A Moment-SOS Hierarchy for Robust Polynomial Matrix Inequality Optimization with SOS-Convexity[*]

Feng Guo[†] and Jie Wang[‡]

Abstract. We study a class of polynomial optimization problems with a robust polynomial matrix inequality constraint for which the uncertainty set is defined also by a polynomial matrix inequality (including robust polynomial semidefinite programs as a special case). Under certain SOS-convexity assumptions, we construct a hierarchy of moment-SOS relaxations for this problem to obtain convergent upper bounds of the optimal value by solving a sequence of semidefinite programs. To this end, we apply the Positivstellensatz for polynomial matrices and its dual matrix-valued moment theory to a conic reformulation of the problem. Most of the nice features of the moment-SOS hierarchy for the scalar polynomial optimization are generalized to the matrix case. In particular, the finite convergence of the hierarchy can be also certified if the flat extension condition holds. To extract global minimizers in this case, we develop a linear algebra approach to recover the representing matrix-valued measure for the corresponding truncated matrix-valued moment problem. As an application, we use this hierarchy to solve the problem of minimizing the smallest eigenvalue of a polynomial matrix subject to a polynomial matrix inequality. Finally, if SOS-convexity is replaced by convexity, we can still approximate the optimal value as closely as desired by solving a sequence of semidefinite programs, and certify global optimality in case that certain flat extension conditions hold true.

Key words. polynomial optimization, polynomial matrix inequality, robust optimization, moment-SOS hierarchy, semidefinite programming

AMS subject classifications. 90C23, 90C17, 90C22, 90C26

1. Introduction.

Polynomial optimization problems with polynomial matrix inequality (PMI) constraints have a wide range of applications in many fields. In particular, as special cases of PMIs, linear or bilinear matrix inequality constrained problems appear frequently in most synthesis problems for linear systems in optimal control. Due to estimation errors or lack of information, the data of real-world optimization problems often involve uncertainty. Hence, robust optimization is an appropriate modeling paradigm for some safety-critical applications with little tolerance for failure [6].

In this paper, we study the following robust PMI optimization problem under data uncertainty in the PMI constraint:

$$
\text{(RPMIO)} \qquad \begin{cases} f^\star := \inf\limits_{\mathbf{y}\in\mathcal{Y}} f(\mathbf{y}) \\ \qquad\quad \text{s.t. } \mathcal{Y}\subseteq\mathbb{R}^\ell,\ P(\mathbf{y},\mathbf{x})\succeq 0,\ \forall \mathbf{x}\in\mathcal{X}\subset\mathbb{R}^n, \end{cases}
$$

[*]Submitted to the editors DATE.

Funding: FG was supported by the NSFC under grant 11571350. JW was supported by the NSFC grant 12201618.

[†]School of Mathematical Sciences, Dalian University of Technology (fguo@dlut.edu.cn).

[‡]Academy of Mathematics and Systems Science, CAS (wangjie212@amss.ac.cn, https://wangjie212.github.io/jiewang/).

where $\mathbf{y} = (y_1, \ldots, y_\ell)$ is the decision variables constrained in a basic semialgebraic set

$$\mathcal{Y} = \{\mathbf{y} \in \mathbb{R}^\ell \mid \theta_1(\mathbf{y}) \geq 0, \ldots, \theta_s(\mathbf{y}) \geq 0\}, \tag{1.1}$$

$\mathbf{x} = (x_1, \ldots, x_n)$ is the uncertain parameters belonging to some uncertainty set

$$\mathcal{X} \coloneqq \{\mathbf{x} \in \mathbb{R}^n \mid G(\mathbf{x}) \succeq 0\}, \tag{1.2}$$

defined by a $q \times q$ symmetric polynomial matrix $G(\mathbf{x})$, $f \in \mathbb{R}[\mathbf{y}]$ is a polynomial function in $\mathbf{y}$, and $P(\mathbf{y}, \mathbf{x})$ is an $m \times m$ symmetric polynomial matrix in $\mathbf{y}$ and $\mathbf{x}$. That is, the PMI constraint in (RPMIO) depends polynomially on the decision variable $\mathbf{y}$ and the uncertain parameter $\mathbf{x}$. We make the following assumptions on (RPMIO).

Assumption 1. (i) $f(\mathbf{y}), -\theta_1(\mathbf{y}), \ldots, -\theta_s(\mathbf{y})$ are SOS-convex (Definition 2.4); (ii) $-P(\mathbf{y}, \mathbf{x})$ is PSD-SOS-convex (Definition 2.6) in $\mathbf{y}$ for all $\mathbf{x} \in \mathcal{X}$; (iii) $\mathcal{X}$ is compact.

To highlight the modeling power of (RPMIO) under Assumption 1, let us name a few problems from different fields which can be modelled as an instance of (RPMIO). First, note that robust polynomial semidefinite programs (SDP) are linear (RPMIO). A basic problem in interval computations is to estimate intervals of confidence for the components of a given vector-valued function when its variables range in a product of intervals. Assume that the function is given by polynomials and we seek an ellipsoid of confidence for its components. Then this problem can be modelled as a linear instance of (RPMIO). In the context of risk management, the robust correlation stress testing where data uncertainty arises due to untimely recording of portfolio holdings can be formulated as a robust least square SDP which is a special case of (RPMIO) [35]. In optimal control, many problems for systems of ordinary differential equations can be posed as convex optimization problems with matrix inequality constraints which should hold on a prescribed portion of the state space [10, 24, 45]. If the involved functions in the differential equations are polynomials, these problems often take the form of linear (RPMIO). Moreover, the deterministic PMI optimization problem of minimizing a polynomial function $h(\mathbf{x})$ over the set $\mathcal{X}$ studied in [23] can be formulated as a linear instance of (RPMIO), which also implies that (RPMIO) is in general NP-hard, even in the linear case.

For deterministic PMI optimization problems, Kojima [30] proposed SOS relaxations by utilizing a penalty function and a generalized Lagrangian dual, and Henrion and Lasserre [23] gave a hierarchy of moment relaxations allowing to detect finite convergence and to extract global minimizers. Recently, there has been increasing interest in studying robust polynomial optimization problems; see e.g. [34, 26, 25, 11]. However, robust PMI constraints are not considered in these work. Since (RPMIO) under Assumption 1 is computationally intractable, there is very little work on how to solve or even approximate it, though some asymptotically exact approaches are available for linear (RPMIO) treated as a special case of the robust SDP problem [5]. In particular, Ohara and Sasaki [42], Bliman [7, 8] proposed approaches for the robust SDP problem based on the Kalman–Yakubovich–Popov lemma. Scherer and Hol [45] established a Positivstellensatz for polynomial matrices and derived matrix SOS relaxations for the linear case of (RPMIO). Oishi [43] gave an approach for the robust SDP problem based on a division of the parameter region. A computationally tractable hierarchy of inner and outer approximations for the robust SDP problem was proposed by Louca and Bitar [37]. Li etc. [35] gave a simple variant of the spectral projected gradient method to solve robust least square SDP problems.

Before introducing our main contributions, we would like to point out that the matrix SOS relaxations [45] for the linear case of (RPMIO) cannot be straightforwardly extended to (RPMIO) with SOS-convexity. Moreover, the dual moment facet of the matrix SOS relaxations was not investigated in [45]. This motivates us to establish a moment-SOS hierarchy for (RPMIO) with SOS-convexity in full generality and extend its nice features from the scalar case to the matrix case.

Let us denote by $\mathbb{S}[\mathbf{x}]^m$ the cone of $m \times m$ symmetric real polynomial matrices in $\mathbf{x}$ and by $\mathcal{P}^m(\mathcal{X})$ its subcone consisting of polynomial matrices which are positive semidefinite (PSD) on $\mathcal{X}$. By Haviland's theorem for polynomial matrices (Theorem 2.14), the dual cone of $\mathcal{P}^m(\mathcal{X})$ consists of tracial $\mathcal{X}$-moment functionals on $\mathbb{S}[\mathbf{x}]^m$ (Definition 2.13), while justifying the membership of a linear funcitonal on $\mathbb{S}[\mathbf{x}]^m$ to the dual cone of $\mathcal{P}^m(\mathcal{X})$ amounts to the matrix-valued $\mathcal{X}$-moment problem (Definition 2.16). Therefore, to explore the dual aspect of the matrix sums-of-squares (SOS) relaxations for (RPMIO), we need to invoke the results on the matrix-valued moment problem. For a given multi-indexed sequence of real $m \times m$ symmetric matrices $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n}$, the matrix-valued $\mathcal{X}$-moment problem asks if there exists a PSD matrix-valued representing measure Φ supported on $\mathcal{X}$ such that $S_{\boldsymbol{\alpha}} = \int_{\mathcal{X}} \mathbf{x}^{\boldsymbol{\alpha}} \mathrm{d}\Phi(\mathbf{x})$ for all $\boldsymbol{\alpha} \in \mathbb{N}^n$. We refer the reader to [29] for a thorough introduction on the history and background about the matrix-valued moment problem. For the scalar moment problem ($m = 1$), due to Haviland's theorem and Putinar's Positivstellensatz, the representing measure is guaranteed by the PSDness of the associated moment matrices and localizing matrices. Based on this, Lasserre [31] proposed the moment-SOS hierarchy for the scalar polynomial optimization and established its asymptotic convergence. For the truncated scalar moment problem, Curto and Fialkow [14] gave the celebrated flat extension condition on the moment matrix as a sufficient condition for the existence of a representing measure, which allows to detect finite convergence of Lasserre's hierarchy and extract global minimizers [22]. Recently, Kimsey and Trachana [29] obtained a flat extension theorem for the truncated matrix-valued moment problem. Unlike the scalar case, to the best of our knowledge, there is very little work in the literature to link the theory of matrix-valued moments to PMI optimization. As the first attempt to connect these two subjects, we aim to construct a moment-SOS hierarchy for (RPMIO) with SOS-convexity by combining Scherer-Hol's Positivstellensatz with the matrix-valued moment theory.

Contributions. Our main contributions are summarized as follows:

1. As a first contribution, we provide a solution to the truncated matrix-valued $\mathcal{X}$-moment problem using the flat extension condition. Furthermore, we develop a linear algebra procedure to retrieve a finitely atomic representing measure whenever the flat extension condition holds, which extends Henrion-Lasserre's algorithm to the matrix case.
2. We establish a moment-SOS hierarchy for (RPMIO) with SOS-convexity. To achieve this, we first reformulate (RPMIO) to a conic optimization problem via the Lagrange dual theory, and then replace the conic constraints with more tractable matrix qua-

dratic module constraints or matrix-valued pseudo-moment cone constraints. This yields a convergent sequence of upper bounds on the optimum of (RPMIO). More importantly, we show that if the flat extension condition holds, then finite convergence occurs and we can extract a globally optimal solution $\mathbf{y}^\star$ of (RPMIO) as well as the points $\mathbf{x} \in \Delta(\mathbf{y}^\star)$ and the corresponding vectors $\mathbf{v}$, where

$$\Delta(\mathbf{y}^\star) := \{\mathbf{x} \in \mathcal{X} \mid \exists \mathbf{v} \in \mathbb{R}^m \quad \text{s.t.} \quad P(\mathbf{y}^\star, \mathbf{x})\mathbf{v} = 0\} \tag{1.3}$$

is the index set of constraints active at $\mathbf{y}^\star$.

3. For the linear case of (RPMIO), we show that the dual problem is exactly the generalized matrix-valued moment problem and our SOS relaxations recover the matrix SOS relaxations proposed by Scherer and Hol [45]. As an application, we provide a solution to the problem of minimizing the smallest eigenvalue of a polynomial matrix over a set defined by a PMI.
4. In case that the SOS-convexity assumption of (RPMIO) is weakened to convexity, we also provide a sequence of SDPs that can approximate the optimal value of (RPMIO) as closely as desired. Moreover, finite convergence can be detected via certain flat extension conditions.

The rest of the paper is organized as follows. We first recall some preliminaries in Section 2. Then, we consider the truncated matrix-valued $\mathcal{X}$-moment problem in Section 3 and propose a linear algebra procedure for the representing measure retrieve. In Section 4, we construct a moment-SOS hierarchy for (RPMIO) with SOS-convexity, and treat particularly the linear case. An extension of the proposed approach to the general convex case is discussed in Section 5. Conclusions are given in Section 6.

2. Preliminaries.

We collect some notation and basic concepts which will be used in this paper. We denote by $\mathbf{x}$ (resp., $\mathbf{y}$) the n-tuple (resp., ℓ-tuple) of variables $(x_1, \ldots, x_n)$ (resp., $(y_1, \ldots, y_\ell)$). The symbol $\mathbb{N}$ (resp., $\mathbb{R}$, $\mathbb{R}_+$) denotes the set of nonnegative integers (resp., real numbers, nonnegative real numbers). Denote by $\mathbb{R}^m$ (resp. $\mathbb{R}^{l\times q}$, $\mathbb{S}^m$) the m-dimensional real vector (resp. $l \times q$ real matrix, $m \times m$ symmetric real matrix) space. For $\mathbf{v} \in \mathbb{R}^m$ (resp., $N \in \mathbb{R}^{l\times q}$), the symbol $\mathbf{v}^\intercal$ (resp., $N^\intercal$) denotes the transpose of $\mathbf{v}$ (resp., N). For a matrix $N \in \mathbb{R}^{m\times m}$, $\operatorname{tr}(N)$ denotes its trace. For two matrices N_1 and N_2, $N_1 \otimes N_2$ denotes the Kronecker product of N_1 and N_2. The notation I_m denotes the $m \times m$ identity matrix. For any $t \in \mathbb{R}$, $\lceil t \rceil$ denotes the smallest integer that is not smaller than t. For $\mathbf{u} \in \mathbb{R}^m$, $\|\mathbf{u}\|$ denotes the standard Euclidean norm of $\mathbf{u}$. For a vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$, let $|\boldsymbol{\alpha}| = \alpha_1 + \cdots + \alpha_n$. For a set A, we use $|A|$ to denote its cardinality. For $k \in \mathbb{N}$, let $\mathbb{N}^n_k = \{\boldsymbol{\alpha} \in \mathbb{N}^n \mid |\boldsymbol{\alpha}| \leq k\}$ and $|\mathbb{N}^n_k| = \binom{n+k}{k}$ be its cardinality. For variables $\mathbf{x} \in \mathbb{R}^n$ and $\boldsymbol{\alpha} \in \mathbb{N}^n$, $\mathbf{x}^{\boldsymbol{\alpha}}$ denotes the monomial $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. Let $\mathbb{R}[\mathbf{x}]$ (resp. $\mathbb{S}[\mathbf{x}]^m$) denote the set of real polynomials (resp. $m \times m$ symmetric real polynomial matrices) in $\mathbf{x}$. For $h \in \mathbb{R}[\mathbf{x}]$, we denote by $\nabla_{\mathbf{x}}(h)$ its gradient vector and by $\nabla_{\mathbf{xx}}(h)$ its Hessian matrix. For $h \in \mathbb{R}[\mathbf{x}]$, we denote by $\deg(h)$ its (total) degree. For $k \in \mathbb{N}$, denote by $\mathbb{R}[\mathbf{x}]_k$ the set of polynomials in $\mathbb{R}[\mathbf{x}]$ of degree up to k. For a $\mathbb{R}$-vector space A, denote by A^* the dual space of linear functionals from A to $\mathbb{R}$. Given a cone $B \subset A$, its dual cone is $B^* = \{L \in A^* \mid L(b) \geq 0, \ \forall b \in B\}$.

2.1. A positivstellensatz for polynomial matrices. We recall the Positivstellensatz for polynomial matrices obtained in [45]. For an $l_1 \times l_2$ polynomial matrix $T(\mathbf{x}) = [T_{ij}(\mathbf{x})]$, denote

$$\deg(T) \coloneqq \max\{\deg(T_{ij}) \mid i = 1, \ldots, l_1, j = 1, \ldots, l_2\}.$$

A polynomial matrix $\Sigma(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^q$ is said to be a sum-of-squares (SOS) if there exists an $l \times q$ polynomial matrix $T(\mathbf{x})$ for some $l \in \mathbb{N}$ such that $\Sigma(\mathbf{x}) = T(\mathbf{x})^\intercal T(\mathbf{x})$. For $d \in \mathbb{N}$, denote by $u_d(\mathbf{x})$ the canonical basis of $\mathbb{R}[\mathbf{x}]_d$, i.e.,

$$u_d(\mathbf{x}) \coloneqq [1,\ x_1,\ x_2,\ \cdots,\ x_n,\ x_1^2,\ x_1x_2,\ \cdots,\ x_n^d]^\intercal, \tag{2.1}$$

whose cardinality is $|\mathbb{N}_d^n| = \binom{n+d}{d}$. With $d = \deg(T)$, we can write $T(\mathbf{x})$ as

$$T(\mathbf{x}) = Q(u_d(\mathbf{x}) \otimes I_q) \text{ with } Q = [Q_1, \ldots, Q_{|\mathbb{N}_d^n|}], \quad Q_i \in \mathbb{R}^{l \times q},$$

where Q is the vector of coefficient matrices of $T(\mathbf{x})$ with respect to $u_d(\mathbf{x})$. Hence, $\Sigma(\mathbf{x})$ is an SOS with respect to $u_d(\mathbf{x})$ if there exists some $Q \in \mathbb{R}^{l \times q|\mathbb{N}_d^n|}$ satisfying

$$\Sigma(\mathbf{x}) = T(\mathbf{x})^\intercal T(\mathbf{x}) = (u_d(\mathbf{x}) \otimes I_q)^\intercal (Q^\intercal Q)(u_d(\mathbf{x}) \otimes I_q).$$

We thus have the following result.

Proposition 2.1. [45, Lemma 1] *A polynomial matrix $\Sigma(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^q$ is an SOS with respect to the monomial basis $u_d(\mathbf{x})$ if and only if there exists $Z \in \mathbb{S}_+^{q|\mathbb{N}_d^n|}$ such that*

$$\Sigma(\mathbf{x}) = (u_d(\mathbf{x}) \otimes I_q)^\intercal Z (u_d(\mathbf{x}) \otimes I_q).$$

Define the bilinear mapping

$$(\cdot,\cdot)_m \colon \mathbb{R}^{mq \times mq} \times \mathbb{R}^{q \times q} \to \mathbb{R}^{m \times m}, \quad (A, B)_m = \operatorname{tr}_m\left(A^\intercal (I_m \otimes B)\right),$$

with

$$\operatorname{tr}_m(C) \coloneqq \begin{pmatrix} \operatorname{tr}(C_{11}) & \cdots & \operatorname{tr}(C_{1m}) \\ \vdots & \ddots & \vdots \\ \operatorname{tr}(C_{m1}) & \cdots & \operatorname{tr}(C_{mm}) \end{pmatrix} \quad \text{for } C \in \mathbb{R}^{mq \times mq}, C_{jk} \in \mathbb{R}^{q \times q}.$$

Remark 2.2. Note that $(A, B)_1$ is just the standard inner product $\langle A, B\rangle = \operatorname{tr}(A^\intercal B)$. Moreover, we have $(A, B)_m \succeq 0$ if $A \succeq 0$ and $B \succeq 0$ ([45]).

Assumption 2. *For the defining matrix $G(\mathbf{x})$ of $\mathcal{X}$ in (1.2), there exists $r \in \mathbb{R}$ and an SOS polynomial matrix $\Sigma(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^q$ such that*

$$r^2 - \|\mathbf{x}\|^2 - \langle \Sigma(\mathbf{x}), G(\mathbf{x})\rangle \text{ is an SOS.}$$

Theorem 2.3. [45, Corollary 1] *Let Assumption 2 hold and $F(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^m$ be positive on $\mathcal{X}$. Then there exist SOS polynomial matrices $\Sigma_0(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^m$ and $\Sigma(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^{mq}$ such that*

$$F(\mathbf{x}) = \Sigma_0(\mathbf{x}) + (\Sigma(\mathbf{x}), G(\mathbf{x}))_m.$$

For each $k \in \mathbb{N}$, we define the k-th *truncated matrix quadratic module* $\mathcal{Q}_k^m(G)$ associated with $G(\mathbf{x})$ by

$$\mathcal{Q}_k^m(G) := \left\{ \Sigma_0(\mathbf{x}) + (\Sigma(\mathbf{x}), G(\mathbf{x}))_m \;\middle|\; \begin{array}{l} \Sigma_0 \in \mathbb{S}[\mathbf{x}]^m, \Sigma \in \mathbb{S}[\mathbf{x}]^{mq}, \\ \Sigma_0, \Sigma_1 \text{ are SOS}, \\ \deg(\Sigma_0), \deg((\Sigma, G)_m) \leq 2k \end{array} \right\},$$

and define the *matrix quadratic module* by

$$\mathcal{Q}^m(G) = \bigcup_{k \in \mathbb{N}} \mathcal{Q}_k^m(G).$$

By Proposition 2.1, checking membership in $\mathcal{Q}_k^m(G)$ can be written as an SDP.

2.2. Convexity and SOS-convexity. To derive a moment-SOS hierarchy for (RPMIO), we need to impose the following SOS-convexity on (RPMIO).

Definition 2.4. [21] *A polynomial $h \in \mathbb{R}[\mathbf{y}]$ is* SOS-convex *if its Hessian*

$$\nabla_{\mathbf{yy}} h(\mathbf{y}) = H(\mathbf{y})^\intercal H(\mathbf{y}),$$

for some polynomial matrix $H(\mathbf{y})$.

While checking the convexity of a polynomial is generally NP-hard [1], the SOS-convexity can be justified numerically by solving a SDP problem [21]. Ahmadi and Parrilo [2] proved that the set of convex polynomials and the set of SOS-convex polynomials in $\mathbb{R}[\mathbf{y}]$ up to degree k coincide if and only if $\ell = 1$ or $k = 2$ or $(\ell, k) = (2, 4)$.

Recall the definition of the PSD-convexity of a polynomial matrix.

Definition 2.5. *We say that a polynomial matrix $Q(\mathbf{y}) \in \mathbb{S}[\mathbf{y}]^m$ is* PSD-convex *if*

$$tQ(\mathbf{y}^{(1)}) + (1-t)Q(\mathbf{y}^{(2)}) \succeq Q(t\mathbf{y}^{(1)} + (1-t)\mathbf{y}^{(2)})$$

holds for any $\mathbf{y}^{(1)}, \mathbf{y}^{(2)} \in \mathbb{R}^\ell$ and $t \in (0, 1)$.

Nie [41] gave an extension of SOS-convexity to polynomial matrices.

Definition 2.6. [41] *We say that a polynomial matrix $Q(\mathbf{y}) \in \mathbb{S}[\mathbf{y}]^m$ is* PSD-SOS-convex *if for every $\mathbf{v} \in \mathbb{R}^m$, there exists a polynomial matrix $F_{\mathbf{v}}(\mathbf{y})$ in $\mathbf{y}$ such that*

$$\nabla_{\mathbf{yy}}(\mathbf{v}^\intercal Q(\mathbf{y})\mathbf{v}) = F_{\mathbf{v}}(\mathbf{y})^\intercal F_{\mathbf{v}}(\mathbf{y}).$$

In other words, $Q(\mathbf{y})$ is PSD-SOS-convex if and only if $\mathbf{v}^\intercal Q(\mathbf{y})\mathbf{v}$ is an SOS-convex polynomial for each $\mathbf{v} \in \mathbb{R}^m$. Clearly, if $Q(\mathbf{y})$ is PSD-SOS-convex, then it is PSD-convex, but not vice versa.

The PSD-SOS-convexity condition of a polynomial matrix $Q(\mathbf{y}) \in \mathbb{S}[\mathbf{y}]^m$ requires checking the Hessian $\nabla_{\mathbf{yy}}(\mathbf{v}^\intercal Q(\mathbf{y})\mathbf{v})$ for every $\mathbf{v} \in \mathbb{R}^m$, which is very hard in practice. A stronger condition is given in [41].

Definition 2.7. [41] *We say that a polynomial matrix $Q(\mathbf{y}) \in \mathbb{S}[\mathbf{y}]^m$ is uniformly PSD-SOS-convex if there exists a polynomial matrix $F(\mathbf{v}, \mathbf{y})$ in $(\mathbf{v}, \mathbf{y})$ such that*

$$\nabla_{\mathbf{yy}}(\mathbf{v}^\intercal Q(\mathbf{y})\mathbf{v}) = F(\mathbf{v}, \mathbf{y})^\intercal F(\mathbf{v}, \mathbf{y}). \tag{2.2}$$

Clearly, if $Q(\mathbf{y}) \in \mathbb{S}[\mathbf{y}]^m$ is uniformly PSD-SOS-convex, then it is PSD-SOS-convex. Checking the existence of $F(\mathbf{v}, \mathbf{y})$ in (2.2) can be converted to an SDP feasibility problem. Moreover, we have the following proposition.

Proposition 2.8. *A polynomial matrix* $Q(\mathbf{y}) = \sum_{\boldsymbol{\alpha} \in \mathrm{supp}(Q)} Q_{\boldsymbol{\alpha}} \mathbf{y}^{\boldsymbol{\alpha}} \in \mathbb{S}[\mathbf{y}]^m$ *is uniformly PSD-SOS-convex if* $\sum_{\boldsymbol{\alpha} \in \mathrm{supp}(Q)} Q_{\boldsymbol{\alpha}} \otimes \nabla_{\mathbf{yy}} \mathbf{y}^{\boldsymbol{\alpha}}$ *is an SOS polynomial matrix.*

Proof. Observe that

$$\begin{aligned}\nabla_{\mathbf{yy}}(\mathbf{v}^\intercal Q(\mathbf{y}) \mathbf{v}) &= \sum_{i,j=1}^{m} \left(\sum_{\boldsymbol{\alpha} \in \mathrm{supp}(Q)} [Q_{\boldsymbol{\alpha}}]_{ij} \nabla_{\mathbf{yy}} \mathbf{y}^{\boldsymbol{\alpha}} \right) v_i v_j \\ &= (\mathbf{v} \otimes I_\ell)^\intercal \left(\sum_{\boldsymbol{\alpha} \in \mathrm{supp}(Q)} Q_{\boldsymbol{\alpha}} \otimes \nabla_{\mathbf{yy}} \mathbf{y}^{\boldsymbol{\alpha}} \right) (\mathbf{v} \otimes I_\ell).\end{aligned}$$

If there exists a polynomial matrix $T(\mathbf{y})$ such that

$$\left(\sum_{\boldsymbol{\alpha} \in \mathrm{supp}(Q)} Q_{\boldsymbol{\alpha}} \otimes \nabla_{\mathbf{yy}} \mathbf{y}^{\boldsymbol{\alpha}} \right) = T(\mathbf{y})^\intercal T(\mathbf{y}),$$

we have (2.2) by letting $F(\mathbf{v}, \mathbf{y}) = T(\mathbf{y})(\mathbf{v} \otimes I_\ell)$. ∎

Corollary 2.9. *A quadratic polynomial matrix*

$$Q(\mathbf{y}) = C + \sum_{i=1}^{\ell} L_i y_i + \sum_{i,j=1}^{\ell} Q_{ij} y_i y_j \in \mathbb{S}[\mathbf{y}]^m,$$

where $C, L_i, Q_{ij} \in \mathbb{S}^m$ *and* $Q_{ij} = Q_{ji}$*, is uniformly PSD-SOS-convex if the* $m\ell \times m\ell$ *matrix* $[Q_{ij}]_{1 \le i,j \le \ell}$ *is PSD.*

Proof. It is clear that $\sum_{\boldsymbol{\alpha} \in \mathrm{supp}(Q)} \nabla_{\mathbf{yy}} \mathbf{y}^{\boldsymbol{\alpha}} \otimes Q_{\boldsymbol{\alpha}} = 2[Q_{ij}]_{1 \le i,j \le \ell}$ which implies that the matrix $\sum_{\boldsymbol{\alpha} \in \mathrm{supp}(Q)} Q_{\boldsymbol{\alpha}} \otimes \nabla_{\mathbf{yy}} \mathbf{y}^{\boldsymbol{\alpha}} \in \mathbb{S}^{\ell m \times \ell m}$ is PSD. Hence, $Q(\mathbf{y})$ is uniformly PSD-SOS-convex by Proposition 2.8. ∎

2.3. Matrix-valued measures. Now we recall some background on the theory of matrix-valued measures, which is crucial for our subsequent development. For more details, the reader is referred to [19, 20, 17, 16]. Denote by $B(\mathcal{X})$ the smallest σ-algebra generated from the open subsets of $\mathcal{X}$ and by $\mathfrak{m}(\mathcal{X})$ the set of all finite Borel measures on $\mathcal{X}$. A measure $\phi \in \mathfrak{m}(\mathcal{X})$ is *positive* if $\phi(\mathcal{A}) \ge 0$ for all $\mathcal{A} \in B(\mathcal{X})$. Denote by $\mathfrak{m}_+(\mathcal{X})$ the set of all finite positive Borel measures on $\mathcal{X}$. The support $\mathrm{supp}(\phi)$ of a Borel measure $\phi \in \mathfrak{m}(\mathcal{X})$ is the (unique) smallest closed set $\mathcal{A} \in B(\mathcal{X})$ such that $\phi(\mathcal{X} \setminus \mathcal{A}) = 0$.

Definition 2.10. *Let* $\phi_{ij} \in \mathfrak{m}(\mathcal{X})$*,* $i, j = 1, \ldots, m$*. The* $m \times m$ *matrix-valued measure* Φ *on* $\mathcal{X}$ *is defined as the matrix-valued function* $\Phi \colon B(\mathcal{X}) \to \mathbb{R}^{m \times m}$ *with*

$$\Phi(\mathcal{A}) := [\phi_{ij}(\mathcal{A})] \in \mathbb{R}^{m \times m}, \quad \forall \mathcal{A} \in B(\mathcal{X}).$$

If $\phi_{ij} = \phi_{ji}$ for all $i, j = 1, \ldots, m$, we call Φ a symmetric matrix-valued measure. If $\mathbf{v}^\intercal \Phi(\mathcal{A})\mathbf{v} \geq 0$ holds for all $\mathcal{A} \in B(\mathcal{X})$ and for all column vectors $\mathbf{v} \in \mathbb{R}^m$, we call Φ a PSD matrix-valued measure. The set

$$\operatorname{supp}(\Phi) \coloneqq \bigcup_{i,j=1}^{m} \operatorname{supp}(\phi_{ij})$$

is called the support of the matrix-valued measure Φ.

We denote by $\mathfrak{M}^m(\mathcal{X})$ (resp. $\mathfrak{M}^m_+(\mathcal{X})$) the set of all $m \times m$ (resp. PSD) symmetric matrix-valued measures on $\mathcal{X}$.

Definition 2.11. *Let $\Phi = [\phi_{ij}] \in \mathfrak{M}^m(\mathcal{X})$. A function $h\colon \mathcal{X} \to \mathbb{R}$ is called Φ-measurable if h is ϕ_{ij}-measurable for every $i, j = 1, \ldots, m$. The matrix-valued integral of h with respect to the measure Φ is defined by*

$$\int_{\mathcal{X}} h(\mathbf{x})\mathrm{d}\Phi(\mathbf{x}) \coloneqq \left[\int_{\mathcal{X}} h(\mathbf{x})\mathrm{d}\phi_{ij}(\mathbf{x})\right]_{i,j=1,\ldots,m} \in \mathbb{R}^{m\times m}.$$

Definition 2.12. *A* finitely atomic matrix-valued measure $\Phi \in \mathfrak{M}^m_+(\mathcal{X})$ *is a matrix-valued measure of form*

$$\Phi = \sum_{i=1}^{r} W_i \delta_{\mathbf{x}^{(i)}}, \tag{2.3}$$

where $W_i \in \mathbb{S}^m_+, i = 1, \ldots, r$, $\mathbf{x}^{(i)}$'s are distinct points in $\mathcal{X}$ and $\delta_{\mathbf{x}^{(i)}}$ denotes the Dirac measure centered at $\mathbf{x}^{(i)}$.

Clearly, for a finitely atomic matrix-valued measure $\Phi \in \mathfrak{M}^m_+(\mathcal{X})$ and a Φ-measurable function $h\colon \mathcal{X} \to \mathbb{R}$, it holds that

$$\int_{\mathcal{X}} h(\mathbf{x})\mathrm{d}\Phi(\mathbf{x}) = \sum_{i=1}^{r} W_i h(\mathbf{x}^{(i)}).$$

Definition 2.13. *We call a linear functional $\mathscr{L}\colon \mathbb{S}[\mathbf{x}]^m \to \mathbb{R}$ a tracial $\mathcal{X}$-moment functional if there exists a matrix-valued measure $\Phi \in \mathfrak{M}^m_+(\mathcal{X})$ such that*

$$\operatorname{supp}(\Phi) \subseteq \mathcal{X} \quad \textit{and} \quad \mathscr{L}(F) = \int_{\mathcal{X}} \operatorname{tr}\left(F(\mathbf{x})\mathrm{d}\Phi(\mathbf{x})\right), \quad \forall F(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^m. \tag{2.4}$$

The matrix-valued measure $\Phi \in \mathfrak{M}^m_+(\mathcal{X})$ is called a representing measure of $\mathscr{L}$.

For any $\Phi \in \mathfrak{M}^m_+(\mathcal{X})$, denote $\mathscr{L}_\Phi\colon \mathbb{S}[\mathbf{x}]^m \to \mathbb{R}$ the associated tracial $\mathcal{X}$-moment functional. For $F(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^m$, we write $F(\mathbf{x}) = \sum_{\boldsymbol{\alpha}\in\operatorname{supp}(F)} F_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}}$, where $F_{\boldsymbol{\alpha}}$ is the coefficient matrix of $\mathbf{x}^{\boldsymbol{\alpha}}$ in $F(\mathbf{x})$ and

$$\operatorname{supp}(F) \coloneqq \{\boldsymbol{\alpha} \in \mathbb{N}^n \mid \mathbf{x}^{\boldsymbol{\alpha}} \text{ appears in some } F_{ij}(\mathbf{x})\}.$$

We have, if $\mathscr{L}_\Phi\colon \mathbb{S}[\mathbf{x}]^m \to \mathbb{R}$ a tracial $\mathcal{X}$-moment functional, then

$$\mathscr{L}_\Phi(F) = \sum_{i=1}^m \sum_{j=1}^m \int_{\mathcal{X}} F_{ij}(\mathbf{x}) \mathrm{d}\phi_{ij}(\mathbf{x}) = \sum_{\boldsymbol{\alpha}\in \mathrm{supp}(F)} \mathrm{tr}\left(F_{\boldsymbol{\alpha}} \int_{\mathcal{X}} \mathbf{x}^{\boldsymbol{\alpha}} \mathrm{d}\Phi(\mathbf{x})\right). \tag{2.5}$$

We define the convex cones

$$\mathcal{L}^m(\mathcal{X}) := \{\mathscr{L}\colon \mathbb{S}[\mathbf{x}]^m \to \mathbb{R} \mid \mathscr{L} \text{ is a tracial } \mathcal{X}\text{-moment functional}\},$$

and

$$\mathcal{P}^m(\mathcal{X}) := \{F(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^m \mid F(\mathbf{x}) \succeq 0, \quad \forall \mathbf{x} \in \mathcal{X}\}. \tag{2.6}$$

Theorem 2.14 (Haviland's theorem for polynomial matrices). [12, Theorem 3] *For a linear functional $\mathscr{L}\colon \mathbb{S}[\mathbf{x}]^m \to \mathbb{R}$, $\mathscr{L} \in \mathcal{L}^m(\mathcal{X})$ if and only if $\mathscr{L}(F) \geq 0$ for all $F(\mathbf{x}) \in \mathcal{P}^m(\mathcal{X})$.*

Proposition 2.15. *The cones $\mathcal{L}^m(\mathcal{X})$ and $\mathcal{P}^m(\mathcal{X})$ are dual to each other, i.e., $\mathcal{L}^m(\mathcal{X}) = \mathcal{P}^m(\mathcal{X})^*$ and $\mathcal{P}^m(\mathcal{X}) = \mathcal{L}^m(\mathcal{X})^*$.*

Proof. Due to (2.5), it is clear that $\mathcal{L}^m(\mathcal{X}) \subseteq \mathcal{P}^m(\mathcal{X})^*$ and $\mathcal{P}^m(\mathcal{X}) \subseteq \mathcal{L}^m(\mathcal{X})^*$. Theorem 2.14 implies that $\mathcal{L}^m(\mathcal{X}) \supseteq \mathcal{P}^m(\mathcal{X})^*$. So we only need prove $\mathcal{P}^m(\mathcal{X}) \supseteq \mathcal{L}^m(\mathcal{X})^*$. Suppose that there exists a polynomial matrix $F(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^m$ such that $F(\mathbf{x}) \in \mathcal{L}^m(\mathcal{X})^*$ but $F(\mathbf{x}) \notin \mathcal{P}^m(\mathcal{X})$. Then there exists a point $\bar{\mathbf{x}} \in \mathcal{X}$ and a vector $\mathbf{v} \in \mathbb{R}^m$ such that $\mathbf{v}^\intercal F(\bar{\mathbf{x}})\mathbf{v} < 0$. Now define a linear functional $\mathscr{L}_{\bar{\Phi}} \in \mathcal{L}^m(\mathcal{X})$ with $\bar{\Phi} := \mathbf{v}\mathbf{v}^\intercal \delta_{\bar{\mathbf{x}}} \in \mathfrak{M}_+^m(\mathcal{X})$. It holds

$$\mathscr{L}_{\bar{\Phi}}(F) = \mathrm{tr}\left(F(\bar{\mathbf{x}})\mathbf{v}\mathbf{v}^\intercal\right) = \mathbf{v}^\intercal F(\bar{\mathbf{x}})\mathbf{v} < 0,$$

a contradiction. ■

2.4. The matrix-valued $\mathcal{X}$-moment problem. Let $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n}$ be a multi-indexed sequence of symmetric matrices in $\mathbb{S}^m$.

Definition 2.16. *[28] For a non-empty closed set $\mathcal{X} \subseteq \mathbb{R}^n$, the sequence $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n} \subset \mathbb{S}^m$ is called a matrix-valued $\mathcal{X}$-moment sequence if there exists a matrix-valued measure $\Phi \in \mathfrak{M}_+^m(\mathcal{X})$ such that*

$$\mathrm{supp}(\Phi) \subseteq \mathcal{X} \quad \textit{and} \quad S_{\boldsymbol{\alpha}} = \int_{\mathcal{X}} \mathbf{x}^{\boldsymbol{\alpha}} \mathrm{d}\Phi(\mathbf{x}),\ \forall \boldsymbol{\alpha} \in \mathbb{N}^n. \tag{2.7}$$

The matrix-valued measure $\Phi \in \mathfrak{M}_+^m(\mathcal{X})$ satisfying (2.7) is called a representing measure of $\mathbf{S}$.

For a given sequence $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n} \subset \mathbb{S}^m$, we can define a linear functional $\mathscr{L}_{\mathbf{S}}\colon \mathbb{S}[\mathbf{x}]^m \to \mathbb{R}$ in the following way:

$$\mathscr{L}_{\mathbf{S}}(F) := \sum_{\boldsymbol{\alpha}\in\mathrm{supp}(F)} \mathrm{tr}\left(F_{\boldsymbol{\alpha}} S_{\boldsymbol{\alpha}}\right),\ \forall F(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^m.$$

We call $\mathscr{L}_{\mathbf{S}}$ the *Riesz functional* associated to the sequence $\mathbf{S}$. Clearly, $\mathbf{S}$ is a matrix-valued $\mathcal{X}$-moment sequence if and only if $\mathscr{L}_{\mathbf{S}}$ is a tracial $\mathcal{X}$-moment functional.

Definition 2.17. *Given a sequence* $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n} \subset \mathbb{S}^m$*, the associated* moment matrix $M(\mathbf{S})$ *is the block matrix whose block rows and block columns are indexed by* $\mathbb{N}^n$ *and the* $(\boldsymbol{\alpha},\boldsymbol{\beta})$*-th block entry is* $S_{\boldsymbol{\alpha}+\boldsymbol{\beta}}$ *for all* $\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbb{N}^n$*. For* $G \in \mathbb{S}[\mathbf{x}]^q$*, the* localizing matrix $M(G\mathbf{S})$ *associated to* $\mathbf{S}$ *and* G *is the block matrix whose block rows and block columns are indexed by* $\mathbb{N}^n$ *and the* $(\boldsymbol{\alpha},\boldsymbol{\beta})$*-th block entry is* $\sum_{\boldsymbol{\gamma}\in\mathrm{supp}(G)} S_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}} \otimes G_{\boldsymbol{\gamma}}$ *for all* $\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbb{N}^n$*. Let* $d\in\mathbb{N}$*. The* d*-th order moment matrix* $M_d(\mathbf{S})$ *(resp. localizing matrix* $M_d(G\mathbf{S})$*) is the submatrix of* $M(\mathbf{S})$ *(resp.* $M(G\mathbf{S})$*) whose block row and block column are both indexed by* $\mathbb{N}^n_d$*.*

The following proposition can be easily verified from the definitions.

Proposition 2.18. *Let* $\Sigma_0(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^m$ *and* $\Sigma(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^{mq}$ *be SOS polynomial matrices such that*

$$\Sigma_0(\mathbf{x}) = (u_d(\mathbf{x})\otimes I_m)^\intercal Z_0 (u_d(\mathbf{x})\otimes I_m) \text{ and } \Sigma(\mathbf{x}) = (u_d(\mathbf{x})\otimes I_{mq})^\intercal Z (u_d(\mathbf{x})\otimes I_{mq}),$$

with $Z_0 \in \mathbb{S}_+^{m|\mathbb{N}^n_d|}$ *and* $Z \in \mathbb{S}_+^{mq|\mathbb{N}^n_d|}$*. Then for a sequence* $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n} \subset \mathbb{S}^m$*, it holds that*

$$\mathscr{L}_{\mathbf{S}}(\Sigma_0) = \mathrm{tr}\,(Z_0 M_d(\mathbf{S})) \text{ and } \mathscr{L}_{\mathbf{S}}((\Sigma, G)_m) = \mathrm{tr}\,(Z M_d(G\mathbf{S}))\,.$$

Let $d_G := \lceil \deg(G)/2 \rceil$. For each integer $k \geq d_G$, we define the sets

$$\mathcal{M}^m_k(G) := \{\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n_{2k}} \subset \mathbb{S}^m \mid M_k(\mathbf{S}) \succeq 0,\ M_{k-d_G}(G\mathbf{S}) \succeq 0\},$$

and let

$$\mathcal{M}^m(G) := \bigcap_{k\geq d_G} \mathcal{M}^m_k(G),$$

which are all convex cones. Checking membership in $\mathcal{M}^m_k(G)$ can be written as an SDP. Moreover, by Proposition 2.18, $\mathcal{M}^m_k(G)$ is the dual cone of $\mathcal{Q}^m_k(G)$.

For a given $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n} \subset \mathbb{S}^m$, the matrix-valued $\mathcal{X}$-moment problem asks when there exists a matrix-valued measure $\Phi \in \mathfrak{M}^m_+(\mathcal{X})$ satisfying the conditions in (2.7), which is addressed in the following theorem.

Theorem 2.19. *[12, Theorems 5 and 6] Let Assumption 2 hold. Given a sequence* $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n} \subset \mathbb{S}^m$*,* $\mathbf{S}$ *is a matrix-valued* $\mathcal{X}$*-moment sequence if and only if* $\mathbf{S} \in \mathcal{M}^m(G)$*.*

Proof. Suppose that $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n} \subset \mathbb{S}^m$ is a matrix-valued $\mathcal{X}$-moment sequence. By Definition 2.13 and Remark 2.2, $\mathscr{L}_{\mathbf{S}}(\Sigma_0) \geq 0$ and $\mathscr{L}_{\mathbf{S}}((\Sigma, G)_m) \geq 0$ for all SOS polynomial matrices $\Sigma_0(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^m$ and $\Sigma(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^{mq}$. Then by Proposition 2.18, $\mathbf{S} \in \mathcal{M}^m(G)$. The converse part can be obtained by combining Remark 2.2, Theorems 2.3, 2.14 and Proposition 2.18. ∎

Remark 2.20. If $m = 1$, we use the notation $\mathcal{Q}(G)$ (resp., $\mathcal{Q}_k(G)$, $\mathcal{M}(G)$, $\mathcal{M}_k(G)$) instead of $\mathcal{Q}^1(G)$ (resp., $\mathcal{Q}^1_k(G)$, $\mathcal{M}^1(G)$, $\mathcal{M}^1_k(G)$) for simplicity. For a set of polynomials $H(\mathbf{x}) = \{h_1(\mathbf{x}),\ldots,h_s(\mathbf{x})\} \subset \mathbb{R}[\mathbf{x}]$, by slightly abusing notation, we use $\mathcal{Q}^m(H)$, $\mathcal{Q}^m_k(H)$, $\mathcal{M}^m(H)$, $\mathcal{M}^m_k(H)$ to denote the related sets associated with the diagonal matrix $\mathrm{diag}\, h_1(\mathbf{x}),\ldots,h_s(\mathbf{x})$. Then, when $m = 1$, Theorems 2.3 and 2.19 recover Putinar's Positivstellensatz [44] and its dual aspect for the basic semi-algebraic set

$$\{\mathbf{x}\in\mathbb{R}^n \mid h_1(\mathbf{x}) \geq 0,\ldots,h_s(\mathbf{x}) \geq 0\}.$$

3. The flat extension condition and matrix-valued measure recovery.

3.1. The truncated matrix-valued $\mathcal{X}$-moment problem. Recently, Kimsey and Trachana [29] obtained a flat extension theorem which provides a solution to the truncated matrix-valued moment problem.

Theorem 3.1. [29, Theorem 6.2] (flat extension) *For a truncated sequence $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n_{2k}} \subset \mathbb{S}^m$, the following statements are equivalent:*

(i) $\mathbf{S}$ *admits an atomic representing measure* $\Phi = \sum_{i=1}^r W_i\delta_{\mathbf{x}^{(i)}}$ *with* $W_i \in \mathbb{S}^m_+$, $\mathbf{x}^{(i)} \in \mathbb{R}^n$ *and* $\sum_{i=1}^r \operatorname{rank}(W_i) = \operatorname{rank}(M_k(\mathbf{S}))$*;*

(ii) $M_k(\mathbf{S}) \succeq 0$ *and* $\mathbf{S}$ *admits an extension* $\tilde{\mathbf{S}} = (\tilde{S}_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n_{2k+2}}$ *such that* $M_{k+1}(\tilde{\mathbf{S}}) \succeq 0$ *and* $\operatorname{rank}(M_k(\mathbf{S})) = \operatorname{rank}(M_{k+1}(\tilde{\mathbf{S}}))$.

When $m = 1$, Theorem 3.1 recovers the celebrated flat extension theorem of Curto and Fialkow [14]. There is also a constrained version of the result of Curto and Fialkow [15]. As a matrix version of their result, we next extend Theorem 3.1 to matrix-valued measures supported on $\mathcal{X}$, which provides a solution to the truncated matrix-valued $\mathcal{X}$-moment problem.

Theorem 3.2. *Given a truncated sequence $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n_{2k}} \subset \mathbb{S}^m$, the following statements are equivalent:*

(i) $\mathbf{S}$ *admits an atomic representing measure* $\Phi = \sum_{i=1}^r W_i\delta_{\mathbf{x}^{(i)}}$ *with* $W_i \in \mathbb{S}^m_+$, $\mathbf{x}^{(i)} \in \mathcal{X}$ *and* $\sum_{i=1}^r \operatorname{rank}(W_i) = \operatorname{rank}(M_k(\mathbf{S}))$*;*

(ii) $M_k(\mathbf{S}) \succeq 0$ *and* $\mathbf{S}$ *admits an extension* $\tilde{\mathbf{S}} = (\tilde{S}_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n_{2(k+d_G)}}$ *such that* $M_{k+d_G}(\tilde{\mathbf{S}}) \succeq 0$, $M_k(G\tilde{\mathbf{S}}) \succeq 0$ *and* $\operatorname{rank}(M_k(\mathbf{S})) = \operatorname{rank}(M_{k+d_G}(\tilde{\mathbf{S}}))$.

Proof. (i)⇒(ii). It is implied by Theorems 2.19 and 3.1.

(ii)⇒(i). By Theorem 3.1, $\mathbf{S}$ admits a finitely atomic representing measure $\Phi = \sum_{i=1}^r W_i\delta_{\mathbf{x}^{(i)}}$ with $W_i \in \mathbb{S}^m_+$, $\mathbf{x}^{(i)} \in \mathbb{R}^n$ and $\sum_{i=1}^r \operatorname{rank}(W_i) = \operatorname{rank}(M_k(\mathbf{S}))$. We need to prove $\mathbf{x}^{(i)} \in \mathcal{X}$ for $i = 1, \ldots, r$.

By Theorem 3.1, we can extend $\tilde{\mathbf{S}}$ to an infinite sequence $\hat{\mathbf{S}} = (\hat{S}_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n}$ such that $M(\hat{\mathbf{S}}) \succeq 0$ and $\operatorname{rank}(M(\hat{\mathbf{S}})) = \operatorname{rank}(M_k(\mathbf{S}))$. For simplicity, in the following we will still use the symbol $\mathbf{S}$ to denote $\hat{\mathbf{S}}$.

For a column vector of polynomials $H(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]^m$, we write $H(\mathbf{x}) = \sum_{\boldsymbol{\gamma}\in\operatorname{supp}(H)} H_{\boldsymbol{\gamma}}\mathbf{x}^{\boldsymbol{\gamma}}$ with $H_{\boldsymbol{\gamma}} \in \mathbb{R}^m$. We define a subspace $\mathcal{I}_{\mathbf{S}}$ of $\mathbb{R}[\mathbf{x}]^m$ associated with $\mathbf{S}$ by

$$\mathcal{I}_{\mathbf{S}} := \left\{ H(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]^m \;\middle|\; \sum_{\boldsymbol{\gamma}\in\operatorname{supp}(H)} S_{\boldsymbol{\alpha}+\boldsymbol{\gamma}}H_{\boldsymbol{\gamma}} = 0,\ \forall \boldsymbol{\alpha} \in \mathbb{N}^n \right\}.$$

For any $H(\mathbf{x}) \in \mathcal{I}_{\mathbf{S}}$, we have $H(\mathbf{x}^{(i)})^\intercal W_i H(\mathbf{x}^{(i)}) = 0$ for all $i = 1, \ldots, r$. In fact, as $H(\mathbf{x}) \in \mathcal{I}_{\mathbf{S}}$, it holds that

$$0 = \sum_{\boldsymbol{\alpha}\in\operatorname{supp}(H)} \sum_{\boldsymbol{\beta}\in\operatorname{supp}(H)} H_{\boldsymbol{\alpha}}^\intercal S_{\boldsymbol{\alpha}+\boldsymbol{\beta}} H_{\boldsymbol{\beta}} = \sum_{i=1}^r H(\mathbf{x}^{(i)})^\intercal W_i H(\mathbf{x}^{(i)}).$$

As W_i's are PSD, it implies that

$$H(\mathbf{x}^{(i)})^\intercal W_i H(\mathbf{x}^{(i)}) = 0 \quad \text{and} \quad W_i H(\mathbf{x}^{(i)}) = 0. \tag{3.1}$$

Consider the quotient space over $\mathbb{R}$

$$\mathbb{R}[\mathbf{x}]^m/\mathcal{I}_{\mathbf{S}} \coloneqq \{H + \mathcal{I}_{\mathbf{S}} \mid H \in \mathbb{R}[\mathbf{x}]^m\}$$

consisting of equivalence classes modulo $\mathcal{I}_{\mathbf{S}}$. Let $t \coloneqq \operatorname{rank}(M(\mathbf{S})) = \operatorname{rank}(M_k(\mathbf{S}))$. Let $\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(t)} \in \mathbb{N}^n_k$ (not necessarily distinct) and standard basis (column) vectors $\mathbf{e}^{(1)}, \ldots, \mathbf{e}^{(t)}$ (not necessarily distinct) of $\mathbb{R}^m$ be such that

$$\left\{\operatorname{col}\left((S_{\boldsymbol{\alpha}+\boldsymbol{\beta}^{(1)}})_{\boldsymbol{\alpha}\in\mathbb{N}^n}\right)\mathbf{e}^{(1)}, \ldots, \operatorname{col}\left((S_{\boldsymbol{\alpha}+\boldsymbol{\beta}^{(t)}})_{\boldsymbol{\alpha}\in\mathbb{N}^n}\right)\mathbf{e}^{(t)}\right\} \tag{3.2}$$

is a set of t linearly independent column vectors of $M(\mathbf{S})$ and hence forms a basis of the column space of $M(\mathbf{S})$. Here, $\operatorname{col}\left((S_{\boldsymbol{\alpha}+\boldsymbol{\beta}^{(i)}})_{\boldsymbol{\alpha}\in\mathbb{N}^n}\right)$ denotes the column vector with entries $(S_{\boldsymbol{\alpha}+\boldsymbol{\beta}^{(i)}})_{\boldsymbol{\alpha}\in\mathbb{N}^n}$. We claim that the set

$$\left\{\mathbf{x}^{\boldsymbol{\beta}^{(1)}}\mathbf{e}^{(1)} + \mathcal{I}_{\mathbf{S}},\ \ldots,\ \mathbf{x}^{\boldsymbol{\beta}^{(t)}}\mathbf{e}^{(t)} + \mathcal{I}_{\mathbf{S}}\right\} \tag{3.3}$$

forms a basis of $\mathbb{R}[\mathbf{x}]^m/\mathcal{I}_{\mathbf{S}}$. To see this, first note that the elements in (3.3) are linearly independent as the elements in (3.2) are linearly independent. Then, it is sufficient to prove that for arbitrary $\boldsymbol{\gamma} \in \mathbb{N}^n$ and $j \in \mathbb{N}$ with $1 \le j \le t$, the element $\mathbf{x}^{\boldsymbol{\gamma}}\mathbf{e}^{(j)} + \mathcal{I}_{\mathbf{S}}$ can be written as a linear combination of elements in (3.3). This is indeed true since the column vector $\operatorname{col}\left((S_{\boldsymbol{\alpha}+\boldsymbol{\gamma}})_{\boldsymbol{\alpha}\in\mathbb{N}^n}\right)\mathbf{e}^{(j)}$ can be written as a linear combination of elements in (3.2). For any $H(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]^m$, we write $H(\mathbf{x}) = H^{(0)}(\mathbf{x}) + H^{(1)}(\mathbf{x})$ where $H^{(0)}$ is the residue of H modulo $\mathcal{I}_{\mathbf{S}}$ w.r.t. the basis (3.3) and $H^{(1)} \in \mathcal{I}_{\mathbf{S}}$. Since $\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(t)} \in \mathbb{N}^n_k$, we see that $\deg(H^{(0)}) \le k$.

Let $\{p^{(i)}(\mathbf{x})\}_{i=1}^r$ be the Lagrange interpolation polynomials at $\{\mathbf{x}^{(i)}\}_{i=1}^r$ such that $p^{(i)}(\mathbf{x}^{(i)}) = 1$ and $p^{(i)}(\mathbf{x}^{(j)}) = 0$ for all $j \neq i$. Now we fix an i and prove $\mathbf{x}^{(i)} \in \mathcal{X}$. As $W_i \succeq 0$, there exists a vector $\mathbf{v}^{(i)} \in \mathbb{R}^m$ such that $(\mathbf{v}^{(i)})^\intercal W_i \mathbf{v}^{(i)} > 0$. Let $H_i(\mathbf{x}) = p^{(i)}(\mathbf{x})\mathbf{v}^{(i)} \in \mathbb{R}[\mathbf{x}]^m$. Let us write $H_i = H_i^{(0)} + H_i^{(1)}$ and $H_i^{(0)}(\mathbf{x}) = \sum_{\boldsymbol{\alpha}\in\operatorname{supp}(H_i^{(0)})} H_{i,\boldsymbol{\alpha}}^{(0)}\mathbf{x}^{\boldsymbol{\alpha}}$ with $\operatorname{supp}(H_i^{(0)}) \subseteq \mathbb{N}^n_k$. As $M_k(G\mathbf{S}) \succeq 0$, we have

$$\sum_{\boldsymbol{\alpha}\in\operatorname{supp}(H_i^{(0)})}\ \sum_{\boldsymbol{\beta}\in\operatorname{supp}(H_i^{(0)})} \left((H_{i,\boldsymbol{\alpha}}^{(0)})^\intercal \otimes I_q\right)[M_k(G\mathbf{S})]_{\boldsymbol{\alpha}\boldsymbol{\beta}}\left(H_{i,\boldsymbol{\beta}}^{(0)} \otimes I_q\right) \succeq 0.$$

By the definition of $M_k(G\mathbf{S})$, we have

$$\begin{aligned}
&\sum_{\boldsymbol{\alpha}\in\mathrm{supp}(H_i^{(0)})}\sum_{\boldsymbol{\beta}\in\mathrm{supp}(H_i^{(0)})}\left((H_{i,\boldsymbol{\alpha}}^{(0)})^\intercal\otimes I_q\right)M_k(G\mathbf{S})_{\boldsymbol{\alpha}\boldsymbol{\beta}}\left(H_{i,\boldsymbol{\beta}}^{(0)}\otimes I_q\right)\\
=&\sum_{\boldsymbol{\alpha}\in\mathrm{supp}(H_i^{(0)})}\sum_{\boldsymbol{\beta}\in\mathrm{supp}(H_i^{(0)})}\left((H_{i,\boldsymbol{\alpha}}^{(0)})^\intercal\otimes I_q\right)\left(\sum_{\boldsymbol{\gamma}\in\mathrm{supp}(G)}S_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}}\otimes G_{\boldsymbol{\gamma}}\right)\left(H_{i,\boldsymbol{\beta}}^{(0)}\otimes I_q\right)\\
=&\sum_{\boldsymbol{\gamma}\in\mathrm{supp}(G)}\left(\sum_{\boldsymbol{\alpha}\in\mathrm{supp}(H_i^{(0)})}\sum_{\boldsymbol{\beta}\in\mathrm{supp}(H_i^{(0)})}(H_{i,\boldsymbol{\alpha}}^{(0)})^\intercal S_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}}H_{i,\boldsymbol{\beta}}^{(0)}\right)G_{\boldsymbol{\gamma}}\\
=&\left(\sum_{j=1}^{r}H_i^{(0)}(\mathbf{x}^{(j)})^TW_jH_i^{(0)}(\mathbf{x}^{(j)})\right)G(\mathbf{x}^{(j)})\\
=&(\mathbf{v}^{(i)}-H_i^{(1)}(\mathbf{x}^{(i)}))^TW_i(\mathbf{v}^{(i)}-H_i^{(1)}(\mathbf{x}^{(i)}))G(\mathbf{x}^{(i)})\\
&+\sum_{j\neq i}H_i^{(1)}(\mathbf{x}^{(j)}))^TW_jH_i^{(1)}(\mathbf{x}^{(j)})G(\mathbf{x}^{(j)})\\
=&((\mathbf{v}^{(i)})^\intercal W_i\mathbf{v}^{(i)})G(\mathbf{x}^{(i)}),
\end{aligned}$$

where the last equality is due to (3.1) and $H_i^{(1)}\in\mathcal{I}$. As $(\mathbf{v}^{(i)})^\intercal W_i\mathbf{v}^{(i)}>0$, we have $G(\mathbf{x}^{(i)})\succeq 0$, which implies $\mathbf{x}^{(i)}\in\mathcal{X}$. ■

3.2. Matrix-valued measure recovery. For a truncated sequence $\mathbf{S}=(S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}_{2k}^n}\subset\mathbb{S}^m$ with $k\geq d_G$, suppose that $M_k(\mathbf{S})\succeq 0$, $M_{k-d_G}(G\mathbf{S})\succeq 0$ and $\mathrm{rank}(M_k(\mathbf{S}))=\mathrm{rank}(M_{k-d_G}(\mathbf{S}))$. By Theorem 3.2, $\mathbf{S}$ admits a finitely atomic representing measure $\Phi=\sum_{i=1}^r W_i\delta_{\mathbf{x}^{(i)}}$ with $W_i\in\mathbb{S}_+^m$, $\mathbf{x}^{(i)}\in\mathcal{X}$ and $\sum_{i=1}^r\mathrm{rank}(W_i)=\mathrm{rank}(M_k(\mathbf{S}))$. In theory, it was shown in [29] that the points $\{\mathbf{x}^{(i)}\}_i$ can be computed via the intersecting zeros of the determinants of matrix-valued polynomials describing the flat extension, which, however, is not easy to compute in practice. In this subsection, we provide a linear algebra procedure to extract $\mathbf{x}^{(i)}\in\mathcal{X}$ and $W_i\in\mathbb{S}^m$, which is an extension of the approach proposed in [22] for the scalar case.

From the definition of $M_k(\mathbf{S})$, it holds

$$M_k(\mathbf{S})=\sum_{i=1}^r(u_k(\mathbf{x}^{(i)})\otimes I_m)W_i(u_k(\mathbf{x}^{(i)})\otimes I_m)^\intercal=VV^\intercal,$$

where $u_k(\mathbf{x}^{(i)})$ is defined in (2.1) and

$$V=\left[(u_k(\mathbf{x}^{(1)})\otimes I_m)\sqrt{W_1},\ldots,(u_k(\mathbf{x}^{(r)})\otimes I_m)\sqrt{W_r}\right].$$

Let $M_k(\mathbf{S})=\widetilde{V}\widetilde{V}^\intercal$ be a Cholesky decomposition of $M_k(\mathbf{S})$ with $\widetilde{V}\in\mathbb{R}^{m|\mathbb{N}_k^n|\times t}$ and $t=\mathrm{rank}(M_k(\mathbf{S}))$. Notice that V and $\widetilde{V}$ span the same column space. We will recover $\mathbf{x}^{(i)}$ by suitable column operations on $\widetilde{V}$.

Let us write $\sqrt{W_i} = [\mathbf{w}^{(i,1)}, \ldots, \mathbf{w}^{(i,m_i)}]$ with $\mathbf{w}^{(i,j)} \in \mathbb{R}^m$ and $\sum_{i=1}^r m_i = t$. Then each column of V is of form $u_k(\mathbf{x}^{(i)}) \otimes \mathbf{w}^{(i,j)}$ and can be generated by the columns of $\widetilde{V}$. Now we treat the entries in the vectors $\mathbf{w}^{(i,j)}$ as variables and denote it by $\mathbf{w} = (w_1, \ldots, w_m)$. Then, the rows in V correspond to the monomials

$$v_k(\mathbf{x}, \mathbf{w}) = [\mathbf{w},\ x_1\mathbf{w},\ x_2\mathbf{w},\ \cdots,\ x_n\mathbf{w},\ x_1^2\mathbf{w},\ x_1x_2\mathbf{w},\ \cdots,\ x_n^k\mathbf{w}]^\intercal.$$

Reduce the matrix $\widetilde{V}$ to the column echelon form U:

$$U = \begin{bmatrix} 1 & & & & \\ \star & & & & \\ 0 & 1 & & & \\ 0 & 0 & 1 & & \\ \star & \star & \star & & \\ & \vdots & & \ddots & \\ 0 & 0 & 0 & \cdots & 1 \\ \star & \star & \star & \cdots & \star \\ & \vdots & & & \vdots \\ \star & \star & \star & \cdots & \star \end{bmatrix}.$$

From the rows of U where the pivot elements locate, we obtain a (column) monomial basis $b_k(\mathbf{x}, \mathbf{w})$ which consists of t monomials in $v_k(\mathbf{x}, \mathbf{w})$ such that

$$v_k(\mathbf{x}, \mathbf{w}) = U b_k(\mathbf{x}, \mathbf{w}) \tag{3.4}$$

holds at each pair of $(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})$, $j = 1, \ldots, m_i$, $i = 1, \ldots, r$. Note that each monomial $\mathbf{x}^{\boldsymbol{\alpha}} w_j$ in $b_k(\mathbf{x}, \mathbf{w})$ satisfies $|\boldsymbol{\alpha}| \le k - d_G$ since $\operatorname{rank}(M_k(\mathbf{S})) = \operatorname{rank}(M_{k-d_G}(\mathbf{S}))$.

Proposition 3.3. *The vectors $b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})$, $j = 1, \ldots, m_i$, $i = 1, \ldots, r$, are linearly independent.*

Proof. Case 1: $r - 1 \le k$. Suppose on the contrary that $b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})$, $j = 1, \ldots, m_i$, $i = 1, \ldots, r$ are linearly dependent. Then there exist constants $c_{i,j}$'s, not all zeros, such that

$$\sum_{i=1}^r \sum_{j=1}^{m_i} c_{i,j} b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)}) = 0.$$

Because $\mathbf{x}^{(i)}$'s are distinct, we can construct the Lagrange interpolation polynomials $p^{(i)}(\mathbf{x})$'s at $\mathbf{x}^{(i)}$'s such that $p^{(i)}(\mathbf{x}^{(i)}) = 1$ and $p^{(i)}(\mathbf{x}^{(j)}) = 0$ for all $i \ne j$. Now we fix an i' with $1 \le i' \le r$, and consider the column vector of polynomials $\mathbf{w}p^{(i')}(\mathbf{x}) \in \mathbb{R}[\mathbf{x}, \mathbf{w}]$. As $\deg(p^{(i)}) = r - 1 \le k$, due to (3.4), there exists a coefficient matrix $\Xi \in \mathbb{R}^{m \times t}$ such that $\mathbf{w}p^{(i')}(\mathbf{x}) = \Xi b_k(\mathbf{x}, \mathbf{w})$ holds at each pair of $(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})$, $j = 1, \ldots, m_i$, $i = 1, \ldots, r$. Then, we have

$$0 = \Xi \left(\sum_{i=1}^r \sum_{j=1}^{m_i} c_{i,j} b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)}) \right) = \sum_{i=1}^r \sum_{j=1}^{m_i} c_{i,j} \mathbf{w}^{(i,j)} p^{(i')}(\mathbf{x}^{(i)}) = \sum_{j=1}^{m_{i'}} c_{i',j} \mathbf{w}^{(i',j)}.$$

As $\mathbf{w}^{(i',j)}$'s are linearly independent, we have $c_{i',j} = 0$ for all $j = 1, \ldots, m_i$. This leads to a contradiction since i' is arbitrary.

Case 2: $r - 1 > k$. According to Theorem 3.1, $\mathbf{S}$ admits a flat extension $\tilde{\mathbf{S}} = (\tilde{S}_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n_{2r-2}}$ such that $M_{r-1}(\tilde{\mathbf{S}}) \succeq 0$ and $\operatorname{rank}(M_k(\mathbf{S})) = \operatorname{rank}(M_{r-1}(\tilde{\mathbf{S}}))$. By repeating the previous arguments on $M_{r-1}(\tilde{\mathbf{S}})$, we still obtain a column echelon form $\tilde{U}$ and a monomial basis $b_{r-1}(\mathbf{x}, \mathbf{w})$ such that $v_{r-1}(\mathbf{x}, \mathbf{w}) = \tilde{U} b_{r-1}(\mathbf{x}, \mathbf{w})$ holds at all pairs of $(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})$'s. Note that as $M_{r-1}(\tilde{\mathbf{S}})$ is a flat extension of $M_k(\mathbf{S})$, it is easy to see that the basis $b_{r-1}(\mathbf{x}, \mathbf{w})$ is identical with $b_k(\mathbf{x}, \mathbf{w})$. Now as in Case 1, we can show that $b_{r-1}(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})$ and hence $b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})$, $j = 1, \ldots, m_i$, $i = 1, \ldots, r$, are linearly independent. ■

Recall that each monomial $\mathbf{x}^{\boldsymbol{\alpha}} w_j$ in $b_k(\mathbf{x}, \mathbf{w})$ satisfies $|\boldsymbol{\alpha}| \le k - d_G < k$. Hence, for each $q = 1, \ldots, n$, we can extract from U the $t \times t$ multiplication matrix N_q such that $N_q b_k(\mathbf{x}, \mathbf{w}) = x_q b_k(\mathbf{x}, \mathbf{w})$ holds at each pair of $(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})$, $j = 1, \ldots, m_i$, $i = 1, \ldots, r$.

Following [13], we build a random combination of multiplication matrices $N = \sum_{q=1}^n c_q N_q$, where $c_q \ge 0$ and $\sum_{q=1}^n c_q = 1$. Let $N = ATA^\intercal$ be the ordered Schur decomposition of N, where $A = [a_1, \ldots, a_t]$ is an orthogonal matrix with $A^\intercal A = I_t$ and T is upper-triangular with eigenvalues of N being sorted increasingly along the diagonal.

Proposition 3.4. *Suppose that the constants c_q's are chosen such that $h(\mathbf{x}) = \sum_{q=1}^n c_q x_q$ takes distinct values on $\mathbf{x}^{(i)}$, $i = 1, \ldots, r$. Then the set of points*

$$\{(a_1^\intercal N_1 a_1, \ldots, a_1^\intercal N_n a_1),\ \ldots,\ (a_t^\intercal N_1 a_t, \ldots, a_t^\intercal N_n a_t)\} \tag{3.5}$$

is exactly $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(r)}\}$ and each $\mathbf{x}^{(i)}$ appears $m_i = \operatorname{rank}(W_i)$ times.

Proof. Since $N_q b_k(\mathbf{x}, \mathbf{w}) = x_q b_k(\mathbf{x}, \mathbf{w})$ holds at each pair of $(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})$, $j = 1, \ldots, m_i$, $i = 1, \ldots, r$. It is easy to see that

$$h(\mathbf{x}^{(i)}) b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)}) = \sum_{q=1}^n c_q x_q^{(i)} b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)}) = N b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)}),$$

for each $j = 1, \ldots, m_i$, $i = 1, \ldots, r$. In other words, for each $i = 1, \ldots, r$, $h(\mathbf{x}^{(i)})$ is an eigenvalue of N and $b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})$, $j = 1, \ldots, m_i$ are the associated eigenvectors. By Proposition 3.3, the t vectors $b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})$, $j = 1, \ldots, m_i$, $i = 1, \ldots, r$ are linearly independent. Therefore, $\{h(\mathbf{x}^{(1)}), \ldots, h(\mathbf{x}^{(r)})\}$ is exactly the set of eigenvalues of N, and $\{b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)}), j = 1, \ldots, m_i\}$ spans the eigenspace of N associated with $h(\mathbf{x}^{(i)})$. So, we can divide the set $\{a_1, \ldots, a_t\}$ into r groups $\mathcal{A}_1, \ldots, \mathcal{A}_r$ with $|\mathcal{A}_i| = m_i$, such that $\mathcal{A}_i$ spans the eigenspace of N associated with $h(\mathbf{x}^{(i)})$. Now fix an i and a vector $a \in \mathcal{A}_i$. There exist weights $\lambda_1, \ldots, \lambda_{m_i} \in \mathbb{R}$ such that

$$a = \sum_{j=1}^{m_i} \lambda_j b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)}).$$

Then, for each $q = 1, \dots, n$, it holds

$$\begin{aligned} a^\intercal N_q a &= \left(\sum_{j=1}^{m_i} \lambda_j b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})\right)^\intercal N_q \left(\sum_{j=1}^{m_i} \lambda_j b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})\right) \\ &= \left(\sum_{j=1}^{m_i} \lambda_j b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})\right)^\intercal \left(\sum_{j=1}^{m_i} \lambda_j x_q^{(i)} b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})\right) \\ &= x_q^{(i)} a^\intercal a = x_q^{(i)}. \end{aligned}$$

Hence, $(a^\intercal N_1 a, \dots, a^\intercal N_n a) = \mathbf{x}^{(i)}$. The conclusion then follows. ■

Once the points $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(r)}$ are obtained, let

$$\Lambda \coloneqq [u_k(\mathbf{x}^{(1)}), \dots, u_k(\mathbf{x}^{(r)})] \otimes I_m \in \mathbb{R}^{m|\mathbb{N}_k^n| \times mr},$$

and we have

$$M_k(\mathbf{S}) = \Lambda \ \operatorname{diag} W_1, \dots, W_r \ \Lambda^\intercal. \tag{3.6}$$

Notice that the first m columns of $\operatorname{diag} W_1, \dots, W_r \Lambda^\intercal$ is exactly $[W_1, \dots, W_r]^\intercal$. By comparing the first m columns of both sides of (3.6), we get

$$\operatorname{col}(\{S_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathbb{N}_k^n}) = \Lambda [W_1, \dots, W_r]^\intercal. \tag{3.7}$$

Assume that Λ has mr independent rows (see Remark 3.5) and let $\mathcal{R}$ be the index set of these rows. Denote by $\Lambda_{\mathcal{R}}$ (resp. $M_{\mathcal{R}}(\mathbf{S})$) the $mr \times mr$ (resp. $mr \times m$) submatrix of Λ (resp. $\operatorname{col}(\{S_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathbb{N}_k^n})$) whose rows are indexed by $\mathcal{R}$. Then by extracting the rows indexed $\mathcal{R}$ from both sides of (3.7), we have

$$M_{\mathcal{R}}(\mathbf{S}) = \Lambda_{\mathcal{R}} [W_1, \dots, W_r]^\intercal.$$

Hence, the matrices W_i's can be retrieved by

$$[W_1, \dots, W_r]^\intercal = \Lambda_{\mathcal{R}}^{-1} M_{\mathcal{R}}(\mathbf{S}).$$

We provide an example illustrating the above procedure in Appendix A.

Remark 3.5. It is clear that if

$$\operatorname{rank}([u_k(\mathbf{x}^{(1)}), \dots, u_k(\mathbf{x}^{(r)})]) = r, \tag{3.8}$$

then Λ must have mr independent rows. As $\mathbf{x}^{(i)}$'s are distinct, by considering the Lagrange interpolation polynomials at $\mathbf{x}^{(i)}$'s, we know that (3.8) always holds if $k \geq r - 1$. If $k < r - 1$, then it is possible that (3.8) fails and we may need to consider flat extensions of $M_k(\mathbf{S})$ to recover the weights W_i's.

4. A moment-SOS hierarchy for (RPMIO) with SOS-convexity. In this section, we first reformulate (RPMIO) as a conic optimization problem, based on which we can then derive a moment-SOS hierarchy whose optima monotonically converge to the optimum of (RPMIO). Furthermore, the results in Section 3 enable us to detect finite convergence of the moment-SOS hierarchy and to extract optimal solutions.

4.1. A conic reformulation. For simplicity, we write

$$P(\mathbf{y},\mathbf{x}) = \sum_{\boldsymbol{\alpha}\in\mathbb{N}^n} P_{\boldsymbol{\alpha}}(\mathbf{y})\mathbf{x}^{\boldsymbol{\alpha}} = \sum_{\boldsymbol{\beta}\in\mathbb{N}^\ell} P_{\boldsymbol{\beta}}(\mathbf{x})\mathbf{y}^{\boldsymbol{\beta}},$$

where $P_{\boldsymbol{\alpha}}(\mathbf{y}) \in \mathbb{S}[\mathbf{y}]^m$ (resp. $P_{\boldsymbol{\beta}}(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^m$) is the coefficient matrix of $\mathbf{x}^{\boldsymbol{\alpha}}$ (resp. $\mathbf{y}^{\boldsymbol{\beta}}$) with $P(\mathbf{x},\mathbf{y})$ being regarded as a polynomial matrix in $\mathbb{S}[\mathbf{x}]^m$ (resp. $\mathbb{S}[\mathbf{y}]^m$). For a linear functional $\mathscr{L}\colon \mathbb{S}[\mathbf{x}]^m \to \mathbb{R}$, we denote

$$\mathscr{L}(P(\mathbf{y},\mathbf{x})) \coloneqq \sum_{\boldsymbol{\beta}\in\mathbb{N}^\ell} \mathscr{L}(P_{\boldsymbol{\beta}}(\mathbf{x}))\mathbf{y}^{\boldsymbol{\beta}} \in \mathbb{R}[\mathbf{y}],$$

and for a linear functional $\mathscr{H}\colon \mathbb{R}[\mathbf{y}] \to \mathbb{R}$, we denote

$$\mathscr{H}(P(\mathbf{y},\mathbf{x})) \coloneqq \sum_{\boldsymbol{\beta}\in\mathbb{N}^\ell} P_{\boldsymbol{\beta}}(\mathbf{x})\mathscr{H}(\mathbf{y}^{\boldsymbol{\beta}}) \in \mathbb{S}[\mathbf{x}]^m.$$

To obtain a conic reformulation of (RPMIO), we need to assume the Slater condition to hold.

Assumption 3. *The Slater condition holds for* (RPMIO)*, i.e., there exists* $\bar{\mathbf{y}} \in \mathcal{Y}$ *such that* $\theta_i(\bar{\mathbf{y}}) > 0$ *for all* $i = 1,\ldots,s$*, and* $P(\bar{\mathbf{y}},\mathbf{x}) \succ 0$ *for all* $\mathbf{x} \in \mathcal{X}$*.*

Proposition 4.1. *Under Assumptions 1 and 3, there exists a finitely atomic matrix-valued measure* $\Phi^\star \in \mathfrak{M}_+^m(\mathcal{X})$ *such that*

$$f^\star = \inf_{\mathbf{y}\in\mathcal{Y}} f(\mathbf{y}) - \mathscr{L}_{\Phi^\star}(P(\mathbf{y},\mathbf{x})).$$

If $\mathbf{y}^\star$ *is an optimal solution to* (RPMIO)*, then* $\mathscr{L}_{\Phi^\star}(P(\mathbf{y}^\star,\mathbf{x})) = 0$*.*

Proof. For any $\mathbf{v} \in \mathbf{V} \coloneqq \{\mathbf{v} \in \mathbb{R}^m \mid \sum_{i=1}^m v_i^2 = 1\}$ and $\mathbf{x} \in \mathcal{X}$, it is easy to see that the function $-\mathbf{v}^\intercal P(\cdot,\mathbf{x})\mathbf{v}$ is convex in $\mathbf{y}$. Then, (RPMIO) can be equivalently reformulated as the convex semi-infinite program:

$$f^\star = \inf_{\mathbf{y}\in\mathcal{Y}} f(\mathbf{y}) \quad \text{s.t. } \mathbf{v}^\intercal P(\mathbf{y},\mathbf{x})\mathbf{v} \geq 0,\ \forall(\mathbf{x},\mathbf{v}) \in \mathcal{X}\times\mathbf{V}. \tag{4.1}$$

Let $(\mathbf{x}^{(0)},\mathbf{v}^{(0)}),(\mathbf{x}^{(1)},\mathbf{v}^{(1)}),\ldots,(\mathbf{x}^{(\ell)},\mathbf{v}^{(\ell)})$ be $\ell+1$ arbitrary points in $\mathcal{X}\times\mathbf{V}$. By Assumption 3, there exists $\bar{\mathbf{y}} \in \mathcal{Y}$ such that $P(\bar{\mathbf{y}},\mathbf{x}^{(i)}) \succ 0$ for all $i = 0,1,\ldots,\ell$. Hence, it holds $(\mathbf{v}^{(i)})^\intercal P(\bar{\mathbf{y}},\mathbf{x}^{(i)})\mathbf{v}^{(i)} > 0$ for all $i = 0,1,\ldots,\ell$. Notice that $\mathcal{X}\times\mathbf{V}$ is compact in $\mathbb{R}^n\times\mathbb{R}^m$. By applying [9, Theorem 4.1] to (4.1), we can find ℓ points $(\tilde{\mathbf{x}}^{(1)},\tilde{\mathbf{v}}^{(1)}),\ldots,(\tilde{\mathbf{x}}^{(\ell)},\tilde{\mathbf{v}}^{(\ell)}) \in \mathcal{X}\times\mathbf{V}$, and $\lambda_1,\ldots,\lambda_\ell > 0$ such that

$$f^\star = \inf_{\mathbf{y}\in\mathcal{Y}} f(\mathbf{y}) - \sum_{i=1}^{\ell} \lambda_i(\tilde{\mathbf{v}}^{(i)})^\intercal P(\mathbf{y},\tilde{\mathbf{x}}^{(i)})\tilde{\mathbf{v}}^{(i)}.$$

Define a finitely atomic matrix-valued measure

$$\Phi^\star \coloneqq \sum_{i=1}^{\ell} \lambda_i\tilde{\mathbf{v}}^{(i)}(\tilde{\mathbf{v}}^{(i)})^\intercal\delta_{\tilde{\mathbf{x}}^{(i)}} \in \mathfrak{M}_+^m(\mathcal{X}).$$

Then, it holds that

$$f^\star = \inf_{\mathbf{y}\in\mathcal{Y}} f(\mathbf{y}) - \mathscr{L}_{\Phi^\star}(P(\mathbf{y},\mathbf{x})).$$

If $\mathbf{y}^\star$ is an optimal solution to (RPMIO), then $P(\mathbf{y}^\star,\mathbf{x}) \succeq 0$ on $\mathcal{X}$ and hence

$$f^\star \le f(\mathbf{y}^\star) - \mathscr{L}_{\Phi^\star}(P(\mathbf{y}^\star,\mathbf{x})) \le f(\mathbf{y}^\star) = f^\star,$$

which implies that $\mathscr{L}_{\Phi^\star}(P(\mathbf{y}^\star,\mathbf{x})) = 0$. ■

Remark 4.2. From the above proof, one can see that Proposition 4.1 remains true if the Slater condition is weakened as: For any $\ell+1$ points $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(\ell)} \in \mathcal{X}$, there exists $\bar{\mathbf{y}} \in \mathcal{Y}$ such that $P(\bar{\mathbf{y}}, \mathbf{x}^{(i)}) \succ 0$ for all $i = 0, 1, \ldots, \ell$.

For any measure $\mu \in \mathfrak{m}_+(\mathcal{Y})$, we define an associated linear functional $\mathscr{H}_\mu\colon \mathbb{R}[\mathbf{y}] \to \mathbb{R}$ by $\mathscr{H}_\mu(h) = \int_{\mathcal{Y}} h(\mathbf{y})\mathrm{d}\mu(\mathbf{y})$ for all $h \in \mathbb{R}[\mathbf{y}]$. Let us consider the following conic optimization problem:

$$\text{(4.2)}\qquad \begin{cases} \tilde{f} \coloneqq \sup\limits_{\rho,\Phi}\ \rho \\ \quad\ \text{s.t.}\ f(\mathbf{y}) - \rho - \mathscr{L}_\Phi(P(\mathbf{y},\mathbf{x})) \in \mathcal{P}(\mathcal{Y}), \\ \qquad\quad \rho \in \mathbb{R},\ \ \Phi \in \mathfrak{M}_+^m(\mathcal{X}), \end{cases}$$

whose dual reads as

$$\text{(4.3)}\qquad \begin{cases} \hat{f} \coloneqq \inf\limits_{\mu}\ \mathscr{H}_\mu(f) \\ \quad\ \text{s.t.}\ \mu \in \mathfrak{m}_+(\mathcal{Y}),\ \ \mathscr{H}_\mu(1) = 1, \\ \qquad\quad \mathscr{H}_\mu(P(\mathbf{y},\mathbf{x})) \in \mathcal{P}^m(\mathcal{X}), \end{cases}$$

with $\mathcal{P}^m(\mathcal{X})$ being defined in (2.6).

Theorem 4.3. *Under Assumptions 1 and 3, it holds $\tilde{f} = \hat{f} = f^\star$.*

Proof. Let $\Phi^\star \in \mathfrak{M}_+^m(\mathcal{X})$ be the finitely atomic matrix-valued measure given in Proposition 4.1. Then $(f^\star, \Phi^\star)$ is feasible to (4.2) due to Proposition 4.1. Thus, $\tilde{f} \ge f^\star$. By the weak duality, we have $\hat{f} \ge \tilde{f}$ and remain to show $\hat{f} \le f^\star$.

Let $(\mathbf{y}^{(k)})_{k\in\mathbb{N}}$ be a minimizing sequence of (RPMIO). Then, for any $\varepsilon > 0$, there exists $k_\varepsilon \in \mathbb{N}$ such that $\mathbf{y}^{(k_\varepsilon)}$ is feasible to (RPMIO) and $f(\mathbf{y}^{(k_\varepsilon)}) \le f^\star + \varepsilon$. The Dirac measure $\delta_{\mathbf{y}^{(k_\varepsilon)}}$ centered at $\mathbf{y}^{(k_\varepsilon)}$ is feasible to (4.3). Therefore, $\hat{f} \le \mathscr{H}_{\delta_{\mathbf{y}^{(k_\varepsilon)}}}(f) = f(\mathbf{y}^{(k_\varepsilon)}) \le f^\star + \varepsilon$. As $\varepsilon > 0$ is arbitrary, we have $\hat{f} \le f^\star$ as desired. ■

Proposition 4.4. *Under Assumptions 1 and 3, suppose that $(\rho^\star, \Phi^\star)$ is an optimal solution to (4.2) such that $\Phi^\star = \sum_{i=1}^r W_i \delta_{\mathbf{x}^{(i)}} \in \mathfrak{M}_+^m(\mathcal{X})$ for some $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(r)} \in \mathcal{X}$ and $W_1, \ldots, W_r \in \mathbb{S}_+^m$. If $\mathbf{y}^\star$ is an optimal solution to (RPMIO), then for any decomposition $W_i = \sum_{k=1}^{m_i} \mathbf{v}^{(i,k)}(\mathbf{v}^{(i,k)})^\intercal$, $\mathbf{v}^{(i,k)} \in \mathbb{R}^m$, $i = 1, \ldots, r$, it holds that*

$$P(\mathbf{y}^\star, \mathbf{x}^{(i)})\mathbf{v}^{(i,k)} = 0, \quad k = 1, \ldots, m_i,\ i = 1, \ldots, r.$$

Proof. By Theorem 4.3, we have $\rho^\star = f^\star$ and hence

$$f(\mathbf{y}) - f^\star - \mathscr{L}_{\Phi^\star}(P(\mathbf{y},\mathbf{x})) = f(\mathbf{y}) - f^\star - \sum_{i=1}^{r} \operatorname{tr}\left((P(\mathbf{y},\mathbf{x}^{(i)}))W_i\right) \in \mathcal{P}(\mathcal{Y}).$$

Since $\mathbf{y}^\star \in \mathcal{Y}$, we have

$$f(\mathbf{y}^\star) - f^\star - \sum_{i=1}^{r} \operatorname{tr}\left((P(\mathbf{y}^\star,\mathbf{x}^{(i)}))W_i\right) = -\sum_{i=1}^{r} \operatorname{tr}\left((P(\mathbf{y}^\star,\mathbf{x}^{(i)}))W_i\right) \geq 0.$$

As $P(\mathbf{y}^\star,\mathbf{x}^{(i)})$ and W_i are both PSD, it follows

$$0 = \sum_{i=1}^{r} \operatorname{tr}\left((P(\mathbf{y}^\star,\mathbf{x}^{(i)}))W_i\right) = \sum_{i=1}^{r}\sum_{k=1}^{m_i} (\mathbf{v}^{(i,k)})^\intercal P(\mathbf{y}^\star,\mathbf{x}^{(i)})\mathbf{v}^{(i,k)}.$$

The PSDness of $P(\mathbf{y}^\star,\mathbf{x}^{(i)})$ implies that $P(\mathbf{y}^\star,\mathbf{x}^{(i)})\mathbf{v}^{(i,k)} = 0$ for all $k = 1,\ldots,m_i$, $i = 1,\ldots,r$. ■

Remark 4.5. It is easy to verify that the conic reformulation in this section and the results in Propositions 4.1, 4.4 and Theorem 4.3 remain true if "SOS-convexity" in Assumption 1 is weaken to "convexity" (i.e., if Assumption 1 is replaced by Assumption 5 in Section 5).

4.2. A moment-SOS hierarchy. Let $\Theta \coloneqq \{\theta_1,\ldots,\theta_s\} \subset \mathbb{R}[\mathbf{y}]$ collect the description polynomials of the semialgebraic set $\mathcal{Y}$. Moreover, let

$$\begin{aligned} k_{\mathbf{y}} &\coloneqq \max\{\deg(f),\ \deg(\theta_1),\ldots,\deg(\theta_s),\ \deg_{\mathbf{y}}(P_{ij}),\ i,j=1,\ldots,m\},\\ k_{\mathbf{x}} &\coloneqq \max\{\deg_{\mathbf{x}}(P_{ij}),\ i,j=1,\ldots,m,\ \deg(G)\}. \end{aligned}$$

Proposition 4.6. *Under Assumptions 1 and 3, there exists a finitely atomic matrix-valued measure $\Phi^\star \in \mathfrak{M}_+^m(\mathcal{X})$ such that*

$$f(\mathbf{y}) - f^\star - \mathscr{L}_{\Phi^\star}(P(\mathbf{y},\mathbf{x})) \in \mathcal{Q}_{\lceil k_{\mathbf{y}}/2\rceil}(\Theta).$$

Proof. By Proposition 4.1, there exists a finitely atomic matrix-valued measure $\Phi^\star \in \mathfrak{M}_+^m(\mathcal{X})$ such that for all $\mathbf{y} \in \mathcal{Y}$,

$$f(\mathbf{y}) - f^\star - \mathscr{L}_{\Phi^\star}(P(\mathbf{y},\mathbf{x})) \geq 0.$$

Suppose that $\Phi^\star = \sum_{i=1}^r W_i\delta_{\mathbf{x}^{(i)}} \in \mathfrak{M}_+^m(\mathcal{X})$ for some $\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(r)} \in \mathcal{X}$ and $W_1,\ldots,W_r \in \mathbb{S}_+^m$. For each $i = 1,\ldots,r$, let $W_i = \sum_{k=1}^{m_i} \mathbf{v}^{(i,k)}(\mathbf{v}^{(i,k)})^\intercal$ for some $\mathbf{v}^{(i,k)} \in \mathbb{R}^m$. Then,

$$\begin{aligned} f(\mathbf{y}) - f^\star - \mathscr{L}_{\Phi^\star}(P(\mathbf{y},\mathbf{x})) &= f(\mathbf{y}) - f^\star - \sum_{i=1}^{r} \operatorname{tr}\left((P(\mathbf{y},\mathbf{x}^{(i)}))W_i\right)\\ &= f(\mathbf{y}) - f^\star - \sum_{i=1}^{r}\sum_{k=1}^{m_i} (\mathbf{v}^{(i,k)})^\intercal P(\mathbf{y},\mathbf{x}^{(i)})\mathbf{v}^{(i,k)}. \end{aligned}$$

By Assumption 1 and the definition of PSD-SOS-convexity, $f(\mathbf{y}) - f^\star - \mathscr{L}_{\Phi^\star}(P(\mathbf{y},\mathbf{x}))$ is SOS-convex. Let $\mathbf{y}^\star$ be an optimal solution to (RPMIO). By Proposition 4.1, we have

$$f(\mathbf{y}^\star) - f^\star - \mathscr{L}_{\Phi^\star}(P(\mathbf{y}^\star,\mathbf{x})) = 0.$$

Then by Assumptions 1, 3 and [33, Theorem 3.3], the conclusion follows. ■

For each $k \geq \lceil k_{\mathbf{x}}/2 \rceil$, by replacing the cones $\mathcal{P}(\mathcal{Y})$ and $\mathfrak{M}_+^m(\mathcal{X})$ in (4.2) with the more tractable cones $\mathcal{Q}_{\lceil k_{\mathbf{y}}/2\rceil}(\Theta)$ and $\mathcal{M}_k^m(G)$ respectively, we obtain the following SDP:

$$(4.4)\qquad \begin{cases} f_k^{\text{primal}} := \sup\limits_{\rho,\mathbf{S}} \ \rho \\ \qquad\qquad \text{s.t. } f(\mathbf{y}) - \rho - \mathscr{L}_{\mathbf{S}}(P(\mathbf{y},\mathbf{x})) \in \mathcal{Q}_{\lceil k_{\mathbf{y}}/2\rceil}(\Theta), \\ \qquad\qquad\qquad \rho \in \mathbb{R},\ \mathbf{S} \in \mathcal{M}_k^m(G), \end{cases}$$

It follows from Proposition 4.6 that for each $k \geq \lceil k_{\mathbf{x}}/2 \rceil$, (4.4) is an SDP relaxation of (4.2), and hence gives an upper bound on the optimum of (RPMIO). The dual of (4.4) reads as

$$(4.5)\qquad \begin{cases} f_k^{\text{dual}} := \inf\limits_{\mathbf{s}} \ \mathscr{H}_{\mathbf{s}}(f) \\ \qquad\qquad \text{s.t. } \mathbf{s} \in \mathcal{M}_{\lceil k_{\mathbf{y}}/2\rceil}(\Theta),\ \mathscr{H}_{\mathbf{s}}(1) = 1, \\ \qquad\qquad\qquad \mathscr{H}_{\mathbf{s}}(P(\mathbf{y},\mathbf{x})) \in \mathcal{Q}_k^m(G), \end{cases}$$

where the linear functional $\mathscr{H}_{\mathbf{s}}\colon \mathbb{R}[\mathbf{y}]_{2\lceil k_{\mathbf{y}}/2\rceil} \to \mathbb{R}$ is defined by $\mathscr{H}_{\mathbf{s}}(h) = \sum_{\boldsymbol{\alpha}\in\text{supp}(h)} h_{\boldsymbol{\alpha}} s_{\boldsymbol{\alpha}}$ for $h \in \mathbb{R}[\mathbf{y}]_{2\lceil k_{\mathbf{y}}/2\rceil}$. It is clear that both sequences $(f_k^{\text{primal}})_{k\geq\lceil k_{\mathbf{x}}/2\rceil}$ and $(f_k^{\text{dual}})_{k\geq\lceil k_{\mathbf{x}}/2\rceil}$ are monotonically non-increasing. We call (4.4)–(4.5) the *moment-SOS hierarchy* for (RPMIO) and call k the *relaxation order*.

Before proving zero duality gap and asymptotic convergence of the moment-SOS hierarchy for (RPMIO), we collect several preliminary results.

We write $\mathbf{e} = (\mathbf{e}_1, \ldots, \mathbf{e}_\ell)$ for the standard basis of $\mathbb{R}^\ell$ and let $\mathbf{s}_{\mathbf{e}} = (s_{\mathbf{e}_1}, \ldots, s_{\mathbf{e}_\ell})$ for any feasible point $\mathbf{s}$ of (4.5).

Proposition 4.7. *Suppose that $Q(\mathbf{y}) \in \mathbb{S}[\mathbf{y}]^m$ is PSD-SOS-convex. Let $\mathbf{s} = (s_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^\ell_{2\lceil \deg(Q)/2\rceil}}$ satisfy $s_{\mathbf{0}} = 1$ and $M_{\lceil \deg(Q)/2\rceil}(\mathbf{s}) \succeq 0$. Then $\mathscr{H}_{\mathbf{s}}(Q) \succeq Q(\mathbf{s}_{\mathbf{e}})$.*

Proof. As $\mathbf{v}^\intercal Q(\mathbf{y})\mathbf{v}$ is SOS-convex in $\mathbf{y}$ for all $\mathbf{v} \in \mathbb{R}^m$, by [33, Theorem 2.6], it holds

$$\mathbf{v}^\intercal \mathscr{H}_{\mathbf{s}}(Q(\mathbf{y}))\mathbf{v} = \mathscr{H}_{\mathbf{s}}(\mathbf{v}^\intercal Q(\mathbf{y})\mathbf{v}) \geq \mathbf{v}^\intercal Q(\mathbf{s}_{\mathbf{e}})\mathbf{v},$$

for all $\mathbf{v} \in \mathbb{R}^m$. Hence we have $\mathscr{H}_{\mathbf{s}}(Q) \succeq Q(\mathbf{s}_{\mathbf{e}})$. ■

Corollary 4.8. *Suppose that $-\theta_1(\mathbf{y}), \ldots, -\theta_s(\mathbf{y})$ are SOS-convex and $-P(\mathbf{y},\mathbf{x})$ is PSD-SOS-convex in $\mathbf{y}$ for all $\mathbf{x} \in \mathcal{X}$. If $\mathbf{s}$ is feasible to (4.5), then $\mathbf{s}_{\mathbf{e}} \in \mathcal{Y}$ and is feasible to (RPMIO).*

Proof. By the extended Jensen's inequality for SOS-convex polynomials [33, Theorem 2.6], it holds $\theta_i(\mathbf{s}_{\mathbf{e}}) \geq \mathscr{H}_{\mathbf{s}}(\theta_i) \geq 0$ for $i = 1, \ldots, s$, which implies $\mathbf{s}_{\mathbf{e}} \in \mathcal{Y}$. By Proposition 4.7, for every $\mathbf{x} \in \mathcal{X}$,

$$P(\mathbf{s}_{\mathbf{e}},\mathbf{x}) \succeq \mathscr{H}_{\mathbf{s}}(P(\mathbf{y},\mathbf{x})) \succeq 0.$$

So $\mathbf{s}_{\mathbf{e}}$ is feasible to (RPMIO). ■

Lemma 4.9. *Suppose that $\mathcal{X}$ has non-empty interior. Then for each $k \geq d_G$, there exists $\mathbf{S}^\circ \in \mathscr{M}_k^m(G)$ such that $M_k(\mathbf{S}^\circ) \succ 0$ and $M_{k-d_G}(G\mathbf{S}^\circ) \succ 0$.*

Proof. We only prove that there exists $\mathbf{S}^\circ \in \mathscr{M}_k^m(G)$ such that $M_{k-d_G}(G\mathbf{S}^\circ) \succ 0$. Similar arguments apply also to $M_k(\mathbf{S}^\circ)$. Suppose on the contrary that the conclusion is false. Let $\Phi \in \mathfrak{M}_+^m(\mathcal{X})$ be such that $\Phi = \operatorname{diag} \phi, \dots, \phi$ where ϕ is the probability measure with uniform distribution on $\mathcal{X}$, and $\mathbf{S}^\circ = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}_{2k}^n}$ where each $S_{\boldsymbol{\alpha}} = \int_{\mathcal{X}} \mathbf{x}^{\boldsymbol{\alpha}} \mathrm{d}\Phi(\mathbf{x})$. Now fix a nonzero vector $\mathbf{v} \in \mathbb{R}^{mq|\mathbb{N}_{k-d_G}^n|}$ such that $\mathbf{v}^\intercal M_{k-d_G}(G\mathbf{S}^\circ)\mathbf{v} = 0$. Let

$$\Sigma(\mathbf{x}) = (u_{k-d_G}(\mathbf{x}) \otimes I_{mq})^\intercal \mathbf{v}\mathbf{v}^\intercal (u_{k-d_G}(\mathbf{x}) \otimes I_{mq}).$$

Then by Proposition 2.18, it holds

$$\mathscr{L}_{\mathbf{S}^\circ}((\Sigma, G)_m) = \operatorname{tr}(\mathbf{v}\mathbf{v}^\intercal M_{k-d_G}(G\mathbf{S}^\circ)) = \mathbf{v}^\intercal M_{k-d_G}(G\mathbf{S}^\circ)\mathbf{v} = 0.$$

For each $i = 1, \dots, mq$, let $\mathbf{v}^{(i)}$ be the subvector of $\mathbf{v}$ whose entries are indexed by

$$i,\ mq+i,\ 2mq+i,\ \dots, (|\mathbb{N}_{k-d_G}^n| - 1)mq + i,$$

and $T_i(\mathbf{x}) = (\mathbf{v}^{(i)})^\intercal u_{k-d_G}(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$. Then,

$$[T_1(\mathbf{x}), \dots, T_{mq}(\mathbf{x})] = \mathbf{v}^\intercal (u_{k-d_G}(\mathbf{x}) \otimes I_{mq}) \quad \text{and} \quad \Sigma(\mathbf{x}) = T(\mathbf{x})^\intercal T(\mathbf{x}).$$

For each $j = 1, \dots, m$, let

$$H_j(\mathbf{x}) = [T_{(j-1)q+1}(\mathbf{x}), \dots, T_{jq}(\mathbf{x})] \in \mathbb{R}[\mathbf{x}]^q.$$

Then,

$$(\Sigma, G)_m = [H_i(\mathbf{x})^\intercal G(\mathbf{x}) H_j(\mathbf{x})]_{i,j=1,\dots,m}$$

and

$$0 = \mathscr{L}_{\mathbf{S}^\circ}((\Sigma, G)_m) = \int_{\mathcal{X}} \sum_{j=1}^m H_j(\mathbf{x})^\intercal G(\mathbf{x}) H_j(\mathbf{x}) \mathrm{d}\phi(\mathbf{x}).$$

As $H_j(\mathbf{x})^\intercal G(\mathbf{x}) H_j(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathcal{X}$, we have

$$\int_{\mathcal{X}} H_j(\mathbf{x})^\intercal G(\mathbf{x}) H_j(\mathbf{x}) \mathrm{d}\phi(\mathbf{x}) = 0, \quad \forall j = 1, \dots, m.$$

Let $\mathcal{O}$ be an open and bounded subset of $\mathcal{X}$. Then, there exists $\lambda > 0$ such that $G(\mathbf{x}) \succeq \lambda I_m$ on $\mathcal{O}$, and for each $j = 1, \dots, m$,

$$\begin{aligned} 0 &= \int_{\mathcal{X}} H_j(\mathbf{x})^\intercal G(\mathbf{x}) H_j(\mathbf{x}) \mathrm{d}\phi(\mathbf{x}) \\ &\geq \int_{\mathcal{O}} H_j(\mathbf{x})^\intercal G(\mathbf{x}) H_j(\mathbf{x}) \mathrm{d}\phi(\mathbf{x}) \\ &\geq \lambda \int_{\mathcal{O}} H_j(\mathbf{x})^\intercal H_j(\mathbf{x}) \mathrm{d}\phi(\mathbf{x}) \\ &= \lambda \int_{\mathcal{O}} \sum_{i=1}^q T_{(j-1)q+i}(\mathbf{x})^2 \mathrm{d}\phi(\mathbf{x}). \end{aligned}$$

Since $\mathcal{O}$ is open, we have $T_i(\mathbf{x}) \equiv 0$ for each $i = 1, \ldots, mq$. We then conclude $\mathbf{v} = 0$, yielding a contradiction. ■

Lemma 4.10. *[27, Lemma 3] Suppose that there exists a ball constraint $\sum_{i=1}^{\ell} y_i^2 \leq b^2$ in the description of $\mathcal{Y}$. Let $t \geq \max\{\lceil \deg(\theta_j)/2 \rceil,\ j = 1, \ldots, s\}$ be an integer. Let $\mathbf{s} = (s_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}_{2t}^{\ell}}$ satisfy $\mathbf{s} \in \mathcal{M}_t(\Theta)$ and $\mathscr{H}_{\mathbf{s}}(1) = 1$. Then*

$$\|\mathbf{s}\| \leq \sqrt{\binom{\ell + t}{\ell} \sum_{i=0}^{t} b^{2i}}.$$

Assumption 4. (i) $\theta_1(\mathbf{y}) = b^2 - \sum_{i=1}^{\ell} y_i^2$ for some $b > 0$; (ii) $\mathcal{X}$ has non-empty interior.

We are now in a position to establish asymptotic convergence of the moment-SOS hierarchy (4.4)–(4.5).

Theorem 4.11. *Under Assumptions 1–4, the following are true:*

(i) For each $k \geq \lceil k_{\mathbf{x}}/2 \rceil$, $f_k^{\text{primal}} = f_k^{\text{dual}}$;

(ii) $f_k^{\text{primal}} \searrow f^\star$ and $f_k^{\text{dual}} \searrow f^\star$ as $k \to \infty$;

(iii) For any convergent subsequence $(\mathbf{s}_{\mathbf{e}}^{(k_i,\star)})_i$ (always exists) of $(\mathbf{s}_{\mathbf{e}}^{(k,\star)})_k$ where $\mathbf{s}^{(k,\star)}$ is a minimizer of (4.5), $\lim_{i\to\infty} \mathbf{s}_{\mathbf{e}}^{(k_i,\star)}$ is a global minimizer of (RPMIO). Consequently, if the set of optimal solutions of (RPMIO) is a singleton, then $\lim_{k\to\infty} \mathbf{s}_{\mathbf{e}}^{(k,\star)}$ is the unique global minimizer.

Proof. (i). Fix $k \geq \lceil k_{\mathbf{x}}/2 \rceil$. Let $\mathbf{S}^\circ$ be given in Lemma 4.9. As there is a ball constraint in defining $\mathcal{Y}$ by Assumption 4, by the proof of [38, Corollary 3.6], there exists $\lambda \in \mathbb{R}$ such that

$$f(\mathbf{y}) - \lambda - \mathscr{L}_{\mathbf{S}^\circ}(P(\mathbf{y}, \mathbf{x})) \in \mathcal{Q}^\circ_{\lceil k_{\mathbf{y}}/2 \rceil}(\Theta),$$

where $\mathcal{Q}^\circ_{\lceil k_{\mathbf{y}}/2 \rceil}(\Theta)$ denotes the interior of $\mathcal{Q}_{\lceil k_{\mathbf{y}}/2 \rceil}(\Theta)$. It follows that $(\lambda, \mathbf{S}^\circ)$ is a strictly feasible point of (4.4), which implies $f_k^{\text{primal}} = f_k^{\text{dual}}$.

(ii). Let $\mathbf{y}^\star$ be a minimizer of (RPMIO) and $\bar{\mathbf{y}}$ be the Slater point given in Assumption 3. Because f, $\mathcal{Y}$ are convex by Assumption 1, we can choose $0 < t < 1$ such that $\mathbf{y}' := t\mathbf{y}^\star + (1-t)\bar{\mathbf{y}} \in \mathcal{Y}$ and $f(\mathbf{y}') \leq f^\star + \varepsilon$ for an arbitrary $\varepsilon > 0$. As $-P(\mathbf{y}, \mathbf{x})$ is PSD-SOS-convex in $\mathbf{y}$ for every $\mathbf{x} \in \mathcal{X}$, it holds

$$P(\mathbf{y}', \mathbf{x}) \succeq tP(\mathbf{y}^\star, \mathbf{x}) + (1-t)P(\bar{\mathbf{y}}, \mathbf{x}) \succ 0,$$

for all $\mathbf{x} \in \mathcal{X}$. Let $\mathbf{s}' = (s'_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^{\ell}_{2\lceil k_{\mathbf{y}}/2 \rceil}}$ with $s'_{\boldsymbol{\alpha}} = (\mathbf{y}')^{\boldsymbol{\alpha}}$. By Theorem 2.3, there exists $k^{(\varepsilon)} \in \mathbb{N}$ such that $\mathbf{s}'$ is feasible to (4.5) for all $k \geq k^{(\varepsilon)}$. Therefore, $f_k^{\text{dual}} \leq f^\star + \varepsilon$ for all $k \geq k^{(\varepsilon)}$. As $\varepsilon > 0$ is arbitrary, we have $f_k^{\text{dual}} \searrow f^\star$ and $f_k^{\text{primal}} \searrow f^\star$ as $k \to \infty$.

(iii). Let $\mathbf{s}^{(k,\star)} = (s_{\boldsymbol{\alpha}}^{(k,\star)})_{\boldsymbol{\alpha} \in \mathbb{N}^{\ell}_{2\lceil k_{\mathbf{y}}/2 \rceil}}$ be a minimizer of (4.5) for each $k \geq \lceil k_{\mathbf{x}}/2 \rceil$. By Lemma 4.10, there exists a subsequence $(\mathbf{s}^{(k_i,\star)})_i$ of $(\mathbf{s}^{(k,\star)})_k$ and $\mathbf{s}^\star = (s^\star_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^{\ell}_{2\lceil k_{\mathbf{y}}/2 \rceil}}$ such that $\lim_{i\to\infty} s_{\boldsymbol{\alpha}}^{(k_i,\star)} = s^\star_{\boldsymbol{\alpha}}$ for all $\boldsymbol{\alpha}$. As the feasible set of (RPMIO) is closed, by Corollary 4.8, $\mathbf{s}^\star_{\mathbf{e}}$ is feasible to (RPMIO). Moreover, as $f(\mathbf{y})$ is SOS-convex, by (ii) and Proposition 4.7, it holds that

$$f^\star = \mathscr{H}_{\mathbf{s}^\star}(f) \geq f(\mathbf{s}^\star_{\mathbf{e}}),$$

which indicates that $\mathbf{s}_{\mathbf{e}}^{\star}$ is a global minimizer of (RPMIO). ■

The next theorem allows us to detect finite convergence of the moment-SOS hierarchy (4.4)–(4.5) and to extract an optimal solution whenever the flat extension condition is satisfied.

Theorem 4.12. *Suppose that Assumptions 1, 3 and 4 hold. If the following flat extension condition*

$$\exists \lceil k_{\mathbf{x}}/2 \rceil \le t \le k \;\; s.t. \;\; \operatorname{rank}(M_t(\mathbf{S}^{(k,\star)})) = \operatorname{rank}(M_{t-d_G}(\mathbf{S}^{(k,\star)})) \tag{4.6}$$

holds for some $k \ge \lceil k_{\mathbf{x}}/2 \rceil$, where $(f_k^{\text{primal}}, \mathbf{S}^{(k,\star)})$ and $\mathbf{s}^{(k,\star)}$ are optimal solutions to (4.4) and (4.5), respectively, then

(i) $\mathbf{S}^{(k,\star)}$ admits a representing measure $\Phi^\star = \sum_{i=1}^r W_i \delta_{\mathbf{x}^{(i)}} \in \mathfrak{M}_+^m(\mathcal{X})$ for some points $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(r)} \in \mathcal{X}$ and $W_1, \ldots, W_r \in \mathbb{S}_+^m$;

(ii) $f_k^{\text{primal}} = f_k^{\text{dual}} = f^\star$;

(iii) $\mathbf{s}_{\mathbf{e}}^{(k,\star)}$ is an optimal solution to (RPMIO);

(iv) For any decomposition $W_i = \sum_{l=1}^{m_i} \mathbf{v}^{(i,l)}(\mathbf{v}^{(i,l)})^\intercal$, $\mathbf{v}^{(i,l)} \in \mathbb{R}^m$, $i = 1, \ldots, r$, it holds that

$$P\left(\mathbf{s}_{\mathbf{e}}^{(k,\star)}, \mathbf{x}^{(i)}\right) \mathbf{v}^{(i,l)} = 0, \quad l = 1, \ldots, m_i, i = 1, \ldots, r.$$

Proof. (i). This follows from Theorem 3.2.

(ii). As $(f_k^{\text{primal}}, \mathbf{S}^{(k,\star)})$ is feasible to (4.4), by (i), it holds

$$f(\mathbf{y}) - f_k^{\text{primal}} - \mathscr{L}_{\mathbf{S}^{(k,\star)}}(P(\mathbf{y}, \mathbf{x})) = f(\mathbf{y}) - f_k^{\text{primal}} - \sum_{i=1}^r \operatorname{tr}\left(W_i P(\mathbf{y}, \mathbf{x}^{(i)})\right) \ge 0, \tag{4.7}$$

for all $\mathbf{y} \in \mathcal{Y}$. Let $\mathbf{y}^\star$ be a global minimizer of (RPMIO). Noting $\sum_{i=1}^r \operatorname{tr}\left(W_i P(\mathbf{y}^\star, \mathbf{x}^{(i)})\right) \ge 0$, then by (4.7), we have

$$f_k^{\text{primal}} \le f(\mathbf{y}^\star) - \sum_{i=1}^r \operatorname{tr}\left(W_i P(\mathbf{y}^\star, \mathbf{x}^{(i)})\right) \le f(\mathbf{y}^\star) = f^\star.$$

By Theorem 4.11 (i), it follows $f_k^{\text{primal}} = f_k^{\text{dual}} = f^\star$.

(iii). By Corollary 4.8, $\mathbf{s}_{\mathbf{e}}^{(k,\star)}$ is feasible to (RPMIO). Since $f(\mathbf{y})$ is SOS-convex, by (ii) and Proposition 4.7, it holds

$$f(\mathbf{s}_{\mathbf{e}}^{(k,\star)}) \le \mathscr{H}_{\mathbf{s}^{(k,\star)}}(f) = f_k^{\text{dual}} = f^\star,$$

which implies that $\mathbf{s}_{\mathbf{e}}^{(k,\star)}$ is a global minimizer of (RPMIO).

(iv). By (ii) and Theorem 4.3, $(f_k^{\text{primal}}, \mathbf{S}^{(k,\star)})$ is an optimal solution to (4.2). It follows from (iii) and Proposition 4.4. ■

Remark 4.13. From the proof of Theorem 4.12 (ii), we can see that the flat extension condition (4.6) implies $f_k^{\text{primal}} = f^\star$ even without Assumption 4. If $f_k^{\text{dual}} = f_k^{\text{primal}}$ in this case, then the statements (iii) and (iv) in Theorem 4.12 are still true.

For all numerical examples in the rest of this paper, we use the software Yalmip [36] to build the SDPs and call the SDP solver Mosek [40] to solve them[1]. To check the flat extension condition, we use the Matlab command rank with tolerance 10^{-3}.

Example 4.14. *Consider the following instance of* (RPMIO)*:*

$$f^\star := \inf_{\mathbf{y}\in\mathbb{R}^2} f(\mathbf{y}) \quad s.t. \quad P(\mathbf{y},\mathbf{x}) \succeq 0,\ \forall \mathbf{x}\in\mathcal{X} := \{\mathbf{x}\in\mathbb{R}^2 \mid G(\mathbf{x})\succeq 0\}, \tag{4.8}$$

where

$$P(\mathbf{y},\mathbf{x}) = \begin{pmatrix} 1-(x_1y_1-x_2y_2)^2 & 2(x_2y_1+x_1y_2) \\ 2(x_2y_1+x_1y_2) & 1 \end{pmatrix}$$

and

$$G(\mathbf{x}) = \begin{pmatrix} 1-x_1 & x_2 & 0 & 0 \\ x_2 & 1+x_1 & 0 & 0 \\ 0 & 0 & x_1^2+x_2^2-1 & 0 \\ 0 & 0 & 0 & x_1x_2 \end{pmatrix}.$$

It is clear that

$$\mathcal{X} = \{\mathbf{x}\in\mathbb{R}^2 \mid x_1^2+x_2^2=1,\ x_1x_2\geq 0\}.$$

Geometrically, the feasible region is constructed by rotating the shape in the $\mathbf{y}$*-plane defined by* $y_1^2+4y_2^2\leq 1$ *continuously around the origin by* 90° *clockwise and then taking the common area of these shapes in this process.*

By Corollary 2.9, it is easy to check that $-P(\mathbf{y},\mathbf{x})$ *is PSD-SOS-convex in* $\mathbf{y}$ *for every* $\mathbf{x}\in\mathcal{X}$*. Consider the following objective functions* $f_1(\mathbf{y}) = (y_1-1)^2+(y_2-1)^2$ *and* $f_2(\mathbf{y}) = (1+y_1)^2+(1-y_2)^2$*, respectively. Clearly, both* f_1 *and* f_2 *are SOS-convex.*

For f_1*, the optimal solution is* $(\sqrt{2}/4,\sqrt{2}/4)\approx(0.3536,0.3536)$ *and the optimal value is* $2(\sqrt{2}/4-1)^2\approx 0.8358$*. We solve the SDP relaxations* (4.4) *and* (4.5) *with* $k=3$*, obtaining* $f_3^{\text{primal}} = f_3^{\text{dual}} = 0.8358$*, and the rank condition* (4.6) *is satisfied with* $t=2$*. Then by Theorem 4.12 and Remark 4.13, global optimality is reached and a minimizer* $\mathbf{s}_{\mathbf{e}}^{(3,\star)} = (0.3536,0.3536)$ *can be extracted.*

For f_2*, the optimal solution is* $(-\sqrt{5}/5,\sqrt{5}/5)\approx(-0.4472,0.4472)$ *and the optimal value is* $2(\sqrt{5}/5-1)^2\approx 0.6111$*. Solving* (4.4) *and* (4.5) *with* $k=3$*, we get* $f_3^{\text{primal}} = f_3^{\text{dual}} = 0.6111$ *and the rank condition* (4.6) *is satisfied with* $t=2$*. So global optimality is reached and a minimizer* $\mathbf{s}_{\mathbf{e}}^{(3,\star)} = (-0.4472,0.4472)$ *can be extracted.*

4.3. The linear case and the generalized matrix-valued moment problem. If $\mathcal{Y}=\mathbb{R}^\ell$, $f(\mathbf{y})$ and $P(\mathbf{x},\mathbf{y})$ are affine in $\mathbf{y}$, then (RPMIO) becomes the robust polynomial semidefinite program (RPSDP) which was studied in [45]. In this case, we will see that the dual problem (4.5) of the moment-SOS hierarchy recovers the matrix SOS relaxation for (RPSDP) proposed in [45]. Meanwhile, the primal problem (4.4) allows to detect finite convergence and extract optimal solutions.

[1] The script is available at https://github.com/wangjie212/PMOptimization.

4.3.1. Robust polynomial semidefinite programming. Consider the robust polynomial semidefinite programming problem which is a special case of (RPMIO):

$$\text{(RPSDP)} \qquad \tau^\star := \inf_{\mathbf{y}\in\mathbb{R}^\ell} \mathbf{c}^\intercal \mathbf{y} \quad \text{s.t.} \quad \sum_{i=1}^{\ell} P_i(\mathbf{x})y_i - P_0(\mathbf{x}) \succeq 0, \ \forall \mathbf{x} \in \mathcal{X} \subset \mathbb{R}^n,$$

where $\mathbf{c} = (c_1, \ldots, c_\ell) \in \mathbb{R}^\ell$ and $P_i(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^m$, $i = 0, 1, \ldots, \ell$.

Applying the conic reformulation (4.2) to (RPSDP) with $\mathcal{Y} = \mathbb{R}^\ell$, we obtain

$$\text{(4.9)} \qquad \begin{cases} \sup\limits_{\rho,\Phi} \rho \\ \text{s.t. } \mathbf{c}^\intercal \mathbf{y} - \rho - \sum\limits_{i=1}^{\ell} \mathscr{L}_\Phi(P_i(\mathbf{x}))y_i + \mathscr{L}_\Phi(P_0(\mathbf{x})) \in \mathcal{P}(\mathbb{R}^\ell), \\ \qquad \rho \in \mathbb{R}, \ \Phi \in \mathfrak{M}_+^m(\mathcal{X}). \end{cases}$$

For any $\rho \in \mathbb{R}$ and $\Phi \in \mathfrak{M}_+^m(\mathcal{X})$, it holds

$$\begin{aligned} &\mathbf{c}^\intercal \mathbf{y} - \rho - \sum_{i=1}^{\ell} \mathscr{L}_\Phi(P_i(\mathbf{x}))y_i + \mathscr{L}_\Phi(P_0(\mathbf{x})) \\ =&\sum_{i=1}^{\ell} \left(c_i - \int_{\mathcal{X}} P_i(\mathbf{x}) \mathrm{d}\Phi(\mathbf{x}) \right) y_i - \rho + \int_{\mathcal{X}} P_0(\mathbf{x}) \mathrm{d}\Phi(\mathbf{x}). \end{aligned}$$

Thus, if (ρ, Φ) is feasible to (4.9), then we necessarily have

$$c_i = \int_{\mathcal{X}} P_i(\mathbf{x}) \mathrm{d}\Phi(\mathbf{x}), \ i = 1, \ldots, \ell, \quad \text{and} \quad \rho \leq \int_{\mathcal{X}} P_0(\mathbf{x}) \mathrm{d}\Phi(\mathbf{x}),$$

and (4.9) can be rewritten as

$$\text{(4.10)} \qquad \sup_{\Phi\in\mathfrak{M}_+^m(\mathcal{X})} \int_{\mathcal{X}} P_0(\mathbf{x}) \mathrm{d}\Phi(\mathbf{x}) \quad \text{s.t.} \quad \int_{\mathcal{X}} P_i(\mathbf{x}) \mathrm{d}\Phi(\mathbf{x}) = c_i, \ i = 1, \ldots, \ell.$$

Remark 4.15. Note that the generalized moment problem extensively studied by Lasserre in [32] is a special case of (4.10) with $m = 1$ and $G(\mathbf{x})$ being a diagonal matrix.

The dual of (4.9) reads as

$$\text{(4.11)} \qquad \begin{cases} \inf\limits_{\mu} \mathbf{c}^\intercal \mathscr{H}_\mu(\mathbf{y}) \\ \text{s.t. } \mu \in \mathfrak{m}_+(\mathbb{R}^\ell), \ \mathscr{H}_\mu(1) = 1, \\ \qquad \sum\limits_{i=1}^{\ell} P_i(\mathbf{x}) \mathscr{H}_\mu(y_i) - P_0(\mathbf{x}) \in \mathcal{P}^m(\mathcal{X}). \end{cases}$$

By identifying $\mathscr{H}_\mu(\mathbf{y})$ with $\mathbf{y}$, we see that (4.11) is actually equivalent to (RPSDP).

For each $k \geq \lceil k_{\mathbf{x}}/2 \rceil$, replacing the cone $\mathfrak{M}_+^m(\mathcal{X})$ in (4.10) by $\mathcal{M}_k^m(G)$, we obtain an SDP relaxation:

$$(4.12) \qquad \begin{cases} \tau_k^{\text{primal}} := \sup\limits_{\mathbf{S} \in \mathcal{M}_k^m(G)} \mathscr{L}_{\mathbf{S}}(P_0(\mathbf{x})) \\ \qquad\qquad \text{s.t. } \mathscr{L}_{\mathbf{S}}(P_i(\mathbf{x})) = c_i, \; i = 1, \ldots, \ell, \end{cases}$$

whose dual is

$$(4.13) \qquad \begin{cases} \tau_k^{\text{dual}} := \inf\limits_{\mathbf{y} \in \mathbb{R}^\ell} \mathbf{c}^\intercal \mathbf{y} \\ \qquad\qquad \text{s.t. } \sum\limits_{i=1}^{\ell} P_i(\mathbf{x}) y_i - P_0(\mathbf{x}) \in \mathcal{Q}_k^m(G). \end{cases}$$

Note that when $\mathcal{X}$ is compact, Assumption 1 is satisfied for (RPSDP). Moreover, if $\mathcal{X}$ has non-empty interior, then (4.12) is strictly feasible by Lemma 4.9 and hence there is no duality gap between (4.12) and (4.13). Then, by Theorems 4.11 and 4.12, we get the following theorems.

Theorem 4.16. *Under Assumptions 2–3, suppose that $\mathcal{X}$ is compact and has non-empty interior. Then, the following are true:*

(i) $\tau_k^{\text{primal}} = \tau_k^{\text{dual}}$ *for all* $k \geq \lceil k_{\mathbf{x}}/2 \rceil$*;*

(ii) $\tau_k^{\text{primal}} \searrow \tau^\star$ *and* $\tau_k^{\text{dual}} \searrow \tau^\star$ *as* $k \to \infty$*;*

(iii) For any convergent subsequence $(\mathbf{y}^{(k_i,\star)})_i$ *of* $(\mathbf{y}^{(k,\star)})_k$ *where* $\mathbf{y}^{(k,\star)}$ *is an optimal solution to* (4.13), $\lim_{i\to\infty} \mathbf{y}^{(k_i,\star)}$ *is an optimal solution to* (RPSDP).

Theorem 4.17. *Under Assumption 3, suppose that $\mathcal{X}$ is compact and has non-empty interior. If the following flat extension condition*

$$(4.14) \qquad \exists \lceil k_{\mathbf{x}}/2 \rceil \leq t \leq k \text{ s.t. } \operatorname{rank}(M_t(\mathbf{S}^{(k,\star)})) = \operatorname{rank}(M_{t-d_G}(\mathbf{S}^{(k,\star)})),$$

holds for some $k \geq \lceil k_{\mathbf{x}}/2 \rceil$*, where* $\mathbf{S}^{(k,\star)}$ *and* $\mathbf{y}^{(k,\star)}$ *are optimal solutions to* (4.12) *and* (4.13) *respectively, then*

(i) $\mathbf{S}^{(k,\star)}$ *admits a representing measure* $\Phi^\star = \sum_{i=1}^r W_i \delta_{\mathbf{x}^{(i)}} \in \mathfrak{M}_+^m(\mathcal{X})$ *for some points* $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(r)} \in \mathcal{X}$ *and* $W_1, \ldots, W_r \in \mathbb{S}_+^m$*;*

(ii) $\tau_k^{\text{primal}} = \tau_k^{\text{dual}} = \tau^\star$*;*

(iii) $\mathbf{y}^{(k,\star)}$ *is an optimal solution to* (RPSDP)*;*

(iv) For any decomposition $W_i = \sum_{l=1}^{m_i} \mathbf{v}^{(i,l)}(\mathbf{v}^{(i,l)})^\intercal$, $\mathbf{v}^{(i,l)} \in \mathbb{R}^m$, $i = 1, \ldots, r$, *it holds that*

$$\left(\sum_{i=1}^{\ell} P_i(\mathbf{x}^{(i)}) y_i^{(k,\star)} - P_0(\mathbf{x}^{(i)}) \right) \mathbf{v}^{(i,l)} = 0, \quad l = 1, \ldots, m_i, \; i = 1, \ldots, r.$$

Example 4.18. *Consider the following instance of* (RPSDP)*:*

$$(4.15) \qquad \begin{cases} \tau^\star := \inf\limits_{\mathbf{y} \in \mathbb{R}} \mathbf{y} \\ \qquad \text{s.t. } P(\mathbf{y}, \mathbf{x}) = \begin{pmatrix} \mathbf{y} & h(\mathbf{x}) \\ h(\mathbf{x}) & \mathbf{y} \end{pmatrix} \succeq 0, \; \forall \mathbf{x} \in \mathcal{X} := \{\mathbf{x} \in \mathbb{R}^2 \mid G(\mathbf{x}) \succeq 0\}, \end{cases}$$

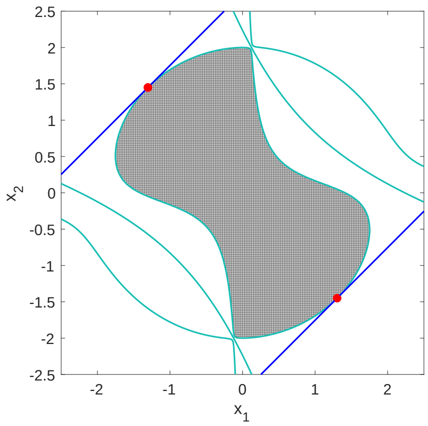

Figure 1. *The set $\mathcal{X}$ (gray), the lines $x_2 - x_1 = \pm 2.7535$ (blue), the points $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ (red) in Example 4.18.*

where $h(\mathbf{x}) = x_2 - x_1$ and

$$G(\mathbf{x}) = \begin{pmatrix} 1 - 4x_1x_2 & x_1 \\ x_1 & 4 - x_1^2 - x_2^2 \end{pmatrix}.$$

The matrix $G(\mathbf{x})$ is borrowed from [23].

Clearly, it holds $\tau^\star = \sup_{\mathbf{x}\in\mathcal{X}} |h(\mathbf{x})|$. Consider the SDP relaxation (4.12) for (4.15). If the condition (4.14) holds, then according to Theorem 4.17, $\tau_k^{\mathrm{primal}} = \tau^\star$ and the minimizer $\mathbf{S}^{(k,\star)}$ of (4.12) admits a representing measure $\Phi^\star = \sum_{i=1}^r W_i \delta_{\mathbf{x}^{(i)}} \in \mathfrak{M}_+^m(\mathcal{X})$ for some $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(r)} \in \mathcal{X}$ and $W_1, \ldots, W_r \in \mathbb{S}_+^m$. Moreover, by Theorem 4.17, for each $i = 1, \ldots, r$,

$$\det(P(\tau^\star, \mathbf{x}^{(i)})) = (\tau^\star)^2 - h(\mathbf{x}^{(i)})^2 = 0.$$

So each $\mathbf{x}^{(i)}$ is a minimizer of $\sup_{\mathbf{x}\in\mathcal{X}} |h(\mathbf{x})|$.

Solving the SDP relaxation (4.12) with $k = 2$, we get $\tau_2^{\mathrm{primal}} = 2.7535$ and the rank condition (4.14) is satisfied with

$$\mathrm{rank}(M_2(\mathbf{S}_2^\star)) = \mathrm{rank}(M_1(\mathbf{S}_2^\star)) = 2.$$

By Theorem 4.17, global optimality is reached. By the procedure described in Section 3.2, we recover the representing measure $\Phi^\star = \sum_{i=1}^2 W_i \delta_{\mathbf{x}^{(i)}}$ of $\mathbf{S}_2^\star$ with

$$\mathbf{x}^{(1)} = (-1.3038, 1.4496), \quad \mathbf{x}^{(2)} = (1.3038, -1.4496),$$

and

$$W_1 = \begin{bmatrix} 0.2500 & 0.2780 \\ 0.2780 & 0.2500 \end{bmatrix}, \; W_2 = \begin{bmatrix} 0.2500 & -0.2780 \\ -0.2780 & 0.2500 \end{bmatrix}.$$

See the illustration of the results in Figure 1.

4.3.2. An application: minimizing the smallest eigenvalue of a polynomial matrix. Consider the problem of minimizing the smallest eigenvalue of a polynomial matrix:

$$\lambda^\star := \inf_{\mathbf{x}\in\mathbb{R}^n} \lambda_{\min}(F(\mathbf{x})) \quad \text{s.t.} \quad \mathbf{x}\in\mathcal{X} := \{\mathbf{x}\in\mathbb{R}^n \mid G(\mathbf{x})\succeq 0\}, \tag{4.16}$$

where $F(\mathbf{x})\in\mathbb{S}[\mathbf{x}]^m$, $\lambda_{\min}(F(\mathbf{x}))$ denotes the smallest eigenvalue of $F(\mathbf{x})$ and $G(\mathbf{x})\in\mathbb{S}[\mathbf{x}]^q$. The motivations for studying this problem come from many different fields. For example, in the global optimization method, $\boldsymbol{\alpha}$BB, for general constrained nonconvex problems, a convex relaxation of the original nonconvex problem is constructed. To underestimate nonconvex terms of generic structure in the involved nonconvex functions, one needs to compute a parameter α which amounts to minimizing the smallest eigenvalue of the corresponding Hessian matrix over a product of intervals. If the involved functions are polynomials, then the problem can be formulated as (4.16) [4, 39]. For another example from optimal control, the stabilisability radius of a continuous-time system described by a state-space equation is defined as the norm of the smallest perturbation that makes the system unstabilisable. To compute such a radius, one needs to minimize the smallest eigenvalue of a bivariate polynomial matrix over a half disc on a plane, which is a special case of the problem (4.16) [18].

Clearly, the problem (4.16) is equivalent to

$$\sup_{\lambda\in\mathbb{R}} \lambda \quad \text{s.t.} \quad F(\mathbf{x})-\lambda I_m\succeq 0, \quad \forall \mathbf{x}\in\mathcal{X}, \tag{4.17}$$

which is a special case of (RPSDP).

For each $k\geq\max\{\lceil\deg(F)/2\rceil,\lceil\deg(G)/2\rceil\}$, the k-th SDP relaxation of (4.16) can be written as

$$\lambda_k^{\text{primal}} := \inf_{\mathbf{S}\in\mathcal{M}_k^m(G)} \mathscr{L}_{\mathbf{S}}(F(\mathbf{x})) \quad \text{s.t.} \quad \mathscr{L}_{\mathbf{S}}(I_m)=1, \tag{4.18}$$

with dual

$$\lambda_k^{\text{dual}} := \sup_{\lambda\in\mathbb{R}} \lambda \quad \text{s.t.} \quad F(\mathbf{x})-\lambda I_m\in\mathcal{Q}_k^m(G). \tag{4.19}$$

The dual problem (4.19) is considered in [46] when $\mathcal{X}$ is the n-dimensional boolean hypercube $\{0,1\}^n$. The significance of the primal problem (4.18) is that we can detect finite convergence and extract optimal solutions if certain flat extension condition holds.

Clearly, the Slater condition holds for (4.17) if $\mathcal{X}$ is compact. Then, by Theorems 4.16–4.17, we deduce the following theorems.

Theorem 4.19. *Under Assumption 2, suppose that $\mathcal{X}$ is compact and has non-empty interior, then the following are true:*

(i) $\lambda_k^{\text{dual}}=\lambda_k^{\text{primal}}$ *for all* $k\geq\max\{\lceil\deg(F)/2\rceil,\lceil\deg(G)/2\rceil\}$*;*

(ii) $\lambda_k^{\text{primal}}\searrow\lambda^\star$ *and* $\lambda_k^{\text{dual}}\searrow\lambda^\star$ *as* $k\to\infty$*.*

Theorem 4.20. *Suppose that $\mathcal{X}$ is compact and has non-empty interior. If the following flat extension condition*

$$\exists\lceil k_{\mathbf{x}}/2\rceil\leq t\leq k \text{ s.t. } \operatorname{rank}(M_t(\mathbf{S}_k^\star))=\operatorname{rank}(M_{t-d_G}(\mathbf{S}_k^\star)) \tag{4.20}$$

holds for some $k\geq\max\{\lceil\deg(F)/2\rceil,\lceil\deg(G)/2\rceil\}$*, where* $\mathbf{S}_k^\star$ *is an optimal solution to* (4.18)*, then the following are true:*

(i) $\mathbf{S}_k^\star$ *admits a representing measure* $\Phi^\star = \sum_{i=1}^r W_i \delta_{\mathbf{x}^{(i)}} \in \mathfrak{M}_+^m(\mathcal{X})$ *for some points* $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(r)} \in \mathcal{X}$ *and* $W_1, \dots, W_r \in \mathbb{S}_+^m$ *with* $\sum_{i=1}^r \operatorname{tr}(W_i) = 1$*;*

(ii) $\lambda_k^{\mathrm{dual}} = \lambda_k^{\mathrm{primal}} = \lambda^\star$*;*

(iii) *For any decomposition* $W_i = \sum_{l=1}^{m_i} \mathbf{v}^{(i,l)} (\mathbf{v}^{(i,l)})^\intercal$, $\mathbf{v}^{(i,l)} \in \mathbb{R}^m$, $i = 1, \dots, r$, *it holds*

$$F(\mathbf{x}^{(i)}) \mathbf{v}^{(i,l)} = \lambda^\star \mathbf{v}^{(i,l)}, \quad l = 1, \dots, m_i, i = 1, \dots, r.$$

That is, the smallest eigenvalue $\lambda^\star$ *of* $F(\mathbf{x})$ *over* $\mathcal{X}$ *is attained at* $\mathbf{x}^{(i)}$ *with* $\mathbf{v}^{(i,l)}$ *being the corresponding eigenvectors.*

Example 4.21. *Consider the following instance of* (4.16)*:*

$$\lambda^\star := \inf_{\mathbf{x} \in \mathbb{R}^n} \quad \lambda_{\min}(F(\mathbf{x})) \quad s.t. \quad \mathbf{x} \in \mathcal{X} := \{\mathbf{x} \in \mathbb{R}^n \mid G(\mathbf{x}) \succeq 0\}, \tag{4.21}$$

where $F(\mathbf{x}) = Q \operatorname{diag} f_1, f_2, f_3 Q^\intercal$ *for some* $f_1, f_2, f_3 \in \mathbb{R}[\mathbf{x}]$ *and* $Q \in \mathbb{R}^{3\times 3}$ *with* $Q^\intercal Q = I_3$*. Clearly,* $\lambda^\star = \inf_{\mathbf{x} \in \mathcal{X}} \{f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x})\}$*.*

We let Q *be the orthogonal matrix*

$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \end{pmatrix},$$

$G(\mathbf{x})$ *be the polynomial matrix in* (4.18)*, and*

$$\begin{aligned} f_1(\mathbf{x}) &= -x_1^2 - x_2^2, \\ f_2(\mathbf{x}) &= -\frac{1}{4}(x_1+1)^2 - \frac{1}{4}(x_2-1)^2, \\ f_3(\mathbf{x}) &= -\frac{1}{4}(x_1-1)^2 - \frac{1}{4}(x_2+1)^2. \end{aligned}$$

It is easy to check that $\lambda^\star = \inf_{\mathbf{x} \in \mathcal{X}} \{f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x})\} = -4$ *which is achieved by* f_1 *at* $(0, \pm 2)$*. The eigenvector space of* $F(0, \pm 2)$ *associated with the eigenvalue* -4 *consists of all vectors of the form* $(c,\ 0,\ c)^\intercal$, $c \in \mathbb{R}$*.*

Solving the SDP relaxation (4.18) *with* $k = 2$*, we get* $\lambda_2^{\mathrm{primal}} = -4.0000$ *and the rank condition* (4.20) *is satisfied with*

$$\operatorname{rank}(M_2(\mathbf{S}_2^\star)) = \operatorname{rank}(M_1(\mathbf{S}_2^\star)) = 2.$$

By Theorem 4.20, global optimality is reached. By the procedure described in Section 3.2, we recover the representing measure $\Phi^\star = \sum_{i=1}^2 W_i \delta_{\mathbf{x}^{(i)}}$ *of* $\mathbf{S}_2^\star$ *with*

$$\mathbf{x}^{(1)} = (-0.0000, -2.0000), \quad \mathbf{x}^{(2)} = (0.0000, 2.0000),$$

and

$$W_1 = \begin{bmatrix} 0.2511 & -0.0006 & 0.2508 \\ -0.0006 & -0.0003 & 0.0006 \\ 0.2508 & 0.0006 & 0.2511 \end{bmatrix}, \quad W_2 = \begin{bmatrix} 0.2489 & 0.0006 & 0.2492 \\ 0.0006 & 0.0003 & -0.0006 \\ 0.2492 & -0.0006 & 0.2489 \end{bmatrix}.$$

Up to some numerical errors, both W_1 and W_2 have the approximate decomposition:

$$(0.5000,\ -0.0012,\ 0.4994)^\intercal(0.5000,\ -0.0012,\ 0.4994).$$

By Theorem 4.20, $(0.5000,\ -0.0012,\ 0.4994)^\intercal$ *is an eigenvector of* $F(\mathbf{x}^{(i)})$, $i=1,2$, *associated with the eigenvalue* -4.0000. *In fact, it corresponds to the vector* $(c,\ 0,\ c)^\intercal$ *with* $c=0.5$.

5. Extension to the general convex case. In this section, we provide an extension of the proposed approach to (RPMIO) with general convexity. The following assumptions are obtained from Assumption 1 by replacing "SOS-convex" with "convex".

Assumption 5. (i) $f(\mathbf{y}), -\theta_1(\mathbf{y}), \ldots, -\theta_s(\mathbf{y})$ are convex; (ii) $-P(\mathbf{y},\mathbf{x})$ is PSD-convex in $\mathbf{y}$ for all $\mathbf{x}\in\mathcal{X}$; (iii) $\mathcal{X}$ is compact.

For each $k \geq \max\{\lceil k_{\mathbf{x}}/2\rceil, \lceil k_{\mathbf{y}}/2\rceil\}$, by replacing the cones $\mathcal{P}(\mathcal{Y})$ and $\mathfrak{M}_+^m(\mathcal{X})$ in (4.2) with $\mathcal{Q}_k(\Theta)$ and $\mathcal{M}_k^m(G)$ respectively, we obtain the following SDP problem

$$\begin{cases} f_k^{\mathrm{primal}} := \sup\limits_{\rho,\mathbf{S}}\ \rho \\ \qquad\qquad \text{s.t. } f(\mathbf{y}) - \rho - \mathscr{L}_{\mathbf{S}}(P(\mathbf{y},\mathbf{x})) \in \mathcal{Q}_k(\Theta), \\ \qquad\qquad\quad \rho\in\mathbb{R},\ \mathbf{S}\in\mathcal{M}_k^m(G), \end{cases} \tag{5.1}$$

whose dual reads as

$$\begin{cases} f_k^{\mathrm{dual}} := \inf\limits_{\mathbf{s}}\ \mathscr{H}_{\mathbf{s}}(f) \\ \qquad\qquad \text{s.t. } \mathbf{s}\in\mathcal{M}_k(\Theta),\ \mathscr{H}_{\mathbf{s}}(1)=1, \\ \qquad\qquad\quad \mathscr{H}_{\mathbf{s}}(P(\mathbf{y},\mathbf{x}))\in\mathcal{Q}_k^m(G). \end{cases} \tag{5.2}$$

We have the following theorem; see Appendix B for the proof.

Theorem 5.1. *Under Assumptions 2–5, the following are true:*
(i) $f_k^{\mathrm{primal}} = f_k^{\mathrm{dual}}$ *for each* $k\geq \max\{\lceil k_{\mathbf{x}}/2\rceil, \lceil k_{\mathbf{y}}/2\rceil\}$;
(ii) $\lim_{k\to\infty} f_k^{\mathrm{primal}} = \lim_{k\to\infty} f_k^{\mathrm{dual}} = f^\star$;
(iii) For any convergent subsequence $(\mathbf{s}_{\mathbf{e}}^{(k_i,\star)})_i$ *(always exists) of* $(\mathbf{s}_{\mathbf{e}}^{(k,\star)})_k$ *where* $\mathbf{s}^{(k,\star)}$ *is a minimizer of* (5.2), $\lim_{i\to\infty}\mathbf{s}_{\mathbf{e}}^{(k_i,\star)}$ *is a global minimizer of* (RPMIO). *Consequently, if the optimal solution set of* (RPMIO) *is a singleton, then* $\lim_{k\to\infty}\mathbf{s}_{\mathbf{e}}^{(k,\star)}$ *is the unique global minimizer.*

Remark 5.2. Since $\mathcal{Q}_k(\Theta)$ is an inner approximation for $\mathcal{P}(\mathcal{Y})$ while $\mathcal{M}_k^m(G)$ is an outer approximation for $\mathfrak{M}_+^m(\mathcal{X})$, the convergence of f_k^{primal} and f_k^{dual} to $f^\star$ is not from one side.

As for the SOS-convex case, finite convergence of (5.1)–(5.2) can be detected via certain flat extension conditions.

Lemma 5.3 (Jensen's inequality for PSD-convexity). *If a polynomial matrix* $Q(\mathbf{y})\in\mathbb{S}[\mathbf{y}]^m$ *is PSD-convex in* $\mathbf{y}$ *and a sequence* $\mathbf{s}=(s_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^\ell}\subset\mathbb{R}$ *admits a representing probability measure, then* $\mathscr{H}_{\mathbf{s}}(Q(\mathbf{y}))\succeq Q(\mathbf{s}_{\mathbf{e}})$.

Proof. For any $\mathbf{v} \in \mathbb{R}^m$, $\mathbf{v}^\intercal Q(\mathbf{y})\mathbf{v}$ is convex in $\mathbf{y}$. Since $\mathbf{s}$ admits a representing probability measure, by Jensen's inequality for convex functions, it holds

$$\mathbf{v}^\intercal Q(\mathbf{s_e})\mathbf{v} \le \mathscr{H}_{\mathbf{s}}(\mathbf{v}^\intercal Q(\mathbf{y})\mathbf{v}) = \mathbf{v}^\intercal \mathscr{H}_{\mathbf{s}}(Q(\mathbf{y}))\mathbf{v}$$

for any $\mathbf{v} \in \mathbb{R}^m$. Hence, $\mathscr{H}_{\mathbf{s}}(Q(\mathbf{y})) \succeq Q(\mathbf{s_e})$. ■

Let $d_\Theta := \max\{\lceil \deg(\theta_i)/2 \rceil, i = 1, \ldots, s\}$.

Theorem 5.4. *Suppose that Assumptions 3–5 hold. If the following flat extension conditions*

$$\begin{aligned}
&\exists \lceil k_{\mathbf{x}}/2 \rceil \le t_1 \le k \ s.t. \ \operatorname{rank}(M_{t_1}(\mathbf{S}^{(k,\star)})) = \operatorname{rank}(M_{t_1-d_G}(\mathbf{S}^{(k,\star)})),\\
&\exists \lceil k_{\mathbf{y}}/2 \rceil \le t_2 \le k \ s.t. \ \operatorname{rank}(M_{t_2}(\mathbf{s}^{(k,\star)})) = \operatorname{rank}(M_{t_2-d_\Theta}(\mathbf{s}^{(k,\star)}))
\end{aligned} \tag{5.3}$$

hold for some $k \ge \max\{\lceil k_{\mathbf{x}}/2 \rceil, \lceil k_{\mathbf{y}}/2 \rceil\}$*, where* $(\rho_k^\star, \mathbf{S}^{(k,\star)})$ *and* $\mathbf{s}^{(k,\star)}$ *are optimal solutions to (5.1) and (5.2) respectively, then the following are true:*

(i) $\mathbf{S}^{(k,\star)}$ *admits a representing measure* $\Phi^\star = \sum_{i=1}^{r_1} W_i \delta_{\mathbf{x}^{(i)}} \in \mathfrak{M}_+^m(\mathcal{X})$ *for some points* $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(r_1)} \in \mathcal{X}$ *and* $W_1, \ldots, W_{r_1} \in \mathbb{S}_+^m$*;*

(ii) $\mathbf{s}^{(k,\star)}$ *admits a representing probability measure* $\mu^\star = \sum_{j=1}^{r_2} \lambda_j \delta_{\mathbf{y}^{(j)}} \in \mathfrak{m}_+(\mathcal{Y})$ *for some* $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(r_2)} \in \mathcal{Y}$ *and positive real numbers* $\lambda_1, \ldots, \lambda_{r_2}$*;*

(iii) $f_k^{\mathrm{primal}} = f_k^{\mathrm{dual}} = f^\star$*;*

(iv) $\sum_{j=1}^{r_2} \lambda_j \mathbf{y}^{(j)}$ *is an optimal solution to (RPMIO);*

(v) For any decomposition $W_i = \sum_{q=1}^{m_i} \mathbf{v}^{(i,q)}(\mathbf{v}^{(i,q)})^\intercal$, $\mathbf{v}^{(i,q)} \in \mathbb{R}^m$, $i = 1, \ldots, r_1$*, it holds that*

$$P\left(\sum_{j=1}^{r_2} \lambda_j \mathbf{y}^{(j)}, \mathbf{x}^{(i)}\right) \mathbf{v}^{(i,q)} = 0, \quad q = 1, \ldots, m_i, i = 1, \ldots, r_1.$$

Proof. (i)–(ii). They follow from Theorem 3.2.

(iii). By slightly modifying the proof of Theorem 4.11 (i) and Theorem 4.12 (ii), we have $f_k^{\mathrm{primal}} = f_k^{\mathrm{dual}} \le f^\star$. As $\mathbf{s}^{(k,\star)}$ is feasible to (5.2), by Lemma 5.3 and (ii),

$$P\left(\sum_{j=1}^{r_2} \lambda_j \mathbf{y}^{(j)}, \mathbf{x}\right) \succeq \sum_{j=1}^{r_2} \lambda_j (P(\mathbf{y}^{(j)}, \mathbf{x})) = \mathscr{H}_{\mathbf{s}^{(k,\star)}}(P(\mathbf{y}, \mathbf{x})) \succeq 0,$$

for all $\mathbf{x} \in \mathcal{X}$. So, $\sum_{j=1}^{r_2} \lambda_j \mathbf{y}^{(j)}$ is feasible to (RPMIO). Then, as $f(\mathbf{y})$ is convex, by Jensen's inequality for convex functions,

$$f^\star \le f\left(\sum_{j=1}^{r_2} \lambda_j \mathbf{y}^{(j)}\right) \le \sum_{j=1}^{r_2} \lambda_j f(\mathbf{y}^{(j)}) = \mathscr{H}_{\mathbf{s}^{(k,\star)}}(f) = f_k^{\mathrm{dual}}. \tag{5.4}$$

Hence, it holds $f_k^{\mathrm{dual}} = f_k^{\mathrm{primal}} = f^\star$.

(iv). It follows from (iii) and (5.4).

(v). By (iii) and Theorem 4.3, $(f_k^{\mathrm{primal}}, \mathbf{S}^{(k,\star)})$ is an optimal solution to (4.2). Then, the conclusion follows from (iv) and Proposition 4.4. ■

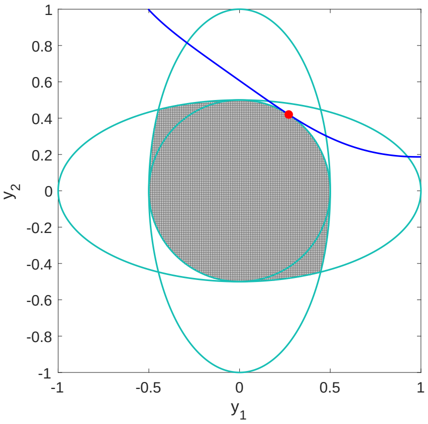

Figure 2. *The feasible set (the gray area), the contour line $f(\mathbf{y}) = 0.4504$ (the blue line) and the extracted minimizer $\mathbf{y}^\star$ (the red dot) in Example 5.6.*

Remark 5.5. From the proofs of Theorem 5.4 (iii), Theorem 4.11 (i) and Theorem 4.12 (ii), we can see that even without Assumption 5, if $f_k^{\text{primal}} = f_k^{\text{dual}}$, then the conclusions of Theorem 5.4 still hold under the flat extension conditions (5.3).

Example 5.6. *Consider the following instance of* (RPMIO)*:*

$$f^\star := \inf_{\mathbf{y}\in\mathbb{R}^2} f(\mathbf{y}) \quad s.t. \quad P(\mathbf{y},\mathbf{x}) \succeq 0, \ \forall \mathbf{x} \in \mathcal{X} := \{\mathbf{x} \in \mathbb{R}^2 \mid G(\mathbf{x}) \succeq 0\},$$

where $P(\mathbf{y},\mathbf{x})$ and $G(\mathbf{x})$ are defined as in Example 4.14. The following polynomial ([3, (5.2)]) is convex but not SOS-convex:

$$\begin{aligned} h(\mathbf{y}) = {} & 89 - 363y_1^4y_2 + \frac{51531}{64}y_2^6 - \frac{9005}{4}y_2^5 + \frac{49171}{16}y_2^4 + 721y_1^2 - 2060y_2^3 \\ & - 14y_1^3 + \frac{3817}{4}y_2^2 + 363y_1^4 - 9y_1^5 + 77y_1^6 + 316y_1y_2 + 49y_1y_2^3 \\ & - 2550y_1^2y_2 - 968y_1y_2^2 + 1710y_1y_2^4 + 794y_1^3y_2 + \frac{7269}{2}y_1^2y_2^2 \\ & - \frac{301}{2}y_1^5y_2 + \frac{2143}{4}y_1^4y_2^2 + \frac{1671}{2}y_1^3y_2^3 + \frac{14901}{16}y_1^2y_2^4 - \frac{1399}{2}y_1y_2^5 \\ & - \frac{3825}{2}y_1^3y_2^2 - \frac{4041}{2}y_1^2y_2^3 - 364y_2 + 48y_1. \end{aligned} \tag{5.5}$$

Let $f(\mathbf{y}) = h(y_1-1, y_2-1)/10000$ which is again convex but not SOS-convex. We have $k_{\mathbf{y}} = 6$ and $k_{\mathbf{x}} = 2$.

Solving the SDPs (5.1) *and* (5.2) *with $k = 3$, we get $f_3^{\text{primal}} = f_3^{\text{dual}} = 0.4504$ and the rank conditions* (5.3) *are satisfied with $t_1 = 3$ and $t_2 = 1$. By Theorem 5.4 and Remark 5.5, global optimality is reached and so $f^\star \approx 0.4504$. The extracted minimizer is $\mathbf{y}^\star = (0.2711, 0.4201)$. We show the feasible set, the contour line $f(\mathbf{y}) = 0.4504$ and the extracted minimizer $\mathbf{y}^\star$ in Figure 2.*

6. Conclusions. We have proposed a moment-SOS hierarchy for the robust PMI optimization problems with SOS-convexity for which asymptotic convergence is established and the flat extension condition is used to detect global optimality. This work generalizes most of the nice features of the moment-SOS hierarchy for the scalar polynomial optimization to the robust PMI optimization, and we would expect to stimulate more applications of robust PMI optimization in different fields (e.g., robust optimization, control theory). For the scalar polynomial optimization, various algebraic structures (e.g., symmetry, sparsity [47, 48, 49]) can be exploited to derive a structured moment-SOS hierarchy with lower computational complexity. A recent work on exploiting sparsity for PMIs could be found in [50]. As a line of further research, we intend to extend such techniques to the robust PMI optimization in future work.

Appendix A. An example for matrix-valued measure recovery. We construct a finitely atomic matrix-valued measure and then recover it from its moment matrix. Let $m = n = k = 2$, $r = 3$,

$$\mathbf{x}^{(1)} = (0.3855, -0.2746), \quad \mathbf{x}^{(2)} = (-0.5863, 0.9648) \quad \mathbf{x}^{(3)} = (0.1130, -0.8247),$$

and

$$W_1 = \begin{bmatrix} 0.6731 & -0.7569 \\ -0.7569 & 0.8512 \end{bmatrix}, \; W_2 = \begin{bmatrix} 0.6399 & 0.5259 \\ 0.5259 & 0.8048 \end{bmatrix}, \; W_3 = \begin{bmatrix} 0.0661 & -0.2294 \\ -0.2294 & 0.7968 \end{bmatrix}.$$

Note that $\text{rank}(W_1) = \text{rank}(W_3) = 1$ and $\text{rank}(W_3) = 2$. We let $\Phi = \sum_{i=1}^{3} W_i \delta_{\mathbf{x}^{(i)}}$. Denote by $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^2_4}$ and $M_2(\mathbf{S})$ the associated truncated moment sequence and moment matrix, respectively. Next, we recover $\mathbf{x}^{(i)}$'s and W_i's from $M_2(\mathbf{S})$ by the procedure described in Section 3.2.

We have $t = \text{rank}(M_2(\mathbf{S})) = \text{rank}(M_1(\mathbf{S})) = 4$. Compute the Cholesky decomposition $M_2(\mathbf{S}) = \widetilde{V}\widetilde{V}^\intercal$ with $\widetilde{V} \in \mathbb{R}^{12\times 4}$ and reduce matrix $\widetilde{V}$ to the column echelon form

$$U = \begin{bmatrix} 1.0000 & 0 & 0 & 0 \\ 0 & 1.0000 & 0 & 0 \\ 0 & 0 & 1.0000 & 0 \\ 0 & 0 & 0 & 1.0000 \\ 0.5775 & 0.3205 & -0.6606 & 0.5467 \\ -1.2518 & -0.8961 & -2.1350 & -3.1739 \\ 0.3025 & 0.0680 & -0.0703 & 0.1160 \\ -0.2657 & -0.0103 & -0.4532 & -0.6038 \\ -0.3451 & -0.0506 & 0.3762 & -0.0862 \\ 0.1975 & -0.1126 & 0.3368 & 0.7727 \\ 0.2682 & -0.1303 & -1.1301 & -0.2222 \\ 0.5088 & 0.8672 & 0.8678 & -0.1086 \end{bmatrix}.$$

The rows of U correspond to the monomials

$$v_2(\mathbf{x}, \mathbf{w}) = [w_1, w_2, x_1w_1, x_1w_2, x_2w_1, x_2w_2, x_1^2w_1, x_1^2w_2, x_1x_2w_1x_1x_2w_2, x_2^2w_1, x_2^2w_2]^\intercal.$$

From U, we can read the generating basis $b_2(\mathbf{x},\mathbf{w}) = [w_1, w_2, x_1w_1, x_1w_2]$ which satisfies that $v_2(\mathbf{v},\mathbf{w}) = Ub_2(\mathbf{x},\mathbf{w})$ holds at each pair of $\mathbf{x}^{(i)}$ and $\mathbf{w}^{(i,j)}$. Moreover, we can get from U the multiplication matricex of x_1 and x_2 with respect to $b_2(\mathbf{x},\mathbf{w})$:

$$N_1 = \begin{bmatrix} 0 & 0 & 1.0000 & 0 \\ 0 & 0 & 0 & 1.0000 \\ 0.3025 & 0.0680 & -0.0703 & 0.1160 \\ -0.2657 & -0.0103 & -0.4532 & -0.6038 \end{bmatrix},$$

and

$$N_2 = \begin{bmatrix} 0.5775 & 0.3205 & -0.6606 & 0.5467 \\ -1.2518 & -0.8961 & -2.1350 & -3.1739 \\ -0.3451 & -0.0506 & 0.3762 & -0.0862 \\ 0.1975 & -0.1126 & 0.3368 & 0.7727 \end{bmatrix}.$$

Build a random combination of multiplication matrices $N = 0.3687N_1 + 0.6313N_2$, and compute the ordered Schur decomposition $N = ATA^\intercal$ with

$$A = \begin{bmatrix} 0.2749 & -0.7785 & -0.3800 & 0.4171 \\ -0.9549 & -0.2787 & -0.1012 & 0.0171 \\ 0.0311 & -0.4340 & 0.9003 & -0.0103 \\ -0.1079 & 0.3577 & 0.1866 & 0.9086 \end{bmatrix}.$$

Compute the 4 points in (3.5):

$$\begin{bmatrix} a_1^\intercal N_1 a_1 \\ a_1^\intercal N_2 a_1 \end{bmatrix} = \begin{bmatrix} 0.1130 \\ -0.8247 \end{bmatrix}, \quad \begin{bmatrix} a_2^\intercal N_1 a_2 \\ a_2^\intercal N_2 a_2 \end{bmatrix} = \begin{bmatrix} 0.3855 \\ -0.2746 \end{bmatrix},$$
$$\begin{bmatrix} a_3^\intercal N_1 a_3 \\ a_3^\intercal N_2 a_3 \end{bmatrix} = \begin{bmatrix} -0.5863 \\ 0.9648 \end{bmatrix}, \quad \begin{bmatrix} a_4^\intercal N_1 a_4 \\ a_4^\intercal N_2 a_4 \end{bmatrix} = \begin{bmatrix} -0.5863 \\ 0.9648 \end{bmatrix}.$$

As we can see, all points $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$ and $\mathbf{x}^{(3)}$ have been recovered. Among the above 4 points, $\mathbf{x}^{(1)}$, $\mathbf{x}^{(3)}$ appear one time and $\mathbf{x}^{(3)}$ appears two times, which corresponds to the ranks of W_1, W_2 and W_3. We compute the matrix Λ and find $\mathcal{R} = \{1,2,3,4,5,6\}$ indexing the $mr = 6$ independent rows in Λ. We have

$$\Lambda_{\mathcal{R}} = \begin{bmatrix} 1.0000 & 0 & 1.0000 & 0 & 1.0000 & 0 \\ 0 & 1.0000 & 0 & 1.0000 & 0 & 1.0000 \\ 0.3855 & 0 & -0.5863 & 0 & 0.1130 & 0 \\ 0 & 0.3855 & 0 & -0.5863 & 0 & 0.1130 \\ -0.2746 & 0 & 0.9648 & 0 & -0.8247 & 0 \\ 0 & -0.2746 & 0 & 0.9648 & 0 & -0.8247 \end{bmatrix},$$

and

$$M_{\mathcal{R}}(\mathbf{S}) = \begin{bmatrix} 1.3791 & -0.4605 \\ -0.4605 & 2.4528 \\ -0.1082 & -0.6261 \\ -0.6261 & -0.0537 \\ 0.3781 & 0.9044 \\ 0.9044 & -0.1145 \end{bmatrix}.$$

Now we compute

$$\Lambda_{\mathcal{R}}^{-1} M_{\mathcal{R}}(\mathbf{S}) = \begin{bmatrix} 0.6731 & -0.7569 \\ -0.7569 & 0.8512 \\ 0.6399 & 0.5259 \\ 0.5259 & 0.8048 \\ 0.0661 & -0.2294 \\ -0.2294 & 0.7968 \end{bmatrix},$$

which corresponds exactly to $[W_1, W_2, W_3]^\intercal$.

Appendix B. Proof of Theorem 5.1.

Proof. (i). It can be justified by replacing $\mathcal{Q}_{\lceil k_{\mathbf{y}}/2 \rceil}(\Theta)$ with $\mathcal{Q}_k(\Theta)$ in the proof of Theorem 4.11 (i).

(ii). Fix an arbitrary $\varepsilon > 0$. By slightly modifying the proof of Theorem 4.11 (ii), we can see that there exists $k_1^{(\varepsilon)} \in \mathbb{N}$ such that $f_k^{\text{dual}} \leq f^\star + \varepsilon$ for all $k \geq k_1^{(\varepsilon)}$.

Let $\Phi^\star \in \mathfrak{M}_+^m(\mathcal{X})$ be the finitely atomic matrix-valued measure in Proposition 4.1 and let $\mathbf{S}^\star = (S_{\boldsymbol{\alpha}}^\star)_{\boldsymbol{\alpha} \in \mathbb{N}_{2k}^n}$ with $S_{\boldsymbol{\alpha}}^\star = \int_{\mathcal{X}} \mathbf{x}^{\boldsymbol{\alpha}} \mathrm{d}\Phi^\star(\mathbf{x})$. Then $\mathbf{S}^\star \in \mathcal{M}_k^m(G)$. Due to Proposition 4.1 and Remark 4.5, it holds

$$f(\mathbf{y}) - (f^\star - \varepsilon) - \mathscr{L}_{\mathbf{S}^\star}(P(\mathbf{y}, \mathbf{x})) > 0, \tag{B.1}$$

for all $\mathbf{y} \in \mathcal{Y}$. By Putinar's Positivstellensatz [44] (see also Theorem 2.3), there exists $k_2^{(\epsilon)} \in \mathbb{N}$, such that for all $k \geq k_2^{(\epsilon)}$, $(f^\star - \varepsilon, \mathbf{S}^\star)$ is feasible to (5.1) and hence $f_k^{\text{primal}} \geq f^\star - \varepsilon$.

As $\varepsilon > 0$ is arbitrary, by the weak duality, we have $\lim_{k\to\infty} f_k^{\text{primal}} = \lim_{k\to\infty} f_k^{\text{dual}} = f^\star$.

(iii). Fix a sequence $(\mathbf{s}^{(k,\star)})_k$ such that $\mathbf{s}^{(k,\star)} = (s_{\boldsymbol{\alpha}}^{(k,\star)})_{\boldsymbol{\alpha} \in \mathbb{N}_{2k}^\ell}$ is a minimizer of (5.2) for each k. For each $\boldsymbol{\alpha} \in \mathbb{N}^\ell$, define

$$N(\boldsymbol{\alpha}) \coloneqq \sqrt{\binom{\ell + \lceil \frac{|\boldsymbol{\alpha}|}{2} \rceil}{\ell} \sum_{i=1}^{\lceil \frac{|\boldsymbol{\alpha}|}{2} \rceil} b^{2i}}.$$

By Lemma 4.10, $|s_{\boldsymbol{\alpha}}^{(k,\star)}| \leq N(\boldsymbol{\alpha})$ for all k and $\boldsymbol{\alpha} \in \mathbb{N}_{2k}^\ell$. Complete each $\mathbf{s}^{(k,\star)}$ with zeros to make it an infinite vector. Then,

$$\left\{ (s_{\boldsymbol{\alpha}}^{(k,\star)})_{\boldsymbol{\alpha} \in \mathbb{N}^\ell} \right\}_k \subset \prod_{\boldsymbol{\alpha} \in \mathbb{N}^\ell} [-N(\boldsymbol{\alpha}), N(\boldsymbol{\alpha})].$$

By Tychonoff's theorem, the product space $\prod_{\boldsymbol{\alpha} \in \mathbb{N}^\ell} [-N(\boldsymbol{\alpha}), N(\boldsymbol{\alpha})]$ is compact in the product topology. Therefore, there exists a subsequence $(\mathbf{s}^{(k_i,\star)})_i$ of $(\mathbf{s}^{(k,\star)})_k$ and $\mathbf{s}^\star = (s_{\boldsymbol{\alpha}}^\star)_{\boldsymbol{\alpha} \in \mathbb{N}^\ell}$ such that $\lim_{i\to\infty} s_{\boldsymbol{\alpha}}^{(k_i,\star)} = s_{\boldsymbol{\alpha}}^\star$ for all $\boldsymbol{\alpha} \in \mathbb{N}^\ell$. By the pointwise convergence, we have (a) $\mathbf{s}^\star \in \mathcal{M}_k(\Theta)$ for all k; (b) $\mathscr{H}_{\mathbf{s}^\star}(1) = 1$; (c) $\mathscr{H}_{\mathbf{s}^\star}(P(\mathbf{y}, \mathbf{x})) \succeq 0$ for all $x \in \mathcal{X}$. By (a), (b), Putinar's Positivstellensatz and Haviland's theorem, $\mathbf{s}^\star$ has a representing probability measure μ supported on $\mathcal{Y}$, i.e., $s_{\boldsymbol{\alpha}}^\star = \int_{\mathcal{Y}} \mathbf{y}^{\boldsymbol{\alpha}} \mathrm{d}\mu(\mathbf{y})$ for all $\boldsymbol{\alpha} \in \mathbb{N}^\ell$. By Lemma 5.3, $P(\mathbf{s}_{\mathbf{e}}^\star, \mathbf{x}) \succeq \mathscr{H}_{\mathbf{s}^\star}(P(\mathbf{y}, \mathbf{x})) \succeq 0$ and $\theta_i(\mathbf{s}_{\mathbf{e}}^\star) \geq \mathscr{H}_{\mathbf{s}^\star}(\theta_i) \geq 0$, $i = 1, \ldots, s$. Hence $\mathbf{s}_{\mathbf{e}}^\star$ is feasible to (RPMIO). Moreover, as $f(\mathbf{y})$ is convex in $\mathbf{y}$, by (ii) and the pointwise convergence,

$$f^\star = \mathscr{H}_{\mathbf{s}^\star}(f) \geq f(\mathbf{s}_{\mathbf{e}}^\star),$$

which indicates that $\mathbf{s}_{\mathbf{e}}^{\star}$ is a minimizer of (RPMIO). ■

REFERENCES

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

[2] A. A. Ahmadi and P. A. Parrilo, *A convex polynomial that is not sos-convex*, Mathematical Programming, 135 (2012), pp. 275–292.

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

[4] I. P. Androulakis, C. D. Maranas, and C. A. Floudas, *αBB: A global optimization method for general constrained nonconvex problems*, Journal of Global Optimization, 7 (1995), pp. 337–363.

[5] A. Ben-Tal, L. E. Ghaoui, and A. Nemirovski, *Robust semidefinite programming*, in Handbook of Semidefinite Programming - Theory, Algorithms, and Applications, H. Wolkowicz, R. Saigal, and L. Vandenberghe, eds., Kluwer Academic Publisher, Boston, 2000, pp. 139–162.

[6] D. Bertsimas, D. B. Brown, and C. Caramanis, *Theory and applications of robust optimization*, SIAM Review, 53 (2011), pp. 464–501.

[7] P.-A. Bliman, *A convex approach to robust stability for linear systems with uncertain scalar parameters*, SIAM Journal on Control and Optimization, 42 (2004), pp. 2016–2042.

[8] P.-A. Bliman, *On robust semidefinite programming*, in Proceedings of the 16th International Symposium on Mathematical Theory of Networks and Systems (MTNS2004), 2004.

[9] J. M. Borwein, *Direct theorems in semi-infinite convex programming*, Mathematical Programming, 21 (1981), pp. 301–318.

[10] G. Chesi, *LMI techniques for optimization over polynomials in control: A survey*, IEEE Transactions on Automatic Control, 55 (2010), pp. 2500–2510.

[11] T. D. Chuong, V. Jeyakumar, G. Li, and D. Woolnough, *Exact dual semi-definite programs for affinely adjustable robust SOS-convex polynomial optimization problems*, Optimization, 71 (2022), pp. 3539–3569.

[12] J. Cimprič and A. Zalar, *Moment problems for operator polynomials*, Journal of Mathematical Analysis and Applications, 401 (2013), pp. 307–316.

[13] R. M. Corless, P. M. Gianni, and B. M. Trager, *A reordered Schur factorization method for zero-dimensional polynomial systems with multiple roots*, in Proceedings of the 1997 International Symposium on Symbolic and Algebraic Computation, ISSAC '97, New York, NY, USA, 1997, Association for Computing Machinery, p. 133–140.

[14] R. E. Curto and L. A. Fialkow, *Solution of the truncated complex moment problem for flat data*, Memoirs of the American Mathematical Society, 119 (1996).

[15] R. E. Curto and L. A. Fialkow, *The truncated complex k-moment problem*, Transactions of the American Mathematical Society, 352 (2000), pp. 2825–2855.

[16] A. P. David Damanik and B. Simon, *The analytic theory of matrix orthogonal polynomials*, Surveys in Approximation Theory, 4 (2008), pp. 1–85.

[17] H. Dette and W. J. Studden, *Quadrature formulas for matrix measures—a geometric approach*, Linear Algebra and its Applications, 364 (2003), pp. 33–64.

[18] B. Dumitrescu, B. C. Şicleru, and R. Ştefan, *Computing the controllability radius: a semi-definite programming approach*, IET Control Theory & Applications, 3 (2009), pp. 654–660(6).

[19] A. J. Duran and P. Lopez-Rodriguez, *Density questions for the truncated matrix moment problem*, Canadian Journal of Mathematics, 49 (1997), p. 708–721.

[20] F. Gamboa, J. Nagel, and A. Rouault, *Sum rules and large deviations for spectral matrix measures*, Bernoulli, 25 (2019), pp. 712 – 741.

[21] J. Helton and J. Nie, *Semidefinite representation of convex sets*, Mathematical Programming, 122 (2010), pp. 21–64.

[22] D. Henrion and J. B. Lasserre, *Detecting global optimality and extracting solutions in gloptipoly*, in Positive Polynomials in Control, D. Henrion and A. Garulli, eds., Springer, Berlin, Heidelberg, 2005, pp. 293–310.

[23] D. Henrion and J. B. Lasserre, *Convergent relaxations of polynomial matrix inequalities and static output feedback*, IEEE Transactions on Automatic Control, 51 (2006), pp. 192–202.
[24] D. Henrion and J.-B. Lasserre, *Inner approximations for polynomial matrix inequalities and robust stability regions*, IEEE Transactions on Automatic Control, 57 (2012), pp. 1456–1467.
[25] L. Huang, D. Liu, and Y. Fang, *Convergence of an SDP hierarchy and optimality of robust convex polynomial optimization problems*, Annals of Operations Research, 320 (2023), pp. 33–59.
[26] V. Jeyakumar, G. Li, and J. Vicente-Pérez, *Robust SOS-convex polynomial optimization problems: exact SDP relaxations*, Optimization Letters, 9 (2015), pp. 1–18.
[27] C. Josz and D. Henrion, *Strong duality in Lasserre's hierarchy for polynomial optimization*, Optimization Letters, 10 (2016), pp. 3–10.
[28] D. P. Kimsey, *Matrix-valued moment problems*, PhD thesis, Drexel University, Philadelphia, PA, 2011.
[29] D. P. Kimsey and M. Trachana, *On a solution of the multidimensional truncated matrix-valued moment problem*, Milan Journal of Mathematics, 90 (2022), pp. 17–101.
[30] M. Kojima, *Sums of squares relaxations of polynomial semidefinite programs*, Research Report B-397, Dept. of Mathematical and computing Sciences, Tokyo Institute of Technology, Tokyo, Japan, (2003).
[31] J. B. Lasserre, *Global optimization with polynomials and the problem of moments*, SIAM Journal on Optimization, 11 (2001), pp. 796–817.
[32] J. B. Lasserre, *A semidefinite programming approach to the generalized problem of moments*, Mathematical Programming, 112 (2008), pp. 65–92.
[33] J. B. Lasserre, *Convexity in semialgebraic geometry and polynomial optimization*, SIAM Journal on Optimization, 19 (2009), pp. 1995–2014.
[34] J. B. Lasserre, *Min-max and robust polynomial optimization*, Journal of Global Optimization, 51 (2011), pp. 1–10.
[35] G. Li, A. K. C. Ma, and T. K. Pong, *Robust least square semidefinite programming with applications*, Computational Optimization and Applications, 58 (2014), pp. 347–379.
[36] J. Löfberg, *YALMIP : a toolbox for modeling and optimization in MATLAB*, in 2004 IEEE International Conference on Robotics and Automation (IEEE Cat. No.04CH37508), 2004, pp. 284–289.
[37] R. Louca and E. Bitar, *A hierarchy of polyhedral approximations of robust semidefinite programs*, in 2016 IEEE 55th Conference on Decision and Control (CDC), 2016, pp. 7056–7062.
[38] N. H. A. Mai, J. B. Lasserre, V. Magron, and J. Wang, *Exploiting constant trace property in large-scale polynomial optimization*, ACM Trans. Math. Softw., 48 (2022).
[39] C. D. Maranas and C. A. Floudas, *Global minimum potential energy conformations of small molecules*, Journal of Global Optimization, 4 (1994), pp. 135–170.
[40] A. Mosek, *The MOSEK Optimization Software*. https://www.mosek.com/.
[41] J. Nie, *Polynomial matrix inequality and semidefinite representation*, Mathematics of Operations Research, 36 (2011), pp. 398–415.
[42] A. Ohara and Y. Sasaki, *On solvability and numerical solutions of parameter-dependent differential matrix inequality*, in Proceedings of the 40th IEEE Conference on Decision and Control (Cat. No.01CH37228), vol. 4, 2001, pp. 3593–3594.
[43] Y. Oishi, *A region-dividing approach to robust semidefinite programming and its error bound*, in Proceedings of the 2006 American Control Conference, Minneapolis, Minnesota, USA, 2006, pp. 123–129.
[44] M. Putinar, *Positive polynomials on compact semi-algebraic sets*, Indiana University Mathematics Journal, 42 (1993), pp. 969–984.
[45] C. W. Scherer and C. W. J. Hol, *Matrix sum-of-squares relaxations for robust semi-definite programs*, Mathematical Programming, 107 (2006), pp. 189–211.
[46] L. Slot and M. Laurent, *Sum-of-squares hierarchies for binary polynomial optimization*, Mathematical Programming, 197 (2023), pp. 621–660.
[47] H. Waki, S. Kim, M. Kojima, and M. Muramatsu, *Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity*, SIAM Journal on Optimization, 17 (2006), pp. 218–242.
[48] J. Wang, V. Magron, and J.-B. Lasserre, *TSSOS: A moment-SOS hierarchy that exploits term sparsity*, SIAM Journal on Optimization, 31 (2021), pp. 30–58.
[49] J. Wang, V. Magron, J. B. Lasserre, and N. H. A. Mai, *CS-TSSOS: Correlative and term sparsity for large-scale polynomial optimization*, ACM Transactions on Mathematical Software, 48 (2022),

pp. 1–26.
[50] Y. Zheng and G. Fantuzzi, *Sum-of-squares chordal decomposition of polynomial matrix inequalities*, Mathematical Programming, 197 (2023), pp. 71–108.