $\mathbb{Z}_+$ into itself satisfying

$$\lambda(q) \leq \lambda(q+1) \ (q \in \mathbb{Z}_+) \ \text{ and } \lim_{q\to\infty} \lambda(q) = \infty.$$

Let

$$\boldsymbol{\Sigma}^q = \prod_{j=1}^{m} \Sigma(\mathcal{A}^q_j) \ (q \in \mathbb{Z}_+), \quad \boldsymbol{\Xi}^q \;=\; \prod_{i=1}^{n} \Sigma(\mathcal{B}^q_i) \ (q \in \mathbb{Z}_+),$$
$$\boldsymbol{\Sigma}^\infty = \bigcup_{q\in\mathbb{Z}_+} \boldsymbol{\Sigma}^q \ \text{ and } \ \boldsymbol{\Xi}^\infty = \bigcup_{q\in\mathbb{Z}_+} \boldsymbol{\Xi}^q.$$

By construction, we see that

$$(\boldsymbol{\varphi}^{\lambda(q)}, \boldsymbol{\psi}^{\lambda(q)}) \in \boldsymbol{\Sigma}^q \times \boldsymbol{\Xi}^q \ (q \geq q^*).$$

Hence $\boldsymbol{\Sigma}^\infty \times \boldsymbol{\Xi}^\infty$ contains an infinite subsequence $\{(\boldsymbol{\varphi}^{\lambda(q)}, \boldsymbol{\psi}^{\lambda(q)}) \ (q \geq q^*)\}$. Therefore $\boldsymbol{\Sigma} \times \boldsymbol{\Xi} = \boldsymbol{\Sigma}^\infty \times \boldsymbol{\Xi}^\infty$ satisfies the assumption of Theorem 3.1.

We give some examples of $\mathcal{A}^q_j$ $(j = 1,2,\ldots,m,\ q \in \mathbb{Z}_+)$ and $\mathcal{B}^q_i$ $(i = 1,2,\ldots,n,\ q \in \mathbb{Z}_+)$ satisfying the assumption (12).

Example 3.3. For every $j = 1,2,\ldots,m,\ i = 1,2,\ldots,n,\ q \in \mathbb{Z}_+$, let

$$\mathcal{A}^0_j = \{\boldsymbol{0}\},\ \mathcal{A}^1_j = \{\boldsymbol{0}\} \bigcup \mathcal{F}_j,\ \mathcal{A}^{q+1}_j = \left\{\boldsymbol{a} + \boldsymbol{b} : \boldsymbol{a} \in \mathcal{A}^q_j,\ \boldsymbol{b} \in \mathcal{A}^1_j\right\} \ (q \geq 1),$$
$$\mathcal{B}^q_i = \{k\boldsymbol{e}^i : k = 0,1,2,\ldots,(q+1)r\}.$$

Example 3.4. For every $j = 1,2,\ldots,m,\ i = 1,2,\ldots,n,\ q \in \mathbb{Z}_+$, let

$$\mathcal{A}^0_j = \{\boldsymbol{0}\},\ \mathcal{A}^1_j = \{\boldsymbol{0}\} \bigcup \left\{\boldsymbol{e}^k : k \in I_+(\mathcal{F}_j)\right\},$$
$$\mathcal{A}^{q+1}_j = \left\{\boldsymbol{a} + \boldsymbol{b} : \boldsymbol{a} \in \mathcal{A}^q_j,\ \boldsymbol{b} \in \mathcal{A}^1_j\right\} \ (q \geq 1),$$
$$\mathcal{B}^q_i = \{k\boldsymbol{e}^i : k = 0,1,2,\ldots,q+1\} \ (q \in \mathbb{Z}_+).$$

Here $I_+(\mathcal{F}_j) = \{i : a_i > 0 \ \text{ for some } \boldsymbol{a} \in \mathcal{F}_j\}$.

Example 3.5. For every $j = 1, 2, \ldots, m,\ i = 1, 2, \ldots, n,\ q \in \mathbb{Z}_+$, let

$$\mathcal{A}_j^0 = \mathcal{B}_i^0 = \{\mathbf{0}\},\ \mathcal{A}_j^1 = \mathcal{B}_i^1 = \{\mathbf{0}\} \bigcup \left\{\boldsymbol{e}^k : k = 1, 2, \ldots, n\right\},$$
$$\mathcal{A}_j^{q+1} = \mathcal{B}_i^{q+1} = \left\{\boldsymbol{a} + \boldsymbol{b} : \boldsymbol{a} \in \mathcal{A}_j^q,\ \boldsymbol{b} \in \mathcal{A}_j^1\right\}\ (q \geq 1).$$

In all the examples, both $\mathcal{A}_j^q$ and $\mathcal{B}_i^q$ expand monotonically as q increases, and for any $\bar{p} \in \mathbb{Z}_+$ there exists $\bar{q} \in \mathbb{Z}_+$ such that $\varphi_j^p(\boldsymbol{x}) \in \Sigma(\mathcal{A}_j^q)$ and $\psi_i^p(\boldsymbol{x}) \in \Sigma(\mathcal{B}_i^q)$ for all $p \leq \bar{p}$ and $q \geq \bar{q}$. It should be noted that if $\mathcal{F}_j$ is sparse then $\mathcal{A}_j^q$ remains sparse in Example 3.3. The choice of $\mathcal{A}_j^q$ in Example 3.4 may be also reasonable when the number of the indices $I_+(\mathcal{F}_j)$ is smaller than n. When we take $\mathcal{A}_j^q$ and $\mathcal{B}_i^q$ as in Example 3.5, we have

$$\begin{aligned} \Sigma(\mathcal{A}_j^q) &= \Sigma(\mathcal{B}_i^q) \\ &= \left\{ \sum_{i=1}^k \chi_i(\boldsymbol{x})^2 : \begin{array}{l} \chi_i \text{ is a polynomial in } \boldsymbol{x} \in \mathbb{R}^n \text{ with degree } \leq q \\ (i = 1, 2, \ldots, k) \text{ and } k \text{ is any finite positive integer} \end{array} \right\}; \end{aligned}$$

Hence the sparsity of the original POP (1) is destroyed in the Lagrangian duals (6) and (7); the Lagrangian duals involve at least all the monomials with degree $\leq q$, many of which are not contained in the original POP (1) when the POP is sparse. We remark here that the choice of $\mathcal{A}_j^q$ as in Example 3.5 for the Lagrangian dual (7) leads to the dual of Lasserre's SDP relaxation [10] applied to the POP (3). This is discussed briefly in Section 4.5.

4 Numerical methods for generalized Lagrangian duals

In the following five subsections, we discuss numerical methods for generalized Lagrangian duals. The first three subsections include how $L_C^*(\boldsymbol{\Sigma}^\infty \times \boldsymbol{\Xi}^\infty)$ can be approximated numerically, and the fourth section is for the approximation of $L_B^*(\boldsymbol{\Sigma}^\infty)$. We briefly mention the relationship between the proposed methods and Lasserre's SDP relaxation [10] applied to the POP (2) in the last subsection. Throughout this section, we assume that $F \subset C_\rho$, so that $L_C^*(\boldsymbol{\Sigma}^\infty \times \boldsymbol{\Xi}^\infty) = \zeta^*$.

4.1 Approximation of generalized Lagrangian duals

We introduce a sequence of subproblems of the Lagrangian dual (6).

$$\text{maximize} \quad L_C^*(\boldsymbol{\varphi}, \boldsymbol{\psi}) \quad \text{subject to} \quad (\boldsymbol{\varphi}, \boldsymbol{\psi}) \in \boldsymbol{\Sigma}^q \times \boldsymbol{\Xi}^q \tag{13}$$

$(q \in \mathbb{Z}_+)$.

Lemma 4.1.

(a) $L_C^*(\boldsymbol{\Sigma}^q \times \boldsymbol{\Xi}^q) \leq L_C^*(\boldsymbol{\Sigma}^\infty \times \boldsymbol{\Xi}^\infty)$ $(q \in \mathbb{Z}_+)$.

(b) For any $\epsilon > 0$, there exists a nonnegative integer $\bar{q}$ such that

$$L_C^*(\boldsymbol{\Sigma}^\infty \times \boldsymbol{\Xi}^\infty) - \epsilon \leq L_C^*(\boldsymbol{\Sigma}^q \times \boldsymbol{\Xi}^q)\ (q \geq \bar{q}).$$

Proof: By construction, we know that $\boldsymbol{\Sigma}^q \times \boldsymbol{\Xi}^q \subset \boldsymbol{\Sigma}^\infty \times \boldsymbol{\Xi}^\infty$ ($q \in \mathbb{Z}_+$). Thus, (a) follows. Let $\epsilon > 0$. By (d) of Lemma 3.2, there exists $\bar{p}$ such that

$$L_C^*(\boldsymbol{\Sigma}^\infty \times \boldsymbol{\Xi}^\infty) - \epsilon \leq L_C^*(\boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) \ (p \geq \bar{p}).$$

Take $\bar{q} \in \mathbb{Z}_+$ such that $\lambda(q) \geq \bar{p}$ ($q \geq \bar{q}$). Then

$$(\boldsymbol{\varphi}^{\lambda(q)}, \boldsymbol{\psi}^{\lambda(q)}) \in \boldsymbol{\Sigma}^q \times \boldsymbol{\Xi}^q \ \text{ and } \lambda(q) \geq \bar{p} \text{ for every } q \geq \bar{q}.$$

Therefore we obtain that

$$L_C^*(\boldsymbol{\Sigma}^\infty \times \boldsymbol{\Xi}^\infty) - \epsilon \leq L_C^*(\boldsymbol{\varphi}^{\lambda(q)}, \boldsymbol{\psi}^{\lambda(q)}) \leq L_C^*(\boldsymbol{\Sigma}^q \times \boldsymbol{\Xi}^q) \text{ for every } q \geq \bar{q}.$$

∎

4.2 Sums of square relaxations

Let $q \in \mathbb{Z}_+$ be fixed throughout this subsection. We can rewrite the problems in (13) as

$$\left.\begin{array}{ll} \text{maximize} & \zeta \\ \text{subject to} & L_C(\boldsymbol{x}, \boldsymbol{\varphi}, \boldsymbol{\psi}) - \zeta \geq 0 \ (\forall \boldsymbol{x} \in \mathbb{R}^n), \\ & (\boldsymbol{\varphi}, \boldsymbol{\psi}) \in \boldsymbol{\Sigma}^q \times \boldsymbol{\Xi}^q. \end{array}\right\} \tag{14}$$

Note that $\boldsymbol{x} \in \mathbb{R}^n$ is not a vector variable but it serves as an index vector for infinite number of inequality constrains $L_C(\boldsymbol{x}, \boldsymbol{\varphi}, \boldsymbol{\psi}) - \zeta \geq 0$ ($\forall \boldsymbol{x} \in \mathbb{R}^n$). Replacing the inequality constraints $L_C(\boldsymbol{x}, \boldsymbol{\varphi}, \boldsymbol{\psi}) - \zeta \geq 0$ ($\forall \boldsymbol{x} \in \mathbb{R}^n$) by a sum of squares condition $L_C(\boldsymbol{x}, \boldsymbol{\varphi}, \boldsymbol{\psi}) - \zeta \in \overline{\Sigma}$ in (14), we obtain an SOSOP (sums of squares optimization problem).

$$\left.\begin{array}{ll} \text{maximize} & \zeta \\ \text{subject to} & L_C(\boldsymbol{x}, \boldsymbol{\varphi}, \boldsymbol{\psi}) - \zeta = \varphi_0(\boldsymbol{x}) \ (\forall \boldsymbol{x} \in \mathbb{R}^n), \\ & (\boldsymbol{\varphi}, \boldsymbol{\psi}) \in \boldsymbol{\Sigma}^q \times \boldsymbol{\Xi}^q, \ \varphi_0(\boldsymbol{x}) \in \overline{\Sigma}. \end{array}\right\} \tag{15}$$

Let ζ_C^q denote the optimal value of the SOSOP (15);

$$\zeta_C^q = \sup \left\{ \zeta : \begin{array}{l} L_C(\boldsymbol{x}, \boldsymbol{\varphi}, \boldsymbol{\psi}) - \zeta = \varphi_0(\boldsymbol{x}) \ (\forall \boldsymbol{x} \in \mathbb{R}^n), \\ (\boldsymbol{\varphi}, \boldsymbol{\psi}) \in \boldsymbol{\Sigma}^q \times \boldsymbol{\Xi}^q, \ \varphi_0(\boldsymbol{x}) \in \overline{\Sigma} \end{array} \right\}.$$

If $(\zeta, \boldsymbol{\varphi}, \boldsymbol{\psi}, \varphi_0)$ is a feasible solution of the SOSOP (15), then $(\zeta, \boldsymbol{\varphi}, \boldsymbol{\psi})$ is a feasible solution of the problem (14). It follows that $\zeta_C^q \leq L_C^*(\boldsymbol{\Sigma}^q \times \boldsymbol{\Xi}^q)$. Although neither $\zeta_C^q = L_C^*(\boldsymbol{\Sigma}^q \times \boldsymbol{\Xi}^q)$ nor the convergence of ζ_C^q to $L_C^*(\boldsymbol{\Sigma}^\infty \times \boldsymbol{\Xi}^\infty)$ as $q \to \infty$ is guaranteed, we can solve the SOSOP (15) as we show in the next subsection while the problem (14) is difficult to solve in general.

4.3 Reduction to SDPs

Let us fix $q \in \mathbb{Z}_+$ throughout this subsection. We show how to solve the SOSOP (15) as an SDP (semidefinite program). If we rewrite the constraint $(\boldsymbol{\varphi}, \boldsymbol{\psi}) \in \boldsymbol{\Sigma}^q \times \boldsymbol{\Xi}^q$ for each component, we have

$$\varphi_j(\boldsymbol{x}) \in \Sigma(\mathcal{A}_j^q) \ (j = 1, 2, \ldots, m) \ \text{ and } \psi_i(\boldsymbol{x}) \in \Sigma(\mathcal{B}_i^q) \ (i = 1, 2, \ldots, n).$$

Notice that finite supports $\mathcal{A}_j^q$ and $\mathcal{B}_i^q$ are given for generating variable polynomials $\varphi_j(\boldsymbol{x})$ and $\psi_i(\boldsymbol{x})$ $(j = 1, 2, \ldots, m,\ i = 1, 2, \ldots, n)$. But, no finite support is specified for the variable polynomial $\varphi_0(\boldsymbol{x})$. The first step for constructing an SDP is to find an appropriate finite set $\mathcal{G} \subset \mathbb{Z}_+^n$ so that $\varphi_0(\boldsymbol{x})$ can be chosen from $\Sigma(\mathcal{G})$.

To choose such a $\mathcal{G} \subset \mathbb{Z}_+^n$, we focus on the support of the left hand side polynomial $L_C(\boldsymbol{x}, \boldsymbol{\varphi}, \boldsymbol{\psi}) - \zeta$ of the equality constraint in the SOSOP (15). From the support $\mathcal{F}_0$ of the objective polynomial function $f_0(\boldsymbol{x})$, the support of each term $\varphi_j(\boldsymbol{x}) f_j(\boldsymbol{x})$

$$\widehat{\mathcal{A}}_j^q \equiv \left\{ \boldsymbol{a} + \boldsymbol{b} + \boldsymbol{c} : \boldsymbol{a} \in \mathcal{A}_j^q,\ \boldsymbol{b} \in \mathcal{A}_j^q,\ \boldsymbol{c} \in \mathcal{F}_j \right\}$$

$(j = 1, 2, \ldots, m)$ and the support of term $\psi_i(\boldsymbol{x})(\rho^2 - x_i^2)$

$$\widehat{\mathcal{B}}_i^q \equiv \left\{ \boldsymbol{a} + \boldsymbol{b} + \boldsymbol{c} : \boldsymbol{a} \in \mathcal{B}_i^q,\ \boldsymbol{b} \in \mathcal{B}_i^q,\ \boldsymbol{c} \in \{\boldsymbol{0}, 2\boldsymbol{e}^i\} \right\}$$

$(i = 1, 2, \ldots, n)$, we know that the support of $L_C(\boldsymbol{x}, \boldsymbol{\varphi}, \boldsymbol{\psi}) - \zeta$ becomes

$$\mathcal{F}_L = \mathcal{F}_0 \bigcup \{\boldsymbol{0}\} \bigcup \left(\bigcup_{j=1}^{m} \widehat{\mathcal{A}}_j^q \right) \bigcup \left(\bigcup_{i=1}^{n} \widehat{\mathcal{B}}_i^q \right).$$

Here $\{\boldsymbol{0}\}$ stands for the support of the term ζ.

By Theorem 1 of [17], we can use

$$\mathcal{G}^0 \ = \ \left(\text{the convex hull of } \left\{ \boldsymbol{a}/2 : \begin{array}{l} \boldsymbol{a} \in \mathcal{F}_L, \\ \text{every } a_i \ \text{ is even } (i = 1, 2, \ldots, n) \end{array} \right\} \right) \bigcap \mathbb{Z}_+^n.$$

for such a support $\mathcal{G}$ that $\varphi_0(\boldsymbol{x})$ can be chosen from $\Sigma(\mathcal{G})$. We can further apply a method proposed recently by the authors [7] for reducing the size of $\mathcal{G}^0$ to obtain the smallest support $\mathcal{G}^*$ in a class of supports including $\mathcal{G}^0$. See the paper [7] for more details.

Remark 4.2. Even when all $\mathcal{F}_j$ $(j = 0, 1, 2, \ldots, m)$ are sparse, $\mathcal{G}^*$ becomes dense. This is because the method proposed in [7] does not eliminate any integer point in the simplex with vertices $\bar{k}\boldsymbol{e}^i$ $(i = 1, 2, \ldots, n)$ and $\boldsymbol{0}$ contained in

$$\text{the convex hull of } \left\{ \boldsymbol{a}/2 : \ \boldsymbol{a} \in \mathcal{F}_L,\ \text{every } a_i \ \text{ is even } (i = 1, 2, \ldots, n) \ \right\},$$

which has induced $\mathcal{G}^0$ above. Here $\bar{k} = \min_i \max\{k : k\boldsymbol{e}^i \in \mathcal{B}_i^q\}$. Hence the support $\mathcal{G}^*$ does not benefit much from the sparsity of $\mathcal{F}_j$ $(j = 0, 1, 2, \ldots, m)$.

To transform the SOSOP (15) into an SDP, we need some notation and symbols. Let $\mathcal{F} \in \mathbb{Z}_+^n$ be a nonempty finite set. Let $|\mathcal{F}|$ denote the cardinality of $\mathcal{F}$ and $\mathbb{R}(\mathcal{F})$ the $|\mathcal{F}|$-dimensional Euclidean space whose coordinates are indexed by $\boldsymbol{a} \in \mathcal{F}$. Although the order of the coordinates is not relevant in the succeeding discussions, we may assume that the coordinates are arranged according to the lexicographical order. Each element of $\mathbb{R}(\mathcal{F})$ is denoted as $\boldsymbol{v} = (v_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{F})$. We use the symbol $\mathcal{S}(\mathcal{F})_+$ for the set of $|\mathcal{F}| \times |\mathcal{F}|$ symmetric positive semidefinite matrices with coordinates $\boldsymbol{a} \in \mathcal{F}$; each $\boldsymbol{V} \in \mathcal{S}(\mathcal{F})_+$ has elements $V_{\boldsymbol{ab}}$ $(\boldsymbol{a} \in \mathcal{F},\ \boldsymbol{b} \in \mathcal{F})$ such that $V_{\boldsymbol{ab}} = V_{\boldsymbol{ba}}$ and that $\boldsymbol{w}^T \boldsymbol{V} \boldsymbol{w} = \sum_{\boldsymbol{a} \in \mathcal{F}} \sum_{\boldsymbol{b} \in \mathcal{F}} V_{\boldsymbol{ab}} w_{\boldsymbol{a}} w_{\boldsymbol{b}} \geq 0$ for every $\boldsymbol{w} = (w_{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{F}) \in \mathbb{R}(\mathcal{F})$. For every $\boldsymbol{x} \in \mathbb{R}^n$, let $\boldsymbol{u}(\boldsymbol{x}, \mathcal{F}) = (\boldsymbol{x}^{\boldsymbol{a}} : \boldsymbol{a} \in \mathcal{F})$ be a column vector consisting of elements $\boldsymbol{x}^{\boldsymbol{a}}$ $(\boldsymbol{a} \in \mathcal{F})$.

The lemma below is well-known ([2, 7, 13, 15]).

Lemma 4.3. *Let $\mathcal{F}$ be a nonempty finite subset of $\mathbb{Z}^n_+$. A polynomial $\varphi(\boldsymbol{x})$ is contained in $\Sigma(\mathcal{F})$ if and only if there exists a $\boldsymbol{V} \in \mathcal{S}(\mathcal{F})_+$ such that*

$$\varphi(\boldsymbol{x}) = \boldsymbol{u}(\boldsymbol{x}, \mathcal{F})^T \boldsymbol{V} \boldsymbol{u}(\boldsymbol{x}, \mathcal{F}) = \sum_{\boldsymbol{a} \in \mathcal{F}} \sum_{\boldsymbol{b} \in \mathcal{F}} V_{\boldsymbol{ab}} \boldsymbol{x}^{\boldsymbol{a}+\boldsymbol{b}}. \tag{16}$$

Applying Lemma 4.3 to the polynomials $\varphi_j(\boldsymbol{x}) \in \Sigma(\mathcal{A}_j^{(\lambda_j)})$ $(j = 1, 2, \ldots, m)$, $\psi_i(\boldsymbol{x}) \in \Sigma(\mathcal{B}_i^{(\mu_i)})$ $(i = 1, 2, \ldots, n)$ and $\varphi_0(\boldsymbol{x}) \in \Sigma(\mathcal{G}^*)$, we represent as follows:

$$\begin{array}{rcl} \varphi_j(\boldsymbol{x}) & = & \boldsymbol{u}(\boldsymbol{x}, \mathcal{A}_j^q)^T \boldsymbol{V}^j \boldsymbol{u}(\boldsymbol{x}, \mathcal{A}_j^q), \ \boldsymbol{V}^j \in \mathcal{S}(\mathcal{A}_j^q)_+, \\ \psi_i(\boldsymbol{x}) & = & \boldsymbol{u}(\boldsymbol{x}, \mathcal{B}_i^q)^T \boldsymbol{V}^{m+i} \boldsymbol{u}(\boldsymbol{x}, \mathcal{B}_i^q), \ \boldsymbol{V}^{m+i} \in \mathcal{S}(\mathcal{B}_i^q)_+, \\ \varphi_0(\boldsymbol{x}) & = & \boldsymbol{u}(\boldsymbol{x}, \mathcal{G}^*)^T \boldsymbol{V}^0 \boldsymbol{u}(\boldsymbol{x}, \mathcal{G}^*), \ \boldsymbol{V}^0 \in \mathcal{S}(\mathcal{G}^*)_+. \end{array}$$

Substituting these functions in the SOSOP (15) leads to

$$\left.\begin{array}{ll} \text{maximize} & \zeta \\ \text{subject to} & f_0(\boldsymbol{x}) - \sum_{j=1}^m f_j(\boldsymbol{x}) \boldsymbol{u}(\boldsymbol{x}, \mathcal{A}_j^q)^T \boldsymbol{V}^j \boldsymbol{u}(\boldsymbol{x}, \mathcal{A}_j^q) \\ & - \sum_{i=1}^n (\rho - x_i^2) \boldsymbol{u}(\boldsymbol{x}, \mathcal{B}_i^q)^T \boldsymbol{V}^{m+i} \boldsymbol{u}(\boldsymbol{x}, \mathcal{B}_i^q) \\ & - \boldsymbol{u}(\boldsymbol{x}, \mathcal{G}^*)^T \boldsymbol{V}^0 \boldsymbol{u}(\boldsymbol{x}, \mathcal{G}^*) - \zeta = 0 \ (\forall \boldsymbol{x} \in \mathbb{R}^n), \\ & \boldsymbol{V}^j \in \mathcal{S}(\mathcal{A}_j^q)_+ \ (j = 1, 2, \ldots, m), \\ & \boldsymbol{V}^{m+i} \in \mathcal{S}(\mathcal{B}_i^q)_+ \ (i = 1, 2, \ldots, n), \\ & \boldsymbol{V}^0 \in \mathcal{S}(\mathcal{G}^*)_+. \end{array}\right\}$$

Since the left hand side of the equality constraint in the problem above is a polynomial with the support

$$\mathcal{F}_C = \mathcal{F}_0 \bigcup \{\boldsymbol{0}\} \bigcup \left(\bigcup_{j=1}^m \widehat{\mathcal{A}}_j^q \right) \bigcup \left(\bigcup_{i=1}^n \widehat{\mathcal{B}}_i^q \right) \bigcup \left\{ \boldsymbol{a} + \boldsymbol{b} : \boldsymbol{a} \in \mathcal{G}^*, \ \boldsymbol{b} \in \mathcal{G}^* \right\},$$

and the coefficients are linear functions of matrix variable

$$\boldsymbol{V}^j \ (j = 1, 2, \ldots, m), \ \boldsymbol{V}^{m+i} \ (i = 1, 2, \ldots, n), \ \boldsymbol{V}^0$$

and ζ, we can rewrite equality constraint of the problem above as

$$\sum_{\boldsymbol{a} \in \mathcal{F}_C} d(\boldsymbol{a}, \boldsymbol{V}, \zeta) \boldsymbol{x}^{\boldsymbol{a}} = 0,$$

where $d(\boldsymbol{a}, \boldsymbol{V}, \zeta)$ is a linear function in the matrix variables $\boldsymbol{V}^j$ $(j = 0, 1, 2, \ldots, m+n)$ and a real variable ζ for each $\boldsymbol{a} \in \mathcal{F}_C$. This identity needs to be satisfied for all $\boldsymbol{x} \in \mathbb{R}^n$ in the problem, and the equality constraint is equivalent to a system of linear equations

$$d(\boldsymbol{a}, \boldsymbol{V}, \zeta) = 0 \ (\boldsymbol{a} \in \mathcal{F}_C).$$

Consequently, we obtain the following SDP which is equivalent to the SOSOP (15).

$$\left.\begin{array}{ll} \text{maximize} & \zeta \\ \text{subject to} & d(\boldsymbol{a}, \boldsymbol{V}, \zeta) = 0 \ (\boldsymbol{a} \in \mathcal{F}_C), \\ & \boldsymbol{V}^j \in \mathcal{S}(\mathcal{A}_j^q)_+ \ (j = 1, 2, \ldots, m), \\ & \boldsymbol{V}^{m+i} \in \mathcal{S}(\mathcal{B}_i^q)_+ \ (i = 1, 2, \ldots, n), \\ & \boldsymbol{V}^0 \in \mathcal{S}(\mathcal{G}^*)_+. \end{array}\right\} \tag{17}$$

The numerical efficiency of solving an SDP depends largely on its size. In the SDP (17) above, the number of equality constraint and the sizes of matrix variables are determined by the supports $\mathcal{F}_j$ of the polynomial functions $f_j(\boldsymbol{x})$ $(j = 0, 1, \ldots, m)$ in the original POP (1) to be solved and $q \in \mathbb{Z}_+$. When the supports are sparse, the size of the resulting SDP becomes small. As we take a larger $q \in \mathbb{Z}_+$, we can expect to have a more accurate lower bound for the unknown optimal value ζ^* of the POP (1), but the number of equality constraint and the size of the matrix variables increase.

4.4 Sums of squares relaxation of $L^*_B(\boldsymbol{\Sigma}^\infty)$

In this section, we derive a sequence of SOSOPs from the Lagrangian dual (7). The SOSOPs obtained here are less sparse than (15) in the previous section, but their optimal values converge to $L^*_B(\boldsymbol{\Sigma}^\infty)$. Thus, this compensates the theoretical weakness of the SOSOP (15) whose optimal value is not guaranteed to converge to $L^*_C(\boldsymbol{\Sigma}^\infty \times \boldsymbol{\Xi}^\infty)$. Throughout this section, we assume that $F \subset B_\rho$; hence $L^*_C(\boldsymbol{\Sigma}^\infty \times \boldsymbol{\Xi}^\infty) = L^*_B(\boldsymbol{\Sigma}^\infty) = \zeta^*$ is satisfied by Theorem 3.1.

As we have obtained the sequence of SOSOPs (15) from the Lagrangian dual (6), we can similarly derive the following sequence of SOSOPs from the Lagrangian dual (7):

$$\left.\begin{array}{ll} \text{maximize} & \zeta \\ \text{subject to} & L_B(\boldsymbol{x}, \boldsymbol{\varphi}) - \zeta - \varphi_{m+1}(\boldsymbol{x})(\rho^2 - \boldsymbol{x}^T\boldsymbol{x}) = \varphi_0(\boldsymbol{x}) \ (\forall \boldsymbol{x} \in \mathbb{R}^n), \\ & \boldsymbol{\varphi} \in \boldsymbol{\Sigma}^q, \ \varphi_{m+1} \in \Sigma(\mathcal{G}(\tau(q) - 1)), \ \varphi_0 \in \overline{\Sigma}. \end{array}\right\} \tag{18}$$

$(q \in \mathbb{Z}_+)$. Here

$$\begin{aligned} \tau(q) &= \lceil (\text{the degree of } L_B(\boldsymbol{x}, \boldsymbol{\varphi}) \text{ with } \boldsymbol{\varphi} \in \boldsymbol{\Sigma}^q)/2 \rceil, \\ \mathcal{G}(\xi) &= \left\{ \boldsymbol{a} \in \mathbb{Z}^n_+ : \sum_{i=1}^n \leq \xi \right\} \ (\xi = 0, 1, 2, \ldots). \end{aligned}$$

Let ζ^q_B denote the optimal value of the SOSOP (18);

$$\zeta^q_B = \sup \left\{ \zeta : \begin{array}{l} L_B(\boldsymbol{x}, \boldsymbol{\varphi}) - \zeta - \varphi_{m+1}(\boldsymbol{x})(\rho^2 - \boldsymbol{x}^T\boldsymbol{x}) = \varphi_0(\boldsymbol{x}), \\ \boldsymbol{\varphi} \in \boldsymbol{\Sigma}^q, \ \varphi_{m+1} \in \Sigma(\mathcal{G}(\tau(q) - 1)), \ \varphi_0 \in \overline{\Sigma} \end{array} \right\}.$$

Theorem 4.4.

(a) $\zeta^q_B \leq L^*_B(\boldsymbol{\Sigma}^\infty)$ $(q \in \mathbb{Z}_+)$.

(b) In addition to (12), assume that $\mathcal{A}^q_j \subset \mathcal{A}^{q+1}_j$ $(j = 1, 2, \ldots, m, \ q \in \mathbb{Z}_+)$. *For any* $\epsilon > 0$, *there exists a* $\hat{q} \in \mathbb{Z}_+$ *such that* $L^*_B(\boldsymbol{\Sigma}^\infty) - \epsilon \leq \zeta^q_B$ $(q \geq \hat{q})$.

Proof: (a) Let $q \in \mathbb{Z}_+$. Suppose that $(\zeta, \boldsymbol{\varphi}, \varphi_{m+1}, \varphi_0) = (\bar{\zeta}, \bar{\boldsymbol{\varphi}}, \bar{\varphi}_{m+1}, \bar{\varphi}_0)$ is a feasible solution of (18). Then $0 \leq L_B(\boldsymbol{x}, \boldsymbol{\varphi}) - \bar{\zeta}$ for every $\boldsymbol{x} \in B_\rho$; hence $\bar{\zeta} \leq L^*_B(\boldsymbol{\varphi}) \leq L^*_B(\boldsymbol{\Sigma}^\infty)$. This ensures that $\zeta^q_B \leq L^*_B(\boldsymbol{\Sigma}^\infty)$.

(b) It follows from the assumption $\mathcal{A}^q_j \subset \mathcal{A}^{q+1}_j$ $(j = 1, 2, \ldots, m, \ q \in \mathbb{Z}_+)$ that $\boldsymbol{\Sigma}^q \subset \boldsymbol{\Sigma}^{q+1}$ $(q \in \mathbb{Z})$. Let $\epsilon > 0$. By the construction of $L^*_B(\boldsymbol{\Sigma}^\infty)$, there exist $\widehat{q} \in \mathbb{Z}_+$ and $\widehat{\boldsymbol{\varphi}} \in \boldsymbol{\Sigma}^{\widehat{q}}$ such that

$$L^*_B(\boldsymbol{\Sigma}^\infty) - \epsilon/2 \leq L^*_B(\widehat{\boldsymbol{\varphi}}).$$

Then

$$L_B(\boldsymbol{x}, \widehat{\boldsymbol{\varphi}}) - L_B^*(\boldsymbol{\Sigma}^\infty) + \epsilon$$

is a polynomial that is positive in the ball B_ρ. By Lemma 4.1 of [16],

$$L_B(\boldsymbol{x}, \widehat{\boldsymbol{\varphi}}) - L_B^*(\boldsymbol{\Sigma}^\infty) + \epsilon - \widehat{\varphi}_{m+1}(\boldsymbol{x})(\rho^2 - \boldsymbol{x}^T\boldsymbol{x}) \in \overline{\Sigma}$$

for some $\widehat{\omega} \geq \tau(\widehat{q})$ and some $\widehat{\varphi}_{m+1} \in \Sigma(\mathcal{G}(\widehat{\omega} - 1))$. Choose a $\widetilde{q} \in \mathbb{Z}_+$ such that $\widehat{q} \leq \widetilde{q}$ and $\tau(\widetilde{q}) \geq \widehat{\omega}$. Let $\widetilde{\omega} = \tau(\widetilde{q})$. Then

$$\begin{aligned}
\zeta_B^{\widetilde{q}} &= \sup\left\{\zeta : \begin{array}{l} L_B(\boldsymbol{x}, \boldsymbol{\varphi}) - \zeta - \varphi_{m+1}(\boldsymbol{x})(\rho^2 - \boldsymbol{x}^T\boldsymbol{x}) = \varphi_0(\boldsymbol{x}), \\ \boldsymbol{\varphi} \in \boldsymbol{\Sigma}^{\widetilde{q}},\ \varphi_{m+1} \in \Sigma(\mathcal{G}(\tau(\widetilde{q}) - 1)),\ \varphi_0 \in \overline{\Sigma} \end{array}\right\} \\
&\geq \sup\left\{\zeta :\ L_B(\boldsymbol{x}, \widehat{\boldsymbol{\varphi}}) - \zeta - \widehat{\varphi}_{m+1}(\boldsymbol{x})(\rho^2 - \boldsymbol{x}^T\boldsymbol{x}) = \varphi_0(\boldsymbol{x}),\ \varphi_0 \in \overline{\Sigma}\ \right\} \\
&\quad \text{(since } \widehat{\boldsymbol{\varphi}} \in \boldsymbol{\Sigma}^{\widehat{q}} \subset \boldsymbol{\Sigma}^{\widetilde{q}} \text{ and } \widehat{\varphi}_{m+1} \in \Sigma(\mathcal{G}(\tau(\widehat{q}) - 1)) \subset \Sigma(\mathcal{G}(\tau(\widetilde{q}) - 1))) \\
&\geq L_B^*(\boldsymbol{\Sigma}^\infty) - \epsilon.
\end{aligned}$$

∎

Comparing the SOSOP (15) with the SOSOP (18) shows that the latter problem involves the term $\varphi_{m+1}(\boldsymbol{x})(\rho^2 - \boldsymbol{x}^T\boldsymbol{x})$ with $\varphi_{m+1} \in \Sigma(\mathcal{G}(\tau(q) - 1))$ which is a fully dense polynomial with degree $2\tau(q)$; hence we need to prepare a fully dense polynomial variable $\varphi_0 \in \Sigma(\mathcal{G}(\tau(q)))$ in general. If we convert the SOSOP (18) into an SDP as converted the SOSOP (15) into the SDP (17), the polynomial variables $\varphi_{m+1} \in \Sigma(\mathcal{G}(\tau(q) - 1))$ and $\varphi_0 \in \Sigma(\mathcal{G}(\tau(q)))$ induce matrix variables

$$\boldsymbol{V}^{m+1} \in \mathcal{S}(\mathcal{G}(\tau(q) - 1))_+ \text{ and } \boldsymbol{V}^0 \in \mathcal{S}(\mathcal{G}(\tau(q)))_+,$$

respectively. Let $\mathcal{F}_B$ denote the support of the polynomial

$$\begin{array}{l} L_B(\boldsymbol{x}, \boldsymbol{\varphi}) - \zeta - \varphi_{m+1}(\boldsymbol{x})(\rho^2 - \boldsymbol{x}^T\boldsymbol{x}) - \varphi_0(\boldsymbol{x}) \ \text{ with} \\ \boldsymbol{\varphi} \in \overline{\Sigma}^q,\ \varphi_{m+1} \in \Sigma(\mathcal{G}(\tau(q) - 1)) \ \text{ and } \varphi_0 \in \Sigma(\mathcal{G}(\tau(q))). \end{array}$$

Then the resulting SDP formulation of the SOSOP (18) turns out to be

$$\left.\begin{array}{ll} \text{maximize} & \zeta \\ \text{subject to} & d'(\boldsymbol{a}, \boldsymbol{V}, \zeta) = 0 \ (\boldsymbol{a} \in \mathcal{F}_B), \\ & \boldsymbol{V}^j \in \mathcal{S}(\mathcal{A}_j^q)_+ \ (j = 1, 2, \ldots, m), \\ & \boldsymbol{V}^{m+1} \in \mathcal{S}(\mathcal{G}(\tau(q) - 1))_+, \\ & \boldsymbol{V}^0 \in \mathcal{S}(\mathcal{G}(\tau(q)))_+. \end{array}\right\} \quad (19)$$

Here $d'(\boldsymbol{a}, \boldsymbol{V}, \zeta)$ denotes a linear function in the matrix variables $\boldsymbol{V}^j$ $(j = 0, 1, 2, \ldots, m+1)$ and a real variable ζ for each $\boldsymbol{a} \in \mathcal{F}_B$. If the polynomials $f_j(\boldsymbol{x})$ $(j = 0, 1, 2, \ldots, m)$ are sparse, the size of the SDP (19) is larger than the size of the SDP (17); the number $|\mathcal{F}_B|$ of linear equations in the SDP (19) is larger than the number $|\mathcal{F}_C|$ of linear equations in the SDP (17), and the sizes of two matrix variables $\boldsymbol{V}^{m+1}$ and $\boldsymbol{V}^0$ in the SDP (19) are larger than the sizes of the matrix variables in the SDP (17).

4.5 Comparison to Lasserre's SDP relaxation

In this subsection, we briefly mention a modification of Lasserre's SDP relaxation [10] that takes account of the supports $\mathcal{F}_j$ of the polynomials $f_j(\boldsymbol{x})$ $(j = 0, 1, \ldots, m)$ in the POP (3). The modification leads to the dual SDP of (19).

First, we describe the original Lasserre's SDP relaxation [10] applied to the POP (3) according the interpretation given in the paper [6] on the relaxation. Let $\lceil r/2 \rceil \leq N \in \mathbb{Z}_+$ and $\nu_j = N - \lceil r_j/2 \rceil$ $(j = 1, 2, \ldots, m)$. Consider the following polynomial optimization problem on positive semidefinite cones which is equivalent to the POP (1).

$$\left.\begin{array}{ll} \text{minimize} & f_0(\boldsymbol{x}) \\ \text{subject to} & f_j(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x}, \mathcal{G}(\nu_j))\boldsymbol{u}(\boldsymbol{x}, \mathcal{G}(\nu_j))^T \in \mathcal{S}(\mathcal{G}(\nu_j))_+ \ (j = 1, 2, \ldots, m) \\ & (\rho^2 - \boldsymbol{x}^T\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x}, \mathcal{G}(N-1))\boldsymbol{u}(\boldsymbol{x}, \mathcal{G}(N-1))^T \in \mathcal{S}(\mathcal{G}(N-1))_+, \\ & \boldsymbol{u}(\boldsymbol{x}, \mathcal{G}(N))\boldsymbol{u}(\boldsymbol{x}, \mathcal{G}(N))^T \in \mathcal{S}(\mathcal{G}(N))_+. \end{array}\right\} \tag{20}$$

We then apply a linearization to this problem to obtain an SDP relaxation of the POP (1). See the paper [6]

When the modification for the primal approach corresponding to (15) is implemented, the POP (3) is converted into a polynomial optimization problem on positive semidefinite cones:

$$\left.\begin{array}{ll} \text{minimize} & f_0(\boldsymbol{x}) \\ \text{subject to} & f_j(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x}, \mathcal{A}^q_j)\boldsymbol{u}(\boldsymbol{x}, \mathcal{A}^q_j)^T \in \mathcal{S}(\mathcal{A}^q_j)_+ \ (j = 1, 2, \ldots, m), \\ & (\rho^2 - \boldsymbol{x}^T\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x}, \mathcal{G}(\tau(q)-1))\boldsymbol{u}(\boldsymbol{x}, \mathcal{G}(\tau(q)-1))^T \in \mathcal{S}(\mathcal{G}(\tau(q)-1))_+, \\ & \boldsymbol{u}(\boldsymbol{x}, \mathcal{G}(\tau(q)))\boldsymbol{u}(\boldsymbol{x}, \mathcal{G}(\tau(q)))^T \in \mathcal{S}(\mathcal{G}(\tau(q)))_+. \end{array}\right\} \tag{21}$$

We obtain an SDP by applying the linearization to the problem (21). Then the resulting SDP is the dual of (19) derived from the Lagrangian dual (7) of the POP (3). See Section 6 of the paper [6] for more details about the linearization technique and the proof of the fact above.

Notice that N in the problem (20) corresponds to $\tau(q)$ in the problem (21). Using Lemma 4.1 of [16] which we have used to prove for the convergence of ζ^q_B to ζ^*, Lasserre proved the convergence of the optimal value of the SDP derived from the problem (20) to the optimal value ζ^* of the POP (1) as the parameter N tends to ∞ in the paper [10]. An essential difference between the original and the modified Lasserre's SDP relaxations is that the modified relaxation takes account of the supports $\mathcal{F}_j$ of the polynomials $f_j(\boldsymbol{x})$ $(j = 1, \ldots, m)$ in the POP (3). As the supports $\mathcal{F}_j$ $(j = 0, 1, 2, \ldots, m)$ are more sparse, the modified relaxation results in a smaller SDP relaxation. In other words, the SDP (17) derived from the Lagrangian dual (6) utilizes the sparsity of the POP (1) more effectively and therefore, its size is smaller than the SDP (19) derived from the Lagrangian dual (7) of the POP (3).

5 Proof of Lemma 3.2

(a) Suppose that $\sqrt{2}\rho \leq \|\boldsymbol{x}\|_\infty$. Let $p \geq p^*$. If $f_j(\boldsymbol{x}) < 0$ then

$$f_j(\boldsymbol{x})\varphi^p_j(\boldsymbol{x}) = f_j(\boldsymbol{x})\left(1 - f_j(\boldsymbol{x})/\gamma\right)^{2p} < 0.$$

Otherwise, we obtain that

$$\begin{aligned}
0 &\leq f_j(\boldsymbol{x})\varphi_j^p(\boldsymbol{x}) \\
&= f_j(\boldsymbol{x})\left(1 - f_j(\boldsymbol{x})/\gamma\right)^{2p} \\
&\leq \gamma \left\|\boldsymbol{x}/\rho\right\|_\infty^r \max\left\{1, \left(f_j(\boldsymbol{x})/\gamma\right)^{2p}\right\} \\
&\quad \text{(by (10) and } |1 - f_j(\boldsymbol{x})/\gamma| \leq \max\{1,\ |f_j(\boldsymbol{x})/\gamma|\}) \\
&\leq \gamma \left\|\boldsymbol{x}/\rho\right\|_\infty^r \left(\left\|\boldsymbol{x}/\rho\right\|_\infty^r\right)^{2p} \ \text{(by (10))} \\
&= \gamma \left\|\boldsymbol{x}/\rho\right\|_\infty^{r(2p+1)}.
\end{aligned}$$

Hence

$$\sum_{j=1}^m f_j(\boldsymbol{x})\varphi_j^p(\boldsymbol{x}) \leq m\gamma \left\|\boldsymbol{x}/\rho\right\|_\infty^{r(2p+1)}.$$

On the other hand, we see that

$$\begin{aligned}
&\sum_{i=1}^n (\rho^2 - x_i^2)\psi_i^p(\boldsymbol{x}) \\
&= \left((m+2)\gamma/\rho^2\right) \sum_{i=1}^n (\rho^2 - x_i^2)\left(x_i/\rho\right)^{2r(p+1)} \\
&= \left((m+2)\gamma/\rho^2\right) \left(\sum_{x_i^2 \leq \rho^2} (\rho^2 - x_i^2)\left(x_i/\rho\right)^{2r(p+1)} + \sum_{\rho^2 < x_i^2} (\rho^2 - x_i^2)\left(x_i/\rho\right)^{2r(p+1)} \right) \\
&\leq (m+2)\gamma \left(n + \left(1 - \left\|\boldsymbol{x}/\rho\right\|_\infty^2\right) \left\|\boldsymbol{x}/\rho\right\|_\infty^{2r(p+1)} \right) \quad (\text{since } \{i \mid \rho^2 < x_i^2\} \neq \emptyset) \\
&\leq \gamma \left((m+2)n - (m+2)\left\|\boldsymbol{x}/\rho\right\|_\infty^{2r(p+1)} \right) \quad (\text{since } \sqrt{2} \leq \|\boldsymbol{x}/\rho\|_\infty) \\
&\leq \gamma \left((m+2)n - \left\|\boldsymbol{x}/\rho\right\|_\infty^{2r(p^*+1)} - (m+1)\left\|\boldsymbol{x}/\rho\right\|_\infty^{2r(p+1)} \right) \\
&\quad (\text{since } p^* \leq p \text{ and } \sqrt{2} \leq \|\boldsymbol{x}/\rho\|_\infty) \\
&\leq -(m+1)\gamma \left\|\boldsymbol{x}/\rho\right\|_\infty^{2r(p+1)} \\
&\quad (\text{by } \sqrt{2} \leq \|\boldsymbol{x}/\rho\|_\infty \text{ and } (m+2)n - 2^{r(p^*+1)} \leq 0).
\end{aligned}$$

Therefore

$$\begin{aligned}
&L_C(\boldsymbol{x}, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) \\
&= f_0(\boldsymbol{x}) - \sum_{j=1}^m f_j(\boldsymbol{x})\varphi_j^p(\boldsymbol{x}) - \sum_{i=1}^n (\rho^2 - x_i^2)\psi_i^p(x_i) \\
&\geq f_0(\boldsymbol{x}) - m\gamma \left\|\boldsymbol{x}/\rho\right\|_\infty^{r(2p+1)} + (m+1)\gamma \left\|\boldsymbol{x}/\rho\right\|_\infty^{2r(p+1)} \\
&\geq f_0(\boldsymbol{x}) + \gamma \left\|\boldsymbol{x}/\rho\right\|_\infty^{2r(p+1)} \ (\text{since } -m\left\|\boldsymbol{x}/\rho\right\|_\infty^{r(2p+1)} + m\left\|\boldsymbol{x}/\rho\right\|_\infty^{2r(p+1)} \geq 0) \\
&\geq -\gamma \left\|\boldsymbol{x}/\rho\right\|_\infty^r + \gamma \left\|\boldsymbol{x}/\rho\right\|_\infty^{2r(p+1)} \ \text{(by (10))} \\
&= \gamma \left\|\boldsymbol{x}/\rho\right\|_\infty^{2r(p+1)} \left(1 - \left\|\boldsymbol{x}/\rho\right\|_\infty^{-r(2p+1)}\right) \\
&\geq \frac{1}{2} \left\|\boldsymbol{x}/\rho\right\|_\infty^{2r} \left\|\boldsymbol{x}/\rho\right\|_\infty^{2rp} \ (\text{since } \gamma \geq 1, r \geq 1,\ p \geq 1 \text{ and } \|\boldsymbol{x}/\rho\|_\infty \geq \sqrt{2})
\end{aligned}$$

$$\geq \ \|\boldsymbol{x}/\rho\|_\infty^{2rp} \geq \|\boldsymbol{x}/\rho\|_\infty^2 \text{ (since } r \geq 1,\ p \geq 1 \text{ and } \|\boldsymbol{x}/\rho\|_\infty \geq \sqrt{2}).$$

Thus we have shown (a).

(b) Suppose that $\tilde{\boldsymbol{x}} \notin F$ and $\kappa > 0$. If $\|\tilde{\boldsymbol{x}}\|_\infty > \sqrt{2}\rho$, take $\delta > 0$ such that $\|\boldsymbol{x}\|_\infty > \sqrt{2}\rho$ for every $\boldsymbol{x} \in U_\delta(\tilde{\boldsymbol{x}})$. Then the desired result follows from (a). Now assume that $\|\tilde{\boldsymbol{x}}\|_\infty \leq \sqrt{2}\rho$. By (10), we see that

$$|f_j(\tilde{\boldsymbol{x}})/\gamma| \leq 1 \ (j = 0, 1, 2, \dots, m).$$

Since $\tilde{\boldsymbol{x}} \notin F$, $f_k(\tilde{\boldsymbol{x}}) < 0$ also holds for some $k \in \{1, 2, \dots, m\}$. Hence we can take $\epsilon > 0$ and $\delta > 0$ such that

$$f_k(\boldsymbol{x})/\gamma < -\epsilon \ \text{ and } |f_j(\boldsymbol{x})/\gamma| \leq 2 \ (j = 0, 1, 2, \dots, m) \ \text{ for every } \boldsymbol{x} \in U_\delta(\tilde{\boldsymbol{x}}).$$

Let $\tilde{p} \geq p^*$ be a positive integer such that

$$\epsilon(1+\epsilon)^{2\tilde{p}} - 2 - 2m - (m+2)n \geq \kappa/\gamma.$$

Then, for every $\boldsymbol{x} \in U_\delta(\tilde{\boldsymbol{x}})$ and $p \geq \tilde{p}$,

$$\begin{aligned}
L_C(\boldsymbol{x}, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) &= f_0(\boldsymbol{x}) - \sum_{j=1}^m f_j(\boldsymbol{x})\varphi_j^p(\boldsymbol{x}) - \sum_{i=1}^n (\rho^2 - x_i^2)\psi_i^p(\boldsymbol{x}) \\
&\geq -2\gamma - f_k(\boldsymbol{x})\varphi_k^p(\boldsymbol{x}) - \sum_{f_j(\boldsymbol{x})\geq 0} f_j(\boldsymbol{x})\varphi_j^p(\boldsymbol{x}) - \sum_{\rho^2 - x_i^2 \geq 0} (\rho^2 - x_i^2)\psi_i^p(\boldsymbol{x}) \\
&= -2\gamma - f_k(\boldsymbol{x})\left(1 - f_k(\boldsymbol{x})/\gamma\right)^{2p} - \sum_{f_j(\boldsymbol{x})\geq 0} f_j(\boldsymbol{x})\left(1 - f_j(\boldsymbol{x})/\gamma\right)^{2p} \\
&\quad - \left((m+2)\gamma/\rho^2\right) \sum_{\rho^2 - x_i^2 \geq 0} (\rho^2 - x_i^2)\left(x_i/\rho\right)^{2r(p+1)} \\
&\geq -2\gamma + \gamma\epsilon(1+\epsilon)^{2p} - 2m\gamma - (m+2)n\gamma \\
&\geq \gamma\left(\epsilon(1+\epsilon)^{2p} - 2 - 2m - (m+2)n\right) \\
&\geq \kappa.
\end{aligned}$$

(c) Suppose that $\hat{\boldsymbol{x}} \in F$ and $\epsilon > 0$. Then

$$\|\hat{\boldsymbol{x}}\|_\infty \leq \rho \ \text{ and } 0 \leq f_j(\hat{\boldsymbol{x}})/\gamma \leq 1 \ (j = 1, 2, \dots, m).$$

We can take a $\delta > 0$ such that

$$f_j(\hat{\boldsymbol{x}})/\gamma \leq 1.1 \ (j = 1, 2, \dots, m) \text{ for every } \boldsymbol{x} \in U_\delta(\hat{\boldsymbol{x}}),$$

and $\tilde{p} \geq p^*$ such that

$$\begin{aligned}
0 &\leq \eta(1-\eta)^{2\tilde{p}} \leq \epsilon/(2m\gamma) \text{ if } 0 \leq \eta \leq 1.1, \\
0 &\leq (1-\xi)\xi^{r(\tilde{p}+1)} \leq \epsilon/((2m+4)n\gamma) \text{ if } 0 \leq \xi \leq 1.
\end{aligned}$$

Then, for every $\boldsymbol{x} \in U_\delta(\hat{\boldsymbol{x}})$ and $p \geq \tilde{p}$,

$$\begin{aligned}
\Phi_C(\boldsymbol{x}, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) &= -\sum_{j=1}^{m} f_j(\boldsymbol{x})\varphi_j^p(\boldsymbol{x}) - \sum_{i=1}^{n}(\rho^2 - x_i^2)\psi_i^p(\boldsymbol{x}) \\
&\geq -\sum_{f_j(\boldsymbol{x})\geq 0} f_j(\boldsymbol{x})(1 - f_j(\boldsymbol{x})/\gamma)^{2p} \\
&\quad -\left((m+2)\gamma/\rho^2\right) \sum_{\rho^2 - x_i^2 \geq 0} \left(\rho^2 - x_i^2\right)(x_i/\rho)^{2r(p+1)} \\
&\geq -\sum_{f_j(\boldsymbol{x})\geq 0} \gamma\epsilon/(2m\gamma) - (m+2)\gamma \sum_{\rho^2 - x_i^2 \geq 0} \epsilon/((2m+4)n\gamma) \\
&= -\epsilon \sum_{f_j(\boldsymbol{x})\geq 0} 1/(2m) - \sum_{\rho^2 - x_i^2\geq 0} \epsilon/(2n) \\
&\geq -\epsilon.
\end{aligned}$$

Thus we have shown (c).

(d) Define a compact subset A of $\mathbb{R}^n$ by

$$A = \left\{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_\infty \leq \sqrt{2}\rho\right\} \bigcup \left\{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}/\rho\|_\infty^2 \leq \zeta^*\right\}.$$

In view of the property (a), we know that

$$\{\boldsymbol{x} \in \mathbb{R}^n : L_C(\boldsymbol{x}, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) \leq \zeta^*\} \subset A \ (p \geq p^*).$$

By the property (4) with $(\boldsymbol{\varphi}, \boldsymbol{\psi}) = (\boldsymbol{\varphi}^p, \boldsymbol{\psi}^p)$, $\boldsymbol{x} = \boldsymbol{x}^* \in F$ and $\zeta^* = f_0(\boldsymbol{x}^*)$, we also see that

$$L_C(\boldsymbol{x}^*, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) \leq f_0(\boldsymbol{x}^*) = \zeta^* \ (p \geq p^*) \ \text{ and } \boldsymbol{x}^* \in F \subset A.$$

Hence

$$L_C^*(\boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) = \inf\{L_C(\boldsymbol{x}, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) : \boldsymbol{x} \in \mathbb{R}^n\} = \inf\{L_C(\boldsymbol{x}, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) : \boldsymbol{x} \in A\} \ (p \geq p^*).$$

Since A is compact, $L_C(\boldsymbol{x}, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p)$ has a minimizer over $\mathbb{R}^n$ at an $\boldsymbol{x} = \boldsymbol{x}^p \in A$;

$$L_C^*(\boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) = L_C(\boldsymbol{x}^p, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) \ \text{ and } \boldsymbol{x}^p \in A \ (p \geq p^*)$$

Now assume on the contrary that the sequence $\{L_C(\boldsymbol{x}^p, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) \ (p \geq p^*)\}$ does not converge to ζ^*. Then we can take an $\epsilon > 0$ and a subsequence $\{L_C(\boldsymbol{x}^p, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) \ (p \in J)\}$ such that

$$L_C(\boldsymbol{x}^p, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) < \zeta^* - \epsilon \ (p \in J). \tag{22}$$

Since the sequence $\{\boldsymbol{x}^p \ (p \in J)\}$ lies in the compact set A, we can take a subsequence $\{\boldsymbol{x}^p \ (p \in K)\}$ with $K \subset J$ which converges to some $\bar{\boldsymbol{x}} \in A$. Since $L_C(\boldsymbol{x}^p, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) \leq \zeta^*$ $(p \in K)$, we know by the property (b) that $\bar{\boldsymbol{x}} \in F$. By the property (c), there exists a $\tilde{p} \geq p^*$ such that

$$f_0(\boldsymbol{x}^p) - \epsilon/2 \leq L_C(\boldsymbol{x}^p, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) \leq \zeta^* \text{ if } p \in K \ \text{ and } p \geq \tilde{p}.$$

Taking the limit along the subsequence $\{\boldsymbol{x}^p \ (p \in K)\}$, the left hand side $f_0(\boldsymbol{x}^p) - \epsilon/2$ of the inequality above converges to $f_0(\bar{\boldsymbol{x}}) - \epsilon/2 \geq \zeta^* - \epsilon/2$. Hence

$$\zeta^* - \epsilon \leq L_C(\boldsymbol{x}^p, \boldsymbol{\varphi}^p, \boldsymbol{\psi}^p) \leq \zeta^* \text{ for every sufficiently large } p \in K.$$

This contradicts to the assumption (22). Therefore, (d) follows.

6 Preliminary numerical results

We provide an illustrative example of structured and sparse POPs and show how the choice of SOS polynomials in SOS relaxations can enhance the efficiency of the proposed relaxations greatly while preserving the effectiveness.

As mentioned in Remark 4.2, the support set $\mathcal{G}^*$ in the proposed SOS relaxation of (2) becomes dense even for sparse $\mathcal{F}_j$ $(j = 0, 1, 2, \ldots, m)$ of the POP (2) because a polynomial $\varphi_0(\boldsymbol{x})$ is determined from the support of $L_C(\boldsymbol{x}, \varphi) - \xi$. The convergence result shown in Section 4 is based on this choice of φ_0. In practical implementation of the proposed SOS relaxations, however, it may be more important to obtain a good lower bound with relatively small size SDP relaxations. We show the formulation of SOS relaxation presented in this paper can be easily adapted in practice with the following example. The aim of the illustrative example is not to propose a practical method for general structured and sparse POPs, but to show how the SOS relaxation with convergent property can be modified for a specific problem in practice.

We consider an example

$$\left.\begin{array}{ll} \text{minimize} & f_0(\boldsymbol{x}) \equiv \displaystyle\sum_{i=1}^{n-1} f_{0i}(x_i, x_{i+1}) \\ \text{subject to} & f_{ij}(x_i, x_{i+1}) \geq 0 \ (i = 1, 2, \ldots, n-1, \ j = 1, 2, \ldots, m). \end{array}\right\} \quad (23)$$

Here $m \in \{1, \ n, \ n^2\}$, each $f_{0i}(x_i, x_{i+1})$ denotes a (fully dense) polynomial with degree 6 in two variables x_i, and x_{i+1} whose coefficients are chosen randomly from the interval $(-1, 1)$ $(i = 1, 2, \ldots, n-1)$, and each $f_{ij}(x_i, x_{i+1})$ denotes a polynomial in two variables x_i and x_{i+1} of the form

$$1 - \left(x_i^\ell, x_{i+1}\right) \left(\frac{1}{\lambda_1^2} \begin{pmatrix} -a_2 \\ a_1 \end{pmatrix} (-a_2, a_1) + \frac{1}{\lambda_2^2} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} (a_1, a_2) \right) \begin{pmatrix} x_i^\ell \\ x_{i+1} \end{pmatrix}$$

for some $\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ chosen from the unit circle, $\lambda_1, \ \lambda_2$ chosen randomly from the interval $(0.5, 2)$ and $\ell \in \{1, \ 3\}$ $(i = 1, 2, \ldots, n-1, j = 1, 2, \ldots, m)$. When $\ell = 1$, each constraint $f_{ij}(x_i, x_{i+1}) \geq 0$ forms an ellipsoid in the (x_i, x_{i+1}) space with the center at the origin; if $\lambda_1 > \lambda_2$, $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ corresponds to the major axis and $\begin{pmatrix} -a_2 \\ a_1 \end{pmatrix}$ the minor axis.

Let us derive three relaxations of (23): the dual of Lasserre's SDP relaxation, the SOS relaxation presented in Section 4.2, and a practical version of the SOS relaxation. If we want to have the POP (23) in the form of (2) and to follow the theory described so far literally, the redundant inequalities

$$1 - x_i^2 \geq 0 \ (i = 1, 2, \ldots, n)$$

need to be added to the POP (23). However, for simplicity of discussion, we consider the problem without these inequalities. Notice that if these inequalities are added, stronger relaxations for the three relaxations result in. As far as the size of the relaxations is concerned, adding the inequalities increases the size of all three relaxations. The biggest increase in the size occurs in case of the dual of Lasserre's SDP relaxation given in (24). We also note that all the SOS relaxations presented below remain effective without the inequalities.

Define the Lagrangian function

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{\varphi}) &= f_0(\boldsymbol{x}) - \sum_{i=1}^{n-1} \sum_{j=1}^{m} \varphi_{ij}(\boldsymbol{x}) f_{ij}(x_i, x_{i+1}) \text{ for every } \boldsymbol{x} \in \mathbb{R}^n \\ & \qquad \text{and every } \boldsymbol{\varphi} \equiv (\varphi_{ij} \ (i = 1, 2, \ldots, n-1, j = 1, 2, \ldots, m)) \in \overline{\Sigma}^{m(n-1)}. \end{aligned}$$

For every $i = 1, 2, \ldots, n-1$ and $q = 0, 1, \ldots$, let

$$\begin{aligned} \mathcal{A}_i^q &= \left\{ \mu \boldsymbol{e}^i + \nu \boldsymbol{e}^{i+1} : \mu \in \mathbb{Z}_+, \ \nu \in \mathbb{Z}_+, \ \mu + \nu \leq q \right\}, \\ \mathcal{A}_0^q &= \left\{ \boldsymbol{a} \in \mathbb{Z}_+^n : \sum_{i=1}^{n} a_i \leq q \right\}. \end{aligned}$$

Then we have two types of SOS relaxations. The one is

$$\left. \begin{array}{ll} \text{maximize} & \zeta \\ \text{subject to} & L(\boldsymbol{x}, \boldsymbol{\varphi}) - \zeta = \varphi_0(\boldsymbol{x}) \ (\forall \boldsymbol{x} \in \mathbb{R}^n), \\ & \varphi_{ij} \in \Sigma(\mathcal{A}_0^q) \ (i = 1, 2, \ldots, n-1, j = 1, 2, \ldots, m), \\ & \varphi_0 \in \Sigma(\mathcal{A}_0^{q+\ell}), \end{array} \right\} \tag{24}$$

which corresponds to the dual of Lasserre's SDP relaxation applied to the POP (23). The other is

$$\left. \begin{array}{ll} \text{maximize} & \zeta \\ \text{subject to} & L(\boldsymbol{x}, \boldsymbol{\varphi}) - \zeta = \varphi_0(\boldsymbol{x}) \ (\forall \boldsymbol{x} \in \mathbb{R}^n), \\ & \varphi_{ij} \in \Sigma(\mathcal{A}_i^q) \ (i = 1, 2, \ldots, n-1, j = 1, 2, \ldots, m), \\ & \varphi_0 \in \Sigma(\mathcal{A}_0^{q+\ell}), \end{array} \right\} \tag{25}$$

which exploits the sparsity of the constraint inequalities of (23). In both relaxations, we take nonnegative integers q and ℓ such that $q + \ell \geq 3$; hence $q = 2, 3, \ldots$ if $\ell = 1$, and $q = 0, 1, 2, \ldots$ if $\ell = 3$.

The SDP relaxation (24) uses

$$\begin{aligned} & m(n-1) \text{ copies of support sets } \mathcal{A}_0^q \text{ of size } \#\mathcal{A}_0^q = \binom{n+q}{n}, \\ & \text{a support set } \mathcal{A}_0^{q+\ell} \text{ of size } \#\mathcal{A}_0^{q+\ell} = \binom{n+q+\ell}{n}, \end{aligned}$$

while the SDP relaxation (25) uses

$$\begin{aligned} & m \text{ copies of support sets } \mathcal{A}_i^q \text{ of size } \#\mathcal{A}_i^q = \binom{2+q}{2} \ (i = 1, 2, \ldots, n-1), \\ & \text{a support set } \mathcal{A}_0^{q+\ell} \text{ of size } \#\mathcal{A}_0^{q+\ell} = \binom{n+q+\ell}{n}. \end{aligned}$$

The two SOS relaxations (24) and (25) share the support set $\mathcal{A}_0^{q+\ell}$ of size $\binom{n+q+\ell}{n}$. The difference between them lies in the support sets $\mathcal{A}_0^q$ and $\mathcal{A}_i^q$. We can see that the size of the SOS relaxation (25) is smaller than the size of the SOS relaxation (24). When q is fixed, the advantage of the SOS relaxation (25) in the size of the problem over the SOS relaxation

(24) becomes larger as m increases. This will be shown in Tables 1, 2 and 3. In the case of m fixed, the common support set $\mathcal{A}_0^{q+\ell}$ dominates all other support sets in both SOS relaxations in terms of size. As a result, the advantage from the size when increasing q is not as much as the case of q fixed.

Exploiting the structure of polynomials may improve the performance of the SOS relaxation (25) of the POP (23). We focus on " tridiagonal structure" of the support of the left hand side polynomial $L(\boldsymbol{x}, \boldsymbol{\varphi}) - \zeta$ of the equality constraint of the SOS relaxation (25), where φ_{ij} is assumed to be chosen from $\Sigma(\mathcal{A}_i^q)$. Specifically, the support of the polynomial $L(\boldsymbol{x}, \boldsymbol{\varphi}) - \zeta$ is covered by $\bigcup_{i=1}^{n-1} \left(\mathcal{A}_i^{q+\ell} + \mathcal{A}_i^{q+\ell} \right)$. Here we assume that $q + \ell \geq 3$. From this, we can expect that the polynomial $L(\boldsymbol{x}, \boldsymbol{\varphi}) - \zeta$ is represented as sums of squares of polynomials, each of which has a single support from $\mathcal{A}_i^{q+\ell}$ $(i = 1, 2, \ldots, n-1)$. We replace $\varphi_0(\boldsymbol{x})$ and $\varphi_0 \in \Sigma(\mathcal{A}_0^{q+\ell})$ by $\sum_{i=1}^{n-1} \psi_i(\boldsymbol{x})$ and $\psi_i \in \Sigma(\mathcal{A}_i^{q+\ell})$ $(i = 1, 2, \ldots, n-1)$ in the SOS relaxation (25), respectively, to obtain a new SOS relaxation of the POP (23):

$$\left.\begin{array}{ll} \text{maximize} & \zeta \\ \text{subject to} & L(\boldsymbol{x}, \boldsymbol{\varphi}) - \zeta = \displaystyle\sum_{i=1}^{n-1} \psi_i(\boldsymbol{x}) \ (\forall \boldsymbol{x} \in \mathbb{R}^n), \\ & \varphi_{ij} \in \Sigma(\mathcal{A}_i^q) \ (i = 1, 2, \ldots, n-1, j = 1, 2, \ldots, m), \\ & \psi_i \in \Sigma(\mathcal{A}_i^{q+\ell}) \ (i = 1, 2, \ldots, n-1). \end{array}\right\} \tag{26}$$

It should be noted that the size of every support set in the SOS relaxation (26) is independent of the dimension n of the POP (23). When m and q are fixed, the total size of support sets in the SOS relaxation (26) grows linearly with the dimension n while the growth rate of the total size of support sets in the SOS relaxation (24) as well as that in the SOS relaxation (25) are of $O(n^{q+\ell})$. This shows that the SOS relaxation (26) has a considerable computational advantage in solving the POP (23) of large dimension n.

The numerical experiment was done using SDPA 6.0 [21] on Pentium IV (XEON) 2.4 GHz with 6GB memory, and the optimal values of the POP (23) with $m \in \{1,\ n,\ n^2\}$, $\ell \in \{1,\ 3\}$, and $n \in \{4,\ 5,\ 6,\ 7\}$ were computed by GloptiPoly [3]. Tables 1, 2 and 3 show numerical results from the three SOS relaxations (24), (25) and (26) of the POP (23) with $m \in \{1,\ n,\ n^2\}$, $\ell = 1$ and dimension $n \in \{4,\ 5,\ 6,\ 7\}$. We observe that:

- All the SOS relaxations (24), (25) and (26) attain optimal values of the POP (23) with the lowest order $q = 2$.
- The SOS relaxation (25) requires less cpu time than the SOS relaxation (24), and the difference in cpu time becomes larger as m increases.

Table 4 shows numerical results from the three SOS relaxations (24), (25) and (26) of the POP (23) with $m = n$, $\ell = 3$ and dimension $n \in \{3,\ 4,\ 5,\ 6\}$. In this case:

- The SOS relaxations (25) and (26) obtain optimal values of the POP (23) or their lower bounds of (almost) the same quality as the SOS relaxation (24).
- The SOS relaxation (25) spends less cpu time than the SOS relaxation (24), but the difference is small.

- When $n = 6$, the SOS relaxation (24) with the order $q = 2$ attains the optimal value, but the other two SOS relaxations with the same order $q = 2$ provide only lower bounds for the optimal value. It should also be noted that the optimal value is obtained by the SOS relaxation (26) for the order $q = 3$.

In all cases reported in Tables 1, 2, 3 and 4:

- The SOS relaxation (26) has a clear advantage over the other SOS relaxations.

Table 1: Numerical results on the POP (23) with $m = 1$, $\ell = 0$ and $q = 2$

POP (23)	cpu time in seconds		
n	relaxation (24)	relaxation (25)	relaxation (26)
4	0.6	0.4	0.1
5	4.9	2.2	0.1
6	22.3	21.5	0.1
7	153.8	98.6	0.2

Table 2: Numerical results on the POP (23) with $m = n$, $\ell = 0$ and $q = 2$

POP (23)	cpu time in seconds		
n	relaxation (24)	relaxation (25)	relaxation (26)
4	1.6	0.5	0.1
5	11.2	2.6	0.2
6	83.3	13.2	0.4
7	607.1	64.4	0.5

Table 3: Numerical results on the POP (23) with $m = n^2$, $\ell = 0$ and $q = 2$

POP (23)	cpu time in seconds		
n	relaxation (24)	relaxation (25)	relaxation (26)
4	6.6	1.0	0.5
5	83.7	5.1	1.1
6	717.1	23.2	2.7
7	7402.5	135.6	4.1

Table 4: Numerical results on the POP (23) with $m = n$ and $\ell = 3$

POP (23)		cpu time in seconds (optimal value)		
n (optimal value)	q	relaxation (24)	relaxation (25)	relaxation (26)
3 (-1.782266)	0	0.1 (-148.0654)	0.1 (-148.0654)	0.1 (-148.0654)
	1	0.4 (-1.872454)	0.4 (-1.888884)	0.1 (-1.888890)
	2	1.9 (-1.782266)	1.6 (-1.782266)	0.2 (-1.782266)
4 (-2.244005)	0	0.4 (-129.5713)	0.4 (-129.5713)	0.1 (-129.5713)
	1	11.7 (-2.277639)	5.6 (-2.277639)	0.2 (-2.277844)
	2	46.2 (-2.244005)	36.1 (-2.244005)	0.4 (-2.244005)
5 (-3.848386)	0	2.6 (-120.1503)	2.5 (-120.2150)	0.1 (-120.1503)
	1	65.3 (-3.888779)	61.7 (-3.888779)	0.3 (-3.892605)
	2	787.1 (-3.848386)	644.6 (-3.848386)	0.8 (-3.848386)
6 (-3.531009)	0	13.3 (-120.2150)	13.4 (-120.2150)	0.1 (-120.2168)
	1	500.4 (-3.603920)	469.2 (-3.696910)	0.5 (-3.698462)
	2	11,912.4 (-3.531009)	11,718.1 (-3.535911)	1.4 (-3.537123)
	3	can't solve	can't solve	2.8 (-3.531009)

7 Concluding discussions

Considering two types of POPs (2) and (3) obtained from different characterizations of the feasible region of the POP (1), we have proposed a sequence of SOS relaxations from generalized Lagrangian duals of each POP. We have also provided a theoretical foundation on the convergence of the sequence of SOS relaxations to the optimal value of (3). The sequence of SOS relaxations have been transformed into a sequence of SDP relaxations to solve the POP (1) computationally. We have shown that the sequence of SDP relaxations derived here has primal-dual relationship with the one obtained after modifying Lasserre's SDP relaxation [10].

Theoretically, we have proved that the SOS relaxation of the Lagrangian dual of (3) attains the optimal value ζ^* of the POP (3). But there remains a gap between the Lagrangian dual of (2) and its SOS relaxation in Section 4.2; the former attains ζ^* but the latter is not guaranteed to attain ζ^*. Thus it is interesting to prove or disprove that the SOS relaxation of the Lagrangian dual of (2) attains ζ^*. This will be a subject of future study.

The size of the SOS relaxation or the SDP relaxation obtained from the Lagrangian dual approach by exploiting sparsity is smaller than the size of Lasserre's SDP relaxation. This is of course a nice feature, but this may not necessarily mean that the former SDP relaxation is as effective as the latter in practice. To attain an approximation to the optimal value ζ^* of the POP (1) with as high accuracy as the one from Lasserre's SDP relaxation, we may need higher degree SOS polynomials in our dual approach, which makes the size of the resulting SDP relaxation larger.

One of the advantages of the proposed method is that we have much flexibility in implementation of the SOS relaxation and the SDP relaxation of the POP (1); sets of Lagrangian multiplier SOS polynomials satisfying the assumption of Theorem 3.1 can be freely chosen to strengthen the resulting relaxations. We have presented an illustrative example of how the framework of the proposed SOS relaxation can be used to have a practical SOS relaxation exploiting a structured sparsity. Numerical results of the example have indicated that it is possible to drastically improve computational efficiency of SOS relaxations by making proper heuristic choices of supports, depending on problems.

In addition to the preliminary numerical results in Section 6, numerical experiments on the SDP relaxation with heuristically chosen supports were performed for various types of polynomial optimization problems with certain types of sparsity, but not included in this paper because we believe that the discussion of heuristics is beyond the scope of this paper. The main purpose of this paper has been proposing general methods for sparsity in SDP relaxations for polynomial optimization and introducing Lagrangian dual and penalty function approaches into SDP relaxations for polynomial optimization. Although the numerical results supported the claim that the SOS relaxations could improve the efficiency, it would be necessary to address issues such as (i) a reasonable definition of structured sparsity in polynomial optimization problems, (ii) technical details of heuristic choices of supports, and (iii) extensive numerical experiments on various problems with structured sparsity, These will consist of a paper on practical performance of heuristics, which we hope to present in near future.

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] M. D. Choi, T. Y. Lam, and B. Reznick, "Sums of squares of real polynomials", *Proceedings of Symposia in Pure Mathematics*, **58** 2 (1995) 103-126.

[3] D. Henrion and J. B. Lasserre, "GloptiPoly: Global optimization over polynomials with Matlab and SeDuMi", Laboratoire d'Analyse et d'Architecture des Syst'emes, Centre National de la Recherche Scientifique, 7 Avenue du Colonel Roche, 31 077 Toulouse, cedex 4, France, February 2002.

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

[5] S. Kim and M. Kojima, "Exact solutions of some nonconvex quadratic optimization problems via SDP and SOCP relaxations", *Computational Optimization and Applications*, **26** (2003) 143-154.

[6] M. Kojima, S. Kim and H. Waki, "A general framework for convex relaxation of polynomial optimization problems over cones", *Journal of Operations Research Society of Japan*, **46** 2 (2003) 125-144.

[7] M. Kojima, S. Kim and H. Waki, "Sparsity in sums of squares of polynomials", Research Report B-391, Dept. of Mathematical and computing Sciences, Tokyo Institute of Technology, Meguro, Tokyo 152-8552, June 2003. Revised August 2003.

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

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

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

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

[12] R. D. C. Monteiro, "First- and second-order methods for semidefinite programming", *Mathematical Programming*, **97** (2003) 209–244.

[13] P. A. Parrilo, "Semidefinite programming relaxations for semialgebraic problems", *Mathematical Programming*, **96** (2003) 293–320.

[14] S. Prajna, A. Papachristodoulou and P. A. Parrilo, "SOSTOOLS: Sum of Squares Optimization Toolbox for MATLAB – User's Guide", Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125 USA, 2002.

[15] V. Powers and T. Wörmann, "An algorithm for sums of squares of real polynomials", *Journal of Pure and Applied Algebra*, **127** (1998) 99-104.

[16] M. Putinar, "Positive polynomials on compact semi-algebraic sets", *Indiana University Mathematics Journal*, **42** (1993) 969–984.

[17] B. Reznick, "Extremal psd forms with few terms", *Duke Mathematical Journal*, **45** (1978) 363-374.

[18] 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.

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

[21] M. Yamashita, K. Fujisawa and M. Kojima, "Implementation and evaluation of SDPA 6.0 (SemiDefinite Programming Algorithm 6.0)", *Optimization Methods and Software*, **18** (2003) 491-505.