Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

A Moment-Sum-of-Squares Hierarchy for Robust Polynomial Matrix Inequality Optimization with Sum-of-Squares Convexity

Feng Guo, Jie Wang

To cite this article:
Feng Guo, Jie Wang (2024) A Moment-Sum-of-Squares Hierarchy for Robust Polynomial Matrix Inequality Optimization with Sum-of-Squares Convexity. Mathematics of Operations Research

Published online in Articles in Advance 23 Jul 2024

. https://doi.org/10.1287/moor.2023.0361

MATHEMATICS OF OPERATIONS RESEARCH
https://pubsonline.informs.org/journal/moor
Articles in Advance, pp. 1–28
ISSN 0364-765X (print), ISSN 1526-5471 (online)

A Moment-Sum-of-Squares Hierarchy for Robust Polynomial Matrix Inequality Optimization with Sum-of-Squares Convexity

Feng Guo,[a] **Jie Wang**[b,*]

[a] School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China; [b] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

*Corresponding author

Contact: fguo@dlut.edu.cn, https://orcid.org/0000-0003-3106-2653 (FG); wangjie212@amss.ac.cn, https://orcid.org/0000-0002-9681-1451 (JW)

Received: November 27, 2023
Revised: April 22, 2024
Accepted: May 24, 2024
Published Online in Articles in Advance: July 23, 2024

MSC2020 Subject Classifications: Primary: 90C23, 90C17; secondary: 90C22, 90C26

https://doi.org/10.1287/moor.2023.0361

Abstract. We study a class of polynomial optimization problems with a robust polynomial matrix inequality (PMI) constraint where the uncertainty set itself is also defined by a PMI. These can be viewed as matrix generalizations of semi-infinite polynomial programs because they involve actually infinitely many PMI constraints in general. Under certain sum-of-squares (SOS)-convexity assumptions, we construct a hierarchy of increasingly tight moment-SOS relaxations for solving such problems. Most of the nice features of the moment-SOS hierarchy for the usual polynomial optimization are extended to this more complicated setting. In particular, asymptotic convergence of the hierarchy is guaranteed, and finite convergence can be certified if some flat extension condition holds true. To extract global minimizers, we provide a linear algebra procedure for recovering a finitely atomic matrix-valued measure from truncated matrix-valued moments. As an application, we are able to solve the problem of minimizing the smallest eigenvalue of a polynomial matrix subject to a PMI constraint. 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. Finally, an extension to the nonconvexity setting is provided under a rank 1 condition. To obtain the above-mentioned results, techniques from real algebraic geometry, matrix-valued measure theory, and convex optimization are employed.

Funding: This work was supported by the National Key Research and Development Program of China [Grant 2023YFA1009401] and the National Natural Science Foundation of China [Grants 11571350 and 12201618].

Keywords: polynomial matrix optimization • polynomial matrix inequality • robust optimization • moment-SOS hierarchy • semidefinite relaxation

1. Introduction

Polynomial optimization problems with polynomial matrix inequality (PMI) constraints have a wide range of applications in many fields (Aravanis et al. [5], Henrion and Lasserre [21], Henrion and Lasserre [22], Ichihara [24], Klep and Nie [29], Pozdyayev [44]). In particular, as special cases of PMIs, linear or bilinear matrix inequality constrained problems appear frequently in many synthesis problems for linear systems in optimal control (VanAntwerp and Braatz [49]). Because of 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 (Bertsimas et al. [6]).

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

$$\begin{cases} f^\star := \inf\limits_{\mathbf{y}\in\mathcal{Y}} & f(\mathbf{y}) \\ \qquad\quad \text{s.t.} & \mathcal{Y}\subseteq\mathbb{R}^\ell,\ P(\mathbf{y},\mathbf{x})\succeq 0,\ \ \forall\mathbf{x}\in\mathcal{X}\subseteq\mathbb{R}^n, \end{cases} \tag{1}$$

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

$$\mathcal{Y} := \{\mathbf{y}\in\mathbb{R}^\ell \mid \theta_1(\mathbf{y})\geq 0,\ldots,\theta_s(\mathbf{y})\geq 0\}; \tag{2}$$

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

$$\mathcal{X} := \{\mathbf{x} \in \mathbb{R}^n \,|\, G(\mathbf{x}) \succeq 0\} \tag{3}$$

defined by a $q \times q$ symmetric polynomial matrix $G(\mathbf{x})$; and $P(\mathbf{y}, \mathbf{x})$ is an $m \times m$ symmetric polynomial matrix in $\mathbf{y}$ and $\mathbf{x}$ (i.e., $P(\mathbf{y}, \mathbf{x})$ depends polynomially on the decision variable $\mathbf{y}$ and the uncertain parameter $\mathbf{x}$). This problem essentially has infinitely many PMI constraints corresponding to different points of $\mathcal{X}$. We assume that the set of optimizers of (1) is nonempty and make the following sum-of-squares (SOS)-convexity assumptions on (1).

Assumption 1. (i) *The polynomials* $f(\mathbf{y})$, $-\theta_1(\mathbf{y}), \ldots, -\theta_s(\mathbf{y})$ *are SOS convex* (*Definition* 1); (ii) *the polynomial matrix* $-P(\mathbf{y}, \mathbf{x})$ *is PSD-SOS convex* (*Definition* 3) *in* $\mathbf{y}$ *for all* $\mathbf{x} \in \mathcal{X}$; *and* (iii) *the set* $\mathcal{X}$ *is compact*.

To highlight the modeling power of (1) under Assumption 1, let us name a few problems from different fields that can be modeled as an instance of (1). First of all, we note that if $\mathcal{Y} = \mathbb{R}^\ell$ and $f(\mathbf{y})$, $P(\mathbf{y}, \mathbf{x})$ depend affinely on $\mathbf{y}$, then we retrieve the robust polynomial semidefinite program (SDP) considered in Scherer and Hol [46]. 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 modeled as a robust polynomial SDP (hence, an instance of (1)). 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 (Chesi [8], Henrion and Lasserre [22], Scherer and Hol [46]). If the involved functions in the differential equations are polynomials, these problems often take the form of robust polynomial SDPs. In the context of risk management, the robust correlation stress testing where data uncertainty arises because of untimely recording of portfolio holdings can be formulated as a robust least square SDP fitting into the form of (1) (Li et al. [35]). Moreover, the deterministic PMI optimization problem of maximizing a polynomial function $h(\mathbf{x})$ subject to a PMI constraint $G(\mathbf{x}) \succeq 0$ (recall that multiple PMI constraints can be easily merged into one by diagonal augmentation) can be also formulated as an instance of (1):

$$\inf_{y \in \mathbb{R}} y \quad \text{s.t.} \;\; y - h(\boldsymbol{x}) \geq 0, \;\; \forall \boldsymbol{x} \in \mathcal{X}. \tag{4}$$

Because the deterministic PMI optimization problem (including the usual polynomial optimization problem as a special case) is already difficult to solve in general, solving (1) (even under Assumption 1) is more challenging.

For deterministic PMI optimization problems, Kojima [30] proposed SOS relaxations by utilizing a penalty function and a generalized Lagrangian dual, and subsequently, Henrion and Lasserre [21] gave a hierarchy of moment relaxations allowing them to detect finite convergence and to extract global minimizers. Recently, there has been increasing interest in studying robust polynomial optimization problems; see, for example, Chuong et al. [9], Huang et al. [23], Jeyakumar et al. [25], and Lasserre [34]. However, robust PMI constraints are not considered in these works. For robust polynomial SDPs, Scherer and Hol [46] provided a hierarchy of matrix SOS relaxations by establishing Putinar's style Positivstellensätz for polynomial matrices. There are also approaches for solving other types of robust SDPs (Li et al. [35], Louca and Bitar [37], Oishi [43]). As far as the authors know, there is *little* work on how to solve or even approximate (1) in general.

Before introducing our main contributions, we would like to point out that the matrix SOS relaxations in Scherer and Hol [46] cannot be straightforwardly extended to handle (1) because of the *nonlinearity* of $P(\mathbf{y}, \mathbf{x})$ in $\mathbf{y}$. Also, the dual-*moment* facet of the matrix SOS relaxations for (1) was not investigated in Scherer and Hol [46] or elsewhere. This motivates us to establish a moment-SOS hierarchy for (1) (under Assumption 1) in full generality.

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), the dual cone of $\mathcal{P}^m(\mathcal{X})$ consists of tracial $\mathcal{X}$-moment functionals on $\mathbb{S}[\mathbf{x}]^m$ (Definition 6) while justifying the membership of a linear functional 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 7). Therefore, to explore the dual aspect of the matrix SOS relaxations for (1), 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 Kimsey and Trachana [28] for a thorough introduction on the history and background about the matrix-valued moment problem. For the scalar moment problem ($m = 1$), because of Haviland's theorem and Putinar's Positivstellensätz, the representing measure is guaranteed by the PSD-ness 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 [12] gave the celebrated flat extension condition (FEC) on the moment matrix as a sufficient condition for the existence of a representing measure, which allowed them to detect finite convergence of Lasserre's hierarchy and extract global minimizers (Henrion and Lasserre [20]). Recently, Kimsey and Trachana [28] obtained a flat extension theorem for the truncated matrix-valued moment problem. Very recently, Nie [42, chapter 10.3.2] investigated the truncated matrix-valued K-moment problem where K is defined by polynomial equalities and PMIs. He gave a solution to this problem using an FEC on the *block rank* of the associated moment matrix. We would like to point out that the FEC on the block rank is stronger than on the usual rank (see Remark 2). In the rest of this paper, unless stated otherwise, the abbreviation FEC means the flat extension condition on the usual rank of the associated moment matrix. Unlike the scalar case, to the best of our knowledge, there is *little* work in the literature to link the theory of matrix-valued moments to PMI optimization. In this paper, we aim to construct a moment-SOS hierarchy for (1) with SOS convexity by employing Scherer–Hol's Positivstellensätz and the matrix-valued measure theory.

1.1. Contributions

Our main contributions are summarized as follows.

1. As the first contribution, we provide a solution to the truncated matrix-valued $\mathcal{X}$-moment problem using the FEC. Furthermore, we develop a linear algebra procedure for retrieving a finitely atomic representing measure whenever the FEC holds, which extends Henrion–Lasserre's algorithm to the matrix setting. We remark that those extensions from the scalar setting to the matrix setting are not straightforward applications of existing results. Indeed, the matrix-valued moment matrix is a multi-indexed *block* matrix, and the kernel of a truncated PSD moment matrix is a linear subspace rather than a real radical ideal, making the matrix setting more involved.

2. We establish a moment-SOS hierarchy for solving (1) with SOS convexity. Here, the main difficulty stems from the nonlinearity of $P(\mathbf{y},\mathbf{x})$ in $\mathbf{y}$. To overcome this obstacle, we reformulate (1) to a conic optimization problem via the Lagrange dual theory, and then, we replace the conic constraints with more tractable matrix quadratic module constraints or matrix-valued pseudomoment cone constraints. This yields a sequence of upper bounds on the optimum of (1) with guaranteed asymptotic convergence. More importantly, we show that if the FEC holds, then finite convergence occurs, and we can extract a globally optimal solution $\mathbf{y}^\star$ of (1) as well as the points $\mathbf{x}\in\Delta(\mathbf{y}^\star)$ and the corresponding vectors $\mathbf{v}$, where

$$\Delta(\mathbf{y}^\star) := \{\mathbf{x}\in\mathcal{X}\,|\,\exists\mathbf{v}\in\mathbb{R}^m \text{ s.t. } P(\mathbf{y}^\star,\mathbf{x})\mathbf{v}=0\} \tag{5}$$

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

3. For the linear case of (1), 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 [46]. As a complement to the matrix SOS relaxations in Scherer and Hol [46], the dual matrix-valued moment relaxation allows us to detect finite convergence and to extract optimal solutions. 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 (1) is weakened to convexity, we also provide a sequence of SDPs that can approximate the optimal value of (1) as closely as desired. Moreover, finite convergence can be detected via certain FECs. For the general nonconvex case, we give, respectively, conditions under which a lower bound on the optimum, an upper bound on the optimum, or a globally optimal solution can be retrieved.

Although the moment-SOS hierarchy provides a powerful framework for scalar polynomial optimization, extending it to handle robust PMI constraints is a nontrivial task even in the SOS-convex case. A straightforward idea would be scalarizing the PMI constraint via the equivalence relation: an $m\times m$ polynomial matrix $P\succeq 0$ if and only if $\mathbf{z}^\top P\mathbf{z}\geq 0$ for all $\mathbf{z}\in\mathbb{R}^m$. However, this approach bears two drawbacks. (a) It necessitates m auxiliary variables, which increase the problem size, and (b) the information on the structure (e.g., sparsity) of the polynomial matrix becomes less transparent. By contrast, our approach natively treats (1) at the matrix level, thus preserving its matrix structure and allowing the theory of matrix-valued measures to come into play. The information on matrix structures can be exploited to develop a structured moment-SOS hierarchy in order to tackle large-scale application problems (which will be addressed in a follow-up paper).

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 retrieving the representing measure. In Section 4, we construct a moment-SOS hierarchy for solving (1) with SOS convexity and treat some special cases. Extensions of the results to the general convex case and the nonconvex case are 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 $\mathbf{x}$ (respectively, $\mathbf{y}$) the n-tuple (respectively, ℓ-tuple) of variables $(x_1,\dots,x_n)$ (respectively, $(y_1,\dots,y_\ell)$). The symbol $\mathbb{N}$ (respectively, $\mathbb{R}$) denotes the set of nonnegative integers (respectively, real numbers). Denote by $\mathbb{R}^m$ (respectively, $\mathbb{R}^{l_1\times l_2}$, $\mathbb{S}^m$) the m-dimensional real vector (respectively, $l_1\times l_2$ real matrix, $m\times m$ symmetric real matrix) space. For $\mathbf{v}\in\mathbb{R}^m$ (respectively, $N\in\mathbb{R}^{l_1\times l_2}$), the symbol $\mathbf{v}^\top$ (respectively, $N^\top$) denotes the transpose of $\mathbf{v}$ (respectively, N). For a matrix $N\in\mathbb{R}^{m\times m}$, $\mathrm{tr}(N)$ 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. For two matrices N_1 and N_2 of the same size, $\langle N_1,N_2\rangle$ denotes the inner product $\mathrm{tr}(N_1^\top N_2)$ 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 $\boldsymbol{\alpha}=(\alpha_1,\dots,\alpha_n)\in\mathbb{N}^n$, 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 $\mathbb{N}^n_k:=\{\boldsymbol{\alpha}\in\mathbb{N}^n \mid |\boldsymbol{\alpha}|\le k\}$ 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 $\mathbb{R}[\mathbf{x}]$ (respectively, $\mathbb{S}[\mathbf{x}]^m$) denote the set of real polynomials (respectively, $m\times m$ 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 polynomial $f(\mathbf{x})\in\mathbb{R}[\mathbf{x}]$, if there exist polynomials $f_1(\mathbf{x}),\dots,f_t(\mathbf{x})$ such that $f(\mathbf{x})=\sum_{i=1}^t f_i(\mathbf{x})^2$; then, we call $f(\mathbf{x})$ an SOS polynomial. For an $\mathbb{R}$-vector space A, denote by A^* the dual space of linear functionals from A to $\mathbb{R}$. Given a cone $B\subseteq A$, its dual cone is $B^*:=\{L\in A^* \mid L(b)\ge 0,\ \forall b\in B\}$.

2.1. SOS Polynomial Matrices, SOS Convexity, and Positivstellensatz for Polynomial Matrices

For an $l_1\times l_2$ polynomial matrix $T(\mathbf{x})=[T_{ij}(\mathbf{x})]$, define

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

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

$$u_d(\mathbf{x}):=[1, x_1, x_2, \dots, x_n, x_1^2, x_1x_2, \dots, x_n^d]^\top, \tag{6}$$

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

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

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 matrix with respect to $u_d(\mathbf{x})$ if there exists some $Q\in\mathbb{R}^{l\times q|\mathbb{N}^n_d|}$ satisfying

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

We thus have the following result.

Proposition 1 (Scherer and Hol [46, Lemma 1]). *A polynomial matrix $\Sigma(\mathbf{x})\in\mathbb{S}[\mathbf{x}]^q$ is an SOS matrix with respect to the monomial basis $u_d(\mathbf{x})$ if and only if there exists $Z\in\mathbb{S}_+^{q|\mathbb{N}^n_d|}$ such that*

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

Now, let us recall some basic concepts about SOS convexity.

Definition 1 (Helton and Nie [19]). A polynomial $h\in\mathbb{R}[\mathbf{y}]$ is *SOS convex* if its Hessian $\nabla_{\mathbf{yy}}h(\mathbf{y})$ is an SOS matrix.

Although checking the convexity of a polynomial is generally nondeterministic polynomial time (NP) hard (Ahmadi et al. [2]), the SOS convexity can be justified numerically by solving an SDP by Proposition 1. Recall Definition 2.

Definition 2. 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 [40] gave an extension of SOS convexity to polynomial matrices.

Definition 3 (Nie [40]). 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

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

In other words, $Q(\mathbf{y})$ is PSD-SOS convex if and only if $\mathbf{v}^\top Q(\mathbf{y})\mathbf{v}$ 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 requires checking the Hessian $\nabla_{\mathbf{yy}}(\mathbf{v}^\top Q(\mathbf{y})\mathbf{v})$ for every $\mathbf{v} \in \mathbb{R}^m$, which is hard in general. See Appendix A for a stronger and easier-to-check condition called uniform PSD-SOS convexity given in Nie [40].

We next recall the Positivstellensätz for polynomial matrices obtained in Scherer and Hol [46]. Define the bilinear mapping

$$(\cdot,\cdot)_m : \mathbb{R}^{mq\times mq} \times \mathbb{R}^{q\times q} \to \mathbb{R}^{m\times m}, \quad (A,B)_m = \mathrm{tr}_m(A^\top(I_m \otimes B)),$$

with

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

Assumption 2. *For the defining matrix $G(\mathbf{x})$ of $\mathcal{X}$ in (3), there exists $r \in \mathbb{R}$ and an SOS polynomial matrix $\Sigma(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^q$ such that $r^2 - \|\mathbf{x}\|^2 - \langle\Sigma(\mathbf{x}), G(\mathbf{x})\rangle$ is an SOS.*

Theorem 1 (Scherer and Hol [46, Corollary 1]). *Let Assumption 2 hold and $F(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^m$ be positive definite on $\mathcal{X}$. Then, there exist SOS polynomial matrices $\Sigma_0(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^m$ and $\Sigma_1(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^{mq}$ such that*

$$F(\mathbf{x}) = \Sigma_0(\mathbf{x}) + (\Sigma_1(\mathbf{x}), G(\mathbf{x}))_m.$$

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

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

and we define the *matrix quadratic module* by $\mathcal{Q}^m(G) := \cup_{k\in\mathbb{N}} \mathcal{Q}_k^m(G)$. By Proposition 1, checking membership in $\mathcal{Q}_k^m(G)$ can be accomplished with an SDP.

2.2. 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 Damanik et al. [14], Dette and Studden [15], Duran and Lopez-Rodriguez [17], and Gamboa et al. [18]. 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 4. 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 $\Phi : B(\mathcal{X}) \to \mathbb{R}^{m\times m}$ 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 $\mathbf{v}^\top\Phi(\mathcal{A})\mathbf{v} \ge 0$ 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 $\mathrm{supp}(\Phi) := \cup_{i,j=1}^m \mathrm{supp}(\phi_{ij})$ is called the support of the matrix-valued measure Φ. A function $h : \mathcal{X} \to \mathbb{R}$ 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

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

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

Definition 5. A *finitely atomic PSD* matrix-valued measure $\Phi \in \mathfrak{M}^m_+(\mathcal{X})$ is a matrix-valued measure of form $\Phi = \sum_{i=1}^r W_i \delta_{\mathbf{x}^{(i)}}$, where $W_i \in \mathbb{S}^m_+$, $\mathbf{x}^{(i)}$'s are distinct points in $\mathcal{X}$, and $\delta_{\mathbf{x}^{(i)}}$ denotes the Dirac measure centered at $\mathbf{x}^{(i)}$, $i = 1, \ldots, r$.

Definition 6. A linear functional $\mathscr{L} : \mathbb{S}[\mathbf{x}]^m \to \mathbb{R}$ is called a tracial $\mathcal{X}$-moment functional if there exists a matrix-valued measure $\Phi \in \mathfrak{M}^m_+(\mathcal{X})$ such that

$$\operatorname{supp}(\Phi) \subseteq \mathcal{X},\ \mathscr{L}(F) = \int_{\mathcal{X}} \operatorname{tr}(F(\mathbf{x})\mathrm{d}\Phi(\mathbf{x})) = \sum_{i,j} \int_{\mathcal{X}} F_{ij}(\mathbf{x})\mathrm{d}\phi_{ij}(\mathbf{x}), \quad \forall F(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^m. \tag{7}$$

The matrix-valued measure $\Phi \in \mathfrak{M}^m_+(\mathcal{X})$ is called a *representing measure* of $\mathscr{L}$, and we write $\mathscr{L}_\Phi$ for $\mathscr{L}$ to indicate the associated measure.

Now, we define the convex cones

$$\mathcal{L}^m(\mathcal{X}) := \{\mathscr{L} : \mathbb{S}[\mathbf{x}]^m \to \mathbb{R} \,|\, \mathscr{L} \text{ is a tracial } \mathcal{X}\text{-moment functional}\} \tag{8}$$

and

$$\mathcal{P}^m(\mathcal{X}) := \{F(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^m \,|\, F(\mathbf{x}) \succeq 0,\ \ \forall \mathbf{x} \in \mathcal{X}\}. \tag{9}$$

The following theorem is a matrix version of Haviland's theorem (Cimprič and Zalar [10], Schmüdgen [47]).

Theorem 2 (Haviland's Theorem for Polynomial Matrices (Cimprič and Zalar [10, Theorem 3])). *A linear functional $\mathscr{L}$ is a tracial $\mathcal{X}$-moment functional if and only if $\mathscr{L}(F) \geq 0$ for all $F(\mathbf{x}) \in \mathcal{P}^m(\mathcal{X})$.*

Proposition 2 (Nie [42, Proposition 10.3.3]). *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 $\mathcal{P}^m(\mathcal{X}) = \mathcal{L}^m(\mathcal{X})^*$).*

2.3. 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 7 (Kimsey [27]). For a nonempty closed set $\mathcal{X} \subseteq \mathbb{R}^n$, the sequence $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n} \subseteq \mathbb{S}^m$ is called a matrix-valued $\mathcal{X}$-moment sequence if there exists a matrix-valued measure $\Phi = [\phi_{ij}] \in \mathfrak{M}^m_+(\mathcal{X})$ such that

$$\operatorname{supp}(\Phi) \subseteq \mathcal{X} \quad \text{and} \quad S_{\boldsymbol{\alpha}} = \int_{\mathcal{X}} \mathbf{x}^{\boldsymbol{\alpha}} \mathrm{d}\Phi(\mathbf{x}),\ \ \forall \boldsymbol{\alpha} \in \mathbb{N}^n. \tag{10}$$

The measure $\Phi \in \mathfrak{M}^m_+(\mathcal{X})$ satisfying (10) is called a *representing measure* of $\mathbf{S}$.

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

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

where $F_{\boldsymbol{\alpha}}$ is the coefficient matrix of $\mathbf{x}^{\boldsymbol{\alpha}}$ in $F(\mathbf{x})$ and

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

We call $\mathscr{L}_{\mathbf{S}}$ the *Riesz functional* associated with 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 8. Given a sequence $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n} \subseteq \mathbb{S}^m$, the associated *moment matrix* $M(\mathbf{S})$ is the block matrix whose block row and block column 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 with $\mathbf{S}$ and G is the block matrix whose block row and block column 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 $\boldsymbol{\alpha},\boldsymbol{\beta} \in \mathbb{N}^n$. For $d \in \mathbb{N}$, the dth-order moment matrix $M_d(\mathbf{S})$ (respectively, localizing matrix $M_d(G\mathbf{S})$) is the submatrix of $M(\mathbf{S})$ (respectively, $M(G\mathbf{S})$) 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 3. *Let $\Sigma_0(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^m$ and $\Sigma_1(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^{mq}$ be SOS matrices such that*

$$\Sigma_0(\mathbf{x}) = (u_d(\mathbf{x}) \otimes I_m)^\top Z_0 (u_d(\mathbf{x}) \otimes I_m) \text{ and } \Sigma_1(\mathbf{x}) = (u_d(\mathbf{x}) \otimes I_{mq})^\top Z_1 (u_d(\mathbf{x}) \otimes I_{mq}),$$

with $Z_0 \in \mathbb{S}_+^{m|\mathbb{N}^n_d|}$ and $Z_1 \in \mathbb{S}_+^{mq|\mathbb{N}^n_d|}$. Then, for a sequence $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n} \subseteq \mathbb{S}^m$, it holds that

$$\mathscr{L}_{\mathbf{S}}(\Sigma_0) = \mathrm{tr}(Z_0 M_d(\mathbf{S})) \text{ and } \mathscr{L}_{\mathbf{S}}((\Sigma_1, G)_m) = \mathrm{tr}(Z_1 M_d(G\mathbf{S})).$$

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

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

and let $\mathcal{M}^m(G) := \cap_{k\geq d_G} \mathcal{M}_k^m(G)$, which are all convex cones. Checking membership in $\mathcal{M}_k^m(G)$ can be accomplished with an SDP. Moreover, by Proposition 3, $\mathcal{M}_k^m(G)$ is the dual cone of $\mathcal{Q}_k^m(G)$.

For a given $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n} \subseteq \mathbb{S}^m$, the matrix-valued $\mathcal{X}$-moment problem asks when there exists a matrix-valued measure $\Phi \in \mathfrak{M}_+^m(\mathcal{X})$ satisfying (10). A necessary condition can be derived from Proposition 3.

Corollary 1. *If $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n} \subseteq \mathbb{S}^m$ has a matrix-valued representing measure $\Phi \in \mathfrak{M}_+^m(\mathcal{X})$, then $\mathbf{S} \in \mathcal{M}^m(G)$.*

Proof. For any $k \in \mathbb{N}$, PSD matrices $Z_0 \in \mathbb{S}_+^{m|\mathbb{N}^n_k|}$, and $Z_1 \in \mathbb{S}_+^{mq|\mathbb{N}^n_k|}$, let Σ_0 and Σ_1 be the SOS polynomial matrices defined in Proposition 3. Then, for any $\mathbf{x} \in \mathcal{X}$, we have $\Sigma_0(\mathbf{x}) \succeq 0$ and $(\Sigma_1(\mathbf{x}), G(\mathbf{x}))_m \succeq 0$ (see Scherer and Hol [46]). Hence,

$$\mathrm{tr}(Z_0 M_k(\mathbf{S})) = \mathscr{L}_{\mathbf{S}}(\Sigma_0) = \int_{\mathcal{X}} \mathrm{tr}(\Sigma_0(\mathbf{x}) \mathrm{d}\Phi(\mathbf{x})) \geq 0$$

and

$$\mathrm{tr}(Z_1 M_k(G\mathbf{S})) = L_{\mathbf{S}}((\Sigma_1, G)_m) = \int_{\mathcal{X}} \mathrm{tr}((\Sigma_1(\mathbf{x}), G(\mathbf{x}))_m \mathrm{d}\Phi(\mathbf{x})) \geq 0.$$

As Z_0 and Z_1 are arbitrary, we have $M_k(\mathbf{S}) \succeq 0$ and $M_k(G\mathbf{S}) \succeq 0$. □

Moreover, the matrix-valued $\mathcal{X}$-moment problem is addressed in the following theorem.

Theorem 3 (Cimprič and Zalar [10, Theorems 5 and 6]). *Let Assumption 2 hold. Given a sequence $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n} \subseteq \mathbb{S}^m$, $\mathbf{S}$ is a matrix-valued $\mathcal{X}$-moment sequence if and only if $\mathbf{S} \in \mathcal{M}^m(G)$.*

Remark 1. If $m = 1$, we use the notation $\mathcal{Q}(G)$ (respectively, $\mathcal{Q}_k(G)$, $\mathcal{M}(G)$, $\mathcal{M}_k(G)$) instead of $\mathcal{Q}^1(G)$ (respectively, $\mathcal{Q}_k^1(G)$, $\mathcal{M}^1(G)$, $\mathcal{M}_k^1(G)$) for simplicity. For a set of polynomials $H(\mathbf{x}) = \{h_1(\mathbf{x}), \ldots, h_s(\mathbf{x})\} \subseteq \mathbb{R}[\mathbf{x}]$, by slightly abusing notation, we use $\mathcal{Q}^m(H)$, $\mathcal{Q}_k^m(H)$, $\mathcal{M}^m(H)$, $\mathcal{M}_k^m(H)$ 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 1 and 3 recover Putinar's Positivstellensätz (Putinar [45]) and its dual aspect for the basic semialgebraic set $\{\mathbf{x} \in \mathbb{R}^n \,|\, h_1(\mathbf{x}) \geq 0, \ldots, h_s(\mathbf{x}) \geq 0\}$.

3. The FEC and Matrix-Valued Measure Recovery

3.1. The Truncated Matrix-Valued $\mathcal{X}$-Moment Problem

Recently, Kimsey and Trachana [28] obtained a flat extension theorem, which provides a solution to the truncated matrix-valued moment problem.

Theorem 4 (Flat Extension (Kimsey and Trachana [28, Theorem 6.2])). *For a truncated sequence $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n_{2k}} \subseteq \mathbb{S}^m$, the following statements are equivalent.*

i. *The sequence $\mathbf{S}$ 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 $\sum_{i=1}^r \mathrm{rank}(W_i) =$* $\mathrm{rank}(M_k(\mathbf{S}))$.

ii. *The matrix* $M_k(\mathbf{S}) \succeq 0$, *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 4 recovers the celebrated flat extension theorem of Curto and Fialkow [12]. There is also a version of the result by Curto and Fialkow [13] that characterizes a truncated real sequence having a representing measure supported on a prescribed semialgebraic subset of $\mathbb{R}^n$. We next extend Theorem 4 to matrix-valued measures supported on $\mathcal{X}$, which provides a solution to the truncated matrix-valued $\mathcal{X}$-moment problem.

Theorem 5. *Given a truncated sequence* $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n_{2k}} \subseteq \mathbb{S}^m$, *the following statements are equivalent.*

i. *The sequence* $\mathbf{S}$ *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. *The matrix* $M_k(\mathbf{S}) \succeq 0$, *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 Corollary 1 and Theorem 4.

(ii) ⇒ (i). By Theorem 4, $\mathbf{S}$ 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 $\sum_{i=1}^r \operatorname{rank}(W_i) = \operatorname{rank}(M_k(\mathbf{S}))$. We need to prove $\mathbf{x}^{(i)} \in \mathcal{X}$ for $i = 1, \ldots, r$. By Theorem 4, 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 $H(\mathbf{x}^{(i)})^\top W_i H(\mathbf{x}^{(i)}) = 0$ for all $i = 1, \ldots, r$. In fact, as $H(\mathbf{x}) \in \mathcal{I}_{\mathbf{S}}$, it holds that

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

As W_i's are PSD, it implies that

$$H(\mathbf{x}^{(i)})^\top W_i H(\mathbf{x}^{(i)}) = 0 \quad \text{and} \quad W_i H(\mathbf{x}^{(i)}) = 0. \tag{11}$$

Consider the quotient space $\mathbb{R}[\mathbf{x}]^m/\mathcal{I}_{\mathbf{S}} := \{H + \mathcal{I}_{\mathbf{S}} \mid H \in \mathbb{R}[\mathbf{x}]^m\}$ over $\mathbb{R}$ consisting of equivalence classes modulo $\mathcal{I}_{\mathbf{S}}$. Let $t := \operatorname{rank}(M(\mathbf{S})) = \operatorname{rank}(M_k(\mathbf{S}))$. 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

$$\{\operatorname{col}((S_{\boldsymbol{\alpha}+\boldsymbol{\beta}^{(1)}})_{\boldsymbol{\alpha}\in\mathbb{N}^n})\mathbf{e}^{(1)}, \ldots, \operatorname{col}((S_{\boldsymbol{\alpha}+\boldsymbol{\beta}^{(t)}})_{\boldsymbol{\alpha}\in\mathbb{N}^n})\mathbf{e}^{(t)}\} \tag{12}$$

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, $\operatorname{col}((S_{\boldsymbol{\alpha}+\boldsymbol{\beta}^{(i)}})_{\boldsymbol{\alpha}\in\mathbb{N}^n})$ denotes the block-column vector with infinite $m \times m$ block entries $(S_{\boldsymbol{\alpha}+\boldsymbol{\beta}^{(i)}})_{\boldsymbol{\alpha}\in\mathbb{N}^n}$. We claim that the set

$$\{\mathbf{x}^{\boldsymbol{\beta}^{(1)}}\mathbf{e}^{(1)} + \mathcal{I}_{\mathbf{S}}, \ldots, \mathbf{x}^{\boldsymbol{\beta}^{(t)}}\mathbf{e}^{(t)} + \mathcal{I}_{\mathbf{S}}\} \tag{13}$$

forms a basis of $\mathbb{R}[\mathbf{x}]^m/\mathcal{I}_{\mathbf{S}}$. To see this, first note that the elements in (13) are linearly independent as the elements in (12) 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 (13). This is indeed true because the column vector $\operatorname{col}((S_{\boldsymbol{\alpha}+\boldsymbol{\gamma}})_{\boldsymbol{\alpha}\in\mathbb{N}^n})\mathbf{e}^{(j)}$ can be written as a linear combination of elements in (12). 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 $\mathcal{I}_{\mathbf{S}}$ with respect to the basis (13) and $H^{(1)} \in \mathcal{I}_{\mathbf{S}}$. Because $\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(t)} \in \mathbb{N}^n_k$, we see that $\deg(H^{(0)}) \le k$.

Let $\{p^{(i)}(\mathbf{x})\}_{i=1}^r$ be the Lagrange interpolation polynomials at the points $\{\mathbf{x}^{(i)}\}_{i=1}^r$ such that $p^{(i)}(\mathbf{x}^{(i)}) = 1$ and $p^{(i)}(\mathbf{x}^{(j)}) = 0$ for all $j \ne i$. 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 $(\mathbf{v}^{(i)})^\top W_i \mathbf{v}^{(i)} > 0$. 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

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

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

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

where the second-to-last equality is because of the fact that $p^{(i)}(\mathbf{x}^{(i)})=1$ and $p^{(i)}(\mathbf{x}^{(j)})=0$ for all $j\neq i$ and where the last equality is because of (11) and $H_i^{(1)}\in\mathcal{I}$. As $(\mathbf{v}^{(i)})^\top W_i\mathbf{v}^{(i)}>0$, we have $G(\mathbf{x}^{(i)})\succeq 0$, which implies $\mathbf{x}^{(i)}\in\mathcal{X}$ as desired. □

Remark 2. The *block rank* $\mathrm{rank}_{bl}(\mathbf{S})$ of $M_k(\mathbf{S})$ is defined as the maximal number of linearly independent block columns (with $M_k(\mathbf{S})$ being regarded as a block matrix of entries $S_{\boldsymbol{\alpha}+\boldsymbol{\beta}}$). If the FEC $\mathrm{rank}(M_k(\mathbf{S}))=\mathrm{rank}(M_{k+d_G}(\tilde{\mathbf{S}}))$ in Theorem 5 is replaced by $\mathrm{rank}_{bl}(M_k(\mathbf{S}))=\mathrm{rank}_{bl}(M_{k+d_G}(\tilde{\mathbf{S}}))$, Nie [42, theorem 10.3.5] proved that $\mathbf{S}$ admits an atomic representing measure. We point out that the FEC on the block rank is stronger than that on the usual rank, and so, Theorem 5 is more general. In fact, it is easy to see that $\mathrm{rank}(M_k(\mathbf{S}))=\mathrm{rank}(M_{k+d_G}(\tilde{\mathbf{S}}))$ holds if $\mathrm{rank}_{bl}(M_k(\mathbf{S}))=\mathrm{rank}_{bl}(M_{k+d_G}(\tilde{\mathbf{S}}))$. However, the converse is not necessarily true. For example, letting $m=2$, $n=2$, $k=0$, and $d_G=1$, consider the sequence $\mathbf{S}=(I_2)$ and its extension $\tilde{\mathbf{S}}=(I_2,J,J,I_2,I_2,I_2)$, where

$$I_2=\begin{bmatrix}1&0\\0&1\end{bmatrix}\quad\text{and}\quad J=\begin{bmatrix}0&1\\1&0\end{bmatrix}.$$

We have $\mathrm{rank}(M_0(\mathbf{S}))=\mathrm{rank}(M_1(\tilde{\mathbf{S}}))=2$, but $1=\mathrm{rank}_{bl}(M_0(\mathbf{S}))\neq\mathrm{rank}_{bl}(M_1(\tilde{\mathbf{S}}))=2$.

3.2. Matrix-Valued Measure Recovery

For a truncated sequence $\mathbf{S}=(S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n_{2k}}\subseteq\mathbb{S}^m$ 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 5, $\mathbf{S}$ 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 Kimsey and Trachana [28] 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. In this subsection, inspired by Henrion and Lasserre [20], we provide a linear algebra procedure for extracting $\mathbf{x}^{(i)}\in\mathcal{X}$ and $W_i\in\mathbb{S}^m$, which is much cheaper to implement.

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

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

where $u_k(\mathbf{x}^{(i)})$ is defined in (6). Letting $m_i=\mathrm{rank}(W_i)$, as W_i is PSD, we have the decomposition $W_i=\sum_{j=1}^{m_i}\mathbf{w}^{(i,j)}$

$(\mathbf{w}^{(i,j)})^\top$ for some $\mathbf{w}^{(i,j)} \in \mathbb{R}^m$. Then, we can write $M_k(\mathbf{S}) = VV^\top$ with

$$V = [(u_k(\mathbf{x}^{(1)}) \otimes I_m)[\mathbf{w}^{(1,1)}, \ldots, \mathbf{w}^{(1,m_1)}], \ldots, (u_k(\mathbf{x}^{(r)}) \otimes I_m)[\mathbf{w}^{(r,1)}, \ldots, \mathbf{w}^{(r,m_r)}]].$$

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

Note that 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 $\tilde{V}$. 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

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

Reduce the matrix $\tilde{V}$ 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

$$v_k(\mathbf{x}, \mathbf{w}) = Ub_k(\mathbf{x}, \mathbf{w}) \tag{14}$$

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

Proposition 4. *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 the points $\mathbf{x}^{(i)}$'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 $i \neq j$. Now, we fix an i' with $1 \le i' \le r$, and we consider the column vector of polynomials $\mathbf{w}p^{(i')}(\mathbf{x}) \in \mathbb{R}[\mathbf{x}, \mathbf{w}]^m$. As $\deg(p^{(i)}) = r - 1 \le k$, because of (14), 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 $(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})$, $j = 1, \ldots, m_i$, $i = 1, \ldots, r$. 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 because i' can be arbitrarily chosen.

Case 2. $r - 1 > k$. According to Theorem 4, $\mathbf{S}$ admits a flat extension $\tilde{\mathbf{S}} = (\tilde{S}_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}_{2r-2}^n}$ 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 can 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 $(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})$. Because $M_{r-1}(\tilde{\mathbf{S}})$ 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 to $b_k(\mathbf{x}, \mathbf{w})$. 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 $l = 1, \ldots, n$, we can extract from U the $t \times t$ multiplication matrix N_l such that $N_l b_k(\mathbf{x}, \mathbf{w}) = x_l b_k(\mathbf{x}, \mathbf{w})$ holds at each pair $(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)})$, $j = 1, \ldots, m_i$, $i = 1, \ldots, r$.

Following Corless et al. [11], we build a random combination of multiplication matrices $N = \sum_{l=1}^n c_l N_l$, where $c_l > 0$ and $\sum_{l=1}^n c_l = 1$. Let $N = ATA^\top$ be the ordered Schur decomposition of N, where $A = [a_1, \ldots, a_t]$ is an orthogonal matrix with $A^\top A = I_t$ and T is upper triangular with eigenvalues of N being sorted increasingly along the diagonal.

Proposition 5. *Suppose that the constants c_l's are chosen such that $h(\mathbf{x}) = \sum_{l=1}^n c_l x_l$ takes distinct values on $\mathbf{x}^{(i)}$, $i = 1, \ldots, r$. Then, the set of points*

$$\{(a_1^\top N_1 a_1, \ldots, a_1^\top N_n a_1), \ \ldots, \ (a_t^\top N_1 a_t, \ldots, a_t^\top N_n a_t)\} \tag{15}$$

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

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

$$h(\mathbf{x}^{(i)}) b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)}) = \sum_{l=1}^n c_l x_l^{(i)} b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)}) = N b_k(\mathbf{x}^{(i)}, \mathbf{w}^{(i,j)}),$$

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 4, 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 $l = 1, \ldots, n$, it holds

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

Hence, $(a^\top N_1 a, \ldots, a^\top N_n a) = \mathbf{x}^{(i)}$. The conclusion then follows. □

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

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

and we have

$$M_k(\mathbf{S}) = \Lambda \operatorname{diag}(W_1, \ldots, W_r) \Lambda^\top. \tag{16}$$

Notice that the first m columns of $\operatorname{diag}(W_1, \ldots, W_r)\Lambda^\top$ are exactly $[W_1, \ldots, W_r]^\top$. By comparing the first m columns of both sides of (16), we get

$$\operatorname{col}(\{S_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathbb{N}_k^n}) = \Lambda [W_1, \ldots, W_r]^\top. \tag{17}$$

Assume that Λ has mr independent rows (see Remark 3), and let $\mathcal{R}$ be the index set of these rows. Denote by $\Lambda_{\mathcal{R}}$ (respectively, $M_{\mathcal{R}}(\mathbf{S})$) the $mr \times mr$ (respectively, $mr \times m$) submatrix of Λ (respectively, $\operatorname{col}(\{S_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathbb{N}_k^n})$) whose rows are indexed by $\mathcal{R}$. Then, by extracting the rows indexed $\mathcal{R}$ from both sides of (17), we have $M_{\mathcal{R}}(\mathbf{S}) = \Lambda_{\mathcal{R}} [W_1, \ldots, W_r]^\top$. Hence, the matrices W_i's can be retrieved by $[W_1, \ldots, W_r]^\top = \Lambda_{\mathcal{R}}^{-1} M_{\mathcal{R}}(\mathbf{S})$. We provide an example illustrating the above procedure in Appendix B.

Remark 3. It is clear that if

$$\operatorname{rank}([u_k(\mathbf{x}^{(1)}), \ldots, u_k(\mathbf{x}^{(r)})]) = r, \tag{18}$$

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 (18) always holds if $k \ge r - 1$. If $k < r - 1$, then it is possible that (18) fails, and we may need to consider flat extensions of $M_k(\mathbf{S})$ to recover the weights W_i's. Inspired by Nie [41], we propose the following heuristic method for finding such a flat extension. Consider the linear matrix-valued moment optimization problem:

$$\min_{\tilde{\mathbf{S}}} \sum_{\boldsymbol{\alpha}\in\mathbb{N}^n_{2d}} \operatorname{tr}(R_{\boldsymbol{\alpha}}\tilde{S}_{\boldsymbol{\alpha}}) \,\text{s.t.}\, \tilde{\mathbf{S}} = (\tilde{S}_\alpha)_{\alpha\in\mathbb{N}^n_{2d}} \in \mathcal{M}^m_d(G),\ \tilde{S}_{\boldsymbol{\alpha}} = S_{\boldsymbol{\alpha}},\ \ \forall \boldsymbol{\alpha}\in\mathbb{N}^n_{2k},$$

for some $d>k$ and $(R_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n_{2d}} \subseteq \mathbb{S}^m$. By randomly choosing the matrices $R_{\boldsymbol{\alpha}}$, one may expect that the optimum of the above problem is achieved at an extreme point of the cone $\mathcal{M}^m_d(G)$, which might admit a finitely atomic matrix-valued representing measure and hence, provide a flat extension of $M_k(\mathbf{S})$. We leave a detailed study on this issue in the future.

To conclude this section, we would like to remark that the results on the truncated matrix-valued $\mathcal{X}$-moment problem and the procedure for matrix-valued measure recovery are not direct generalizations of the scalar case. First, the matrix-valued moment problem cannot be reduced to the scalar case via certain scalarizing procedures as far as we know. For instance, it is not valid to solve the matrix-valued moment problem by separately considering the sequence of the entries at the same position in the matrices $\{S_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha}\in\mathbb{N}^n_{2k}}$. This is because the off-diagonal entries of a PSD matrix-valued measure are not necessarily positive scalar measures, making the theory of the scalar moment problem inapplicable here. One may also consider the scalar sequence $\{\operatorname{tr}(S_{\boldsymbol{\alpha}})\}_{\boldsymbol{\alpha}\in\mathbb{N}^n_{2k}}$ that, however, contains only partial information on the matrix-valued moments. It is thus impossible to recover the matrix-valued measure in this way. Second, extending the theory of the scalar moment problem to the matrix-valued case requires carefully addressing additional complexities and intrinsic challenges present in the matrix-valued setting. In particular, the proof of Theorem 5 is conducted in the quotient space $\mathbb{R}[\mathbf{x}]^m/\mathcal{I}_{\mathbf{S}}$ because of the matrix-valued nature, which involves a subtle construction of the polynomial vectors H_i. In addition, the linear algebra procedure presented in Section 3.2 is the first efficient approach for extracting matrix-valued measures from the matrix-valued moment matrix satisfying FEC. In comparison with the scalar counterpart in Henrion and Lasserre [20], this procedure demands several delicate techniques tailored to the matrix-valued setting to handle the variables introduced by both the points $\mathbf{x}^{(i)}$ and matrices W_i, making the justification of its correctness much more involved.

4. A Moment-SOS Hierarchy for (1) with SOS Convexity

In this section, we first reformulate (1) as a conic optimization problem, based on which we can then derive a moment-SOS hierarchy whose optima monotonically converge to the optimum of (1). 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 $P_{\boldsymbol{\alpha}}(\mathbf{y}) \in \mathbb{S}[\mathbf{y}]^m$ (respectively, $P_{\boldsymbol{\beta}}(\mathbf{x}) \in \mathbb{S}[\mathbf{x}]^m$) is the coefficient matrix of $\mathbf{x}^{\boldsymbol{\alpha}}$ (respectively, $\mathbf{y}^{\boldsymbol{\beta}}$) with $P(\mathbf{x},\mathbf{y})$ being regarded as a polynomial matrix in $\mathbb{S}[\mathbf{x}]^m$ (respectively, $\mathbb{S}[\mathbf{y}]^m$). For a linear functional $\mathscr{L}: \mathbb{S}[\mathbf{x}]^m \to \mathbb{R}$, we let

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

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

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

To obtain a conic reformulation of (1), we need to assume the Slater condition to hold.

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

Proposition 6. *Under Assumptions* 1 *and* 3*, there exists a finitely atomic matrix-valued measure* $\Phi^\star \in \mathfrak{M}^m_+(\mathcal{X})$ *such that*

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

If $\mathbf{y}^\star$ *is an optimal solution to* (1)*, then* $\mathscr{L}_{\Phi^\star}(P(\mathbf{y}^\star,\mathbf{x})) = 0$.

Proof. For any $\mathbf{v} \in \mathbf{V} := \{\mathbf{v} \in \mathbb{R}^m \mid \sum_{i=1}^m v_i^2 = 1\}$ and $\mathbf{x} \in \mathcal{X}$, by Assumption 1(ii), we have that the function $-\mathbf{v}^\top P(\cdot, \mathbf{x})\mathbf{v}$ is convex in $\mathbf{y}$. Then, (1) can be equivalently reformulated as the following convex semi-infinite program under Assumption 1:

$$f^\star = \inf_{\mathbf{y}\in\mathcal{Y}} f(\mathbf{y}) \quad \text{s.t. } \mathbf{v}^\top P(\mathbf{y}, \mathbf{x})\mathbf{v} \geq 0, \;\; \forall (\mathbf{x}, \mathbf{v}) \in \mathcal{X} \times \mathbf{V}. \tag{19}$$

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 $(\mathbf{v}^{(i)})^\top P(\bar{\mathbf{y}}, \mathbf{x}^{(i)})\mathbf{v}^{(i)} > 0$ for all $i = 0, 1, \ldots, \ell$. As $\mathcal{X} \times \mathbf{V}$ is compact in $\mathbb{R}^n \times \mathbb{R}^m$ (Assumption 1(iii)), recall by Borwein [7, theorem 4.1] that if for any arbitrary $\ell+1$ points in $\mathcal{X} \times \mathbf{V}$, the corresponding constraints in (19) are strictly satisfied by some point from $\mathcal{Y}$, then (19) is reducible. In other words, there exist ℓ points $(\tilde{\mathbf{x}}^{(1)}, \tilde{\mathbf{v}}^{(1)}), \ldots, (\tilde{\mathbf{x}}^{(\ell)}, \tilde{\mathbf{v}}^{(\ell)}) \in \mathcal{X} \times \mathbf{V}$ such that

$$f^\star = \inf_{\mathbf{y}\in\mathcal{Y}} f(\mathbf{y}) \quad \text{s.t. } (\tilde{\mathbf{v}}^{(i)})^\top P(\mathbf{y}, \tilde{\mathbf{x}}^{(i)})\tilde{\mathbf{v}}^{(i)} \geq 0, \; i = 1, \ldots, \ell.$$

Then, by the Lagrange multiplier theorem, we can find $\lambda_1, \ldots, \lambda_\ell > 0$ such that

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

Define a finitely atomic matrix-valued measure

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

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 (1), then $P(\mathbf{y}^\star, \mathbf{x}) \succeq 0$ on $\mathcal{X}$, and hence,

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

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

Remark 4. From the above proof, one can see that Proposition 6 remains true if the Slater condition is weakened. 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 $\mathscr{H}_\mu : \mathbb{R}[\mathbf{y}] \to \mathbb{R}$ 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:

$$\begin{cases} \tilde{f} := \sup\limits_{\rho, \Phi} & \rho \\ \qquad\quad \text{s.t.} & f(\mathbf{y}) - \rho - \mathscr{L}_\Phi(P(\mathbf{y}, \mathbf{x})) \in \mathcal{P}(\mathcal{Y}), \\ & \rho \in \mathbb{R}, \, \Phi \in \mathfrak{M}_+^m(\mathcal{X}), \end{cases} \tag{20}$$

whose dual reads as

$$\begin{cases} \hat{f} := \inf\limits_{\mu} & \mathscr{H}_\mu(f) \\ \qquad\quad \text{s.t.} & \mu \in \mathfrak{m}_+(\mathcal{Y}), \, \mathscr{H}_\mu(1) = 1, \\ & \mathscr{H}_\mu(P(\mathbf{y}, \mathbf{x})) \in \mathcal{P}^m(\mathcal{X}), \end{cases} \tag{21}$$

with $\mathcal{P}^m(\mathcal{X})$ being defined in (9).

Theorem 6. *Under Assumptions 1 and 3, it holds that* $\tilde{f} = \hat{f} = f^\star$.

Proof. Let $\Phi^\star \in \mathfrak{M}_+^m(\mathcal{X})$ be the finitely atomic matrix-valued measure given in Proposition 6. Then, $(f^\star, \Phi^\star)$ is feasible to (20) because of Proposition 6. Thus, $\tilde{f} \geq f^\star$. By the weak duality, we have $\hat{f} \geq \tilde{f}$. It remains to show $\hat{f} \leq f^\star$.

Let $(\mathbf{y}^{(k)})_{k\in\mathbb{N}}$ be a minimizing sequence of (1). Then, for any $\varepsilon > 0$, there exists $k_\varepsilon \in \mathbb{N}$ such that $\mathbf{y}^{(k_\varepsilon)}$ is feasible to (1) and $f(\mathbf{y}^{(k_\varepsilon)}) \leq f^\star + \varepsilon$. The Dirac measure $\delta_{\mathbf{y}^{(k_\varepsilon)}}$ centered at $\mathbf{y}^{(k_\varepsilon)}$ is feasible to (21). Therefore, $\hat{f} \leq \mathscr{H}_{\delta_{\mathbf{y}^{(k_\varepsilon)}}}(f) = f(\mathbf{y}^{(k_\varepsilon)}) \leq f^\star + \varepsilon$. As $\varepsilon > 0$ is arbitrary, we have $\hat{f} \leq f^\star$ as desired. □

Remark 5. It is easy to verify that the conic reformulation (20) and (21) for (1) and the results in Proposition 6 and Theorem 6 remain true if "SOS convex" in Assumption 1(i) is weakened to "convex" and "PSD-SOS convex" in Assumption 1(ii) is weakened to "PSD convex" (i.e., if Assumption 1 is replaced by Assumption 5 in Section 5).

Remark 6. Before we proceed, let us highlight the significance of the conic reformulation (20) and (21) and explain how it could provide insights in solving (1). Essentially, through the reformulation (20) and (21), the complexity of (1) is transferred to the conic constraints in (20) and (21), which involve the cone $\mathcal{P}(\mathcal{Y})$ of nonnegative polynomials on the set $\mathcal{Y}$, the cone $\mathfrak{M}_+^m(\mathcal{X})$ of PSD matrix-valued measures supported on $\mathcal{X}$, and their dual cones. Although the exact representations of these cones are unavailable in general case, there are various (matrix-valued) Positivstellensätze and the dual-moment (matrix-valued) theories, which can provide tractable approximations for these cones. Hence, by replacing these cones with suitable approximations, the reformulation (20) and (21) offers us a unified framework to derive tractable relaxations for (1).

4.2. A Moment-SOS Hierarchy

Let $\Theta := \{\theta_1, \ldots, \theta_s\} \subseteq \mathbb{R}[\mathbf{y}]$ collect the description polynomials of the semialgebraic set $\mathcal{Y}$. Moreover, let

$$k_{\mathbf{y}} := \max\{\deg(f), \deg(\theta_1), \ldots, \deg(\theta_s), \deg_{\mathbf{y}}(P_{ij}), i,j = 1,\ldots,m\},$$

$$k_{\mathbf{x}} := \max\{\deg_{\mathbf{x}}(P_{ij}), i,j = 1,\ldots,m, \deg(G)\}.$$

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 (20) 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 relaxation for (1):

$$\begin{cases} f_k^{\text{primal}} := \sup\limits_{\rho, \mathbf{S}} & \rho \\ \qquad\qquad \text{s.t.} & f(\mathbf{y}) - \rho - \mathscr{L}_{\mathbf{S}}(P(\mathbf{y},\mathbf{x})) \in \mathcal{Q}_{\lceil k_{\mathbf{y}}/2\rceil}(\Theta), \\ & \rho \in \mathbb{R},\ \mathbf{S} \in \mathcal{M}_k^m(G). \end{cases} \tag{22}$$

The dual of (22) reads as

$$\begin{cases} f_k^{\text{dual}} := \inf\limits_{\mathbf{s}} & \mathscr{H}_{\mathbf{s}}(f) \\ \qquad\qquad \text{s.t.} & \mathbf{s} \in \mathcal{M}_{\lceil k_{\mathbf{y}}/2\rceil}(\Theta),\ \mathscr{H}_{\mathbf{s}}(1) = 1, \\ & \mathscr{H}_{\mathbf{s}}(P(\mathbf{y},\mathbf{x})) \in \mathcal{Q}_k^m(G), \end{cases} \tag{23}$$

where the linear functional $\mathscr{H}_{\mathbf{s}} : \mathbb{R}[\mathbf{y}]_{2\lceil k_{\mathbf{y}}/2\rceil} \to \mathbb{R}$ is defined by $\mathscr{H}_{\mathbf{s}}(h) = \sum_{\boldsymbol{\alpha}\in \text{supp}(h)} h_{\boldsymbol{\alpha}} s_{\boldsymbol{\alpha}}$ for $h \in \mathbb{R}[\mathbf{y}]_{2\lceil k_{\mathbf{y}}/2\rceil}$. We call (22) and (23) the *moment-SOS hierarchy* for (1) with SOS convexity and call k the *relaxation order*.

4.2.1. Convergence Analysis.

Next, we present the convergence analysis of the moment-SOS hierarchy (22) and (23).

Proposition 7. *Let Assumptions 1 and 3 hold. Then, the 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 nonincreasing and provide upper bounds on* $f^\star$.

Proof. By the weak duality $f_k^{\text{primal}} \leq f_k^{\text{dual}}$, we only need to prove that there exists a finitely atomic matrix-valued measure $\Phi^\star \in \mathfrak{M}_+^m(\mathcal{X})$ such that

$$f(\mathbf{y}) - f^\star - \mathscr{L}_{\Phi^\star}(P(\mathbf{y},\mathbf{x})) \in \mathcal{Q}_{\lceil k_{\mathbf{y}}/2\rceil}(\Theta). \tag{24}$$

As Assumptions 1 and 3 hold, by Proposition 6, there exists a finitely atomic matrix-valued measure $\Phi^\star \in \mathfrak{M}_+^m(\mathcal{X})$ such that

$$f^\star = \inf_{\mathbf{y}\in\mathbb{R}^\ell} f(\mathbf{y}) - \mathscr{L}_{\Phi^\star}(P(\mathbf{y},\mathbf{x})) \quad \text{s.t. } \theta_1(\mathbf{y}) \geq 0, \ldots, \theta_s(\mathbf{y}) \geq 0. \tag{25}$$

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 $W_i = \sum_{k=1}^{m_i} \mathbf{v}^{(i,k)}(\mathbf{v}^{(i,k)})^\top$ for some $\mathbf{v}^{(i,k)} \in \mathbb{R}^m$. Then,

$$f(\mathbf{y}) - \mathscr{L}_{\Phi^\star}(P(\mathbf{y},\mathbf{x})) = f(\mathbf{y}) - \sum_{i=1}^r \text{tr}(P(\mathbf{y},\mathbf{x}^{(i)})W_i) = f(\mathbf{y}) - \sum_{i=1}^r \sum_{k=1}^{m_i} (\mathbf{v}^{(i,k)})^\top P(\mathbf{y},\mathbf{x}^{(i)})\mathbf{v}^{(i,k)}.$$

By Assumption 1 and the definition of PSD-SOS convexity, $f(\mathbf{y}) - \mathscr{L}_{\Phi^\star}(P(\mathbf{y},\mathbf{x}))$ is SOS convex. By Proposition 6, each optimal solution to (1) is also optimal to (25). Because the functions $f(\mathbf{y}) - \mathscr{L}_{\Phi^\star}(P(\mathbf{y},\mathbf{x}))$, $-\theta_i(\mathbf{y})$ in (25) are all

SOS convex and the set of optimal solutions of (25) is nonempty, Lasserre [33, theorem 3.3] states that there exists a convex SOS polynomial $\sigma(\mathbf{y}) \in \mathbb{R}[\mathbf{y}]$ and scalars $\lambda_1 \geq 0, \ldots, \lambda_s \geq 0$ such that

$$f(\mathbf{y}) - f^\star - \mathscr{L}_{\Phi^\star}(P(\mathbf{y},\mathbf{x})) = \sigma(\mathbf{y}) + \sum_{j=1}^{s} \lambda_j \theta_j(\mathbf{y}).$$

Clearly, we have $\deg(\sigma) \leq 2\lceil k_{\mathbf{y}}/2 \rceil$, which implies that (24) holds true, and the conclusion follows. □

We next prove the asymptotic convergence of this moment-SOS hierarchy. 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 $\mathbf{s} = (s_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^\ell_{2\lceil k_{\mathbf{y}}/2 \rceil}}$ of (23).

Proposition 8. *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 $\mathbf{v}^\top Q(\mathbf{y})\mathbf{v}$ is SOS convex in $\mathbf{y}$ for all $\mathbf{v} \in \mathbb{R}^m$, by Lasserre [33, theorem 2.6], it holds that

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

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

Corollary 2. *Suppose that $-\theta_1(\mathbf{y}), \ldots, -\theta_s(\mathbf{y})$ are SOS convex and that $-P(\mathbf{y},\mathbf{x})$ is PSD-SOS convex in $\mathbf{y}$ for all $\mathbf{x} \in \mathcal{X}$. If $\mathbf{s}$ is feasible to (23), then $\mathbf{s}_\mathbf{e} \in \mathcal{Y}$, and it is feasible to (1).*

Proof. By the extended Jensen's inequality for SOS-convex polynomials (Lasserre [33, theorem 2.6]), it holds that $\theta_i(\mathbf{s}_\mathbf{e}) \geq \mathscr{H}_\mathbf{s}(\theta_i) \geq 0$ for $i = 1, \ldots, s$, which implies $\mathbf{s}_\mathbf{e} \in \mathcal{Y}$. By Proposition 8, for every $\mathbf{x} \in \mathcal{X}$, we have $P(\mathbf{s}_\mathbf{e}, \mathbf{x}) \succeq \mathscr{H}_\mathbf{s}(P(\mathbf{y},\mathbf{x})) \succeq 0$. So, $\mathbf{s}_\mathbf{e}$ is feasible to (1). □

Theorem 7. *Under Assumptions 1–3, the following are true.*

i. *We have $f_k^{\text{primal}} \searrow f^\star$ and $f_k^{\text{dual}} \searrow f^\star$ as $k \to \infty$.*

ii. *For any convergent subsequence $(\mathbf{s}^{(k_i,\star)})_i$ of $(\mathbf{s}^{(k,\star)})_k$, where each $\mathbf{s}^{(k,\star)}$ is a minimizer of (23), $\lim_{i\to\infty} \mathbf{s}_\mathbf{e}^{(k_i,\star)}$ is a global minimizer of (1). Consequently, if the set of optimal solutions of (1) is a singleton, then $\lim_{k\to\infty} \mathbf{s}_\mathbf{e}^{(k,\star)}$ is the unique global minimizer.*

Proof. (i) Let $\mathbf{y}^\star$ be a minimizer of (1) and $\bar{\mathbf{y}}$ be the Slater point given in Assumption 3. Because $f, \mathcal{Y}$ are convex by Assumption 1(i), 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 Assumption 2 and Theorem 1, there exists $k^{(\varepsilon)} \in \mathbb{N}$ such that $\mathbf{s}'$ is feasible to (23) 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, by Proposition 7 and weak duality, we have $f^\star \leq f_k^{\text{primal}} \leq f_k^{\text{dual}}$, and so, $f_k^{\text{dual}} \searrow f^\star$ and $f_k^{\text{primal}} \searrow f^\star$ as $k \to \infty$.

(ii) Let $\mathbf{s}^\star = (s^\star_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^\ell_{2\lceil k_{\mathbf{y}}/2 \rceil}}$ be 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 (1) is closed, by Corollary 2, $\mathbf{s}^\star_\mathbf{e}$ is feasible to (1). Moreover, as $f(\mathbf{y})$ is SOS convex by Assumption 1(i), by (i) and Proposition 8, it holds that $f^\star = \mathscr{H}_{\mathbf{s}^\star}(f) \geq f(\mathbf{s}^\star_\mathbf{e})$, which indicates that $\mathbf{s}^\star_\mathbf{e}$ is a global minimizer of (1). □

Remark 7. Suppose that there exists a ball constraint $\sum_{i=1}^{\ell} y_i^2 \leq b^2$ for some $b \neq 0$ in the description of $\mathcal{Y}$. Let $t \in \mathbb{N}$ satisfy $t \geq \max\{\lceil \deg(\theta_j)/2 \rceil, j = 1, \ldots, s\}$. Let $\mathbf{s} = (s_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^\ell_{2t}}$ satisfy $\mathbf{s} \in \mathcal{M}_t(\Theta)$ and $\mathscr{H}_\mathbf{s}(1) = 1$. Then, by Josz and Henrion [26, lemma 3], it holds that

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

In this case, the convergent subsequence $(\mathbf{s}^{(k_i,\star)})_i$ of $(\mathbf{s}^{(k,\star)})_k$ in Theorem 7(ii) always exists.

The next theorem allows us to detect finite convergence of the moment-SOS hierarchy (22) and (23) and to extract an optimal solution whenever the FEC is satisfied.

Theorem 8. *Let $(f_k^{\text{primal}}, \mathbf{S}^{(k,\star)})$ and $\mathbf{s}^{(k,\star)}$ be optimal solutions to (22) and (23), respectively, for some $k \geq \lceil k_\mathbf{x}/2 \rceil$. If the following FEC*

$$\exists \lceil k_\mathbf{x}/2 \rceil \leq t \leq k \text{ s.t. } \operatorname{rank}(M_t(\mathbf{S}^{(k,\star)})) = \operatorname{rank}(M_{t-d_G}(\mathbf{S}^{(k,\star)})) \tag{26}$$

holds, then

i. *the sequence* $\mathbf{S}^{(k,\star)}$ *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. *we have* $f_k^{\text{primal}} \le f^\star$ *and the equality holds if Assumptions 1 and 3 are satisfied.*

Consequently, under Condition (26) and Assumptions 1 and 3, if $f_k^{\text{primal}} = f_k^{\text{dual}}$, *then* $f_k^{\text{primal}} = f_k^{\text{dual}} = f^\star$ *and*

iii. *the sequence* $\mathbf{s}_{\mathbf{e}}^{(k,\star)}$ *is an optimal solution to (1).*

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

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

Proof.

i. It follows from Theorem 5.

ii. As $(f_k^{\text{primal}}, \mathbf{S}^{(k,\star)})$ is feasible to (22), thanks to (i), it holds that

$$f(\mathbf{y}) - f_k^{\text{primal}} - \mathscr{L}_{\mathbf{S}^{(k,\star)}}(P(\mathbf{y},\mathbf{x})) = f(\mathbf{y}) - f_k^{\text{primal}} - \sum_{i=1}^r \operatorname{tr}(W_i P(\mathbf{y},\mathbf{x}^{(i)})) \ge 0, \tag{27}$$

for all $\mathbf{y} \in \mathcal{Y}$. Let $\mathbf{y}^\star$ be a global minimizer of (1). Noting $\sum_{i=1}^r \operatorname{tr}(W_i P(\mathbf{y}^\star,\mathbf{x}^{(i)})) \ge 0$, then by (27), we have

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

If Assumptions 1 and 3 hold, then $f_k^{\text{primal}} \ge f^\star$ by Proposition 7. Hence, $f_k^{\text{primal}} = f^\star$.

iii. By Corollary 2, $\mathbf{s}_{\mathbf{e}}^{(k,\star)}$ is feasible to (1). Because $f(\mathbf{y})$ is SOS convex, by (ii) and Proposition 8, it holds that

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

which implies that $\mathbf{s}_{\mathbf{e}}^{(k,\star)}$ is a global minimizer of (1).

iv. By (ii), (iii), and (27), we deduce that $-\sum_{i=1}^r \operatorname{tr}(W_i P(\mathbf{s}_{\mathbf{e}}^{(k,\star)},\mathbf{x}^{(i)})) \ge 0$. As $P(\mathbf{s}_{\mathbf{e}}^{(k,\star)},\mathbf{x}^{(i)})$ and W_i are both PSD, it follows

$$0 = \sum_{i=1}^r \operatorname{tr}(W_i P(\mathbf{s}_{\mathbf{e}}^{(k,\star)},\mathbf{x}^{(i)})) = \sum_{i=1}^r \sum_{l=1}^{m_i} (\mathbf{v}^{(i,l)})^\top P(\mathbf{s}_{\mathbf{e}}^{(k,\star)},\mathbf{x}^{(i)})\mathbf{v}^{(i,l)}.$$

The PSD-ness of $P(\mathbf{s}_{\mathbf{e}}^{(k,\star)},\mathbf{x}^{(i)})$ implies that $P(\mathbf{s}_{\mathbf{e}}^{(k,\star)},\mathbf{x}^{(i)})\mathbf{v}^{(i,l)} = 0$ for all $l = 1,\ldots,m_i$, $i = 1,\ldots,r$. □

4.2.2. Conditions for Strong Duality. Now, we give conditions under which there is no duality gap between (22) and (23).

Assumption 4. (i) *We have* $\theta_1(\mathbf{y}) = b^2 - \sum_{i=1}^\ell y_i^2$ *for some* $b \ne 0$, *and* (ii) *the set* $\mathcal{X}$ *has nonempty interior.*

As a consequence of Mai et al. [38, corollary 3.6], we have the following result.

Lemma 1. *Suppose that Assumption 4*(i) *holds. Then, for any* $k \ge \max\{\lceil k_{\mathbf{x}}/2\rceil, \lceil k_{\mathbf{y}}/2\rceil\}$ *and* $\mathbf{S} \in \mathcal{M}_k^m(G)$, *there exists* $\lambda \in \mathbb{R}$ *such that*

$$f(\mathbf{y}) - \lambda - \mathscr{L}_{\mathbf{S}}(P(\mathbf{y},\mathbf{x})) \in \mathcal{Q}_k^\circ(\Theta),$$

where $\mathcal{Q}_k^\circ(\Theta)$ *denotes the interior of* $\mathcal{Q}_k(\Theta)$.

Lemma 2. *Suppose that Assumption 4*(ii) *holds. Then, for each* $k \ge d_G$, *there exists* $\mathbf{S}^\circ \in \mathcal{M}_k^m(G)$ *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 $\mathbf{S}^\circ \in \mathcal{M}_k^m(G)$ such that $M_{k-d_G}(G\mathbf{S}^\circ) \succ 0$. Similar arguments also apply 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 $\Phi = \operatorname{diag}(\phi,\ldots,\phi)$, 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 $\mathbf{v}^\top M_{k-d_G}(G\mathbf{S}^\circ)\mathbf{v} = 0$. Let

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

Then, by Proposition 3, it holds that

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

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

$$i,\, mq + i,\, 2mq + i,\, \ldots, (|\mathbb{N}^n_{k-d_G}| - 1)mq + i,$$

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

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

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

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

Then,

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

and

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

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

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

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 $j = 1, \ldots, m$,

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

Because $\mathcal{O}$ 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. □

Theorem 9. *Under Assumption 4, we have* $f_k^{\text{primal}} = f_k^{\text{dual}}$ *for each* $k \geq \lceil k_{\mathbf{x}}/2 \rceil$.

Proof. By Lemmas 1 and 2, (22) is strictly feasible, and the conclusion follows. □

4.2.3. An Example. For all numerical examples in the sequel, we use Yalmip (Löfberg [36]) to model SDPs, and then, we rely on Mosek (Andersen and Andersen [3]) to solve them.[1]

Example 1. Consider the following instance of (1):

$$f^\star := \inf_{\mathbf{y} \in \mathbb{R}^2} f(\mathbf{y}) \quad \text{s.t.}\ P(\mathbf{y}, \mathbf{x}) \succeq 0,\ \forall \mathbf{x} \in \mathcal{X} := \{\mathbf{x} \in \mathbb{R}^2 \,|\, G(\mathbf{x}) \succeq 0\}, \tag{28}$$

where

$$P(\mathbf{y}, \mathbf{x}) = \begin{bmatrix} 1 - (x_1 y_1 - x_2 y_2)^2 & 2(x_2 y_1 + x_1 y_2) \\ 2(x_2 y_1 + x_1 y_2) & 1 \end{bmatrix}$$

and

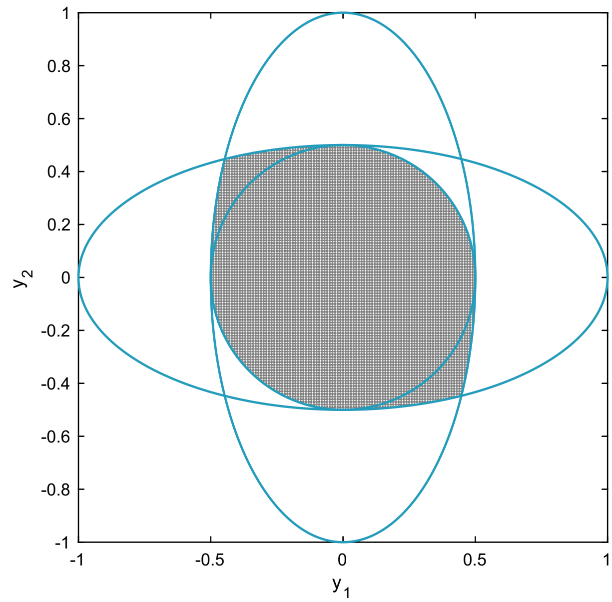
$$G(\mathbf{x}) = \begin{bmatrix} 1 - x_1 & x_2 & 0 & 0 \\ x_2 & 1 + x_1 & 0 & 0 \\ 0 & 0 & x_1^2 + x_2^2 - 1 & 0 \\ 0 & 0 & 0 & x_1 x_2 \end{bmatrix}.$$

It is clear that

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

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. The feasible set (the gray area) is displayed in Figure 1.

Figure 1. (Color online) The feasible set (the gray area) of Problem (28) in Example 1.



By Corollary A.1, 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. The computational results are shown in Table 1.

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 (22) and (23) with $k=1$, 2. For $k=2$, the rank Condition (26) is satisfied with $t=1$ and $f_2^{\text{primal}}=f_2^{\text{dual}}=0.8358$. Then, by Theorem 8, global optimality is certified, and a minimizer $\mathbf{s}_{\mathbf{e}}^{(2,\star)}=(0.3536,0.3536)$ 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 $2(\sqrt{5}/5-1)^2\approx 0.6111$. We solve the SDP relaxations (22) and (23) with $k=1$, 2, 3. As we can see, the global optimality is certified by the rank Condition (26) at $k=3$ with $t=2$, even though it is already achieved at $k=2$. 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 (1) becomes the robust polynomial semidefinite program (29), which was extensively studied in Scherer and Hol [46]. In this case, we will see that the dual Problem (23) recovers the matrix SOS relaxation for (29) proposed in Scherer and Hol [46]. On the other hand, as a complement, the primal Problem (22) allows us to detect finite convergence and extract optimal solutions.

4.3.1. Robust Polynomial Semidefinite Programming. Consider the robust polynomial semidefinite programming problem, which is a special case of (1):

$$\tau^\star := \inf_{\mathbf{y}\in\mathbb{R}^\ell} \mathbf{c}^\top\mathbf{y} \quad \text{s.t.} \quad \sum_{i=1}^{\ell} P_i(\mathbf{x})y_i - P_0(\mathbf{x})\succeq 0,\ \forall \mathbf{x}\in\mathcal{X}\subseteq\mathbb{R}^n, \tag{29}$$

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

Table 1. Computational results for Example 1.

	Results for f_1		Results for f_2		
	$k=1$	$k=2$	$k=1$	$k=2$	$k=3$
f_k^{primal}/time	1.2497/0.61 s	0.8358/0.68 s	1.2499/0.62 s	0.6111/0.67 s	0.6111/1.00 s
f_k^{dual}/time	1.2496/0.52 s	0.8358/0.77 s	1.2497/0.44 s	0.6111/0.77 s	0.6111/2.08 s
FEC (26)	False	True	False	False	True

Applying the conic reformulation (20)–(29) with $\mathcal{Y} = \mathbb{R}^\ell$, we obtain

$$\begin{cases} \sup\limits_{\rho,\Phi} & \rho \\ \text{s.t.} & \mathbf{c}^\top \mathbf{y} - \rho - \sum\limits_{i=1}^{\ell} \mathscr{L}_\Phi(P_i(\mathbf{x}))y_i + \mathscr{L}_\Phi(P_0(\mathbf{x})) \in \mathcal{P}(\mathbb{R}^\ell), \\ & \rho \in \mathbb{R}, \Phi \in \mathfrak{M}_+^m(\mathcal{X}). \end{cases} \tag{30}$$

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

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

Thus, if (ρ,Φ) is feasible to (30), then we necessarily have

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

and (30) can be rewritten as

$$\sup_{\Phi \in \mathfrak{M}_+^m(\mathcal{X})} \int_{\mathcal{X}} P_0(\mathbf{x})\mathrm{d}\Phi(\mathbf{x}) \quad \text{s.t.} \quad \int_{\mathcal{X}} P_i(\mathbf{x})\mathrm{d}\Phi(\mathbf{x}) = c_i,\ i = 1,\dots,\ell. \tag{31}$$

Remark 8. Note that the generalized moment problem extensively studied by Lasserre [32] is a special case of (31) with $m=1$ and $G(\mathbf{x})$ being a diagonal matrix.

The dual of (30) reads as

$$\begin{cases} \inf\limits_{\mu} & \mathbf{c}^\top \mathscr{H}_\mu(\mathbf{y}) \\ \text{s.t.} & \mu \in \mathfrak{m}_+(\mathbb{R}^\ell),\ \mathscr{H}_\mu(1) = 1, \\ & \sum\limits_{i=1}^{\ell} P_i(\mathbf{x})\mathscr{H}_\mu(y_i) - P_0(\mathbf{x}) \in \mathcal{P}^m(\mathcal{X}). \end{cases} \tag{32}$$

By identifying $\mathscr{H}_\mu(\mathbf{y})$ with $\mathbf{y}$, we see that (32) is actually equivalent to (29).

For each $k \ge \lceil k_{\mathbf{x}}/2 \rceil$, replacing the cone $\mathfrak{M}_+^m(\mathcal{X})$ in (31) by $\mathcal{M}_k^m(G)$, we obtain the corresponding moment relaxation:

$$\tau_k^{\text{primal}} := \sup_{\mathbf{S} \in \mathcal{M}_k^m(G)} \mathscr{L}_{\mathbf{S}}(P_0(\mathbf{x})) \quad \text{s.t.} \quad \mathscr{L}_{\mathbf{S}}(P_i(\mathbf{x})) = c_i,\ i = 1,\dots,\ell, \tag{33}$$

whose dual is

$$\tau_k^{\text{dual}} := \inf_{\mathbf{y} \in \mathbb{R}^\ell} \mathbf{c}^\top \mathbf{y} \quad \text{s.t.} \quad \sum_{i=1}^{\ell} P_i(\mathbf{x})y_i - P_0(\mathbf{x}) \in \mathcal{Q}_k^m(G). \tag{34}$$

Note that when $\mathcal{X}$ is compact, Assumption 1 is satisfied for (29). Then, as direct consequences of Theorems 7 and 8, we get the following theorems.

Theorem 10. *Let Assumptions 2 and 3 hold. Then,*

i. *We have $\tau_k^{\text{primal}} \searrow \tau^\star$ and $\tau_k^{\text{dual}} \searrow \tau^\star$ as $k \to \infty$.*

ii. *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 (34), $\lim_{i\to\infty}\mathbf{y}^{(k_i,\star)}$ is an optimal solution to (29).*

Theorem 11. *Let $\mathbf{S}^{(k,\star)}$ and $\mathbf{y}^{(k,\star)}$ be optimal solutions to (33) and (34), respectively, for some $k \ge \lceil k_{\mathbf{x}}/2 \rceil$. If the FEC (26) holds, then*

i. *the sequence $\mathbf{S}^{(k,\star)}$ 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 $W_1,\dots,W_r \in \mathbb{S}_+^m$.*

ii. *we have $\tau_k^{\text{primal}} \le \tau^\star$ and the equality holds if $\mathcal{X}$ is compact and Assumption 3 is satisfied.*

Consequently, under Condition (26), the compactness of $\mathcal{X}$, and Assumption 3, if $\tau_k^{\text{primal}} = \tau_k^{\text{dual}}$, then $\tau_k^{\text{primal}} = \tau_k^{\text{dual}} = \tau^\star$ and

iii. *the point $\mathbf{y}^{(k,\star)}$ is an optimal solution to (29).*

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

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

Remark 9. Note that unlike (22), there is no quadratic module constraint in (33). So, if $\mathcal{X}$ has nonempty interior, then (33) is strictly feasible by Lemma 2, and hence, there is no duality gap between (33) and (34). Besides, we can recover the moment relaxations for deterministic PMI optimization problems by Henrion and Lasserre [21] from (33).

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:

$$\lambda^\star := \inf_{\mathbf{x}\in\mathbb{R}^n} \lambda_{\min}(F(\mathbf{x})) \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{X} := \{\mathbf{x} \in \mathbb{R}^n \,|\, G(\mathbf{x}) \succeq 0\}, \tag{35}$$

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 $\boldsymbol{\alpha}$, 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 (35) (Androulakis et al. [4], Maranas and Floudas [39]). For another example from optimal control, the stability 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 unstabilizable. 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 Problem (35) (Dumitrescu et al. [16]).

Clearly, Problem (35) is equivalent to

$$\sup_{\lambda\in\mathbb{R}} \lambda \quad \text{s.t.} \quad F(\mathbf{x}) - \lambda I_m \succeq 0, \quad \forall \mathbf{x} \in \mathcal{X}, \tag{36}$$

which is a special case of (29).

For each $k \geq \max\{\lceil \deg(F)/2 \rceil, \lceil \deg(G)/2 \rceil\}$, the kth-order primal moment relaxation of (35) reads as

$$\lambda_k^{\text{primal}} := \inf_{\mathbf{S}\in\mathcal{M}_k^m(G)} \mathscr{L}_{\mathbf{S}}(F(\mathbf{x})) \quad \text{s.t.} \quad \mathscr{L}_{\mathbf{S}}(I_m) = 1, \tag{37}$$

with dual

$$\lambda_k^{\text{dual}} := \sup_{\lambda\in\mathbb{R}} \lambda \quad \text{s.t.} \quad F(\mathbf{x}) - \lambda I_m \in \mathcal{Q}_k^m(G). \tag{38}$$

The dual Problem (38) was studied in Slot and Laurent [48], where $\mathcal{X}$ is set to the n-dimensional Boolean hypercube $\{0,1\}^n$.

Clearly, the Slater condition holds for (36) whenever $\mathcal{X}$ is compact. Then, from Theorems 10 and 11, we deduce the following theorems.

Theorem 12. *Let Assumption 2 hold. Then,* $\lambda_k^{\text{primal}} \searrow \lambda^\star$ *and* $\lambda_k^{\text{dual}} \searrow \lambda^\star$ *as* $k \to \infty$.

Theorem 13. *Let* $k \geq \max\{\lceil \deg(F)/2 \rceil, \lceil \deg(G)/2 \rceil\}$ *and* $\mathbf{S}_k^\star$ *be an optimal solution to* (37). *If the FEC* (26) *holds, then*

i. *the sequence* $\mathbf{S}_k^\star$ *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* $W_1, \ldots, W_r \in \mathbb{S}_+^m$ *with* $\sum_{i=1}^r \operatorname{tr}(W_i) = 1$ *(because* $\mathscr{L}_{\mathbf{S}}(I_m) = 1$*)*;

ii. *we have* $\lambda_k^{\text{primal}} \leq \lambda^\star$, *and the equality holds if* $\mathcal{X}$ *is compact.*

Consequently, under Condition (26) *and the compactness of* $\mathcal{X}$, *if* $\lambda_k^{\text{dual}} = \lambda_k^{\text{primal}}$, *then* $\lambda_k^{\text{dual}} = \lambda_k^{\text{primal}} = \lambda^\star$ *and*

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

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

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 2. Consider the following instance of (35), where $F(\mathbf{x}) = Q\,\mathrm{diag}(f_1, f_2, f_3)Q^\top$ for some $f_1, f_2, f_3 \in \mathbb{R}[\mathbf{x}]$ and $Q \in \mathbb{R}^{3\times 3}$ with $Q^\top Q = I_3$. Specifically, we let Q and $G(\mathbf{x})$ be

$$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \end{bmatrix}, \quad G(\mathbf{x}) = \begin{bmatrix} 1 - 4x_1x_2 & x_1 \\ x_1 & 4 - x_1^2 - x_2^2 \end{bmatrix},$$

and

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

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 $[c, 0, c]^\top$, $c \in \mathbb{R}$.

Solving the moment relaxation (37) with $k=2$, we get $\lambda_2^{\mathrm{primal}} = -4.0000$, and the rank Condition (26) is satisfied with $\mathrm{rank}(M_2(\mathbf{S}_2^\star)) = \mathrm{rank}(M_1(\mathbf{S}_2^\star)) = 2$. By Theorem 13, global optimality is reached. Using 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

$$W_1 = \begin{bmatrix} 0.2505 & -0.0000 & 0.2505 \\ -0.0000 & -0.0000 & 0.0000 \\ 0.2505 & 0.0000 & 0.2505 \end{bmatrix}, W_2 = \begin{bmatrix} 0.2495 & 0.0000 & 0.2495 \\ 0.0000 & 0.0000 & -0.0000 \\ 0.2495 & -0.0000 & 0.2495 \end{bmatrix}.$$

Both W_1 and W_2 have the decomposition: $\gamma_i^2[1, 0, 1]^\top[1, 0, 1]$ for some constants $\gamma_1, \gamma_2 \in \mathbb{R}$. Then, by Theorem 13, $[1, 0, 1]^\top$ is an eigenvector of $F(\mathbf{x}^{(i)})$, $i=1, 2$, associated with the eigenvalue -4.0000.

5. Extensions to the General Convex and Nonconvex Settings

In this section, we extend the results of Section 4 to the general convexity and nonconvexity settings. The following assumptions are obtained from Assumption 1 by replacing "SOS convex" with "convex."

Assumption 5. (i) *The polynomials $f(\mathbf{y}), -\theta_1(\mathbf{y}), \ldots, -\theta_s(\mathbf{y})$ are convex;* (ii) *the polynomial matrix $-P(\mathbf{y}, \mathbf{x})$ is PSD convex in $\mathbf{y}$ for all $\mathbf{x} \in \mathcal{X}$; and* (iii) *the set $\mathcal{X}$ 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 (20) with $\mathcal{Q}_k(\Theta)$ and $\mathcal{M}_k^m(G)$, respectively, we obtain the following SDP:

$$\begin{cases} f_k^{\mathrm{primal}} := \sup\limits_{\rho, \mathbf{S}} & \rho \\ \qquad\qquad \text{s.t.} & f(\mathbf{y}) - \rho - \mathscr{L}_{\mathbf{S}}(P(\mathbf{y}, \mathbf{x})) \in \mathcal{Q}_k(\Theta), \\ & \rho \in \mathbb{R}, \mathbf{S} \in \mathcal{M}_k^m(G), \end{cases} \tag{39}$$

whose dual reads as

$$\begin{cases} f_k^{\mathrm{dual}} := \inf\limits_{\mathbf{s}} & \mathscr{H}_{\mathbf{s}}(f) \\ \qquad\qquad \text{s.t.} & \mathbf{s} \in \mathcal{M}_k(\Theta), \mathscr{H}_{\mathbf{s}}(1) = 1, \\ & \mathscr{H}_{\mathbf{s}}(P(\mathbf{y}, \mathbf{x})) \in \mathcal{Q}_k^m(G). \end{cases} \tag{40}$$

Lemma 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 $\mathbf{s} = (s_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^\ell} \subseteq \mathbb{R}$ admits a representing probability measure, then $\mathscr{H}_{\mathbf{s}}(Q(\mathbf{y})) \succeq Q(\mathbf{s}_{\mathbf{e}})$.*

Proof. By the PSD convexity of $Q(\mathbf{y})$, for any $\mathbf{v} \in \mathbb{R}^m$, $\mathbf{v}^\top Q(\mathbf{y})\mathbf{v}$ is convex in $\mathbf{y}$. Because $\mathbf{s}$ admits a representing probability measure, by Jensen's inequality for convex functions, it holds $\mathbf{v}^\top Q(\mathbf{s}_{\mathbf{e}})\mathbf{v} \leq \mathscr{H}_{\mathbf{s}}(\mathbf{v}^\top Q(\mathbf{y})\mathbf{v}) = \mathbf{v}^\top \mathscr{H}_{\mathbf{s}}(Q(\mathbf{y}))\mathbf{v}$ for any $\mathbf{v} \in \mathbb{R}^m$. Hence, $\mathscr{H}_{\mathbf{s}}(Q(\mathbf{y})) \succeq Q(\mathbf{s}_{\mathbf{e}})$. □

We have the following theorem whose proof can be found in Appendix C.

Theorem 14. *Under Assumptions 2, 3, 4(i), and 5, the following are true.*

i. *We have* $\lim_{k\to\infty} f_k^{\text{primal}} = \lim_{k\to\infty} f_k^{\text{dual}} = f^\star$.

ii. *For any convergent subsequence* $(\mathbf{s}_{\mathbf{e}}^{(k_i,\star)})_i$ *(always exists) of* $(\mathbf{s}_{\mathbf{e}}^{(k,\star)})_k$, *where each* $\mathbf{s}^{(k,\star)}$ *is a minimizer of* (40), $\lim_{i\to\infty}\mathbf{s}_{\mathbf{e}}^{(k_i,\star)}$ *is a global minimizer of* (1). *Consequently, if the optimal solution set of* (1) *is a singleton, then* $\lim_{k\to\infty}\mathbf{s}_{\mathbf{e}}^{(k,\star)}$ *is the unique global minimizer.*

Remark 10. Because $\mathcal{Q}_k(\Theta)$ is an inner approximation for $\mathcal{P}(\mathcal{Y})$, whereas $\mathcal{M}_k^m(G)$ is an outer approximation for $\mathfrak{M}_+^m(\mathcal{X})$, the convergence of $(f_k^{\text{primal}})_k$ and $(f_k^{\text{dual}})_k$ to $f^\star$ might not be monotonical (see Proposition 7).

As in the SOS-convex setting, finite convergence of (39) and (40) can be detected via certain FECs. Let $d_\Theta := \max\{\lceil \deg(\theta_i)/2\rceil, i = 1,\ldots,s\}$.

Theorem 15. *Let* $(\rho_k^\star, \mathbf{S}^{(k,\star)})$ *and* $\mathbf{s}^{(k,\star)}$ *be optimal solutions to* (39) *and* (40), *respectively, for some* $k \ge \max\{\lceil k_{\mathbf{x}}/2\rceil, \lceil k_{\mathbf{y}}/2\rceil\}$. *Consider the following FECs:*

$$\exists \lceil k_{\mathbf{x}}/2\rceil \le t_1 \le k\, s.t.\, \operatorname{rank}(M_{t_1}(\mathbf{S}^{(k,\star)})) = \operatorname{rank}(M_{t_1-d_G}(\mathbf{S}^{(k,\star)})), \tag{41}$$

$$\exists \lceil k_{\mathbf{y}}/2\rceil \le t_2 \le k\; s.t.\; \operatorname{rank}(M_{t_2}(\mathbf{s}^{(k,\star)})) = \operatorname{rank}(M_{t_2-d_\Theta}(\mathbf{s}^{(k,\star)})). \tag{42}$$

If Condition (41) *holds, then*

i. *the sequence* $\mathbf{S}^{(k,\star)}$ *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* $\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(r_1)} \in \mathcal{X}$, $W_1,\ldots,W_{r_1} \in \mathbb{S}_+^m$, *and* $f_k^{\text{primal}} \le f^\star$.

Under Assumption 5, if Condition (42) *holds, then*

ii. *the sequence* $\mathbf{s}^{(k,\star)}$ *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* $\lambda_1,\ldots,\lambda_{r_2}$, *and* $f_k^{\text{dual}} \ge f^\star$.

Consequently, under Conditions (41) *and* (42) *and Assumption 5, if* $f_k^{\text{primal}} = f_k^{\text{dual}}$, *then*

iii. *we have* $f_k^{\text{primal}} = f_k^{\text{dual}} = f^\star$;

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

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

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

Proof. (i) See the proof of Theorem 8(i).

(ii) The first part follows from Theorem 5. For the second part, suppose that Assumption 5 holds. As $\mathbf{s}^{(k,\star)}$ is feasible to (40), by Lemma 3,

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

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

$$f^\star \le f\left(\sum_{j=1}^{r_2}\lambda_j\mathbf{y}^{(j)}\right) \le \sum_{j=1}^{r_2}\lambda_j f(\mathbf{y}^{(j)}) = \mathscr{H}_{\mathbf{s}^{(k,\star)}}(f) = f_k^{\text{dual}}. \tag{43}$$

Hence, it holds that $f_k^{\text{dual}} \ge f^\star$.

(iii) It follows from (i) and (ii).

(iv) It follows from (iii) and (43).

(v) This is similar to the proof of Theorem 8(iv). □

Remark 11. Under Assumption 4, by Lemmas 1 and 2, we can deduce that (39) is strictly feasible, and hence, $f_k^{\text{primal}} = f_k^{\text{dual}}$.

Now, we would like to remark on how and where the stronger SOS-convexity conditions (Assumption 1, (i) and (ii)) contribute to achieving better results when relaxing the conic reformulations (20) and (21) compared with those achieved by using the general convexity conditions (Assumption 5, (i) and (ii)). Under Assumptions 1 and 3, Condition (24) holds true as proven in Proposition 7. Consequently, replacing $\mathcal{P}(\mathcal{Y})$ by the quadratic module $\mathcal{Q}_{\lceil k_{\mathbf{y}}/2\rceil}(\Theta)$ of fixed order in (20) does not change its optimal value. Hence, substituting the outer approximations $\mathcal{M}_k^m(G)$, $k\in\mathbb{N}$, for $\mathfrak{M}_+^m(\mathcal{X})$ in (22) provides monotonically nonincreasing upper bounds on $f^\star$. Moreover, the SOS convexity validates an extension of Jensen's inequality (Proposition 8) (Lasserre [33, theorem 2.6]), which can be used to prove that a feasible point of (23) is also feasible to (1) (Corollary 2), and an optimal solution to (1) can be retrieved from (23) provided that the FEC (26) holds (Theorem 8(iii)). In contrast, under the general convexity conditions, (24) does not necessarily hold, and we have to replace $\mathcal{P}(\mathcal{Y})$ by $\mathcal{Q}_k(\Theta)$ with $k\ge\lceil k_{\mathbf{y}}/2\rceil$ in (39). Then, the convergence of $(f_k^{\text{primal}})_k$ and $(f_k^{\text{dual}})_k$ to $f^\star$ might not be monotonical as explained in Remark 10. Additionally, to extract optimal solutions to (1) from (40) under general convexity conditions, we require the FEC (42), besides the FEC (41), to ensure that Theorem 15(ii) holds.

Example 3. Consider the following instance of (1):

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

where $P(\mathbf{y},\mathbf{x})$ and $G(\mathbf{x})$ are defined as in Example 1. The following polynomial (Ahmadi and Parrilo [1, (5.2)]) is convex but not SOS convex:

$$\begin{aligned} h(\mathbf{y}) = {} & 89-363y_1^4y_2+\frac{51{,}531}{64}y_2^6-\frac{9{,}005}{4}y_2^5+\frac{49{,}171}{16}y_2^4+721y_1^2-2{,}060y_2^3-14y_1^3 \\ & +\frac{3{,}817}{4}y_2^2+363y_1^4-9y_1^5+77y_1^6+316y_1y_2+49y_1y_2^3-2{,}550y_1^2y_2-968y_1y_2^2 \\ & +1{,}710y_1y_2^4+794y_1^3y_2+\frac{7{,}269}{2}y_1^2y_2^2-\frac{301}{2}y_1^5y_2+\frac{2{,}143}{4}y_1^4y_2^2+\frac{1{,}671}{2}y_1^3y_2^3 \\ & +\frac{14{,}901}{16}y_1^2y_2^4-\frac{1{,}399}{2}y_1y_2^5-\frac{3{,}825}{2}y_1^3y_2^2-\frac{4{,}041}{2}y_1^2y_2^3-364y_2+48y_1. \end{aligned} \tag{44}$$

Let $f(\mathbf{y})=h(y_1-1,y_2-1)/10{,}000$, which is again convex but not SOS convex. We have $k_{\mathbf{y}}=6$ and $k_{\mathbf{x}}=2$.

Solving the SDPs (39) and (40) with $k=3$, we get $f_3^{\text{primal}}=f_3^{\text{dual}}=0.4504$, and the rank Conditions (41) and (42) are satisfied with $t_1=1$ and $t_2=3$. By Theorem 15, global optimality is reached, and so, $f^\star\approx 0.4504$. The extracted minimizer is $\mathbf{y}^\star=(0.2711,0.4201)$.

Before ending this section, let us consider the most general case—Problem (1) without any convexity assumption. By Theorem 15(i), if the FEC (41) holds, then we obtain a lower bound $f_k^{\text{primal}}\le f^\star$. Moreover, we have the following theorem.

Theorem 16. *Let $\mathbf{s}^{(k,\star)}$ be a minimizer of (40) for some $k\ge\max\{\lceil k_{\mathbf{x}}/2\rceil,\lceil k_{\mathbf{y}}/2\rceil\}$.*

i. *If $\operatorname{rank}(M_{\lceil k_{\mathbf{y}}/2\rceil}(\mathbf{s}^{(k,\star)}))=1$, then $f_k^{\text{dual}}\ge f^\star$, and the truncation sequence $\hat{\mathbf{s}}^{(k,\star)} := (\mathbf{s}_{\boldsymbol{\alpha}}^{(k,\star)})_{|\boldsymbol{\alpha}|\le 2\lceil k_{\mathbf{y}}/2\rceil}$ of $\mathbf{s}^{(k,\star)}$ admits a Dirac representing measure $\delta_{\mathbf{y}^\star}$ for some $\mathbf{y}^\star\in\mathcal{Y}$.*

ii. *If Condition (41) holds and $\operatorname{rank}(M_{\lceil k_{\mathbf{y}}/2\rceil}(\mathbf{s}^{(k,\star)}))=1$ and moreover, if $f_k^{\text{primal}}=f_k^{\text{dual}}$, then $f_k^{\text{primal}}=f_k^{\text{dual}}=f^\star$, and $\mathbf{y}^\star$ is a minimizer of (1).*

Proof. (i) As $\operatorname{rank}(M_{\lceil k_{\mathbf{y}}/2\rceil}(\mathbf{s}^{(k,\star)}))=1$, $\hat{\mathbf{s}}^{(k,\star)}$ admits a Dirac representing measure $\delta_{\mathbf{y}^{(\star)}}$ for some $\mathbf{y}^\star\in\mathcal{Y}$. Then, as $\mathbf{s}^{(k,\star)}$ is feasible to (40), it holds that

$$P(\mathbf{y}^\star,\mathbf{x})=\mathscr{H}_{\hat{\mathbf{s}}^{(k,\star)}}(P(\mathbf{y},\mathbf{x}))=\mathscr{H}_{\mathbf{s}^{(k,\star)}}(P(\mathbf{y},\mathbf{x}))\succeq 0. \tag{45}$$

So, $\mathbf{y}^\star$ is feasible to (1), and hence, $f_k^{\text{dual}}=\mathscr{H}_{\mathbf{s}^{(k,\star)}}(f)=\mathscr{H}_{\hat{\mathbf{s}}^{(k,\star)}}(f)=f(\mathbf{y}^\star)\ge f^\star$.

ii. It follows from Theorems 15(i) and 16(i). □

Example 4. Consider Example 1, where we let $f(\mathbf{y})=(y_1+1)^2+(y_2-1)^2$ and modify the matrix $P(\mathbf{y},\mathbf{x})$ to

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

Geometrically, the feasible region is constructed by rotating the shape in the $\mathbf{y}$ plane defined by $-y_1^2+4y_2^2\le 1$ continuously around the origin by 90° clockwise and then taking the common area of these shapes in this process. Hence, $-P(\mathbf{y},\mathbf{x})$ is *not* PSD convex in $\mathbf{y}$ for every $\mathbf{x}\in\mathcal{X}$. It is easy to check that $f^\star=2(1-1/\sqrt{3})^2\approx 0.3573$ with a unique minimizer $(-1/\sqrt{3},1/\sqrt{3})\approx(-0.5774,0.5774)$.

Solving the SDPs (39) and (40) with $k=3$, we get $f_3^{\text{primal}} = f_3^{\text{dual}} = 0.3573$. As the rank Condition (41) is satisfied with $t=2$ and $\text{rank}(M_{\lceil k_{\mathbf{y}}/2\rceil}(\mathbf{s}^{(3,\star)})) = 1$, global optimality is reached by Theorem 16(ii), and moreover, a minimizer $\mathbf{y}^\star = (-0.5773, 0.5774)$ is extracted.

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 FEC is used to detect global optimality. Extensions to the general convexity and nonconvexity settings are also provided. As this work generalizes most of the nice features of the moment-SOS hierarchy from the scalar polynomial optimization to the robust PMI optimization, 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 (Waki et al. [50], Wang et al. [51], Wang et al. [52])) can be exploited to derive a structured moment-SOS hierarchy with lower computational complexity. Recent work on exploiting sparsity for PMIs can be found in Zheng and Fantuzzi [53]. As a line of further research, we intend to extend such techniques to the robust PMI optimization in future work.

Acknowledgments

The authors are grateful for the comments and suggestions of the anonymous referees that helped to improve the presentation.

Appendix A. Uniform PSD-SOS Convexity

Definition A.1 (Nie [40]). 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

$$\nabla_{\mathbf{yy}}(\mathbf{v}^\top Q(\mathbf{y})\mathbf{v}) = F(\mathbf{v}, \mathbf{y})^\top F(\mathbf{v}, \mathbf{y}). \tag{A.1}$$

Clearly, if $Q(\mathbf{y}) \in \mathbb{S}[\mathbf{y}]^m$ is uniformly PSD-SOS convex, then it is PSD-SOS convex. Moreover, checking the existence of $F(\mathbf{v}, \mathbf{y})$ in (A.1) can be converted into an SDP feasibility problem. For $Q(\mathbf{y}) \in \mathbb{S}[\mathbf{y}]^m$, we write $Q(\mathbf{y}) = \sum_{\boldsymbol{\alpha}\in\text{supp}(Q)} Q_{\boldsymbol{\alpha}}\mathbf{y}^{\boldsymbol{\alpha}}$, where $Q_{\boldsymbol{\alpha}}$ is the coefficient matrix of $\mathbf{y}^{\boldsymbol{\alpha}}$ in $Q(\mathbf{y})$ and

$$\text{supp}(Q) := \{\boldsymbol{\alpha} \in \mathbb{N}^\ell \mid \mathbf{y}^{\boldsymbol{\alpha}} \text{ appears in some } Q_{ij}(\mathbf{y})\}.$$

Then, we have the following proposition.

Proposition A.1. *A polynomial matrix $Q(\mathbf{y}) = \sum_{\boldsymbol{\alpha}\in\text{supp}(Q)} Q_{\boldsymbol{\alpha}}\mathbf{y}^{\boldsymbol{\alpha}} \in \mathbb{S}[\mathbf{y}]^m$ is uniformly PSD-SOS convex if $\sum_{\boldsymbol{\alpha}\in\text{supp}(Q)} Q_{\boldsymbol{\alpha}} \otimes \nabla_{\mathbf{yy}}\mathbf{y}^{\boldsymbol{\alpha}}$ is an SOS matrix.*

Proof. Observe that

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

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

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

we then obtain (A.1) by letting $F(\mathbf{v}, \mathbf{y}) = T(\mathbf{y})(\mathbf{v} \otimes I_\ell)$. □

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

Appendix B. 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 $\operatorname{rank}(W_1) = \operatorname{rank}(W_3) = 1$ and $\operatorname{rank}(W_3) = 2$. We let $\Phi = \sum_{i=1}^3 W_i \delta_{\mathbf{x}^{(i)}}$. Denote by $\mathbf{S} = (S_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}_4^2}$ 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 $t = \operatorname{rank}(M_2(\mathbf{S})) = \operatorname{rank}(M_1(\mathbf{S})) = 4$. Compute the Cholesky decomposition $M_2(\mathbf{S}) = \tilde{V}\tilde{V}^\top$ with $\tilde{V} \in \mathbb{R}^{12\times 4}$, and reduce matrix $\tilde{V}$ 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

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

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 $\mathbf{x}^{(i)}$ and $\mathbf{w}^{(i,j)}$. Moreover, we can get from U the multiplication matrices 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 $N = ATA^\top$ 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 four points in (15):

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

As we can see, all points $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, and $\mathbf{x}^{(3)}$ have been recovered. Among the above four points, $\mathbf{x}^{(1)}$, $\mathbf{x}^{(3)}$ appear one time, and $\mathbf{x}^{(3)}$ appears two times; they correspond 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 $[W_1, W_2, W_3]^\top$.

Appendix C. Proof of Theorem 14

Proof. (i) Fix an arbitrary $\varepsilon > 0$. Under Assumptions 2, 3, and 5, the proof of Theorem 7(i) implies that there exists $k_1^{(\varepsilon)} \in \mathbb{N}$ such that $f_k^{\text{dual}} \le f^\star + \varepsilon$ for all $k \ge k_1^{(\varepsilon)}$.

Let $\Phi^\star \in \mathfrak{M}_+^m(\mathcal{X})$ be the finitely atomic matrix-valued measure in Proposition 6, and let $\mathbf{S}^\star = (S_{\boldsymbol{\alpha}}^\star)_{\boldsymbol{\alpha} \in \mathbb{N}_{2k}^n}$ with each $S_{\boldsymbol{\alpha}}^\star = \int_{\mathcal{X}} \mathbf{x}^{\boldsymbol{\alpha}} \mathrm{d}\Phi^\star(\mathbf{x})$. Then, $\mathbf{S}^\star \in \mathcal{M}_k^m(G)$. Because of Proposition 6 and Remark 5, it holds that

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

for all $\mathbf{y} \in \mathcal{Y}$. As Assumption 4(i) holds, by the Positivstellensätz of Putinar [45] (see also Theorem 1), there exists $k_2^{(\varepsilon)} \in \mathbb{N}$ such that for all $k \ge k_2^{(\varepsilon)}$, $(f^\star - \varepsilon, \mathbf{S}^\star)$ is feasible to (39), and hence, $f_k^{\text{primal}} \ge f^\star - \varepsilon$.

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

(ii) For each $\boldsymbol{\alpha} \in \mathbb{N}^\ell$, define

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

By Assumption 4(i) and Remark 7, $|s_{\boldsymbol{\alpha}}^{(k,\star)}| \le N(\boldsymbol{\alpha})$ for all k and $\boldsymbol{\alpha} \in \mathbb{N}_{2k}^\ell$. Hence, there always exists a convergent subsequence of $(\mathbf{s}_{\mathbf{e}}^{(k,\star)})_k$. Without loss of generality, we assume that the whole sequence $(\mathbf{s}_{\mathbf{e}}^{(k,\star)})_k$ converges. Complete each $\mathbf{s}^{(k,\star)}$ with zeros to make it an infinite vector. Then,

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

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 $\mathbf{s}^\star = (s^\star_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^\ell}$ such that $\lim_{i\to\infty} s_{\boldsymbol{\alpha}}^{(k_i,\star)} = s^\star_{\boldsymbol{\alpha}}$ 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) $\mathscr{H}_{\mathbf{s}^\star}(1) = 1$, and (c) $\mathscr{H}_{\mathbf{s}^\star}(P(\mathbf{y},\mathbf{x})) \succeq 0$ for all $x \in \mathcal{X}$. By (a), (b), Putinar's Positivstellensätz, and Haviland's theorem, $\mathbf{s}^\star$ has a representing probability measure μ supported on $\mathcal{Y}$ (i.e., $s^\star_{\boldsymbol{\alpha}} = \int_{\mathcal{Y}} \mathbf{y}^{\boldsymbol{\alpha}} d\mu(\mathbf{y})$ for all $\boldsymbol{\alpha}\in\mathbb{N}^\ell$). By Lemma 3, $P(\mathbf{s}^\star_{\mathbf{e}}, \mathbf{x}) \succeq \mathscr{H}_{\mathbf{s}^\star}(P(\mathbf{y},\mathbf{x})) \succeq 0$ and $\theta_i(\mathbf{s}^\star_{\mathbf{e}}) \geq \mathscr{H}_{\mathbf{s}^\star}(\theta_i) \geq 0$, $i = 1, \ldots, s$. Hence, $\mathbf{s}^\star_{\mathbf{e}}$ is feasible to (1). Moreover, as $f(\mathbf{y})$ is convex in $\mathbf{y}$, by (i) and the pointwise convergence,

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

which indicates that $\mathbf{s}^\star_{\mathbf{e}}$ is a minimizer of (1). □

Endnote

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

References

[1] Ahmadi AA, Parrilo PA (2013) A complete characterization of the gap between convexity and SOS-convexity. *SIAM J. Optim.* 23(2):811–833.
[2] Ahmadi AA, Olshevsky A, Parrilo PA, Tsitsiklis JN (2013) NP-hardness of deciding convexity of quartic polynomials and related problems. *Math. Programming* 137(1):453–476.
[3] Andersen ED, Andersen KD (2000) The Mosek interior point optimizer for linear programming: An implementation of the homogeneous algorithm. Frenk H, Roos K, Terlaky T, Zhang S, eds. *High Performance Optimization*, Applied Optimization, vol. 33 (Springer, New York), 197–232.
[4] Androulakis IP, Maranas CD, Floudas CA (1995) αBB: A global optimization method for general constrained nonconvex problems. *J. Global Optim.* 7(4):337–363.
[5] Aravanis C, Aspman J, Korpas G, Marecek J (2022) Polynomial matrix inequalities within tame geometry. Preprint, submitted June 8, https://arxiv.org/abs/2206.03941.
[6] Bertsimas D, Brown DB, Caramanis C (2011) Theory and applications of robust optimization. *SIAM Rev.* 53(3):464–501.
[7] Borwein JM (1981) Direct theorems in semi-infinite convex programming. *Math. Programming* 21(1):301–318.
[8] Chesi G (2010) LMI techniques for optimization over polynomials in control: A survey. *IEEE Trans. Automatic Control* 55(11):2500–2510.
[9] Chuong TD, Jeyakumar V, Li G, Woolnough D (2022) Exact dual semi-definite programs for affinely adjustable robust SOS-convex polynomial optimization problems. *Optimization* 71(12):3539–3569.
[10] Cimprič J, Zalar A (2013) Moment problems for operator polynomials. *J. Math. Anal. Appl.* 401(1):307–316.
[11] Corless RM, Gianni PM, Trager BM (1997) A reordered Schur factorization method for zero-dimensional polynomial systems with multiple roots. Küchlin WW, ed. *Proc. 1997 Internat. Sympos. Symbolic Algebraic Comput. ISSAC '97* (Association for Computing Machinery, New York), 133–140.
[12] Curto RE, Fialkow LA (1996) Solution of the truncated complex moment problem for flat data. *Memoirs Amer. Math. Soc.* 119(568).
[13] Curto RE, Fialkow LA (2000) The truncated complex *K*-moment problem. *Trans. Amer. Math. Soc.* 352(6):2825–2855.
[14] Damanik D, Pushnitski A, Simon B (2008) The analytic theory of matrix orthogonal polynomials. *Surveys Approximation Theory* 4:1–85.
[15] Dette H, Studden WJ (2003) Quadrature formulas for matrix measures—A geometric approach. *Linear Algebra Appl.* 364:33–64.
[16] Dumitrescu B, Şicleru BC, Ştefan R (2009) Computing the controllability radius: A semi-definite programming approach. *IET Control Theory Appl.* 3(6):654–660.
[17] Duran AJ, Lopez-Rodriguez P (1997) Density questions for the truncated matrix moment problem. *Canadian J. Math.* 49(4):708–721.
[18] Gamboa F, Nagel J, Rouault A (2019) Sum rules and large deviations for spectral matrix measures. *Bernoulli* 25(1):712–741.
[19] Helton J, Nie J (2010) Semidefinite representation of convex sets. *Math. Programming* 122(1):21–64.
[20] Henrion D, Lasserre JB (2005) Detecting global optimality and extracting solutions in gloptipoly. Henrion D, Garulli A, eds. *Positive Polynomials in Control* (Springer, Berlin), 293–310.
[21] Henrion D, Lasserre JB (2006) Convergent relaxations of polynomial matrix inequalities and static output feedback. *IEEE Trans. Automatic Control* 51(2):192–202.
[22] Henrion D, Lasserre JB (2012) Inner approximations for polynomial matrix inequalities and robust stability regions. *IEEE Trans. Automatic Control* 57(6):1456–1467.
[23] Huang L, Liu D, Fang Y (2023) Convergence of an SDP hierarchy and optimality of robust convex polynomial optimization problems. *Ann. Oper. Res.* 320(1):33–59.
[24] Ichihara H (2009) Optimal control for polynomial systems using matrix sum of squares relaxations. *IEEE Trans. Automatic Control* 54(5): 1048–1053.
[25] Jeyakumar V, Li G, Vicente-Pérez J (2015) Robust SOS-convex polynomial optimization problems: Exact SDP relaxations. *Optim. Lett.* 9(1):1–18.
[26] Josz C, Henrion D (2016) Strong duality in Lasserre's hierarchy for polynomial optimization. *Optim. Lett.* 10(1):3–10.
[27] Kimsey DP (2011) Matrix-valued moment problems. PhD thesis, Drexel University, Philadelphia.

[28] Kimsey DP, Trachana M (2022) On a solution of the multidimensional truncated matrix-valued moment problem. *Milan J. Math.* 90(1): 17–101.
[29] Klep I, Nie J (2020) A matrix Positivstellensatz with lifting polynomials. *SIAM J. Optim.* 30(1):240–261.
[30] Kojima M (2003) Sums of squares relaxations of polynomial semidefinite programs. Research Report No. B-397, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo.
[31] Lasserre JB (2001) Global optimization with polynomials and the problem of moments. *SIAM J. Optim.* 11(3):796–817.
[32] Lasserre JB (2008) A semidefinite programming approach to the generalized problem of moments. *Math. Programming* 112(1):65–92.
[33] Lasserre JB (2009) Convexity in semialgebraic geometry and polynomial optimization. *SIAM J. Optim.* 19(4):1995–2014.
[34] Lasserre JB (2011) Min-max and robust polynomial optimization. *J. Global Optim.* 51(1):1–10.
[35] Li G, Ma AKC, Pong TK (2014) Robust least square semidefinite programming with applications. *Comput. Optim. Appl.* 58:347–379.
[36] Löfberg J (2004) YALMIP: A toolbox for modeling and optimization in MATLAB. *2004 IEEE Internat. Sympos. Comput. Aided Control System Design* (IEEE, Piscataway, NJ), 284–289.
[37] Louca R, Bitar E (2016) A hierarchy of polyhedral approximations of robust semidefinite programs. *2016 IEEE 55th Conf. Decision Control (CDC)* (IEEE, Piscataway, NJ), 7056–7062.
[38] Mai NHA, Lasserre JB, Magron V, Wang J (2022) Exploiting constant trace property in large-scale polynomial optimization. *ACM Trans. Math. Software* 48(4):1–39.
[39] Maranas CD, Floudas CA (1994) Global minimum potential energy conformations of small molecules. *J. Global Optim.* 4(2):135–170.
[40] Nie J (2011) Polynomial matrix inequality and semidefinite representation. *Math. Oper. Res.* 36(3):398–415.
[41] Nie J (2014) The A-truncated K-moment problem. *Foundations Comput. Math.* 14(6):1243–1276.
[42] Nie J (2023) *Moment and Polynomial Optimization* (Society for Industrial and Applied Mathematics, Philadelphia).
[43] Oishi Y (2006) A region-dividing approach to robust semidefinite programming and its error bound. *Proc. 2006 Amer. Control Conf.* (IEEE, Piscataway, NJ), 123–129.
[44] Pozdyayev VV (2014) Atomic optimization. II. Multidimensional problems and polynomial matrix inequalities. *Automatic Remote Control* 75(6):1155–1171.
[45] Putinar M (1993) Positive polynomials on compact semi-algebraic sets. *Indiana Univ. Math. J.* 42(3):969–984.
[46] Scherer CW, Hol CWJ (2006) Matrix sum-of-squares relaxations for robust semi-definite programs. *Math. Programming* 107(1):189–211.
[47] Schmüdgen K (1987) On a generalization of the classical moment problem. *J. Math. Anal. Appl.* 125(2):461–470.
[48] Slot L, Laurent M (2023) Sum-of-squares hierarchies for binary polynomial optimization. *Math. Programming* 197(2):621–660.
[49] VanAntwerp JG, Braatz RD (2000) A tutorial on linear and bilinear matrix inequalities. *J. Process Control* 10(4):363–385.
[50] Waki H, Kim S, Kojima M, Muramatsu M (2006) Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity. *SIAM J. Optim.* 17(1):218–242.
[51] Wang J, Magron V, Lasserre JB (2021) TSSOS: A moment-SOS hierarchy that exploits term sparsity. *SIAM J. Optim.* 31(1):30–58.
[52] Wang J, Magron V, Lasserre JB, Mai NHA (2022) CS-TSSOS: Correlative and term sparsity for large-scale polynomial optimization. *ACM Trans. Math. Software* 48(4):1–26.
[53] Zheng Y, Fantuzzi G (2023) Sum-of-squares chordal decomposition of polynomial matrix inequalities. *Math. Programming* 197(1):71–108.