On semi-infinite systems of convex polynomial inequalities and polynomial optimization problems

Feng Guo[a], Xiaoxia Sun[b]

[a] *School of Mathematical Sciences,*
Dalian University of Technology, Dalian, 116024, China
[b] *School of Mathematics,*
Dongbei University of Finance and Economics, Dalian, 116025, China

Abstract

We consider the semi-infinite system of polynomial inequalities of the form

$$\mathbf{K} := \{x \in \mathbb{R}^m \mid p(x, y) \geq 0, \quad \forall y \in S \subseteq \mathbb{R}^n\},$$

where $p(X, Y)$ is a real polynomial in the variables X and the parameters Y, the index set S is a basic semialgebraic set in $\mathbb{R}^n$, $-p(X, y)$ is convex in X for every $y \in S$. We propose a procedure to construct approximate semidefinite representations of $\mathbf{K}$. These semidefinite representation sets are indexed by two indices which respectively bound the order of some moment matrices and the degree of sums of squares representations of some polynomials in the construction. As two indices increase, these semidefinite representation sets expand and contract, respectively, and can approximate $\mathbf{K}$ as closely as possible under some assumptions. Some special cases when we can fix one of the two indices or both are also investigated. Then, we consider the optimization problem of minimizing a convex polynomial over $\mathbf{K}$. We present an SDP relaxation method for this optimization problem by similar strategies used in constructing approximate semidefinite representations of $\mathbf{K}$. Under certain assumptions, some approximate minimizers of the optimization problem can also be obtained from the SDP relaxations. In some special cases, we show that the SDP relaxation for the optimization problem is exact and all minimizers can be extracted.

Keywords: semi-infinite systems, convex polynomials, semidefinite representations, semidefinite programming relaxations, sum of squares, polynomial optimization
2010 MSC: 65K05, 90C22, 90C34

1. introduction

We consider the following semi-infinite system of polynomial inequalities

$$\mathbf{K} := \{x \in \mathbb{R}^m \mid p(x, y) \geq 0, \quad \forall y \in S \subseteq \mathbb{R}^n\}, \tag{1}$$

where $p(X, Y) \in \mathbb{R}[X, Y] := \mathbb{R}[X_1, \ldots, X_m, Y_1, \ldots, Y_n]$ the polynomial ring in X and Y over the real field and the *index set* S is a basic semialgebraic set defined by

$$S := \{y \in \mathbb{R}^n \mid g_1(y) \geq 0, \ldots, g_s(y) \geq 0\}, \tag{2}$$

where $g_j(Y) \in \mathbb{R}[Y]$, $j = 1, \ldots, s$. Lowercase letters (e.g. x, y) are hereinafter used for denoting points in a space while uppercase letters (e.g. X, Y) for variables. In this paper, we assume that $-p(X, y) \in \mathbb{R}[X]$ is convex for every $y \in S$ and hence $\mathbf{K}$ is a convex set in $\mathbb{R}^m$.

Email addresses: `fguo@dlut.edu.cn` (Feng Guo), `xiaoxiasun@dufe.edu.cn` (Xiaoxia Sun)

We say a convex set C in $\mathbb{R}^m$ is *semidefinitely representable* (or *linear matrices inequality representable*) if there exist some integers l, k and real $k \times k$ symmetric matrices $\{A_i\}_{i=0}^m$ and $\{B_j\}_{j=1}^l$ such that C is identical with

$$\left\{ x \in \mathbb{R}^m \;\middle|\; \exists w \in \mathbb{R}^l, \text{ s.t. } A_0 + \sum_{i=1}^{m} A_i x_i + \sum_{j=1}^{l} B_j w_j \succeq 0 \right\} \tag{3}$$

and (3) is called the *semidefinite representation* (or *linear matrices inequality representation*) of C. Many interesting convex sets are semidefinitely representable, see a collection in Ben-Tal and Nemirovski (2001). Clearly, optimizing a linear function over a semidefinitely representable set can be cast as a semidefinite progamming (SDP) problem, while SDP has an extremely wide area of applicaitons and can be solved by interior-point method to a given accuracy in polynomial time (c.f. Wolkowicz et al. (2000)). Semidefinite representations of convex sets can help us to build SDP relaxations of many computationally intractable optimization problems. Arising from above, one of the basic issues in convex algebraic geometry is to characterize convex sets in $\mathbb{R}^m$ which are semidefinitely representable and give systematic procedures to obtain their semidefinite representations. Clearly, if a set in $\mathbb{R}^m$ is semidefinitely representable, then it is convex and semialgebraic. Conversely, Nemirovski asked in his plenary address at the 2006 ICM that whether each convex semialgebraic set is semidefinitely representable. Yet a negative answer has been recently given by Scheiderer (2018). Hence, it is reasonable to study how to construct approximate semidefinite representations of C, that is a sequence of semidefinite representation sets of the form (3) which converge to C in some sence.

For a given basic semialgebraic set in $\mathbb{R}^m$, Lasserre (2009b) and Gouveia et al. (2010) proposed some methods to construct semidefinite outer approximations of the closure of its convex hull. These appproaches are based on the sums of squares representation of linear functions which are nonnegative on a basic semialgebraic set. If the basic semialgebraic set is compact, these approximations can be made arbitrarily close and become exact under some favorable conditions. Some extensions of these semidefinite approximations to noncompact basic semialgebraic sets are given in Guo et al. (2015). For a convex semialgebraic set, Helton and Nie (2009, 2010) proposed some sufficient conditions, in terms of curvature conditions for the boundary, for its semidefinite representability. These conditions are recently modified and improved by Kriel and Schweighofer (2018).

In this paper, we first consider to construct approximate semidefinite representations of the set $\mathbf{K}$ in (1). The difference of this problem from ones in the literature is that $\mathbf{K}$ is defined by *infinitely* many convex real polynomials. As there is a quantifier in the definition (1), $\mathbf{K}$ is in fact a semialgebraic set by the Tarski-Seidenberg principle (c.f. Bochnak et al. (1998)). Theoretically, $\mathbf{K}$ can be decomposed as a finite union of basic closed semialgebraic sets and hence, as proved in Helton and Nie (2009), the semidefinite approximations of $\mathbf{K}$ can be made by glueing together Lasserre relaxations Lasserre (2009b) of many small pieces of $\mathbf{K}$. However, such a decomposition of $\mathbf{K}$ may not be easily obtained and the approach given in Helton and Nie (2009) is not constructive. These obstacles make the problem studied in this paper nontrivial. As the first contribution in this paper, we propose a procedure to construct approximate semidefinite representations of $\mathbf{K}$. These semidefinite representation sets are indexed by two indices which respectively bound the order of some moment matrices and the degree of sums of squares representations of some polynomials in the construction. As two indices increase, these semidefinite representation sets expand and contract, respectively, and can approximate $\mathbf{K}$ as closely as possible under some assumptions. Some special cases when we can fix one of the two indices or both are also investigated.

In the second part of this paper, we consider the following convex minimization problem

$$(\mathbf{P}) \qquad f^* := \inf_{x \in \mathbf{K}} f(x) \quad \text{where } \mathbf{K} \text{ is defined in (1) and } f(X) \in \mathbb{R}[X] \text{ is convex.}$$

This problem is NP-hard. Indeed, it is obvious that the problem of minimizing a polynomial $h(Y) \in \mathbb{R}[Y]$ over S can be regarded as a special case of $(\mathbf{P})$. As is well known, the polynomial optimization problem is NP-hard even when $n > 1$, $h(Y)$ is a nonconvex quadratic polynomial and $g_j(Y)$'s are linear (c.f. Pardalos and Vavasis (1991)). Hence, a general the problem $(\mathbf{P})$ cannot be expected to be solved in polynomial time unless P=NP.

The problem (**P**) can be seen as a special branch of convex *semi-infinite programming* (SIP), in which the involved functions are not necessarily polynomials. Numerically, SIP problems can be solved by different approaches including, for instance, discretization methods, local reduction methods, exchange methods, simplex-like methods and so on. See Hettich and Kortanek (1993); López and Still (2007); Goberna and López (2017) and the references therein for details. One of main difficulties in numerical treatment of general SIP problems is that the feasibility test of $\bar{u} \in \mathbb{R}^m$ is equivalent to *globally* solve the lower level subproblem of $\min_{y \in S} p(\bar{u}, y)$ which is generally nonlinear and nonconvex. To the best of our knowledge, few of the numerical methods mentioned above are specially designed by exploiting features of polynomial optimization problems. Parpas and Rustem (2009) proposed a discretization-like method to solve minimax polynomial optimization problems, which can be reformulated as semi-infinite polynomial programming (SIPP) problems. Using polynomial approximation and an appropriate hierarchy of SDP relaxations, Lasserre presented an algorithm to solve the generalized SIPP problems in Lasserre (2012). Based on an exchange scheme, an SDP relaxation method for solving SIPP problems was proposed in Wang and Guo (2013). By using representations of nonnegative polynomials in the univariate case, an SDP method was given in Xu et al. (2015) for linear SIPP problems (a special case of (**P**)) with S being closed intervals.

As the second contribution in this paper, we present some SDP relaxation methods for the problem (**P**) by similar strategies used in constructing approximate semidefinite representations of **K**. Under certain assumptions, some approximate minimizers of (**P**) can also be obtained from the SDP relaxations. In some special cases, we show that the SDP relaxation of (**P**) is exact and all minimizers can be extracted.

This paper is organized as follows. In Section 2, we give some notation and preliminaries used in this paper. Approximate semidefinite representations of **K** as well as some examples are proposed in Section 3. We study SDP relaxations of the problem (**P**) in Section 4.

2. Notation and Preliminaries

Here is some notation used in this paper. The symbol $\mathbb{N}$ (resp., $\mathbb{R}$) denotes the set of nonnegative integers (resp., real numbers). For any $t \in \mathbb{R}$, $\lceil t \rceil$ (resp. $\lfloor t \rfloor$) denotes the smallest (resp. largest) integer that is not smaller (resp. larger) than t. For $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, $\|y\|_2$ denotes the standard Euclidean norm of y. For $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$, $\|\alpha\|_1 = \alpha_1 + \cdots + \alpha_n$. For $k \in \mathbb{N}$, denote $\mathbb{N}^n_k = \{\alpha \in \mathbb{N}^n \mid \|\alpha\|_1 \le k\}$ and $|\mathbb{N}^n_k|$ its cardinality. For $y \in \mathbb{R}^n$ and $\alpha \in \mathbb{N}^n$, y^α denotes $y_1^{\alpha_1} \cdots y_n^{\alpha_n}$. $\mathbb{R}[Y] = \mathbb{R}[Y_1, \cdots, Y_n]$ denotes the ring of polynomials in $(Y_1, \cdots, Y_n)$ with real coefficients. For $k \in \mathbb{N}$, denote by $\mathbb{R}[Y]_k$ the set of polynomials in $\mathbb{R}[Y]$ of total degree up to k. For a symmetric matrix W, $W \succeq 0 (\succ 0)$ means that W is positive semidefinite (definite). For two symmetric matrices A, B of the same size, $\langle A, B \rangle$ denotes the inner product of A and B.

We say that the *Slater condition* holds for **K** if there exists $u \in \mathbf{K}$ such that $p(u, y) > 0$ for all $y \in S$ and the point u is called a *Slater point*. Consider the semi-infinite convex polynomial optimization problem (**P**).

Theorem 2.1. (c.f. Borwein (1981); Levin (1969)) *Assume that the Slater condition holds for* **K** *and the index set S is compact. Then for any convex $f(X) \in \mathbb{R}[X]$, there exist points $y_1, \ldots, y_l \in S$ with $l \le n$ such that f^* is equal to the optimal value of the discretization problem*

$$\min_{x \in \mathbb{R}^m} f(x) \quad \text{s.t. } p(x, y_1) \ge 0, \ldots, p(x, y_l) \ge 0. \tag{4}$$

Corollary 2.2. *Suppose that the assumptions in Theorem 2.1 hold. Then for any convex $f[X] \in \mathbb{R}[X]$, there exist points $y_1, \ldots, y_l \in S$ and nonnegative Lagrange multipliers $\lambda_1, \ldots, \lambda_l \in \mathbb{R}$ with $l \le n$ such that the Lagrangian*

$$L_f(X) := f(X) - f^* - \sum_{i=1}^{l} \lambda_i p(X, y_i) \tag{5}$$

is nonnegative on $\mathbb{R}^m$.

Next we recall some background about *sums of squares* (s.o.s) of polynomials and the dual theory of *moment matrices*. A polynomial $\phi(X) \in \mathbb{R}[X]$ is said to be a sum of squares of polynomials if it can be written as $\phi(X) = \sum_{i=1}^{t} \phi_i(X)^2$ for some $\phi_1(X), \ldots, \phi_t(X) \in \mathbb{R}[X]$. The symbols $\Sigma^2[X]$ and $\Sigma^2[Y]$ denote the sets of polynomials that are s.o.s in $\mathbb{R}[X]$ and $\mathbb{R}[Y]$, respectively. Notice that not every nonnegative polynomials can be written as s.o.s, see Reznick (2000). Lasserre and Netzer (2007) gave the following s.o.s approximations of nonnegative polynomials via simple high degree perturbations.

Theorem 2.3. (Lasserre and Netzer, 2007, c.f. Theorem 3.1, 3.2 and Corollary 3.3) *For a given $h \in \mathbb{R}[X]$, the followings are true.*

(i) *For any $r \geq \lceil \deg(h)/2 \rceil$, there exists $\varepsilon_r^* \geq 0$ such that $h + \varepsilon(1 + \sum_{j=1}^{m} X_j^{2r})$ is s.o.s if and only if $\varepsilon \geq \varepsilon_r^*$;*

(ii) *If h is nonnegative on $[-1,1]^m$, then ε_r^* in (i) decreasingly converges to 0 as r tends to ∞;*

(iii) *For any $\varepsilon > 0$, if h is nonnegative on $[-1,1]^m$, then there exists some $r(h,\varepsilon) \in \mathbb{N}$ such that $h + \varepsilon(1 + \sum_{j=1}^{m} X_j^{2r})$ is s.o.s for every $r \geq r(h,\varepsilon)$.*

Moreover, ε_r^* in Theorem 2.3 is computable by solving an SDP problem, see (Lasserre and Netzer, 2007, Theorem 3.1).

Now we consider the cone $\mathcal{P}(S)$ of polynomials in $\mathbb{R}[Y]$ which are nonnegative on S. Let $G := \{g_1, \ldots, g_s\}$ be the set of polynomials that defines the semialgebraic set S (2). We denote by

$$\mathcal{Q}(G) := \left\{ \sum_{j=0}^{s} \sigma_j g_j \;\middle|\; g_0 = 1,\ \sigma_j \in \Sigma^2[Y], j = 0, 1, \ldots, s \right\}$$

the *quadratic module* generated by G and denote by

$$\mathcal{Q}_k(G) := \left\{ \sum_{j=0}^{s} \sigma_j g_j \;\middle|\; g_0 = 1,\ \sigma_j \in \Sigma^2[Y],\ \deg(\sigma_j g_j) \leq 2k, j = 0, 1, \ldots, s \right\}$$

its k-th *quadratic module*. It is clear that if $h \in \mathcal{Q}(G)$, then $h(y) \geq 0$ for any $y \in S$. However, the converse is not necessarily true. Note that checking $h \in \mathcal{Q}_k(G)$ for a fixed $k \in \mathbb{N}$ is an SDP feasibility problem, see Lasserre (2001); Parrilo and Sturmfels (2003).

Definition 2.4. *We say that $\mathcal{Q}(G)$ is Archimedean if there exists $\psi \in \mathcal{Q}(G)$ such that the inequality $\psi(y) \geq 0$ defines a compact set in $\mathbb{R}^n$.*

Note that the Archimedean property implies that S is compact but the converse is not necessarily true. However, for any compact set S we can always force the associated quadratic module to be Archimedean by adding a redundant constraint $M - \|y\|_2^2 \geq 0$ in the description of S for sufficiently large M.

Theorem 2.5. (Putinar, 1993, Putinar's Positivstellensatz) *Suppose that $\mathcal{Q}(G)$ is Archimedean. If a polynomial $h \in \mathbb{R}[Y]$ is positive on S, then $h \in \mathcal{Q}_k(G)$ for some $k \in \mathbb{N}$.*

Consequently, for any $d \in \mathbb{N}$, we have $\mathsf{closure}(\mathcal{Q}(G) \cap \mathbb{R}[Y]_d) = \mathcal{P}(S) \cap \mathbb{R}[X]_d$ if $\mathcal{Q}(G)$ is Archimedean.

For a polynomial $h(Y) = \sum_\alpha h_\alpha Y^\alpha \in \mathbb{R}[Y]$, define the norm

$$\|h\| := \max_\alpha \frac{|h_\alpha|}{\binom{\|\alpha\|_1}{\alpha}}. \tag{6}$$

We have the following result for an estimation of the order k in Theorem 2.5.

Theorem 2.6. (Nie and Schweighofer, 2007, Theorem 6) *Suppose that $\mathcal{Q}(G)$ is Archimedean and $S \subseteq (-\tau_S, \tau_S)^n$ for some $\tau_S > 0$. Then there is some positive $c \in \mathbb{R}$ (depending only on g_j's) such that for all $h \in \mathbb{R}[Y]$ of degree d with $\min_{y \in S} h(y) > 0$, we have $h \in \mathcal{Q}_k(G)$ whenever*

$$k \geq c \exp\left[\left(d^2 n^d \frac{\|h\| \tau_S^d}{\min_{y \in S} h(y)} \right)^c \right].$$

A sequence of real numbers $z := (z_\alpha)_{\alpha\in\mathbb{N}^n} \in \mathbb{R}^{\mathbb{N}^n}$ whose elements are indexed by n-tuples $\alpha \in \mathbb{N}^n$ is called a *moment sequence* and the truncation $(z_\alpha)_{\alpha\in\mathbb{N}^n_{2k}} \in \mathbb{R}^{\mathbb{N}^n_{2k}}$ is called a *truncated moment sequence* up to order $2k$. For $z \in \mathbb{R}^{\mathbb{N}^n}$, if there exists a Borel measure μ on $\mathbb{R}^n$ such that

$$z_\alpha = \int Y^\alpha \mathrm{d}\mu(y),\ \forall \alpha \in \mathbb{N}^n,$$

then we say that z has a *representing measure* μ. A basic problem in the theory of moments concerns the characterization of (infinite or truncated) sequences which have some representing measure. For any moment sequence z, the Riesz functional $\mathscr{L}_z$ on $\mathbb{R}[Y]$ is defined by

$$\mathscr{L}_z\left(\sum_\alpha q_\alpha Y_1^{\alpha_1}\cdots Y_n^{\alpha_n}\right) := \sum_\alpha q_\alpha z_\alpha, \quad \forall q(Y) \in \mathbb{R}[Y]. \tag{7}$$

For bounded moment sequences, we have the following results for the moment problem.

Theorem 2.7. (Berg and Maserick, 1984, Theorem 2.1) *Let $z \in \mathbb{R}^{\mathbb{N}^n}$ be a moment sequence such that $\mathscr{L}_z(h) \geq 0$ for all $h \in \Sigma^2[Y]$. If there exist $a, c > 0$ such that $|z_\alpha| \leq ca^{|\alpha|}$ for every $\alpha \in \mathbb{N}^n$, then z has exactly one representing measure μ on $\mathbb{R}^n$ with support contained in $[-c, c]^n$.*

Denote by $\mathcal{M}_S$ the set of those moment sequences which have some representing measure supported on S in (2). To characterize the elements in $\mathcal{M}_S$, we need to introduce the definitions about moment matrices. The associated k-th *moment matrix* is the matrix $M_k(z)$ indexed by $\mathbb{N}^n_k$, with (α,β)-th entry $z_{\alpha+\beta}$ for $\alpha, \beta \in \mathbb{N}^n_k$. Given a polynomial $h(Y) = \sum_\alpha h_\alpha Y^\alpha$, for $k \geq d_h := \lceil \deg(h)/2 \rceil$, the $(k-d_h)$-th *localizing moment matrix* $M_{k-d_h}(fz)$ is defined as the moment matrix of the *shifted vector* $((hz)_\alpha)_{\alpha\in\mathbb{N}^n_{2(k-d_h)}}$ with $(hz)_\alpha = \sum_\beta h_\beta z_{\alpha+\beta}$. For any $q(Y) \in \mathbb{R}[Y]_k$, let $\mathbf{q}$ denote its column vector of coefficients in the canonical monomial basis of $\mathbb{R}[Y]_k$. From the definition of the localizing moment matrix $M_{k-d_h}(hz)$, it is easy to check that

$$\mathbf{q}^T M_{k-d_h}(hz)\mathbf{q} = \mathscr{L}_z(h(Y)q(Y)^2), \quad \forall q(Y) \in \mathbb{R}[Y]_{k-d_h}. \tag{8}$$

Let

$$d_j := \lceil \deg(g_j)/2 \rceil, \quad \forall\, j = 1, \ldots, s, \quad d_S := \max_j d_j. \tag{9}$$

For any $v \in S$, let $\zeta_{2k,v} := [v^\alpha]_{\alpha\in\mathbb{N}^n_{2k}}$ be the *Zeta vector* of v up to degree $2k$, i.e.,

$$\zeta_{2k,v} = [1 \quad v_1 \quad \cdots \quad v_n \quad v_1^2 \quad v_1v_2 \quad \cdots \quad v_n^{2k}].$$

Then, $M_k(\zeta_{2k,v}) \succeq 0$ and $M_{k-d_j}(g_j\zeta_{2k,v}) \succeq 0$ for $j = 1, \ldots, s$. In fact, let $g_0 = 1$, then for each $j = 0, 1, \ldots, s$,

$$\mathbf{q}^T M_{k-d_j}(g_j\zeta_{2k,v})\mathbf{q} = \mathscr{L}_{\zeta_{2k,v}}(g_j(Y)q(Y)^2) = g_j(v)q(v)^2 \geq 0,\ \forall q(Y) \in \mathbb{R}[Y]_{k-d_j}.$$

Haviland (1935) proved that the dual cone $(\mathcal{P}(S))^* = \mathcal{M}_S$. Hence, in a dual view, Putinar's Positivstellensatz reads

Theorem 2.8. (Putinar, 1993, Putinar's Positivstellensatz) *Suppose that $\mathcal{Q}(G)$ is Archimedean. If $M_k(z) \succeq 0$ and $M_{k-d_j}(g_jz) \succeq 0$ for all $j = 1, \ldots, s$, and all $k = 0, 1, \ldots$, then $z \in \mathcal{M}_S$.*

For a truncated moment sequence $z = (z_\alpha)_{\alpha\in\mathbb{N}^m_{2k}}$, we have the following sufficient condition for $z \in \mathcal{M}_S$.

Condition 2.9. *A truncated moment sequence $z = (z_\alpha)_{\alpha\in\mathbb{N}^m_{2k}}$ satisfies the Rank Condition when*

$$\mathrm{rank} M_{k-d_S}(z) = \mathrm{rank} M_k(z).$$

Theorem 2.10. (Curto and Fialkow, 2005, Theorem 1.1) *Suppose that a truncated moment sequence $z = (z_\alpha)_{\alpha\in\mathbb{N}^m_{2k}}$ satisfies that $M_k(z) \succeq 0$ and $M_{k-d_j}(g_jz) \succeq 0$ for all $j = 1, \ldots, s$, and the Rank Condition 2.9 holds with $r := \mathrm{rank} M_k(z)$, then z has a unique r-atomic measure supported on S.*

To end this section, let us recall a very interesting subclass of convex polynomials in $\mathbb{R}[Y]$ introduced by Helton and Nie (2010).

Definition 2.11. (Helton and Nie (2010)) *A polynomial $h \in \mathbb{R}[Y]$ is s.o.s-convex if its Hessian $\nabla^2 h$ is a s.o.s, i.e., there is some integer r and some matrix polynomial $H \in \mathbb{R}[Y]^{r\times n}$ such that $\nabla^2 h(Y) = H(Y)^T H(Y)$.*

While checking the convexity of a convex polynomial is generally NP-hard (c.f. Ahmadi et al. (2013)), s.o.s-convexity can be checked numerically by solving an SDP, see Helton and Nie (2010). The following result plays a significant role in this paper.

Lemma 2.12. (Helton and Nie, 2010, Lemma 8) *Let $h \in \mathbb{R}[Y]$ be s.o.s-convex. If $h(u) = 0$ and $\nabla h(u) = 0$ for some $u \in \mathbb{R}^n$, then h is s.o.s.*

3. Approximate semidefinite representations of K

As we always assume that the index set S in the definition of $\mathbf{K}$ is compact in this paper, we first show that in generic case a set $\mathbf{K}$ with noncompact index set S can be converted into compact case.

3.1. Noncompact case

In this subsection, we consider the set $\mathbf{K}$ in (1) with noncompact index set S. We used the technique of homogenization proposed in Wang and Guo (2013) to convert a semi-infinite system (1) with general noncompact index set into compact case.

For a polynomial $g(Y) \in \mathbb{R}[Y]$, denote its homogenization by $g^{\rm hom}(\widetilde{Y}) \in \mathbb{R}[\widetilde{Y}]$, where $\widetilde{Y} = (Y_0, Y_1, \ldots, Y_n)$, i.e., $g^{\rm hom}(\widetilde{Y}) = Y_0^{\deg(g)} g(Y/Y_0)$. For the basic semialgebraic set S in (2), define

$$\begin{aligned} \widetilde{S}_> &:= \{\tilde{y} \in \mathbb{R}^{n+1} \mid g_1^{\rm hom}(\tilde{y}) \geq 0, \ \ldots, \ g_s^{\rm hom}(\tilde{y}) \geq 0, \ y_0 > 0, \ \|\tilde{y}\|_2^2 = 1\}, \\ \widetilde{S} &:= \{\tilde{y} \in \mathbb{R}^{n+1} \mid g_1^{\rm hom}(\tilde{y}) \geq 0, \ \ldots, \ g_s^{\rm hom}(\tilde{y}) \geq 0, \ y_0 \geq 0, \ \|\tilde{y}\|_2^2 = 1\}. \end{aligned} \tag{10}$$

Proposition 3.1. (Wang and Guo, 2013, Proposition 4.2) *For any $g(Y) \in \mathbb{R}[Y]$, $g(y) \geq 0$ on S if and only if $g^{\rm hom}(\tilde{y}) \geq 0$ on $\mathsf{closure}(\widetilde{S}_>)$.*

Let $d_Y := \deg_Y(p(X,Y))$ and $p^{\rm hom}(X, \widetilde{Y})$ be the homogenization of $p(X,Y)$ with respect to the variables Y. It follows that the set $\mathbf{K}$ in (1) is equivalent to

$$\{x \in \mathbb{R}^m \mid p^{\rm hom}(x, \tilde{y}) \geq 0, \quad \forall y \in \mathsf{closure}(\widetilde{S}_>)\}.$$

Replacing $\mathsf{closure}(\widetilde{S}_>)$ by the basic semialgebraic set $\widetilde{S}$, we get the following set

$$\widetilde{\mathbf{K}} := \{x \in \mathbb{R}^m \mid p^{\rm hom}(x, \tilde{y}) \geq 0, \quad \forall y \in \widetilde{S}\}.$$

It is obvious that $\widetilde{\mathbf{K}} \subseteq \mathbf{K}$ since $\mathsf{closure}(\widetilde{S}_>) \subseteq \widetilde{S}$.

Definition 3.2. (Nie (2013)) *S is said to be closed at ∞ if $\mathsf{closure}(\widetilde{S}_>) = \widetilde{S}$.*

Remark 3.3. Clearly, $\mathbf{K} = \widetilde{\mathbf{K}}$ when S is closed at ∞. Note that not every set S of form (2) is closed at ∞ even when it is compact (Nie, 2012, Example 5.2). However, it is shown in (Wang and Guo, 2013, Theorem 4.10) that the closedness at ∞ is a *generic* property. Namely, if we consider the space of all coefficients of generators g_j's of all possible sets S of form (2) in the canonical monomial basis of $\mathbb{R}[Y]_d$ with $d = \max_j \deg(g_j)$, coefficients of g_j's of those sets S which are not closed at ∞ are in a Zariski closed set of the space. It follows that $\mathbf{K} = \widetilde{\mathbf{K}}$ for *general* index sets S. Note that $\widetilde{S}_>$ depends only on S, while $\widetilde{S}$ depends not only on S but also on the choice of the inequalities $g_1(y) \geq 0, \ldots, g_s(y) \geq 0$. In some cases, we can add some redundant inequalities in the description of S to force it to be closed at ∞ (c.f. Guo et al. (2015)).

For any polynomial $g(Y) \in \mathbb{R}[Y]$, denote $\hat{g}(Y)$ as its homogeneous part of the highest degree. Define

$$\widehat{S} := \{y \in \mathbb{R}^n \mid \hat{g}_1(y) \geq 0, \ldots, \hat{g}_s(y) \geq 0, \|y\|_2^2 = 1\}. \tag{11}$$

In particular, denote $\hat{p}(X, Y)$ as the homogeneous parts of $p(X, Y)$ with respect to Y of the highest degree d_Y.

Definition 3.4. *We say that the extended Slater condition holds for* $\mathbf{K}$ *if there exists a point* $u \in \mathbb{R}^m$ *of* $\mathbf{K}$ *such that* $p(u, y) > 0$ *for all* $y \in S$ *and* $\hat{p}(u, y) > 0$ *for all* $y \in \widehat{S}$. *We call* u *an extended Slater point of* $\mathbf{K}$.

Proposition 3.5. *The Slater condition holds for* $\widetilde{\mathbf{K}}$ *if and only if the extended Slater condition holds for* $\mathbf{K}$.

Proof. Suppose that u is an extended Slater point of $\mathbf{K}$. For any $\tilde{v} = (v_0, v) \in \widetilde{S}$, we have $v \in \widehat{S}$ if $v_0 = 0$ and $v/v_0 \in S$ otherwise. It is straightforward to verify that the Slater condition also holds for $\widetilde{\mathbf{K}}$ at u.

Suppose that the Slater condition holds for $\widetilde{\mathbf{K}}$ at $u \in \mathbb{R}^m$. For any point $v \in \mathbb{R}^n$, we have $(0, v) \in \widetilde{S}$ if $v \in \widehat{S}$ and $\left(\frac{1}{\sqrt{1+\|v\|_2^2}}, \frac{v}{\sqrt{1+\|v\|_2^2}}\right) \in \widetilde{S}$ if $v \in S$. Then similarly, it implies that the extended Slater condition holds for $\mathbf{K}$ at u. □

As a result of the above arguments, it is reasonable to consider the following assumption in the rest of this paper.

Assumption 3.6. *The set* S *is compact,* $-p(X, y) \in \mathbb{R}[X]$ *is convex for any* $y \in S$ *and the Slater condition holds for* $\mathbf{K}$.

3.2. Approximate semidefinite representations of $\mathbf{K}$

We assume that $\mathbf{K}$ in (1) is compact and a scalar $\tau_{\mathbf{K}}$ such that $\|x\|_2 \leq \tau_{\mathbf{K}}$ for any $x \in \mathbf{K}$ is known. Define $\Theta_r(X) = \sum_{i=1}^m \left(\frac{X_i}{\tau_{\mathbf{K}}}\right)^{2r} \in \mathbb{R}[X]$ for any $r \in \mathbb{N}$. It is clear that $\Theta_r(x) \leq 1$ for any $x \in \mathbf{K}$ and $r \in \mathbb{N}$. For convenience, we write $p(X, Y) = \sum_\alpha p_{X,\alpha}(Y) X^\alpha = \sum_\beta p_{Y,\beta}(X) Y^\beta$, i.e., $p_{X,\alpha}(Y)$ and $p_{Y,\beta}(X)$ denote the coefficients of X^α and Y^β in $p(X, Y)$ regarded as a polynomial in $\mathbb{R}[X]$ and $\mathbb{R}[Y]$, respectively. Denote by $\mathbf{B}$ the unit ball in $\mathbb{R}^m$. Let $d_X = \deg_X(p(X, Y))$ and $d_Y = \deg_Y(p(X, Y))$. Recall the notation d_S in (9) and the Riesz function defined in (7). Let $d_{\mathbf{K}} := \max\{\lceil d_Y/2 \rceil, d_S\}$.

Theorem 3.7. *Suppose that* $\mathbf{K}$ *is compact. For integers* $r \geq \lceil d_X/2 \rceil$ *and* $t \geq d_{\mathbf{K}}$, *define*

$$\Lambda_{r,t} := \left\{ x \in \mathbb{R}^m : \left\{ \begin{array}{l} \exists z = (z_\alpha)_{\alpha \in \mathbb{N}^m_{2r}} \in \mathbb{R}^{\mathbb{N}^m_{2r}}, \sigma, \sigma_j \in \Sigma^2[Y], j = 1, \ldots, s, \\ \text{s.t.} \quad z_0 = 1, \ M_r(z) \succeq 0, \\ \mathscr{L}_z(X_i) = x_i, \ i = 1, \ldots, m, \quad \mathscr{L}_z(\Theta_k) \leq 1, \ k = \lceil d_X/2 \rceil, \ldots, r, \\ \displaystyle\sum_\alpha p_{X,\alpha}(Y) \mathscr{L}_z(X^\alpha) = \sigma + \sum_{j=1}^s \sigma_j g_j, \ \deg(\sigma), \ \deg(\sigma_j g_j) \leq 2t. \end{array} \right. \right\} \tag{12}$$

Then, $\Lambda_{r_2,t} \subseteq \Lambda_{r_1,t}$ *for any* $r_2 > r_1 \geq \lceil d_X/2 \rceil$ *and* $\Lambda_{r,t_2} \supseteq \Lambda_{r,t_1}$ *for any* $t_2 > t_1 \geq d_{\mathbf{K}}$. *If Assumption 3.6 holds, then the followings are true.*

(i) *For any* $\varepsilon > 0$, *there exists an integer* $r(\varepsilon) \geq \lceil d_X/2 \rceil$ *such that for every* $r \geq r(\varepsilon)$ *and* $t \geq d_{\mathbf{K}}$, *it holds that* $\Lambda_{r,t} \subseteq \mathbf{K} + \varepsilon\mathbf{B}$. *If* $\mathcal{Q}(G)$ *is Archimedean, then there exists integer* $t(\varepsilon) \geq d_{\mathbf{K}}$ *such that for every* $r \geq \lceil d_X/2 \rceil$ *and* $t \geq t(\varepsilon)$, *it holds that* $\mathbf{K} \subseteq \Lambda_{r,t} + \varepsilon\mathbf{B}$. *Consequently,* $\Lambda_{r,t}$ *converges to* $\mathbf{K}$ *as* r *and* t *both tend to* ∞;

(ii) *If the Lagrangian* $L_f(X)$ *as defined in (5) is s.o.s for every linear* $f \in \mathbb{R}[X]$, *then* $\mathbf{K} \supseteq \Lambda_{r,t_2} \supseteq \Lambda_{r,t_1}$ *for any* $r \geq \lceil d_X/2 \rceil$, $t_2 > t_1 \geq d_{\mathbf{K}}$. *For any* $\varepsilon > 0$, *if, moreover,* $\mathcal{Q}(G)$ *is Archimedean, then there exists integer* $t(\varepsilon) \geq d_{\mathbf{K}}$ *such that* $\mathbf{K} \subseteq \Lambda_{r,t} + \varepsilon\mathbf{B}$ *for any* $r \geq \lceil d_X/2 \rceil$, $t \geq t(\varepsilon)$. *Consequently,* $\Lambda_{r,t}$ *converges to* $\mathbf{K}$ *as* t *tends to* ∞ *for any* $r \geq \lceil d_X/2 \rceil$.

Proof. For a fixed $x \in \Lambda_{r_2,t}$, there exist $z = (z_\alpha)_{\alpha\in\mathbb{N}^m_{2r_2}} \in \mathbb{R}^{\mathbb{N}^m_{2r_2}}, \sigma, \sigma_j \in \Sigma^2[Y]$ satisfying conditions in (12) for $\Lambda_{r_2,t}$. Let $z' = (z_\alpha)_{\alpha\in\mathbb{N}^m_{2r_1}}$ be the truncation of z. Then, it is clear that z', σ, σ_j satisfy all conditions in (12) for $\Lambda_{r_1,t}$ and thus $x \in \Lambda_{r_1,t}$. Similarly, if $x \in \Lambda_{r,t_1}$, then $x \in \Lambda_{r,t_2}$ for any $t_2 > t_1 \geq d_{\mathbf{K}}$.

(i). Fix an $\varepsilon > 0$ and a point $v \notin \mathbf{K} + \varepsilon\mathbf{B}$. Now we prove that there is some integer $r(\varepsilon)$ that does not depend on v such that $v \notin \Lambda_{r,t}$ for every $r \geq r(\varepsilon)$ and $t \geq d_{\mathbf{K}}$, which implies that $\Lambda_{r,t} \subseteq \mathbf{K} + \varepsilon\mathbf{B}$. By (Lasserre, 2009b, Lemma 5), there exist $a \in \mathbb{R}^m$ and $b = \min_{x\in\mathbf{K}} a^T x$ satisfying $\|a\|_2 = 1$ and $|b| \leq \tau_{\mathbf{K}}$ such that $a^T x - b \geq 0$ for any $x \in \mathbf{K}$ and $a^T v - b < -\varepsilon$. Consider the optimization problem $\min_{x\in\mathbf{K}} a^T x - b$. By Corollary 2.2, the associated Lagrangian $L_{a,b}(X) := a^T X - b - \sum_{j=1}^{l} \lambda_j p(X, y_l)$ as defined in (5) is nonnegative on $\mathbb{R}^m$ for some $y_1, \ldots, y_l \in S$ and nonnegative $\lambda_1, \ldots, \lambda_l \in \mathbb{R}$. In particular, $L_{a,b}$ is nonnegative on $[-\tau_{\mathbf{K}}, \tau_{\mathbf{K}}]^m$. By Theorem 2.3 (iii), there is some integer $r(\varepsilon) \geq \lceil d_X/2 \rceil$ such that for any $r \geq r(\varepsilon)$, it holds that

$$a^T X - b + \frac{\varepsilon}{2}(1 + \Theta_r) = \tilde{\sigma} + \sum_{j=1}^{l} \lambda_j p(X, y_j) \tag{13}$$

for some $\tilde{\sigma} \in \Sigma[X]^2$. As $r \geq r(\varepsilon) \geq \lceil d_X/2 \rceil$, we have $\deg(\tilde{\sigma}) \leq 2r$. Now we show that $r(\varepsilon)$ does not depend on v. According to (Lasserre and Netzer, 2007, Sec. 3.3), $r(\varepsilon)$ depends on ε, the dimension m and the size of a, b, λ_j's and the coefficients $p(X, y_j)$ regarded as polynomials in $\mathbb{R}[X]$. Fix a Slater point $u_0 \in \mathbf{K}$, since $a^T u_0 - b - \sum_{j=1}^{l} \lambda_j p(u_0, y_j) \geq 0$, as proved in (Lasserre, 2009b, Lemma 7), we have

$$0 \leq \lambda_j \leq \frac{a^T u_0 - b}{p(u_0, y_j)} \leq \frac{2\tau_{\mathbf{K}}}{p(u_0, y_j)} \leq \frac{2\tau_{\mathbf{K}}}{\min_{j=1,\ldots,l} p(u_0, y_j)} \leq \frac{2\tau_{\mathbf{K}}}{\min_{y\in S} p(u_0, y)} \leq \frac{2\tau_{\mathbf{K}}}{p^*_{u_0}},$$

where $p^*_{u_0} := \min_{y\in S} p(u_0, y) > 0$ since u_0 is a Slater point and S is compact. Write $p(X, y_j) = \sum_\alpha p_{X,\alpha}(y_j) X^\alpha$, then $p_{X,\alpha}(y_j) \leq \max_\alpha \max_{y\in S} p_{X,\alpha}(y)$. Hence, all a, b, λ_j's and $p_{X,\alpha}(y_j)$'s are uniformly bounded, which means that $r(\varepsilon)$ does not depend on v. For any $r \geq r(\varepsilon)$ and $t \geq d_{\mathbf{K}}$, to the contrary, assume that $v \in \Lambda_{r,t}$. Then, there exist z, σ, σ_j's satisfying the conditions in (12) for $\Lambda_{r,t}$. Let $\mu = \sum_{j=1}^{l} \lambda_j \delta_{y_l}$ where δ_{y_l} denotes the Dirac measure at y_l. As $\deg(\tilde{\sigma}) \leq 2r$, it holds that

$$\begin{aligned}
0 > a^T v - b + \varepsilon &= \mathscr{L}_z(a^T X - b) + \varepsilon \\
&\geq \mathscr{L}_z(a^T X - b) + \frac{\varepsilon}{2}\mathscr{L}_z(1 + \Theta_r) \\
&= \mathscr{L}_z\left(\tilde{\sigma} + \int_S p(X, y) d\mu(y)\right) \\
&= \mathscr{L}_z(\tilde{\sigma}) + \int_S \sum_\alpha p_{X,\alpha}(y) \mathscr{L}_z(X^\alpha) d\mu(y) \\
&= \mathscr{L}_z(\tilde{\sigma}) + \int_S \left(\sigma + \sum_{j=1}^{s} \sigma_j g_j\right) d\mu(y) \geq 0,
\end{aligned} \tag{14}$$

which is a contradiction. Thus, $v \notin \Lambda_{r,t}$ and $\Lambda_{r,t} \subseteq \mathbf{K} + \varepsilon\mathbf{B}$.

Fix a Slater point $u_0 \in \mathbf{K}$. Let $u \in \mathbf{K}$ be arbitrary. Now we first prove that there exist a point $\bar{u} \in \mathbb{R}^m$ and an integer $t(\varepsilon)$ that does not depend on u (in fact, it depends on $\varepsilon, \mathbf{K}, S, u_0, p(X, Y), g_j$'s) such that $\|u - \bar{u}\|_2 \leq \varepsilon$ and $\bar{u} \in \Lambda_{r,t}$ for every $r \geq \lceil d_X/2 \rceil$ and $t \geq t(\varepsilon)$, which implies that $\mathbf{K} \subseteq \Lambda_{r,t} + \varepsilon\mathbf{B}$. If $\|u - u_0\|_2 \leq \varepsilon$, then let $\bar{u} = u_0$; otherwise, let $\lambda = \varepsilon/\|u_0 - u\|_2$ and $\bar{u} = \lambda u_0 + (1 - \lambda)u$, then we have $1 > \lambda \geq \frac{\varepsilon}{2\tau_{\mathbf{K}}}$, $\|u - \bar{u}\|_2 = \lambda\|u_0 - u\|_2 = \varepsilon$ and

$$\begin{aligned}
p(\bar{u}, y) &\geq \lambda p(u_0, y) + (1 - \lambda)p(u, y) \quad &&[\text{as } -p(X, y) \text{ is convex in } X] \\
&\geq \lambda p(u_0, y). &&[\text{as } u \in \mathbf{K}]
\end{aligned}$$

Let $\kappa(\varepsilon) := \min\{\frac{\varepsilon}{2\tau_{\mathbf{K}}}, 1\}$. Then, in either case, it follows that

$$p(\bar{u}, y) \geq \kappa(\varepsilon) p(u_0, y) \geq \kappa(\varepsilon) p^*_{u_0} > 0$$

for any $y \in S$. Write $p(\bar{u}, Y) = \sum_\beta p_{Y,\beta}(\bar{u})Y^\beta \in \mathbb{R}[Y]$. Recall the norm defined in (6), then

$$\|p(\bar{u}, Y)\| = \max_\beta \frac{|p_{Y,\beta}(\bar{u})|}{\binom{\|\beta\|_1}{\beta}} \leq \max_\beta \frac{\max_{x\in\mathbf{K}} |p_{Y,\beta}(x)|}{\binom{\|\beta\|_1}{\beta}} =: N_p.$$

As $\mathbf{K}$ is compact, N_p is well-defined. Note that N_p does not depend on u but only on p and $\mathbf{K}$. By Theorem 2.6, there exists come positive c depending on g_j's such that $p(\bar{u}, Y) \in \mathcal{Q}_t(G)$ whenever

$$t \geq c \exp\left[\left(d_Y^2 n^{d_Y} \frac{N_p \tau_S^{d_Y}}{\kappa(\varepsilon) p_{u_0}^*}\right)^c\right] =: t(\varepsilon).$$

For any $r \geq \lceil d_X/2 \rceil$, set Let $\zeta_{2r,\bar{u}}$ be the Zeta vector of $\bar{u}$ up to degree $2r$. Then, it is clear that $\mathscr{L}_{\zeta_{2r,\bar{u}}}(X_i) = \bar{u}_i$ for $i = 1, \ldots, m$, $\mathscr{L}_{\zeta_{2r,\bar{u}}}(\Theta_k) \leq 1$ for $k = \lceil d_X/2 \rceil, \ldots, r$ and $M_r(\zeta_{2r,\bar{u}}) \succeq 0$. We have $\sum_\alpha p_\alpha(Y)\mathscr{L}_{\zeta_{2r,\bar{u}}}(X^\alpha) = p(\bar{u}, Y)$. It implies that $\bar{u} \in \Lambda_{r,t}$ and thus $\mathbf{K} \subseteq \Lambda_{r,t} + \varepsilon\mathbf{B}$ for every $r \geq \lceil d_X/2 \rceil$ and $t \geq t(\varepsilon)$.

(ii). By (i), we only need to prove $\Lambda_{r,t} \subseteq \mathbf{K}$ for any $r \geq \lceil d_X/2 \rceil$ and $t \geq d_\mathbf{K}$. Fix a point $v \notin \mathbf{K}$. By the Separation Theorem of convex sets, there exist $a \in \mathbb{R}^m$ and $b \in \mathbb{R}$ such that $a^T x - b \geq 0$ for any $x \in \mathbf{K}$ and $a^T v - b < 0$. As proved in (i), there are some $y_1, \ldots, y_l \in S$ and nonnegative $\lambda_1, \ldots, \lambda_l \in \mathbb{R}$ such that $a^T X - b - \sum_{j=1}^l \lambda_j p(X, y_l)$ is nonnagetive on $\mathbb{R}^m$. Since the associated Lagrangian $L_f(X)$ is s.o.s for every linear function f, we have

$$a^T X - b = \tilde{\sigma} + \sum_{j=1}^l \lambda_j p(X, y_j) \tag{15}$$

for some $\tilde{\sigma} \in \Sigma[X]^2$. To the contrary, assume that $v \in \Lambda_{r,t}$. Then, there exist z, σ, σ_j's satisfying the conditions in (12) for $\Lambda_{r,t}$. Define μ as in (i). Like in (14), we get that

$$0 > a^T v - b = \mathscr{L}_z(a^T X - b) = \mathscr{L}_z(\tilde{\sigma}) + \int_S \left(\sigma + \sum_{j=1}^s \sigma_j g_j\right) d\mu(y) \geq 0, \tag{16}$$

which is a contradiction. Thus, $v \notin \Lambda_{r,t}$ and hence $\Lambda_{r,t} \subseteq \mathbf{K}$. □

Remark 3.8. According to the proof, the conclusions (i) and (ii) in Theorem 3.7 are still true if we simplify the condtion $\mathscr{L}_z(\Theta_k) \leq 1,\ k = \lceil d_X/2 \rceil, \ldots, r$ in (12) by $\mathscr{L}_z(\Theta_r) \leq 1$.

According to the proof of Theorem 3.7 (i), the equation

$$\sum_\alpha p_{X,\alpha}(Y)\mathscr{L}_z(X^\alpha) = \sigma + \sum_{j=1}^s \sigma_j g_j \tag{17}$$

in the definition of $\Lambda_{r,t}$ in (12) can be replaced by other representations for positive (nonnegative) polynomials if certain assumptions hold. For instance, if S is compact but $\mathcal{Q}(G)$ is not Archimedean, we can use Schmüdgen's Positivstellensatz Schmüdgen (1991) in the definition of $\Lambda_{r,t}$ to obtain the same results as in Theorem 3.7. Now we consider the case when $m = 1$ and S is a bounded interval. By some representation results of *nonnegative* polynomials in the univariate case, we shall see that analogous approximate semidefinite representations of $\mathbf{K}$ as in Theorem 3.7 can be obtained with some *fixed* order t. Without loss of generality, we can assume that $S = [-1, 1]$. Let

$$[-1, 1] = \{y_1 \in \mathbb{R} \mid g_1(y_1) \geq 0\}, \quad \text{where} \quad g_1(Y_1) = 1 - Y_1^2. \tag{18}$$

Recall the well-known result

Theorem 3.9. (c.f. Powers and Reznick (2000); Laurent (2009)) *Let $h \in \mathbb{R}[Y_1]$ and $h \geq 0$ on $[-1, 1]$, then $h = \sigma + \sigma_1(1 - Y_1^2)$ where $\sigma, \sigma_1 \in \Sigma^2[Y_1]$ and* $\deg(\sigma)$, $\deg(\sigma_1(1 - Y_1^2)) \leq 2\lceil \deg(h)/2 \rceil$.

Theorem 3.10. *Assume that S is in the case of (18) and $\mathbf{K}$ is compact. Let $t_0 = d_{\mathbf{K}}$ and consider the sets Λ_{r,t_0} in (12) for $r \geq \lceil d_X/2 \rceil$. Then, $\mathbf{K} \subseteq \Lambda_{r_2,t_0} \subseteq \Lambda_{r_1,t_0}$ for any $r_2 > r_1 \geq \lceil d_X/2 \rceil$. Suppose that Assumption 3.6 holds, then the followings are true.*

(i) *For any $\varepsilon > 0$, there exists an integer $r(\varepsilon) \geq \lceil d_X/2 \rceil$ such that $\Lambda_{r,t_0} \subseteq \mathbf{K} + \varepsilon \mathbf{B}$ holds for every $r \geq r(\varepsilon)$. Consequently, Λ_{r,t_0} converges to $\mathbf{K}$ as r tends to ∞;*

(ii) *If the Lagrangian $L_f(X)$ as defined in (5) is s.o.s for every linear $f \in \mathbb{R}[X]$. then $\mathbf{K} = \Lambda_{r,t_0}$ for any $r \geq \lceil d_X/2 \rceil$.*

Proof. For any $u \in \mathbf{K}$, let $\zeta_{2r,u}$ be the Zeta vector of u of degree up to $2r$. By Theorem 3.9, there exists $\sigma, \sigma_1 \in \Sigma^2[Y_1]$ such that $\zeta_{2r,u}, \sigma, \sigma_1$ satisfy the conditions in the definition of Λ_{r,t_0} in (12). Hence, $\mathbf{K} \subseteq \Lambda_{r,t_0}$ for any $r \geq \lceil d_X/2 \rceil$.

(i) See the first part of the proof of Theorem 3.7 (i);

(ii) It is clear since $\Lambda_{r,t} \subseteq \mathbf{K}$ for any $r \geq \lceil d_X/2 \rceil$ and $t \geq d_{\mathbf{K}}$ by the proof of Theorem 3.7 (ii). □

Note that $\Lambda_{r,t}$ in (12) is indeed a semidefinite representation set of the form (3) for every $r \geq \lceil d_X/2 \rceil$ and $t \geq d_{\mathbf{K}}$. In fact, for any $t \geq d_{\mathbf{K}}$, let $m_t(Y)$ be the column vector consisting of all the monomials in Y of degree up to t. Let $s(t) = \binom{n+t}{n}$ which is the dimension of $m_t(Y)$. Recall the definitions in (9). There exist positive semidefinite matrices $Z \in \mathbb{R}^{s(t)\times s(t)}$, $Z_j \in \mathbb{R}^{s(t-d_j)\times s(t-d_j)}$, $j = 1, \ldots, s$, such that

$$\sigma(Y) = m_t(Y)^T \cdot Z \cdot m_t(Y), \quad \sigma_j(Y) = m_{t-d_j}(Y)^T \cdot Z_j \cdot m_{t-d_j}(Y), \; j = 1, \ldots, s.$$

For each $\beta \in \mathbb{N}^n_{2t}$, we can find symmetric matrices $C_\beta, C_{j,\beta} \in \mathbb{R}^{s(t-d_j)\times s(t-d_j)}$, $j = 1, \ldots, s$, such that the coefficients of Y^β in σ and $\sigma_j g_j$ are equal to $\langle Z, C_\beta \rangle$ and $\langle Z_j, C_{j,\beta} \rangle$, $j = 0, 1, \ldots, s$, respectively. Write $p(X,Y) = \sum_{\beta \in \mathbb{N}^n_{2t}} p_{Y,\beta}(X) Y^\beta$ and let

$$E_\beta = \mathscr{L}_z(p_{Y,\beta}(X)) - \langle Z, C_\beta \rangle - \sum_{j=1}^{s} \langle Z_j, C_{j,\beta} \rangle$$

for each $\beta \in \mathbb{N}^n_{2t}$. Then, All E_β are linear in z, Z and Z_j's. For each $i = 1, \ldots, m$, let e_i be the vector whose i-th component is 1 and the others are 0. Denote by $\mathcal{M}(z, Z, Z_1, \ldots, Z_s)$ the block diagonal matrix whose diagonal elements are

$$z_0 - 1, \; 1 - z_0, \; M_r(z), \; Z, \; Z_1, \ldots, Z_s, \; 1 - \mathscr{L}_z(\Theta_k), \; k = \lceil d_X/2 \rceil, \ldots, r, \; E_\beta, \; -E_\beta, \; \beta \in \mathbb{N}^m_{2t}.$$

Then, we have the semidefinite representation

$$\Lambda_{r,t} = \{(z_{e_1}, \ldots, z_{e_n}) \in \mathbb{R}^m \mid \mathcal{M}(z, Z, Z_1, \ldots, Z_s) \succeq 0\}.$$

Note that the matrix $\mathcal{M}(z, Z, Z_1, \ldots, Z_j)$ can be easily generated using Yalmip (Löfberg (2004)). For $m = 2$ and 3, we can first generate $\mathcal{M}(z, Z, Z_1, \ldots, Z_j)$ and then use the software package Bermeja Rostalski (2010) to draw the projected spectrahedron $\Lambda_{r,t}$.

Recall Theorem 3.7 (ii). We now strengthen Assumption 3.6 to

Assumption 3.11. *The set S is compact, $-p(X,y) \in \mathbb{R}[X]$ is s.o.s-convex for any $y \in S$ and the Slater condition holds for $\mathbf{K}$.*

Lemma 3.12. *Suppose that Assumption 3.11 holds for $\mathbf{K}$, then for any s.o.s-convex $f \in \mathbb{R}[X]$, the Lagrangian $L_f(X)$ defined in (5) is s.o.s.*

Proof. A Frank-Wolfe type theorem proved in Belousov (1977) states that the discretization problem (4) has a minimizer $u \in \mathbb{R}^m$ even when the feasible set of (4) is noncompact. Since $L_f(X)$ is s.o.s-convex and

$L_f(x) \geq 0 = L_f(u)$ for any $x \in \mathbb{R}^m$, the conclusion follows from KKT optimality conditions for (4) and Lemma 2.12. □

Consequently, if Assumption 3.11 holds, then $L_f(X)$ is s.o.s for every linear function f. In this case, instead of the sets $\Lambda_{r,t}$ in (12), we can define

$$\Lambda_t := \left\{ x \in \mathbb{R}^m : \begin{cases} \exists z = (z_\alpha)_{\alpha \in \mathbb{N}^m_{d_X}} \in \mathbb{R}^{\mathbb{N}^m_{d_X}}, \sigma, \sigma_j \in \Sigma^2[Y], j = 1, \ldots, s, \\ \text{s.t.} \quad z_0 = 1,\ M_{\lfloor d_X/2 \rfloor}(z) \succeq 0,\ \mathscr{L}_z(X_i) = x_i,\ i = 1, \ldots, m, \\ \sum_\alpha p_{X,\alpha}(Y) \mathscr{L}_z(X^\alpha) = \sigma + \sum_{j=1}^s \sigma_j g_j,\ \deg(\sigma),\ \deg(\sigma_j g_j) \leq 2t. \end{cases} \right\} \tag{19}$$

and get the same results as in Theorem 3.7. That is,

Theorem 3.13. *Suppose that Assumption 3.11 holds. Then for Λ_t defined in (19), the followings are true.*

(i) $\mathbf{K} \supseteq \Lambda_{t_1} \supseteq \Lambda_{t_2}$ *for any* $t_1 > t_2 \geq d_{\mathbf{K}}$*;*

(ii) *If* $\mathbf{K}$ *is compact and* $\mathcal{Q}(G)$ *is Archimedean, then for any* $\varepsilon > 0$*, there exists integer* $t(\varepsilon) \geq d_{\mathbf{K}}$ *such that for every* $t \geq t(\varepsilon)$*, it holds that* $\mathbf{K} \subseteq \Lambda_t + \varepsilon\mathbf{B}$*. Consequently,* Λ_t *converges to* $\mathbf{K}$ *as* t *tends to* ∞*.*

(iii) *If* S *is in the case of (18), then* $\mathbf{K} = \Lambda_{t_0}$ *where* $t_0 = d_{\mathbf{K}}$*.*

Proof. (i) Recall the proof of Theorem 3.7 (ii). Note that to show $\Lambda_{r,t} \subseteq \mathbf{K}$ for any $r \geq \lceil d_X/2 \rceil$ and $t \geq d_{\mathbf{K}}$, the constraints $\mathscr{L}_z(\Theta_k) \leq 1$ in definition (12) are redundant. Moreover, it is clear that $\tilde{\sigma} \in \Sigma[X]^2$ in (15) is of degree $\leq 2\lfloor d_X/2 \rfloor$. Hence, we can set $r = \lfloor d_X/2 \rfloor$ and define Λ_t as in (19) to obtain (i);

(ii) See the proof of Theorem 3.7 (ii);

(iii) Similar to the proof of Theorem 3.10, it is clear that $\mathbf{K} \subseteq \Lambda_{t_0}$. Combining (i), $\mathbf{K} = \Lambda_{t_0}$ follows. □

Corollary 3.14. *Assume that the set S is compact, $p(X, y) \in \mathbb{R}[X]$ is linear in X for any $y \in S$ and the Slater condition holds for $\mathbf{K}$. For integer $t \geq d_{\mathbf{K}}$, define*

$$\Lambda_t := \left\{ x \in \mathbb{R}^m : \begin{cases} \exists \sigma, \sigma_j \in \Sigma^2[Y], j = 1, \ldots, s, \\ \text{s.t. } p(x, Y) = \sigma + \sum_{j=1}^s \sigma_j g_j,\ \deg(\sigma),\ \deg(\sigma_j g_j) \leq 2t. \end{cases} \right\}. \tag{20}$$

Then, the statements in Theorem 3.13 hold.

Proof. Clearly, Assumption 3.11 holds. Since $p(X, y) \in \mathbb{R}[X]$ is linear in X for any $y \in S$, it is easy to see that the sets Λ_t defined in (19) and (20) are equal in this case. □

Remark 3.15. Note that we do not require $\mathbf{K}$ to be compact in Theorem 3.13 (i),(iii) and Corollary 3.14.

3.3. Illustrating examples

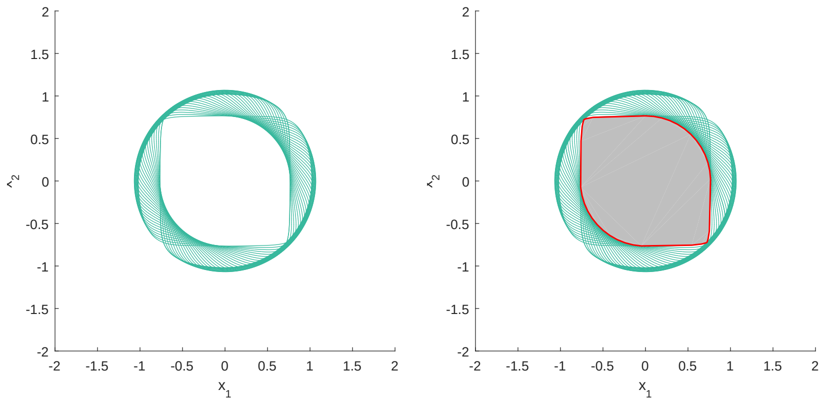
Now we present some illustrating examples. As we shall see, the approximate semidefinite representations defined in this section are very tight for some given sets $\mathbf{K}$.

Example 3.16. *Consider the polynomial*

$$\begin{aligned} f(X_1, X_2, X_3) =& 32X_1^8 + 118X_1^6X_2^2 + 40X_1^6X_3^2 + 25X_1^4X_2^4 - 43X_1^4X_2^2X_3^2 - 35X_1^4X_3^4 + 3X_1^2X_2^4X_3^2 \\ & - 16X_1^2X_2^2X_3^4 + 24X_1^2X_3^6 + 16X_2^8 + 44X_2^6X_3^2 + 70X_2^4X_3^4 + 60X_2^2X_3^6 + 30X_3^8. \end{aligned}$$

It is proved in Ahmadi and Parrilo (2012) that $f(X_1, X_2, 1) \in \mathbb{R}[X_1, X_2]$ is a convex but not s.o.s-convex. Rotate the shape in the (x_1, x_2)-plane defined by $f(x_1, x_2, 1) \leq 100$ continuously around the origin by 90°

Figure 1: The set $\mathbf{K}$ (left) and the semidefinite representation set $\Lambda_{4,4}$ (right) in Example 3.16.



clockwise. Denote by $\mathbf{K}$ *the common area of these shapes in this process. We illustrate* $\mathbf{K}$ *in the left of Figure 1 by making a discrete rotation. In other words, the set* $\mathbf{K}$ *is defined by*

$$\mathbf{K} = \{(x_1, x_2) \in \mathbb{R}^2 \mid p(x_1, x_2, y_1, y_2) \geq 0, \quad \forall y \in S\},$$

where $p(X_1, X_2, Y_1, Y_2) = 100 - f(Y_1X_1 - Y_2X_2, Y_2X_1 + Y_1X_2, 1)$ *and*

$$S = \{(y_1, y_2) \in \mathbb{R}^2 \mid y_1 \geq 0, \ y_2 \geq 0, \ y_1^2 + y_2^2 = 1\}.$$

It is clear that the assumptions in Theorem 3.7 holds for $\mathbf{K}$ *and* $d_X = d_Y = 8$, $d_{\mathbf{K}} = 4$. *By the software* Bermeja, *the semidefinite representation set* $\Lambda_{4,4}$ *as defined in (12) is drawn in gray bounded by the red curve in the right of Figure 1.*

Example 3.17. *Consider the set*

$$\mathbf{K} = \{(x_1, x_2) \in \mathbb{R}^2 \mid p(x_1, x_2, y_1, y_2) \geq 0, \quad \forall y \in S\}$$

where $p(X_1, X_2, Y_1, Y_2) = -X_1^2 - 2Y_2X_1X_2 - Y_1X_2^2 - X_1 - X_2$ *and*

$$S = \{(y_1, y_2) \in \mathbb{R}^2 \mid 1 - y_1 \geq 0, \ 1/2 \geq y_2 \geq -1/2, \ y_1 - y_2^2 \geq 0\}.$$

We illustrate $\mathbf{K}$ *in the left of Figure 2 by using some grid of* S. *The Hessian matrix of* p *with respect to* X_1 *and* X_2 *is*

$$H = -\begin{bmatrix} 2 & 2Y_2 \\ 2Y_2 & 2Y_1 \end{bmatrix} \qquad \text{with} \quad \det(H) = 4Y_2^2 - 4Y_1.$$

Clearly, $-p(X_1, X_2, y_1, y_2)$ *is s.o.s-convex in* (X_1, X_2) *for every* $y \in S$. *We have* $d_X = 2, d_Y = 1$ *and* $d_{\mathbf{K}} = 1$. *The semidefinite representation set* Λ_1 *as defined in (19) is drawn in gray bounded by the red curve in the right of Figure 2.*

Example 3.18. *Consider the ellipse*

$$\mathbf{K} = \{(x_1, x_2) \in \mathbb{R}^2 \mid 2x_1^2 + x_2^2 + 2x_1x_2 + 2x_1 \leq 0\}$$

which can be represented by

$$\{(x_1, x_2) \in \mathbb{R}^2 \mid p(x_1, x_2, y) \geq 0, \ \forall y \in S\}$$

where

$$p(X_1, X_2, Y) = (-Y^4 - 2Y^3 + 3Y^2 + 2Y - 1)X_1 - 2Y(Y^2 - 1)X_2 + 2Y^2$$

and $S = [-1, 1]$ *(See Goberna and López (1998)). As* $p(X_1, X_2, Y)$ *is linear in* X *and* S *is an interval, we have* $\mathbf{K} = \Lambda_2$ *where* Λ_2 *is defined as in (20) by Corollary 3.14. The set* $\mathbf{K}$ *and semidefinite representation set* Λ_2 *are illustrated in Figure 3.*

Figure 2: The set $\mathbf{K}$ (left) and the semidefinite representation set Λ_1 (right) in Example 3.17.

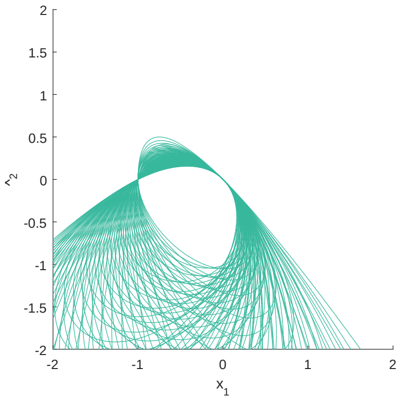
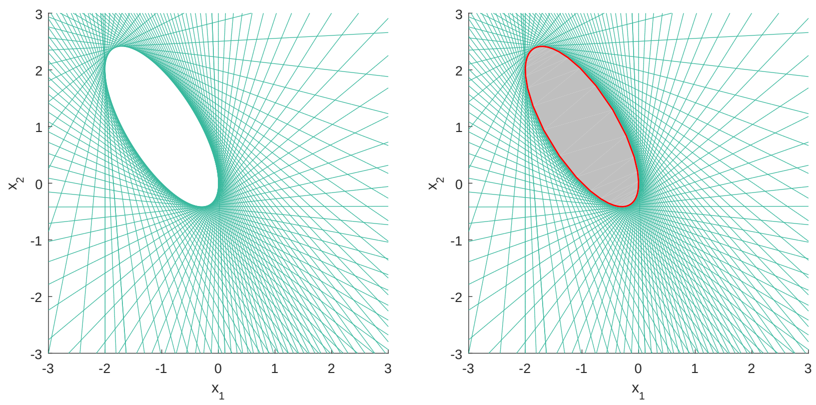


Figure 3: The set $\mathbf{K}$ (left) and the semidefinite representation set Λ_2 (right) in Example 3.18.



4. SDP relaxations of convex semi-infinite polynomial programming

For a convex polynomial $f(X) \in \mathbb{R}[X]$, consider the following convex semi-infinite polynomial programming problem

$$(\mathbf{P}) \qquad f^* := \inf_{x \in \mathbf{K}} \ f(x) \quad \text{where } \mathbf{K} \text{ is defined in (1).}$$

Let $d_P := \max\{\deg(f), d_X\}$ and $\mathscr{M}(S)$ be the set of all (nonnagetive) Borel measures supported on S.

4.1. General case

Consider the case when $\mathbf{K}$ is compact and Assumption 3.6 holds. Recall the Riesz funciton defined in (7). For any integer $r \geq d_P/2$, we first convert $(\mathbf{P})$ to the problem

$$(\mathbf{P}_r) \qquad \begin{cases} f_r^* := \sup\limits_{\rho,\eta,\mu,\sigma} \ \rho - 2\eta \\ \qquad\qquad \text{s.t. } f(X) - \rho + \eta(1+\Theta_r) = \displaystyle\int_S p(X,y)d\mu(y) + \sigma, \\ \qquad\qquad\quad \rho \in \mathbb{R}, \ \eta \geq 0, \ \mu \in \mathscr{M}(S), \ \sigma \in \Sigma^2[X] \cap \mathbb{R}[X]_{2r}, \end{cases} \tag{21}$$

and its dual

$$(\mathbf{P}_r^*) \qquad \begin{cases} \inf\limits_{z} \ \mathscr{L}_z(f) \\ \text{s.t. } z_0 = 1, \ \mathscr{L}_z(\Theta_r) \leq 1, \ M_r(z) \succeq 0, \\ \qquad \displaystyle\sum_{\alpha} p_{X,\alpha}(y)\mathscr{L}_z(X^\alpha) \geq 0, \ \forall y \in S, \\ \qquad z = (z_\alpha)_{\alpha \in \mathbb{N}_{2r}^m} \in \mathbb{R}^{\mathbb{N}_{2r}^m}. \end{cases} \tag{22}$$

Definition 4.1. *We call $z^{(r)}(2r \geq d_P)$ a nearly optimal solution of (22) if $z^{(r)}$ is feasible for (22) and $\lim_{r\to\infty} \mathscr{L}_{z^{(r)}}(f) = \lim_{r\to\infty} \inf \mathbf{P}_r^*$.*

Theorem 4.2. *Suppose that $f(X)$ is convex, $\mathbf{K}$ is compact and Assumption 3.6 holds. Let $z^{(r)}$ be a nearly optimal solution of (22) and $\hat{z}^{(r)} = \{z_\alpha^{(r)} \mid \|\alpha\|_1 = 1\}$.*

- (i) *f_r^* converges to f^* as r tends to ∞;*
- (ii) *f_r^* is attainable in (21) and there is no dual gap between (21) and (22);*
- (iii) *Assume that $\tau_{\mathbf{K}} = 1$ (possibly after scaling). Then, for any convergent subsequence $\{\hat{z}^{(r_i)}\}$ of $\{\hat{z}^{(r)}\}$, $\lim_{i\to\infty} \hat{z}^{(r_i)}$ is a minimizer of $(\mathbf{P})$. Consequently, if x^* is the unique minimizer of $(\mathbf{P})$, then $\lim_{r\to\infty} \hat{z}^{(r)} = x^*$;*
- (iv) *If moreover, the Lagrangian $L_f(X)$ as defined in (5) is s.o.s, then $f_r^* = f^*$ for any $r \geq \lceil d_P/2 \rceil$ and it is also attainable in (22).*

Proof. (i) For any $x \in \mathbf{K}$ and $y \in S$, we have $\Theta_r(x) \leq 1$ and $p(x,y) \geq 0$. Consequently, for any feasible point $(\rho, \eta, \mu, \sigma)$ of (21), it holds that

$$\begin{aligned} f(x) &= \rho - \eta(1+\Theta_r(x)) + \int_S p(x,y)d\mu(y) + \sigma(x) \\ &\geq \rho - \eta(1+\Theta_r(x)) \\ &\geq \rho - 2\eta, \end{aligned}$$

which implies that $f_r^* \leq f^*$.

Conversely, by Corollary 2.2, there exist some $y_1, \ldots, y_l \in S$ and nonnegative Lagrange multipliers $\lambda_1, \ldots, \lambda_l \in \mathbb{R}$ such that

$$f(x) - f^* - \sum_{j=1}^{l} \lambda_j p(x, y_l) = f(x) - f^* - \int_S p(x, y) d\mu(y) \geq 0, \quad \forall x \in \mathbb{R}^m, \tag{23}$$

where $\mu = \sum_{j=1}^{l} \lambda_j \delta_{y_l} \in \mathscr{M}(S)$ and δ_{y_l} is the Dirac measure at y_l. For any fixed $r \in \mathbb{N}$ with $2r \geq d_P$, by Theorem 2.3 (i), there exists a $\varepsilon_r^* \geq 0$ such that

$$f(X) - f^* - \int_S p(X, y) d\mu(y) + \eta(1 + \Theta_r) \in \Sigma^2[X] \cap \mathbb{R}[X]_{2r} \tag{24}$$

if and only if $\eta \geq \varepsilon_r^*$. It means that (21) is feasible and $f_r^* \geq f^* - 2\varepsilon_r^*$. Moreover, by Theorem 2.3 (ii), ε_r^* decreasingly converges to 0 as r tends to ∞. It then follows that f_r^* converges to f^* as r tends to ∞.

(ii) Fix a Slater point u of $\mathbf{K}$. Since S is compact, there exists a neighborhood $\mathcal{O}_u$ of u such that every point in $\mathcal{O}_u$ is a Slater point of $\mathbf{K}$. Let ν be the probability measure with uniform distribution in $\mathcal{O}_u$ and set $z = (z_\alpha)_{\alpha \in \mathbb{N}^m_{2r}}$ where $z_\alpha = \int X^\alpha d\nu$. It is easy to see that z is strictly admissible for (22). The conclusion follows due to the duality theorey in convex optimization.

(iii) For any $2r \geq d_P$, as $\tau_{\mathbf{K}} = 1$ and $\mathscr{L}_{z^{(r)}}(\Theta_r) \leq 1$, it is clear that $\mathscr{L}_{z^{(r)}}(X_i^{2r}) \leq 1$ for all $i = 1, \ldots, n$. Since $z_0^{(r)} = 1$ and $M_r(z^{(r)}) \succeq 0$, we then deduce that $|z_\alpha^{(r)}| \leq 1$ for any $|\alpha| \leq 2r$ by (Lasserre and Netzer, 2007, Lemma 4.1 and 4.3). Complete each $z^{(r)}$ with zeros to make it an infinite vector in $\mathbb{R}^{\mathbb{N}^m}$ indexed in the basis $\{X^\alpha \mid \alpha \in \mathbb{N}^m\}$. Then, it holds that $\{z^{(r)}\} \subseteq [-1, 1]^{\mathbb{N}^m}$.

Let $\{\hat{z}^{(r_i)}\}$ be a convergent subsequence of $\{\hat{z}^{(r)}\}$. By Tychonoff's theorem, there exists a convergent subsequence of the corresponding $\{z^{(r_i)}\}$ in the product topology. Without loss of generality, we assume that the whole sequence $\{z^{(r_i)}\}$ converges as $i \to \infty$. That is, there exists $z^* \in [-1, 1]^{\mathbb{N}^m}$ such that $\lim_{i\to\infty} z_\alpha^{(r_i)} = z_\alpha^*$ holds for all $\alpha \in \mathbb{N}^m$. From the pointwise convergence, we have $\mathscr{L}_{z^*}(h) \geq 0$ for all $h \in \Sigma^2[X]$. As $z^* \in [-1, 1]^{\mathbb{N}^m}$, by Theorem 2.7, z^* has exactly one representing measure ν with support contained in $[-1, 1]^{\mathbb{N}^m}$. Since $z^{(r)}$ is nearly optimal solution of (22), we obtain $\int f d\nu(x) = f^*$ by (i) and (ii). Denote

$$\hat{z}^* := \left(\int X_1 d\nu(x), \ldots, \int X_m d\nu(x) \right).$$

Then, $\lim_{i\to\infty} \hat{z}^{(r_i)} = \hat{z}^*$. For any $\varepsilon > 0$, from the proof of Theorem 3.7 (i) and Remark 3.8, it is easy to see that there exists an integer $r(\varepsilon)$ such that $\hat{z}^{(r_i)} \in \mathbf{K} + \varepsilon\mathbf{B}$ whenever $r_i \geq r(\varepsilon)$. By the pointwise convergence, we deduce that $\hat{z}^* \in \mathbf{K}$. Then, since f is convex, by Jensen's inequality, $f^* \leq f(\hat{z}^*) \leq \int f d\nu(x) = f^*$. Hence, $\hat{z}^*$ is indeed a minimizer of (22).

Assume that x^* is the unique minimizer of (22). We have shown that $\{\hat{z}^{(r)}\}$ is contained in $[-1, 1]^m$ and $\lim_{i\to\infty} \hat{z}^{(r_i)} = x^*$ for any convergent subsequence $\{\hat{z}^{(r_i)}\}$, therefore the whole sequence $\{\hat{z}^{(r)}\}$ converges to x^*.

(iv) Under the assumption, (24) holds with $\eta = 0$ and any $r \geq \lceil d_P/2 \rceil$. Hence, $f_r^* = f^*$ for any $r \geq \lceil d_P/2 \rceil$ by the proof of (i). As $\mathbf{K}$ is compact, suppose that f^* is attainable in (**P**) at a minimizer $x^* \in \mathbf{K}$. Let $z^* \in \mathbb{R}^{\mathbb{N}^m_{2r}}$ be the truncated moment sequence associated with the Dirac measure at x^*, then $f_r^* = f^*$ is attainable in (22) at z^*. □

Recall the definition of d_j in (9). For any $t \geq d_{\mathbf{K}}$, consider the SDP relaxation of (21)

$$\begin{cases} f_{r,t}^{\text{psdp}} := \sup\limits_{\rho,\eta,w,\sigma} \ \rho - 2\eta \\ \qquad\qquad \text{s.t. } f(X) - \rho + \eta(1 + \Theta_r) = \sum\limits_{\beta} p_{Y,\beta}(X) w_\beta + \sigma, \\ \qquad\qquad\quad M_t(w) \succeq 0, \ M_{t-d_j}(g_j w) \succeq 0, \ j = 1, \ldots, s, \\ \qquad\qquad\quad \rho \in \mathbb{R}, \ \eta \geq 0, \ w = (w_\beta)_{\beta \in \mathbb{N}^n_{2t}}, \ \sigma \in \Sigma^2[X] \cap \mathbb{R}[X]_{2r}. \end{cases} \tag{25}$$

Its dual is

$$\begin{cases} f_{r,t}^{\text{dsdp}} := \inf\limits_{z,\sigma,\sigma_j} \ \mathscr{L}_z(f) \\ \qquad\qquad \text{s.t. } z_0 = 1,\ \mathscr{L}_z(\Theta_r) \le 1,\ M_r(z) \succeq 0, \\ \qquad\qquad\quad \sum\limits_{\alpha} p_{X,\alpha}(Y)\mathscr{L}_z(X^\alpha) = \sigma + \sum\limits_{j=1}^{s} \sigma_j g_j,\ \sigma, \sigma_j \in \Sigma^2[Y], \\ \qquad\qquad\quad z = (z_\alpha)_{\alpha\in\mathbb{N}_{2r}^m} \in \mathbb{R}^{\mathbb{N}_{2r}^m},\ \deg(\sigma), \deg(\sigma_j g_j) \le 2t. \end{cases} \tag{26}$$

Theorem 4.3. *For any integer $r \ge d_P/2$, the followings are true.*

(i) *If $\mathcal{Q}(G)$ is Archimedean and the Slater condition holds for $\mathbf{K}$, then $f_{r,t}^{\text{psdp}}$ and $f_{r,t}^{\text{dsdp}}$ decreasingly converge to f_r^* as t tends to ∞;*

(ii) *For some order $t \ge d_{\mathbf{K}}$, if Rank Condition 2.9 holds for w^* in the solution $(\rho^*, \eta^*, w^*, \sigma^*)$ of (25), then $f_{r,t}^{\text{psdp}} = f_r^*$;*

(iii) *If S is in the case of (18), then $f_{r,t_0}^{\text{psdp}} = f_{r,t_0}^{\text{dsdp}} = f_r^*$ where $t_0 = d_{\mathbf{K}}$.*

Proof. (i) For any feasible point $(\rho, \eta, \mu, \sigma)$ of (21), let $w = (w_\beta)_{\beta\in\mathbb{N}_{2t}^n}$ where $w_\alpha = \int Y^\beta d\mu$, then (ρ, η, w, σ) is feasible for (25) and hence $f_{r,t}^{\text{psdp}} \ge f_r^*$ for any $t \ge d_{\mathbf{K}}$. Then by the weak duality and Theorem 4.2, we have $f_r^* \le f_{r,t}^{\text{psdp}} \le f_{r,t}^{\text{dsdp}}$ for any $t \ge d_{\mathbf{K}}$. It is sufficient to prove that $\lim_{t\to\infty} f_{r,t}^{\text{dsdp}} = f_r^*$.

Fixing an arbitrary $\varepsilon > 0$, we show that there is some $t \ge d_{\mathbf{K}}$ such that $0 \le f_{r,t}^{\text{dsdp}} - f_r^* \le \varepsilon$. Fix a Slater point u of $\mathbf{K}$ and let $z' = (z'_\alpha)_{\alpha\in\mathbb{N}_{2r}^m}$ where $z'_\alpha = u^\alpha$. Then z' is feasible for (26) for some $t' \ge d_{\mathbf{K}}$ by Putinar's Positivstellensatz. If $\sum_\alpha f_\alpha z'_\alpha - f_r^* \le \varepsilon$, then $0 \le f_{r,t'}^{\text{dsdp}} - f_r^* \le \varepsilon$. Next, we assume that $\sum_\alpha f_\alpha z'_\alpha - f_r^* > \varepsilon$. Then, we can choose another feasible point $\bar{z}$ of (22) such that $\sum_\alpha f_\alpha z'_\alpha - \sum_\alpha f_\alpha \bar{z}_\alpha > 0$ and $\sum_\alpha f_\alpha \bar{z}_\alpha - f_r^* \le \varepsilon/2$. Let

$$\delta := \frac{\varepsilon}{2\sum_\alpha f_\alpha(z'_\alpha - \bar{z}_\alpha)} \quad \text{and} \quad \hat{z} = (1-\delta)\bar{z} + \delta z'.$$

Then, we have $0 < \delta < 1$ and hence

$$\sum_\alpha p_{X,\alpha}(y)\hat{z}_\alpha = (1-\delta)\sum_\alpha p_{X,\alpha}(y)\bar{z}_\alpha + \delta \sum_\alpha p_{X,\alpha}(y) z'_\alpha > 0, \quad \forall y \in S.$$

Hence, $\hat{z}$ is feasible for (26) for some $\hat{t} \ge d_{\mathbf{K}}$ by Putinar's Positivstellensatz. We have

$$\begin{aligned} f_{r,\hat{t}}^{\text{dsdp}} - f_r^* &\le \sum_\alpha f_\alpha \hat{z}_\alpha - f_r^* \\ &= (1-\delta)\sum_\alpha f_\alpha \bar{z}_\alpha + \delta \sum_\alpha f_\alpha z'_\alpha - f_r^* \\ &= \sum_\alpha f_\alpha \bar{z}_\alpha - f_r^* + \delta \sum_\alpha f_\alpha (z'_\alpha - \bar{z}_\alpha) \\ &\le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

As ε is arbitrary, the conclusion follows.

(ii) Suppose that Rank Condition 2.9 holds for w^* in the solution $(\rho^*, \eta^*, w^*, \sigma^*)$ of (25) at some order $t \ge d_{\mathbf{K}}$. Then, by Theorem 2.10, w admits some measure $\mu^* \in \mathscr{M}(S)$, i.e., $w_\beta^* = \int Y^\beta d\mu^*$ for all $\beta \in \mathbb{N}_{2t}^n$. As $f_r^* \le f_{r,t}^{\text{psdp}}$ and $(\rho^*, \eta^*, \mu^*, \sigma^*)$ is feasible for (21), we conclude that $f_r^* = f_{r,t}^{\text{psdp}}$.

(iii) By the proof of (i), the conclusion follows due to Theorem 3.9 and Theorem 4.2 (ii). □

Corollary 4.4. *Suppose $f(X)$ is convex , $\mathbf{K}$ is compact and Assumption 3.6 holds. Then, for any $\varepsilon > 0$, the followings are true.*

(i) *There exists a* $r(\varepsilon)\in\mathbb{N}$ *such that* $f_{r,t}^{\rm dsdp}\ge f_{r,t}^{\rm psdp}\ge f^*-\varepsilon$ *holds for any* $r\ge r(\varepsilon)$ *and* $t\ge d_{\mathbf{K}}$*;*

(ii) *If* $\mathcal{Q}(G)$ *is Archimedean, then for any* $r\ge\lceil d_P/2\rceil$*, there exists a* $t(r,\varepsilon)\in\mathbb{N}$ *such that* $f_{r,t}^{\rm psdp}\le f_{r,t}^{\rm dsdp}\le f^*+\varepsilon$ *holds for any* $t\ge t(r,\varepsilon)$*;*

(iii) *If* S *is in the case of* (18)*, then* $\lim_{r\to\infty} f_{r,t_0}^{\rm psdp}=\lim_{r\to\infty} f_{r,t_0}^{\rm dsdp}=f^*$*, where* $t_0=d_{\mathbf{K}}$*.*

Proof. (i) It is clear that $f_r^*\le f_{r,t}^{\rm psdp}\le f_{r,t}^{\rm dsdp}$ holds for any $r\ge\lceil d_P/2\rceil$ and $t\ge d_{\mathbf{K}}$. By Theorem 4.2 (i), there exists a $r(\varepsilon)\in\mathbb{N}$ such that $f_r^*\ge f^*-\varepsilon$ holds for any $r\ge r(\varepsilon)$. Thus, (i) follows.

(ii) Due to Theorem 4.3 (i), for any $r\ge\lceil d_P/2\rceil$, there exists a $t(r,\varepsilon)\in\mathbb{N}$ such that $f_{r,t}^{\rm psdp}\le f_{r,t}^{\rm dsdp}\le f_r^*+\varepsilon$ holds for any $t\ge t(r,\varepsilon)$. Then (ii) follows since $f_r^*\le f^*$ for any $r\ge\lceil d_P/2\rceil$ by Theorem 4.2 (i).

(iii) It is clear by Theorem 4.2 (i) and Theorem 4.3 (iii). □

Remark 4.5. (1). Corollary 4.4 shows that we can approximate f^* by $f_{r,t}^{\rm psdp}$ and $f_{r,t}^{\rm dsdp}$ as closely as possible with r and t both large enough; (2). Assume that $\tau_{\mathbf{K}}=1$. By Theorem 4.3 (i), for any $2r\ge d_P$, there exists $t(r)\in\mathbb{N}$ such that $f_{r,t(r)}^{\rm dsdp}\le f_r^*+1/r$. Denote by $(z^{(r,t(r))},\sigma^{(r,t(r))},\sigma_j^{(r,t(r))})$ a minimizer of $f_{r,t(r)}^{\rm dsdp}$, then $\{z^{(r,t(r))}\}$ is a sequence of nearly optimal solutions of (22) and Theorem 4.2 (iii) holds for the corresponding truncated sequence $\{\hat{z}^{(r,t(r))}\}$. In particular, when $(\mathbf{P})$ has an unique minizer x^* and r,t are large enough, we can expect that the truncation $\hat{z}^{(r,t)}$ of any approximate solution $z^{(r,t)}$ of (26) lies in a small neighborhood of x^*.

4.2. S.O.S-Convex case

If Assumption 3.11 holds and $f(X)$ is s.o.s-convex, then the Lagrangian $L_f(X)$ as defined in (5) is s.o.s by Lemma 3.12 and now we can convert $(\mathbf{P})$ to

$$\begin{cases}\sup\limits_{\rho,\mu,\sigma} & \rho\\ \text{s.t.} & f(X)-\rho=\displaystyle\int_S p(X,y)d\mu(y)+\sigma,\\ & \rho\in\mathbb{R},\ \mu\in\mathscr{M}(S),\ \sigma\in\Sigma^2[X]\cap\mathbb{R}[X]_{2\lfloor d_P/2\rfloor},\end{cases}\tag{27}$$

and its dual

$$\begin{cases}\inf\limits_{z}\ \mathscr{L}_z(f)\\ \text{s.t.}\ z_0=1,\ M_{\lfloor d_P/2\rfloor}(z)\succeq 0,\\ \qquad \displaystyle\sum_{\alpha} p_{X,\alpha}(y)\mathscr{L}_z(X^\alpha)\ge 0,\ \forall y\in S,\\ \qquad z=(z_\alpha)_{\alpha\in\mathbb{N}^m_{d_P}}\in\mathbb{R}^{\mathbb{N}^m_{d_P}}.\end{cases}\tag{28}$$

Theorem 4.6. *Assume that* $f(X)$ *is s.o.s-convex and Assumption 3.11 holds, then the followings are true.*

(i) (27) *is solvable with the optimal value equal to* f^* *and there is no dual gap between* (27) *and* (28)*. Moreover, if* $(\mathbf{P})$ *is solvable, then so is* (28)*;*

(ii) *If* z^* *is a minimizer of* (28)*, then* $\hat{z}^*:=\{z^*_\alpha\mid\|\alpha\|_1=1\}$ *is a minimizer of* $(\mathbf{P})$*.*

Proof. (i) Denote by $f^*_{\rm sos}$ the optimal value of (27). Since Assumption 3.6 holds, recalling (23), there exists $\mu\in\mathscr{M}(S)$ such that $L_f(x)=f(x)-f^*-\int_S p(X,y)d\mu\ge 0$ for all $x\in\mathbb{R}^m$. Note that the degree of $L_f(X)$ is even and at most $2\lfloor d_P/2\rfloor$. As L_f is s.o.s, it holds that

$$f(x)-f^*-\int_S p(X,y)d\mu\in\Sigma^2[X]\cap\mathbb{R}[X]_{2\lfloor d_P/2\rfloor},$$

which means that (27) is feasible and $f^*_{\text{sos}} \geq f^*$. For any $x \in \mathbf{K}$ and feasible point (ρ, μ, σ) of (27), it holds that $f(x) - \rho \geq 0$ which implies that $f^*_{\text{sos}} \leq f^*$. Consequently, we have $f^*_{\text{sos}} = f^*$. Since (28) is strictly feasible (see the proof of Theorem 4.2 (ii)), (27) is solvable and there is no dual gap between (27) and (28).

Suppose that f^* is attainable in $(\mathbf{P})$ at a minimizer $x^* \in \mathbf{K}$. Let $z^* \in \mathbb{R}^{\mathbb{N}^m_{d_P}}$ be the truncated moment sequence associated with the Dirac measure at x^*, then f^* is attainable in (28) at z^*.

(ii) By Theorem 3.13 (i) and the proof of Theorem 3.7 (ii), it is easy to see that $\hat{z}^* \in \mathbf{K}$. As $f(X)$ is s.o.s-convex, by (Lasserre, 2009a, Theorem 2.6), the extension of Jensen's inequality $f(\hat{z}^*) \leq \mathscr{L}_{z^*}(f) = f^*$ holds, which implies that $\hat{z}^*$ is a minimizer of $(\mathbf{P})$. □

The corresponding SDP relaxations of (27) and (28) are

$$\begin{cases} f_t^{\text{psdp}} := \sup\limits_{\rho,w,\sigma} \ \rho \\ \qquad \text{s.t. } f(X) - \rho = \sum\limits_{\beta} p_{Y,\beta}(X) w_\beta + \sigma, \\ \qquad\quad M_t(w) \succeq 0, \ M_{t-d_j}(g_j w) \succeq 0, \ j = 1, \ldots, s, \\ \qquad\quad \rho \in \mathbb{R}, \ w = (w_\beta)_{\beta \in \mathbb{N}^n_{2t}}, \ \sigma \in \Sigma^2[X] \cap \mathbb{R}[X]_{2\lfloor d_P/2 \rfloor} \end{cases} \tag{29}$$

and its dual

$$\begin{cases} f_t^{\text{dsdp}} := \inf\limits_{z,\sigma,\sigma_j} \ \mathscr{L}_z(f) \\ \qquad \text{s.t. } z_0 = 1, \ M_{\lfloor d_P/2 \rfloor}(z) \succeq 0, \\ \qquad\quad \sum\limits_{\alpha} p_{X,\alpha}(Y) \mathscr{L}_z(X^\alpha) = \sigma + \sum\limits_{j=1}^{s} \sigma_j g_j, \ \sigma, \sigma_j \in \Sigma^2[Y], \\ \qquad\quad z = (z_\alpha)_{\alpha \in \mathbb{N}^m_{d_P}} \in \mathbb{R}^{\mathbb{N}^m_{d_P}}, \ \deg(\sigma), \deg(\sigma_j g_j) \leq 2t. \end{cases} \tag{30}$$

Theorem 4.7. *Assume that $f(X)$ is s.o.s-convex and Assumption 3.11 holds, then the followings are true.*

(i) *If $\mathcal{Q}(G)$ is Archimedean, then $\lim_{t\to\infty} f_t^{\text{psdp}} = \lim_{t\to\infty} f_t^{\text{dsdp}} = f^*$;*

(ii) *For some order $t \geq d_{\mathbf{K}}$, if Rank Condition 2.9 holds for w^* in the solution (ρ^*, w^*, σ^*) of (29), then $f_t^{\text{psdp}} = f^*$;*

(iii) *Let $\{(z^{(t)}, \sigma^{(t)}, \sigma_j^{(t)})\}$ be a sequence of nearly optimal solutions of (30) and $\hat{z}^{(t)} := \{z_\alpha^{(t)} \mid \|\alpha\|_1 = 1\}$. For any convergent subsequence $\{\hat{z}^{(t_i)}\}$ of $\{\hat{z}^{(t)}\}$, $\lim_{i\to\infty} \hat{z}^{(t_i)}$ is a minimizer of $(\mathbf{P})$. Consequently, if $\{\hat{z}^{(t)}\}$ is bounded and x^* is the unique minimizer of $(\mathbf{P})$, then $\lim_{t\to\infty} \hat{z}^{(t)} = x^*$.*

(iv) *If S is in the case of (18). then $f_{t_0}^{\text{psdp}} = f_{t_0}^{\text{dsdp}} = f^*$ where $t_0 = d_{\mathbf{K}}$. If $(\mathbf{P})$ is solvable, then x^* is a minimizer of $(\mathbf{P})$ if and only if there exists a minimizer $(z^*, \sigma^*, \sigma_j^*)$ of (30) with $t = t_0$ such that $\hat{z}^* := \{z_\alpha^* \mid \|\alpha\|_1 = 1\} = x^*$.*

Proof. (i) and (ii): Similar to the proof of Theorem 4.3.

(iii): Since $f(X)$ is s.o.s-convex, due to the extended Jensen's inequality (Lasserre, 2009a, Theorem 2.6), it holds that $f(\hat{z}^{(t)}) \leq \mathscr{L}_{z^{(t)}}(f)$ and therefore $f(\lim_{t\to\infty} \hat{z}^{(t)}) \leq f^*$. By Theorem 3.13, the sequence $\{\hat{z}^{(t)}\} \subset \mathbf{K}$ and hence $\lim_{t\to\infty} \hat{z}^{(t)} \in \mathbf{K}$. Thus, $\lim_{i\to\infty} \hat{z}^{(t_i)}$ is a minimizer of $(\mathbf{P})$.

(iv): By Theorem 4.6 (i) and the weak duality, it holds that $f^* \leq f_t^{\text{psdp}} \leq f_t^{\text{dsdp}}$ for any $t \geq t_0$. For any $\varepsilon > 0$, there exists a point $x^{(\varepsilon)} \in \mathbf{K}$ such that $f(x^{(\varepsilon)}) \leq f^* + \varepsilon$. Let $z^{(\varepsilon)} \in \mathbb{R}^{\mathbb{N}^m_{d_P}}$ be the truncated moment sequence associated with the Dirac measure at $x^{(\varepsilon)}$. By Theorem 3.9, $z^{(\varepsilon)}$ is feasible to (30) with $t = t_0$, which implies that $f_{t_0}^{\text{dsdp}} \leq f^* + \varepsilon$. Since ε is abitrary, it holds that $f_{t_0}^{\text{psdp}} = f_{t_0}^{\text{dsdp}} = f^*$.

Clearly, we only need to prove the "if" part. Since Assumption 3.11 holds, we have $\hat{z}^* \in \mathbf{K}$ by Theorem 3.13 (iii). As $f(X)$ is s.o.s-convex, due to the extended Jensen's inequality (Lasserre, 2009a, Theorem 2.6), it holds that $f^* \leq f(\hat{z}^*) \leq \mathscr{L}_{z^*}(f) = f^*$. Thus, $\hat{z}^*$ is a minimizer of $(\mathbf{P})$. □

Remark 4.8. Note that we do not require $\mathbf{K}$ to be compact in Theorem 4.6 and 4.7.

4.3. Linear case

Now we consider the case when $f(X), p(X,y)$ are linear in X for every $y \in S$. Let $f(X) = c^T X$ and $p(X,Y) = a(Y)^T X + b(Y)$ for some $c \in \mathbb{R}^m$, $a(Y) \in \mathbb{R}[Y]^m$ and $b(Y) \in \mathbb{R}[Y]$. Then, (**P**) becomes the following linear semi-infinite polynomial programming problem

$$f^* := \inf_{x \in \mathbb{R}^m} \; c^T x \quad \text{s.t. } a(y)^T x + b(y) \geq 0, \; \forall y \in S. \tag{31}$$

As $\lfloor d_P/2 \rfloor = 0$ in (31), the reformulation (28) in this case is just (31). The dual (27) can be written as

$$\begin{cases} \sup\limits_{\mu \in \mathscr{M}(S)} & -\displaystyle\int_S b(y) \mathrm{d}\mu(y) \\ \quad \text{s.t.} & \displaystyle\int_S a_i(y) \mathrm{d}\mu(y) = c_i, \; i = 1, \ldots, m. \end{cases} \tag{32}$$

In fact, according to the proofs of Theorem 4.2 (i) and 4.6 (i), the set $\mathscr{M}(S)$ in (32) can be replaced by the set of atomic measures supported on S. Then, we get

$$\begin{cases} \sup\limits_{\lambda_y} & -\displaystyle\sum_{y \in S} \lambda_y b(y) \\ \text{s.t.} & \displaystyle\sum_{y \in S} \lambda_y a(y) = c, \; \lambda_y \geq 0, \; \forall y \in S, \end{cases} \tag{33}$$

where only finitely many dual variables λ_y, $y \in S$, take positive values. The problem (33) is known as the *Haar dual* problem Charnes et al. (1963) of (31). According to Theorem 4.6 (i), we reproduce the well-known result:

Proposition 4.9. (Charnes et al. (1965)) *If S is compact and the Slater condition holds for $\mathbf{K}$, then (31) and (33) have the same optimal value which is attainable in (33).*

The correspondind SDP relaxations (29) and (30) become

$$\begin{cases} f_t^{\text{psdp}} := \sup\limits_{w \in \mathbb{R}^{\mathbb{N}^n_{2t}}} & -\displaystyle\sum_{\alpha \in \mathbb{N}^n_{2t}} b_\alpha w_\alpha \\ \qquad\qquad \text{s.t.} & \displaystyle\sum_{\alpha \in \mathbb{N}^n_{2t}} a_{i,\alpha} w_\alpha = c_i, \quad i = 1, \ldots, m, \\ & M_t(w) \succeq 0, \; M_{t-d_j}(g_j w) \succeq 0, \; j = 1, \ldots, s, \end{cases} \tag{34}$$

and

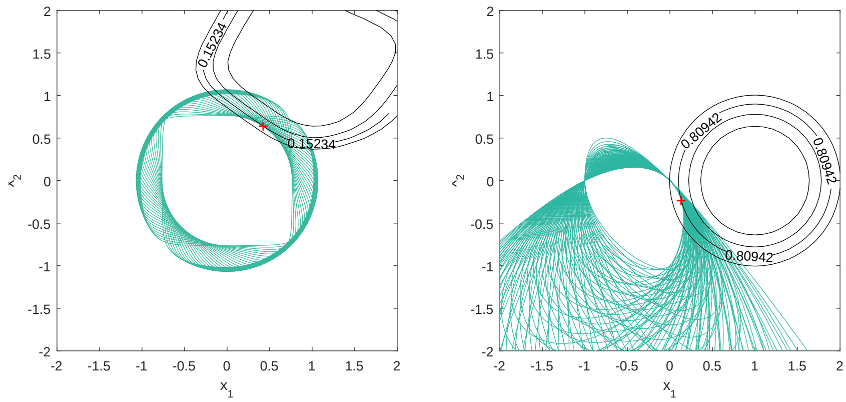
$$\begin{cases} f_t^{\text{dsdp}} := \inf\limits_{x, \sigma, \sigma_j} & c^T x \\ \qquad \text{s.t.} & a(Y)^T x + b(Y) = \sigma + \displaystyle\sum_{j=1}^s \sigma_j g_j, \\ & \sigma, \sigma_j \in \Sigma^2[X], \deg(\sigma), \deg(\sigma_j g_j) \leq 2t, \; j = 1, \ldots, s. \end{cases} \tag{35}$$

It follows from Theorem 4.7 that

Corollary 4.10. *If S is compact and the Slater condition holds for $\mathbf{K}$, then conclusions of Theorem 4.7 hold for (34) and (35).*

Example 4.11. *Now we consider three convex semi-infinite polynomial programming problems using the sets $\mathbf{K}$ defined in Example 3.16, 3.17 and 3.18. Notice that the constraints in the dual SDP relaxations (26), (30) and (35) can be easily generated by* Yalmip*. Hence, we solve the following problems using these corresponding dual SDP relaxations, which can also give us some informations on the minimizers of the problems.*

Figure 4: The sets $\mathbf{K}$ and contoure lines of f in Example 4.11.



1. *Recall the sets* $\mathbf{K}$ *and* S *defined in Example 3.16 where the polynomial* $p(X_1,X_2,y_1,y_2)\in\mathbb{R}[X_1,X_2]$ *is convex but not s.o.s-convex for every* $y\in S$*. Ahmadi and Parrilo (2013) constructed a polynomial*

$$\begin{aligned}\tilde{f}(X_1,X_2)=&89-363X_1^4X_2+\frac{51531}{64}X_2^6-\frac{9005}{4}X_2^5+\frac{49171}{16}X_2^4+721X_1^2-2060X_2^3-14X_1^3+\frac{3817}{4}X_2^2\\&+363X_1^4-9X_1^5+77X_1^6+316X_1X_2+49X_1X_2^3-2550X_1^2X_2-968X_1X_2^2+1710X_1X_2^4\\&+794X_1^3X_2+\frac{7269}{2}X_1^2X_2^2-\frac{301}{2}X_1^5X_2+\frac{2143}{4}X_1^4X_2^2+\frac{1671}{2}X_1^3X_2^3+\frac{14901}{16}X_1^2X_2^4\\&-\frac{1399}{2}X_1X_2^5-\frac{3825}{2}X_1^3X_2^2-\frac{4041}{2}X_1^2X_2^3-364X_2+48X_1.\end{aligned}$$

(*see* (Ahmadi and Parrilo, 2013, (5.2))) *which is convex but not s.o.s-convex. In order to illustrate the efficiency of the SDP relaxations* (26) *better, we shift and scale* $\tilde{f}$ *to define* $f(X_1,X_2):=\tilde{f}(X_1-1,X_2-1)/10000$*, which is still convex but not s.o.s-convex. Then, consider the problem* $\min_{x\in\mathbf{K}} f(x_1,x_2)$*, where* $d_P=d_X=8$*. Letting* $r=t=4$*, we have* $f_{4,4}^{\mathrm{dsdp}}=0.15234$ *and the truncation of SDP relaxation minimizer* $\hat{z}^{(4,4)}=(0.4245,0.6373)$*. To show the accuracy of the solution, we draw some contoure lines of* f*, including* $f(x_1,x_2)=0.15234$*, and mark the point* $\hat{z}^{(4,4)}$ *by red '+' in Figure 4 (left). As we can see, the line* $f(x_1,x_2)=0.15234$ *is almost tangent to* $\mathbf{K}$ *at the point* $\hat{z}^{(4,4)}$*.*

2. *Recall the sets* $\mathbf{K}$ *and* S *defined in Example 3.17. Let* $f(X_1,X_2):=(X_1-1)^2+X_2^2$*, i.e., the square of the distance function of a point to* $(1,0)$*, and consider the problem* $\min_{x\in\mathbf{K}} f(x_1,x_2)$*. Then, the polynomials* $f(X_1,X_2)$ *and* $-p(X_1,X_2,y_1,y_2)$ *for all* $y\in S$ *are s.o.s-convex. As* $d_{\mathbf{K}}=1$*, solving the SDP relaxation* (30) *with* $t=1$*, we get* $f_1^{\mathrm{dsdp}}=0.80942$ *and* $\hat{z}^{(1)}=(0.1311,-0.2335)$*. The corresponding contoures and the minimizer* $\hat{z}^{(1)}$ *are shown in Figure 4 (right).*

3. *Recall the sets* $\mathbf{K}$ *and* S *defined in Example 3.18. Let* $f(X_1,X_2)=X_1+X_2$ *and consider the linear semi-infinite programming problem* $\min_{x\in\mathbf{K}} f(x_1,x_2)$*. As pointed in Example 3.17, the boundary of* $\mathbf{K}$ *is the curve* $g(X_1,X_2):=2X_1^2+X_2^2+2X_1X_2+2X_1=0$*. Hence, the problem is equivalent to*

$$\min_{x\in\mathbb{R}^2} f(x_1,x_2)\quad s.t.\qquad g(x_1,x_2)=0.$$

By the method of Lagrange multipliers, we get $f^*=-1$ *with the minimizer* $(-1,0)$*. Solving the SDP relaxation* (35) *with* $t=2$*, we get* $f_2^{\mathrm{dsdp}}=-1$ *and* $\hat{z}^{(2)}=(-1,-3.6349\times10^{-6})$*.*

Acknowledgments

The first author was supported by the Chinese National Natural Science Foundation under grants 11401074, 11571350. The second author was supported by the Chinese National Natural Science Foundation under grant 11801064, the Foundation of Liaoning Education Committee under grant LN2017QN043.

References

Ahmadi, A., Parrilo, P., 2013. A complete characterization of the gap between convexity and sos-convexity. SIAM Journal on Optimization 23 (2), 811–833.

Ahmadi, A. A., Olshevsky, A., Parrilo, P. A., Tsitsiklis, J. N., 2013. NP-hardness of deciding convexity of quartic polynomials and related problems. Mathematical Programming 137 (1), 453–476.

Ahmadi, A. A., Parrilo, P. A., Oct 2012. A convex polynomial that is not sos-convex. Mathematical Programming 135 (1), 275–292.

Belousov, E., 1977. Introduction to Convex Analysis and Integer Programming. Moscow University Publ.: Moscow.

Ben-Tal, A., Nemirovski, A., 2001. Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications. MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics.

Berg, C., Maserick, P. H., 1984. Exponentially bounded positive definite functions. Illinois Journal of Mathematics 28, 162–179.

Bochnak, J., Coste, M., Roy, M.-F., 1998. Real Algebraic Geometry. Springer.

Borwein, J. M., 1981. Direct theorems in semi-infinite convex programming. Mathematical Programming 21 (1), 301–318.

Charnes, A., Cooper, W. W., Kortanek, K., 1963. Duality in semi-infinite programs and some works of Haar and Carathéodory. Management Science 9 (2), 209–228.

Charnes, A., Cooper, W. W., Kortanek, K., 1965. On representations of semi-infinite programs which have no duality gaps. Management Science 12 (1), 113–121.

Curto, R. E., Fialkow, L. A., 2005. Truncated K-moment problems in several variables. Journal of Operator Theory 54 (1), 189–226.

Goberna, M. Á., López, M. A., 1998. Linear semi-infinite optimization. John Wiley & Sons, Chichester.

Goberna, M. A., López, M. A., Sep 2017. Recent contributions to linear semi-infinite optimization. 4OR 15 (3), 221–264.

Gouveia, J., Parrilo, P., Thomas, R., 2010. Theta bodies for polynomial ideals. SIAM Journal on Optimization 20 (4), 2097–2118.

Guo, F., Wang, C., Zhi, L., 2015. Semidefinite representations of noncompact convex sets. SIAM Journal on Optimization 25 (1), 377–395.

Haviland, E. K., 1935. On the momentum problem for distribution functions in more than one dimension. American Journal of Mathematics 57 (3), 562–568.

Helton, J., Nie, J., 2009. Sufficient and necessary conditions for semidefinite representability of convex hulls and sets. SIAM Journal on Optimization 20 (2), 759–791.

Helton, J., Nie, J., 2010. Semidefinite representation of convex sets. Mathematical Programming 122 (1), 21–64.

Hettich, R., Kortanek, K. O., 1993. Semi-infinite programming: Theory, methods, and applications. SIAM Review 35 (3), 380–429.

Kriel, T., Schweighofer, M., 2018. On the exactness of lasserre relaxations for compact convex basic closed semialgebraic sets. SIAM Journal on Optimization 28 (2), 1796–1816.

Lasserre, J., 2009a. Convexity in semialgebraic geometry and polynomial optimization. SIAM Journal on Optimization 19 (4), 1995–2014.

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

Lasserre, J. B., 2009b. Convex sets with semidefinite representation. Mathematical Programming, Ser. A 120 (2), 457–477.

Lasserre, J. B., 2012. An algorithm for semi-infinite polynomial optimization. TOP 20 (1), 119–129.

Lasserre, J. B., Netzer, T., May 2007. Sos approximations of nonnegative polynomials via simple high degree perturbations. Mathematische Zeitschrift 256 (1), 99–112.

Laurent, M., 2009. Sums of squares, moment matrices and optimization over polynomials. In: Putinar, M., Sullivant, S. (Eds.), Emerging Applications of Algebraic Geometry. Springer, New York, NY, pp. 157–270.

Levin, V. L., 1969. Application of E. Helly's theorem to convex programming, problems of best approximation and related questions. Mathematics of the USSR-Sbornik 8 (2), 235.

Löfberg, J., 2004. YALMIP : a toolbox for modeling and optimization in MATLAB. In: 2004 IEEE International Conference on Robotics and Automation (IEEE Cat. No.04CH37508). pp. 284–289.

López, M., Still, G., 2007. Semi-infinite programming. European Journal of Operational Research 180 (2), 491 – 518.

Nie, J., 2012. Discriminants and nonnegative polynomials. Journal of Symbolic Computation 47 (2), 167–191.

Nie, J., 2013. An exact Jacobian SDP relaxation for polynomial optimization. Mathematical Programming, Ser. A 137 (1-2), 225–255.

Nie, J., Schweighofer, M., 2007. On the complexity of Putinar's positivstellensatz. Journal of Complexity 23 (1), 135 – 150.

Pardalos, P. M., Vavasis, S. A., 1991. Quadratic programming with one negative eigenvalue is NP-hard. Journal of Global Optimization 1 (1), 15–22.

Parpas, P., Rustem, B., 2009. An algorithm for the global optimization of a class of continuous minimax problems. Journal of Optimization Theory and Applications 141 (2), 461–473.

Parrilo, P. A., Sturmfels, B., 2003. Minimizing polynomial functions. In: Algorithmic and quantitative real algebraic geometry. Vol. 60 of DIMACS Series in Discrete Mathematics and Theoretical Computer Science. American Mathematical Society, pp. 83–99.

Powers, V., Reznick, B., 2000. Polynomials that are positive on an interval. Transactions of the American Mathematical Society 352 (10), 4677–4692.

Putinar, M., 1993. Positive polynomials on compact semi-algebraic sets. Indiana University Mathematics Journal 42 (3), 969–984.

Reznick, B., 2000. Some concrete aspects of Hilbert's 17th problem. In: Contemporary Mathematics. Vol. 253. American Mathematical Society, pp. 251–272.

Rostalski, P., 2010. Bermeja - software for convex algebraic geometry, http://math.berkeley.edu/~philipp/cagwiki.

Scheiderer, C., 2018. Spectrahedral shadows. SIAM Journal on Applied Algebra and Geometry 2 (1), 26–44.

Schmüdgen, K., 1991. The K-moment problem for compact semi-algebraic sets. Mathematische Annalen 289 (1), 203–206.

Wang, L., Guo, F., 2013. Semidefinite relaxations for semi-infinite polynomial programming. Computational Optimization and Applications 58 (1), 133–159.

Wolkowicz, H., Saigal, R., Vandenberghe, L., 2000. Handbook of Semidefinite Programming - Theory, Algorithms, and Applications. Kluwer Academic Publisher, Dordrecht.

Xu, Y., Sun, W., Qi, L., 2015. On solving a class of linear semi-infinite programming by SDP method. Optimization 64 (3), 603–616.