Piecewise SOS-Convex Moment Optimization and Applications via Exact Semi-Definite Programs

Q.Y. Huang,[*] V. Jeyakumar[†] and G. Li[‡]

July 3, 2024

Abstract

This paper presents exact Semi-Definite Program (SDP) reformulations for infinite-dimensional moment optimization problems involving a new class of piecewise Sum-of-Squares (SOS)-convex functions and projected spectrahedral support sets. These reformulations show that solving a single SDP finds the optimal value and an optimal probability measure of the original moment problem. This is done by establishing an SOS representation for the non-negativity of a piecewise SOS-convex function over a projected spectrahedron. Finally, as an application and a proof-of-concept illustration, the paper also presents numerical results for the Newsvendor and revenue maximization problems with higher-order moments by solving their equivalent SDP reformulations. These reformulations promise a flexible and efficient approach to solving these models. The main novelty of the present work in relation to the recent research lies in finding the solution to moment problems, for the first time, with piecewise SOS-convex functions from their numerically tractable exact SDP reformulations.

Keywords. Moment Optimization; Sum-of-Squares Convex Polynomials; Piecewise Functions; Generalized moment problems; Semi-Definite Programming

1 Introduction

Consider the generalized moment problem

$$\min_{\mu \in \mathcal{P}_\Omega} \left\{ \mathbb{E}_\mu^\Omega \left[\min_{k=1,\ldots,r} \max_{\ell=1,\ldots,L} g_\ell^k(\omega) \right] \; : \; \mathbb{E}_\mu^\Omega[h_j(\omega)] \le \gamma_j, j = 1, \ldots, J \right\} \tag{P}$$

where g_ℓ^k, $\ell = 1, \ldots, L$, $k = 1, \ldots, r$, and h_j, $j = 1, \ldots, J$, are Sum-of-Squares (SOS)-convex polynomials. The SOS-convexity for polynomials is a numerically tractable relaxation of convexity

---

since checking whether a polynomial is SOS-convex or not can be achieved by solving an SDP [1, 15]. Recent applications of SOS-convexity in optimization may be found in [25, 18]. The set $\Omega \subset \mathbb{R}^m$ is a convex compact projected spectrahedron and $\mathcal{P}_\Omega$ is the set of probability measures supported on Ω. The expectation of a random variable ω with respect to the probability measure $\mu \in \mathcal{P}_\Omega$ is denoted by $\mathbb{E}^\Omega_\mu[\omega]$.

Our model (P) is closely related to generalized moment problems [24] where g^k_ℓ, $\ell = 1, \ldots, L$, $k = 1, \ldots, r$, and h_j, $j = 1, \ldots, J$, are arbitrary real-valued continuous functions and the support set Ω is an arbitrary compact subset of $\mathbb{R}^m$. The dominant technique to reformulate a generalized moment problem as a semi-infinite optimization problem is via the duality theory [32, 36]. However, in general, a semi-infinite optimization problem can be computationally intractable.

The classical moment problem can be recovered from (P) if $L = 1, r = 1$, the random variable is univariate, and the constraints are the first J moments [24]. The multivariate classical moment problem extends to countably many moments, but numerically tractable forms are limited. When $L = 1$, $r = 1$, g^1_1 and h_j, $j = 1, \ldots, J$, are linear functions and the support Ω is a spectrahedron, (P) admits an exact SDP reformulation [17]. When $L = 1$, g^k_1, $k = 1, \ldots, r$, are linear functions with mean and variance constraints, the support Ω is an ellipsoid, it has been shown that (P) admits an exact SDP reformulation [8].

When $L = 1$, $r = 1$, g^1_1 and h_j, $j = 1, \ldots, J$, are any polynomials and the support Ω is a compact semi-algebraic set, the optimal solution for (P) can be found by solving a convergent hierarchy of SDPs [22]. Further, polynomial optimization techniques were employed in [6] to obtain exact SOS reformulations when $\mathscr{P}$ contains probability measures whose density functions are SOS polynomials. A review of generalized moment problems and their applications is available in [7].

The model problem (P) is of great interest in distributionally robust optimization [36], where (P) appears as an uncertainty quantification problem to mitigate risks and uncertainties from the input data [36, 13]. For example, in the Newsvendor model of [33], the cost of ordering represented as a piecewise linear loss function is minimized to meet the uncertain demand. In the portfolio management problem of [8], the profit of the portfolio represented as a piecewise linear utility function is maximized amid random investment returns.

Motivated by the importance of such piecewise functions in optimization and the usefulness of the minimax functions of the form $\min_{k=1,\ldots,r} \max_{\ell=1,\ldots,L} g^k_\ell$ in applications across diverse domains, we introduce a new class of functions, called the piecewise SOS-convex functions. To the best of our knowledge, this notion of piecewise SOS-convex function has so far not been studied in the literature.

A function f on $\mathbb{R}^n$ is *piecewise SOS-convex* if there exist SOS-convex polynomials g^k_ℓ, $\ell = 1, \ldots, L$, $k = 1, \ldots, r$, on $\mathbb{R}^n$ such that $f(v) = \min_{k=1,\ldots,r} \max_{\ell=1,\ldots,L} g^k_\ell(v)$ for all $v \in \mathbb{R}^n$.

Note that a piecewise SOS-convex function is not necessarily a convex or a differentiable function and it covers a broad range of functions that appear in applications across several domains, as outlined below. Details of their piecewise representations are provided in Appendix A.

- The truncated ℓ_1-norm, $f(v) = \min\{1, \varepsilon|v|\}$, $\varepsilon > 0$, which is non-convex and non-smooth, is used extensively in machine-learning regularization [26].

- The piecewise linear function $f(v) = \min_{k=1,\ldots,r} p^\top_k v + q_k$, $p_k \in \mathbb{R}^n$, $q_k \in \mathbb{R}$, $k = 1, \ldots, r$, appears in Newsvendor-type [3, 10], conditional value-at-risk [30], and portfolio selection [8] models,

and it is a useful approximation for various important utility functions [8].

- The class of max-SOS-convex functions $f(v) = \max\limits_{\ell=1,\ldots,L} g_\ell(v)$, where g_ℓ's are SOS-convex polynomials, is studied in [21]. Further, convex quadratics and convex separable polynomials are SOS-convex [1, 19].

- The class of difference-of-convex functions of the form $f(v) = \max\limits_{\ell=1,\ldots,L} f_\ell(v) - \max\limits_{k=1,\ldots,r} (p_k^\top v + q_k)$, where f_ℓ's are SOS-convex polynomials, can be expressed as $f(v) = \min\limits_{k=1,\ldots,r} \max\limits_{\ell=1,\ldots,L} g_\ell^k(v)$ for $g_\ell^k(v) = f_\ell(v) - (p_k^\top v + q_k)$, $\ell = 1,\ldots,L$, $k = 1,\ldots,r$. These functions have attracted applications in feature selection models of machine-learning [18].

- The piecewise convex quadratic function

$$f(v) = \begin{cases} a(v-b)^2 + c, & \text{if } 0 \leq v \leq b, \\ c, & \text{if } v > b. \end{cases}$$

  with $a, b \geq 0$, $c \in \mathbb{R}$, is used to model offer prices from customers [13].

- The Huber loss function with parameter $\varepsilon > 0$,

$$f(v) = \begin{cases} \frac{1}{2}v^2, & \text{if } |v| \leq \varepsilon, \\ \varepsilon|v| - \frac{1}{2}\varepsilon^2, & \text{otherwise,} \end{cases}$$

  is studied in robust statistics [33].

The present study of moment optimization is motivated by two aspects. Firstly, moment problems are numerically challenging due to the presence of infinite-dimensional distributions and multi-dimensional integrals. In response, we examine equivalent Semi-Definite Program (SDP) reformulations of these problems which can be efficiently solved numerically by commonly available software via interior point methods.

Secondly, the study of piecewise SOS-convexity in (P) was stimulated by the computationally attractive features of SOS-convex polynomial optimization and its applications in many areas. It is known that SOS-convex polynomial optimization problems admit numerically favourable SDP reformulations [19, 20, 21, 25].

The main goal of this paper is to provide numerically tractable reformulations for the moment problem involving piecewise SOS-convex functions and show how the solution can be recovered from its SDP reformulation.

Our contributions. Our main contributions are as follows.

(i) Firstly, we introduce the notion of piecewise SOS-convex functions and establish a new SOS representation of the non-negativity of a piecewise SOS-convex function over a compact projected spectrahedron. This result not only generalizes the known representation result of a non-negative SOS-convex polynomial over a spectrahedron [19] but also provides numerically tractable representations of non-negativity for a broad class of functions that are not necessarily convex.

(ii) Secondly, exploiting the representation result, we derive an equivalent numerically tractable SOS optimization reformulation for the generalized moment problem (P). Moreover, under

suitable conditions, we show how to recover an optimal probability measure for (P) by solving a single SDP. This extends the corresponding result of Lasserre [23] on SOS-convex polynomial optimization.

(iii) Finally, as an application and a proof-of-concept illustration, we present numerical results for two important practical models, i.e., the Newsvendor problem [12] with higher-order moments and the revenue maximization problem [13] with a higher-order utility function, by solving their equivalent SDP reformulations.

The novelty of the present work in relation to recent research in SOS-convexity and moment optimization is the derivation of numerically tractable exact SDP reformulations for moment problems involving, *for the first time, piecewise SOS-convex functions*. The main innovation is the combined utilization of powerful tools from real algebraic geometry, convex analysis, and SOS polynomials to produce the exact reformulations. We employ the key computationally favourable features of SOS-convexity within an infinite-dimensional setting and exploit the geometry of the projected spectrahedra to facilitate the SDP reformulations.

Our approach makes use of the piecewise structure of the functions for transforming infinite-dimensional problems into finite-dimensional numerically tractable problems. Moreover, the projected spectrahedron covers a broad class of semi-algebraic sets [25], such as the spectrahedra, ellipsoids, and boxes, used in robust optimization [2].

The paper is organized as follows. Section 2 provides SOS representation results for piecewise SOS-convex functions. Section 3 presents the main theorems for the SOS reformulation of (P) and the optimal solution recovery of (P). Section 4 describes applications to the Newsvendor and revenue maximization models. Section 5 concludes with discussions on potential future work. The appendices provide technical details related to explicit piecewise representations for some piecewise SOS-convex functions (Appendix A), infinite-dimensional conic duality (Appendix B), and SDP representations of SOS optimization problems (Appendix C).

2 Non-Negativity of Piecewise SOS-Convex Functions

In this section, we outline SOS representations for a class of piecewise functions involving SOS-convex polynomials. This will play a vital role in reformulating the generalized moment problem (P) as a numerically tractable SOS optimization problem.

We begin with some preliminaries on polynomials. Denote $\mathbb{R}^m$ the Euclidean space of dimension m and $\mathbb{R}^m_+$ the non-negative orthant of $\mathbb{R}^m$. The standard inner product is $a^\top b$ for $a, b \in \mathbb{R}^m$. Let $\mathbb{R}[v]$ be the space of polynomials with real coefficients over $v \in \mathbb{R}^m$. A polynomial $f \in \mathbb{R}[v]$ is called a Sum-of-Squares (SOS) polynomial if there exist polynomials $f_j \in \mathbb{R}[v], j = 1, \ldots, J$, such that $f = \sum_{j=1}^{J} f_j^2$. For a polynomial $f \in \mathbb{R}[v]$, we use $\deg f$ to denote its degree. We also use $\Sigma_d^2(v)$ to denote the set of all SOS polynomials f of degree at most d with respect to the variable $v \in \mathbb{R}^m$. Next, we recall the definition of SOS-convex polynomials.

Definition 2.1 (SOS-convex polynomial [1, 16]). *A polynomial $f \in \mathbb{R}[v]$ is SOS-convex if its Hessian $H(v)$ is an SOS matrix polynomial, that is, if there exists an $(s \times m)$ polynomial matrix $P(v)$ for some $s \in \mathbb{N}$ such that $H(v) = P(v)^\top P(v)$.*

Several equivalent conditions for SOS-convexity are available in [1]. For instance, $f \in \mathbb{R}[v]$ is

4

SOS-convex whenever the polynomial $g(v_1, v_2) = f(v_1) - f(v_2) - \nabla f(v_2)^\top (v_1 - v_2)$ on $\mathbb{R}^m \times \mathbb{R}^m$ is an SOS polynomial with respect to the variable (v_1, v_2).

Any SOS-convex polynomials are convex polynomials, but the converse is not true [1]. In other words, the class of SOS-convex polynomials is a proper subclass of convex polynomials. Moreover, the class of SOS-convex polynomials covers affine functions, convex quadratics, and convex separable polynomials, while SOS-convex polynomials can be non-quadratic or non-separable in general [19].

Now, we formally define the notion of a piecewise SOS-convex function.

Definition 2.2 (Piecewise SOS-convex function). *We call a function f on $\mathbb{R}^m$ piecewise SOS-convex if there exist SOS-convex polynomials g_ℓ^k, $\ell = 1, \ldots, L$, $k = 1, \ldots, r$, on $\mathbb{R}^m$ such that $f(v) = \min\limits_{k=1,\ldots,r} \max\limits_{\ell=1,\ldots,L} g_\ell^k(v)$ for all $v \in \mathbb{R}^m$.*

The following proposition links non-negative SOS-convex polynomials to SOS polynomials which will be deployed in the tractable representation of piecewise SOS-convex functions (Theorem 2.4).

Proposition 2.3. *(see [16] and [21, Corollary 2.1]). Let $f \in \mathbb{R}[v]$ be a non-negative SOS-convex polynomial. Then, f is an SOS polynomial.*

Denote $e_i \in \mathbb{R}^m$ the i-th standard basis vector of $\mathbb{R}^m$, $i = 1, \ldots, m$. The simplex in $\mathbb{R}^L$ is defined as $\Delta = \{\delta \in \mathbb{R}_+^L : \sum_{\ell=1}^L \delta_\ell = 1\}$. Let $\mathbb{S}^\nu$ be the set of $(\nu \times \nu)$ symmetric matrices. The trace product on $\mathbb{S}^\nu$ is $\mathrm{tr}(MN)$ for any $M, N \in \mathbb{S}^\nu$. A matrix $M \in \mathbb{S}^\nu$ is positive semi-definite, denoted as $M \succeq 0$ (resp. positive definite, denoted as $M \succ 0$), if $x^\top M x \geq 0$ for all $x \in \mathbb{R}^\nu$ (resp. $x^\top M x > 0$ for all $x \in \mathbb{R}^\nu$, $x \neq 0$). Let $\mathbb{S}_+^\nu$ be the cone of symmetric $(\nu \times \nu)$ positive semi-definite matrices.

A projected spectrahedron [15], or a spectrahedral shadow [29], takes the form

$$\Omega = \left\{ v \in \mathbb{R}^m : \exists \xi \in \mathbb{R}^N, F_0 + \sum_{i=1}^m v_i F_i + \sum_{t=1}^N \xi_t M_t \succeq 0 \right\}, \tag{LMI}$$

for some $F_i \in \mathbb{S}^\nu$, $i = 0, 1, \ldots, m$, and $M_t \in \mathbb{S}^\nu$, $t = 1, \ldots, N$. This set will later be used as a support set for probability measures. This set is versatile as any semi-algebraic set $\{v \in \mathbb{R}^m : f_i(v) \leq 0, i = 1, \ldots, n\}$ described by SOS-convex polynomials $f_i, i = 1, \ldots, n$, can be represented as a projected spectrahedron [15]. Moreover, spectrahedra, ellipsoids, and boxes are special forms of projected spectrahedra.

The following representation of a piecewise SOS-convex function extends the result in [19] for an SOS-convex polynomial over a spectrahedron and in [21] for a max-SOS-convex polynomial over a subclass of projected spectrahedron.

Theorem 2.4 (SOS representation for non-negative piecewise SOS-convex functions). *Let $\Omega \subset \mathbb{R}^m$ be given as in (LMI), $g_\ell^k, \ell = 1, \ldots, L, k = 1, \ldots, r$, be SOS-convex polynomials on $\mathbb{R}^m$. Assume that Ω is compact and there exist $\bar{v} \in \mathbb{R}^m$ and $\bar{\xi} \in \mathbb{R}^N$ such that $F_0 + \sum\limits_{i=1}^m \bar{v}_i F_i + \sum\limits_{t=1}^N \bar{\xi}_t M_t \succ 0$. Then, the following statements are equivalent.*

    (i) $v \in \Omega \implies \min\limits_{k=1,\ldots,r} \max\limits_{\ell=1,\ldots,L} g_\ell^k(v) \geq 0$.

(ii) For each $k = 1, \ldots, r$, there exist $Z_k \in \mathbb{S}_+^\nu$ and $\delta^k = (\delta_1^k, \ldots, \delta_L^k) \in \Delta$ such that

$$\begin{cases} \sum_{\ell=1}^{L} \delta_\ell^k g_\ell^k(v) - \mathrm{tr}(Z_k F_0) - \sum_{i=1}^{m} v_i \mathrm{tr}(Z_k F_i) \in \Sigma_d^2(v), \ k = 1, \ldots, r, \\ \mathrm{tr}(Z_k M_t) = 0, \ t = 1, \ldots, N, \ k = 1, \ldots, r. \end{cases}$$

where d is the smallest even integer larger than $\max_{k=1,\ldots,r} \max_{\ell=1,\ldots,L} \deg g_\ell^k$.

Proof. $[(ii) \implies (i)]$. Assume that there exist $Z_k \in \mathbb{S}_+^\nu$ and $\delta^k \in \Delta$, $k = 1, \ldots, r$, such that

$$\begin{cases} \sum_{\ell=1}^{L} \delta_\ell^k g_\ell^k(v) - \mathrm{tr}(Z_k F_0) - \sum_{i=1}^{m} v_i \mathrm{tr}(Z_k F_i) \in \Sigma_d^2(v), \ k = 1, \ldots, r, \\ \mathrm{tr}(Z_k M_t) = 0, \ t = 1, \ldots, N, \ k = 1, \ldots, r. \end{cases}$$

Fix $k \in \{1, \ldots, r\}$. Since an SOS polynomial always takes non-negative values, it follows that

$$\sum_{\ell=1}^{L} \delta_\ell^k g_\ell^k(v) - \mathrm{tr}(Z_k F_0) - \sum_{i=1}^{m} v_i \mathrm{tr}(Z_k F_i) \geq 0, \ \text{for all } v \in \mathbb{R}^m.$$

Further, $\mathrm{tr}(Z_k F_0) + \sum_{i=1}^{m} v_i \mathrm{tr}(Z_k F_i) = \mathrm{tr}(Z_k(F_0 + \sum_{i=1}^{m} v_i F_i + \sum_{t=1}^{N} \xi_t M_t))$ for any $\xi \in \mathbb{R}^N$. This implies that, for all $(v, \xi) \in \mathbb{R}^m \times \mathbb{R}^N$,

$$\sum_{\ell=1}^{L} \delta_\ell^k g_\ell^k(v) \geq \mathrm{tr}\Big(Z_k \Big(F_0 + \sum_{i=1}^{m} v_i F_i + \sum_{t=1}^{N} \xi_t M_t \Big) \Big). \tag{1}$$

Since $(\delta_1^k, \ldots, \delta_L^k) \in \Delta$, we have $\max_{\ell=1,\ldots,L} g_\ell^k(v) \geq \sum_{\ell=1}^{L} \delta_\ell^k g_\ell^k(v)$. Further, for an arbitrary $v \in \Omega$, there exists $\xi \in \mathbb{R}^N$ such that $F_0 + \sum_{i=1}^{m} v_i F_i + \sum_{t=1}^{N} \xi_t M_t \in \mathbb{S}_+^\nu$. Also, since $Z_k \in \mathbb{S}_+^\nu$, we have $\mathrm{tr}\Big(Z_k \Big(F_0 + \sum_{i=1}^{m} v_i F_i + \sum_{t=1}^{N} \xi_t M_t \Big) \Big) \geq 0$. It follows from Eqn. (1) that $\max_{\ell=1,\ldots,L} g_\ell^k(v) \geq 0$ for any $v \in \Omega$. This holds for any $k \in \{1, \ldots, r\}$, so $\min_{k=1,\ldots,r} \max_{\ell=1,\ldots,L} g_\ell^k(v) \geq 0$ for any $v \in \Omega$.

$[(i) \implies (ii)]$. Assume Statement *(i)* is true. Fix $k \in \{1, \ldots, r\}$. We have

$$\min_{v \in \Omega} \max_{\ell=1,\ldots,L} g_\ell^k(v) = \min_{v \in \Omega} \max_{\delta \in \Delta} \sum_{\ell=1}^{L} \delta_\ell g_\ell^k(v) \geq 0. \tag{2}$$

Define a bifunction $P_k : \Omega \times \Delta \to \mathbb{R}$ by $P_k(v, \delta) = \sum_{\ell=1}^{L} \delta_\ell g_\ell^k(v)$. Direct verification shows that $P_k(v, \cdot)$ is linear for each fixed $v \in \Omega$ and $P_k(\cdot, \delta)$ is convex continuous for each fixed $\delta \in \Delta$. Note that both $\Omega \subseteq \mathbb{R}^m$ and $\Delta \subseteq \mathbb{R}^L$ are convex compact sets. The Convex-Concave Minimax Theorem

6

[cf. 35, Theorem 2.10.2] gives us that

$$\min_{v \in \Omega} \max_{\delta \in \Delta} P_k(v, \delta) = \max_{\delta \in \Delta} \min_{v \in \Omega} P_k(v, \delta).$$

Thus, there exists $\delta^k \in \Delta$ such that $\max_{\delta \in \Delta} \min_{v \in \Omega} P_k(v, \delta) = \min_{v \in \Omega} P_k(v, \delta^k) = \min_{v \in \Omega} \sum_{\ell=1}^{L} \delta_\ell^k g_\ell^k(v).$

Observe that

$$\min_{v \in \Omega} \sum_{\ell=1}^{L} \delta_\ell^k g_\ell^k(v) = \min_{v \in \mathbb{R}^m} \left\{ \sum_{\ell=1}^{L} \delta_\ell^k g_\ell^k(v) : F_0 + \sum_{i=1}^{m} v_i F_i + \sum_{t=1}^{N} \xi_t M_t \succeq 0 \text{ for some } \xi \in \mathbb{R}^N \right\}$$

$$= \min_{(v,\xi) \in \mathbb{R}^m \times \mathbb{R}^N} \left\{ \sum_{\ell=1}^{L} \delta_\ell^k g_\ell^k(v) : F_0 + \sum_{i=1}^{m} v_i F_i + \sum_{t=1}^{N} \xi_t M_t \succeq 0 \right\}. \tag{3}$$

By the Lagrangian Duality under the strict feasibility assumption on Ω [35, Theorem 2.9.2], there exists $Z_k \in \mathbb{S}_+^\nu$ such that

$$\min (3) = \min_{(v,\xi) \in \mathbb{R}^m \times \mathbb{R}^N} \left\{ \sum_{\ell=1}^{L} \delta_\ell^k g_\ell^k(v) - \mathrm{tr}(Z_k F_0) - \sum_{i=1}^{m} v_i \mathrm{tr}(Z_k F_i) - \sum_{t=1}^{N} \xi_t \mathrm{tr}(Z_k M_t) \right\}.$$

Hence, Eqn. (2) implies that there exist $Z_k \in \mathbb{S}_+^\nu$ and $\delta^k \in \Delta$ such that

$$\sum_{\ell=1}^{L} \delta_\ell^k g_\ell^k(v) - \mathrm{tr}(Z_k F_0) - \sum_{i=1}^{m} v_i \mathrm{tr}(Z_k F_i) - \sum_{t=1}^{N} \xi_t \mathrm{tr}(Z_k M_t) \geq 0, \tag{4}$$

for all $(v, \xi) \in \mathbb{R}^m \times \mathbb{R}^N$. Letting $\xi = 0 \in \mathbb{R}^N$ in Eqn. (4) gives

$$\sigma_k(v) := \sum_{\ell=1}^{L} \delta_\ell^k g_\ell^k(v) - \mathrm{tr}(Z_k F_0) - \sum_{i=1}^{m} v_i \mathrm{tr}(Z_k F_i) \geq 0, \text{ for all } v \in \mathbb{R}^m.$$

Since σ_k is a sum of finitely many SOS-convex polynomials, σ_k is a non-negative SOS-convex polynomial. By Proposition 2.3, σ_k is thus an SOS polynomial. Moreover, by fixing $v = \widehat{v}$ for some $\widehat{v} \in \mathbb{R}^m$, Eqn. (4) further implies that $\sigma_k(\widehat{v}) - \sum_{t=1}^{N} \xi_t \mathrm{tr}(Z_k M_t) \geq 0$ for any $\xi \in \mathbb{R}^N$ where $\sigma_k(\widehat{v}) \geq 0$ is fixed. This forces $\mathrm{tr}(Z_k M_t) = 0$ for all $t = 1, \ldots, N$. The conclusion now follows because k is chosen arbitrarily from $\{1, \ldots, r\}$. $\qquad \square$

3 Piecewise SOS-Convex Moment Problems

In this section, we focus on reformulating *a generalized moment optimization problem with piecewise SOS-convex functions* as an SDP.

We begin by fixing some terminologies. Let Y and Y' be real topological vector spaces. They are paired if a bilinear form $\langle \cdot, \cdot \rangle : Y' \times Y \to \mathbb{R}$ is defined. The cone generated by any set $S \subseteq Y$ and the interior of S are denoted as $\mathrm{cone}(S)$ and $\mathrm{int}(S)$ respectively. The (positive) polar of a cone

$S \subseteq Y$ is defined as $S^+ = \{y' \in Y' : \langle y', y \rangle \geq 0, \forall y \in S\}$. See Appendix B for details on strong conic duality.

For a non-empty compact set $\Omega \subset \mathbb{R}^m$, $\mathcal{C}_\Omega$ denotes the vector space of all continuous real-valued functions on Ω equipped with the supremum norm. Denote $\mathcal{B}$ the Borel σ-algebra of Ω. *From this section onwards, we let X be the vector space of (finite signed regular) Borel measures on $(\Omega, \mathcal{B})$.* It is known that X can be identified as the topological dual space of $\mathcal{C}_\Omega$, i.e., $X = \mathcal{C}_\Omega^*$. We equip X with the weak* topology. For any point $v \in \Omega$, the Dirac measure $\mathbb{1}_v$ which takes a mass of 1 at the point $v \in \Omega$ and 0 otherwise belongs to X [32].

Consider the following generalized moment problem with piecewise SOS-convex functions:

$$\min_{\mu \in X} \mathbb{E}_\mu^\Omega \left[\min_{k=1,\dots,r} \max_{\ell=1,\dots,L} g_\ell^k(\omega) \right] \tag{P}$$
$$\text{s.t. } \mathbb{E}_\mu^\Omega [h_j(\omega)] \leq \gamma_j, \ j = 1, \dots, J,$$
$$\mathbb{E}_\mu^\Omega[1] = 1, \ \mu \succeq_\mathcal{B} 0,$$

where $\gamma_j \in \mathbb{R}$, $j = 1, \dots, J$, and h_j, $j = 1, \dots, J$, and g_ℓ^k, $\ell = 1, \dots, L$, $k = 1, \dots, r$, are SOS-convex polynomials on $\mathbb{R}^m$. Define $g : \mathbb{R}^m \to \mathbb{R}$, $g(v) = \min_{k=1,\dots,r} \max_{\ell=1,\dots,L} g_\ell^k(v)$ the piecewise SOS-convex function, and note that $g \in \mathcal{C}_\Omega$ and $h_j \in \mathcal{C}_\Omega$, $j = 1, \dots, J$. We let X' be the vector space generated by g, $h_j, j = 1, \dots, J$, and the constant function of 1. Note that $X' \subseteq \mathcal{C}_\Omega$, and X' is equipped with the supremum norm. Following [32], X and X' are paired spaces, and the continuous bilinear form between them is given by

$$\langle f, \mu \rangle = \mathbb{E}_\mu^\Omega[f(\omega)] := \int_\Omega f(\omega) \, d\mu(\omega), \text{ for all } f \in X' \text{ and } \mu \in X,$$

where $\mathbb{E}_\mu^\Omega[f(\omega)]$ refers to the expectation of the random variable $f(\omega)$ with respect to the measure μ supported on Ω. Denote the set of probability measures by $\mathcal{P}_\Omega = \{\mu \in X : \mathbb{E}_\mu^\Omega[1] = 1, \mu \succeq_\mathcal{B} 0\}$, where $\mu \succeq_\mathcal{B} 0$ means that $\mu(A) = \mu(\omega \in A) \geq 0$ for all $\mathcal{B}$-measurable sets A.

The moment problem (P) is known as an uncertainty quantification problem in the distributionally robust optimization literature [14], and the feasible set $\{\mu \in \mathcal{P}_\Omega : \mathbb{E}_\mu^\Omega[h_j(\omega)] \leq \gamma_j, j = 1, \dots, J\}$ is called a moment ambiguity set.

We assume that the support set Ω is a *(convex) projected spectrahedron*:

$$\Omega = \left\{ v \in \mathbb{R}^m : \exists \xi \in \mathbb{R}^N, \ F_0 + \sum_{i=1}^m v_i F_i + \sum_{t=1}^N \xi_t M_t \succeq 0 \right\}, \tag{LMI}$$

for some $F_i \in \mathbb{S}^\nu, i = 0, 1, \dots, m$, and $M_t \in \mathbb{S}^\nu, t = 1, \dots, N$. We associate (P) with the following

SOS optimization problem

$$
\max_{\substack{\lambda \in \mathbb{R}_+^J \times \mathbb{R} \\ Z_k \in \mathbb{S}_+^\nu, \delta^k \in \Delta \\ k=1,\dots,r}} \quad -\sum_{j=1}^{J} \lambda_j \gamma_j - \lambda_{J+1} \tag{D}
$$

$$
\text{s.t.} \quad \sum_{\ell=1}^{L} \delta_\ell^k g_\ell^k(v) + \sum_{j=1}^{J} \lambda_j h_j(v) + \lambda_{J+1} - \mathrm{tr}(Z_k F_0) - \sum_{i=1}^{m} v_i \mathrm{tr}(Z_k F_i) \in \Sigma_d^2(v), \ \ k = 1, \dots, r,
$$

$$
\mathrm{tr}(Z_k M_t) = 0, \ \ t = 1, \dots, N, \ \ k = 1, \dots, r,
$$

where $\delta^k = (\delta_1^k, \dots, \delta_L^k) \in \Delta = \{\delta \in \mathbb{R}_+^L : \sum_{\ell=1}^{L} \delta_\ell = 1\}$, $\Sigma_d^2(v)$ is the set of SOS polynomials of degree at most d with respect to the variable $v \in \mathbb{R}^m$, and d is the smallest even integer with

$$
d \geq \max \left\{ \max_{k=1,\dots,r} \max_{\ell=1,\dots,L} \deg g_\ell^k, \ \max_{j=1,\dots,J} \deg h_j \right\}.
$$

The SOS program (D) can be equivalently rewritten as a Semi-Definite Program (SDP), and thus can be efficiently solved via existing SDP software. We refer the readers to Appendix C for the procedure.

The following useful characterization states that the polar cone of the cone of non-negative measures is the convex cone of functions in X' that are non-negative on Ω. This characterization is known (e.g., [4, Examples 2.37, 2.122], [17, Lemma 3.1]), but we provide the proof here for the sake of self-containment.

Note that, for the convex cone of non-negative measures $C = \{\mu \in X : \mu \succeq_\mathcal{B} 0\}$, we denote $C^+ = \{f \in X' : \langle f, \mu \rangle \geq 0, \forall \mu \in C\}$. Here, C^+ is the intersection of the topological polar cones of C with the subspace X' of the vector space of continuous real-valued functions $\mathcal{C}_\Omega$.

Lemma 3.1 (Characterization of cone of non-negative measures). *Let $\Omega \neq \emptyset$ be any convex compact subset of $\mathbb{R}^m$, $C = \{\mu \in X : \mu \succeq_\mathcal{B} 0\}$, and $f \in X'$. Then, $f \in C^+$ if and only if $\min_{v \in \Omega} f(v) \geq 0$.*

Proof. Suppose that $f \in C^+$. Then, $\langle f, \mu \rangle \geq 0$ for all $\mu \in C$. Note that the Dirac measure $\mathbb{1}_v$ at any point $v \in \Omega$ belongs to C. This gives $f(v) = \langle f, \mathbb{1}_v \rangle \geq 0$ for any $v \in \Omega$. Hence, $\min_{v \in \Omega} f(v) \geq 0$.

Conversely, suppose that $\min_{v \in \Omega} f(v) \geq 0$. Now, for any $\mu \in \mathcal{P}_\Omega = \{\mu \in X : \int_\Omega 1 d\mu(\omega) = 1, \mu \succeq_\mathcal{B} 0\}$, we have

$$
\int_\Omega f(\omega) \, d\mu(\omega) \geq \int_\Omega \left(\min_{v \in \Omega} f(v) \right) d\mu(\omega) = \left(\min_{v \in \Omega} f(v) \right) \int_\Omega 1 d\mu(\omega) = \min_{v \in \Omega} f(v).
$$

This implies $\langle f, \mu \rangle \geq 0$ for any $\mu \in \mathcal{P}_\Omega$. Since $C = \mathrm{cone}(\mathcal{P}_\Omega)$, we see $f \in C^+$. $\quad\square$

Now, we show that (D) is an exact SOS reformulation for (P) in the sense that $\min (\mathrm{P}) = \max (\mathrm{D})$.

Assumption 3.2 (Interior point condition for moment problem).

$$
(\gamma_1, \dots, \gamma_J, 1) \in \mathrm{int} \left\{ \{ (\langle h_1, \mu \rangle, \dots, \langle h_J, \mu \rangle, \langle 1, \mu \rangle) : \mu \in X, \mu \succeq_\mathcal{B} 0 \} + \mathbb{R}_+^J \times \{0\} \right\}.
$$

This interior point condition is commonly used in the moment problem literature and can be satisfied in many situations. See [32, 34] for discussions.

Assumption 3.3. *The projected spectrahedron Ω as in* (LMI) *is compact and there exist $\bar{v} \in \mathbb{R}^m$ and $\bar{\xi} \in \mathbb{R}^N$ such that $F_0 + \sum_{i=1}^m \bar{v}_i F_i + \sum_{t=1}^N \bar{\xi}_t M_t \succ 0$. For the problem* (P), *the polynomials g_ℓ^k, $\ell = 1, \ldots, L$, $k = 1, \ldots, r$, and h_j, $j = 1, \ldots, J$, are SOS-convex polynomials on $\mathbb{R}^m$.*

Theorem 3.4 (**Exact SOS program for piecewise SOS-convex moment optimization**). *Assume that* (P) *admits a minimizer and Assumptions 3.2-3.3 are satisfied. Then, $\min (\mathrm{P}) = \max (\mathrm{D})$.*

Proof. [**Conic duality**]. Let $g(v) = \min_{k=1,\ldots,r} \max_{\ell=1,\ldots,L} g_\ell^k(v)$, $b = (\gamma_1, \ldots, \gamma_J, 1)$, $K = \mathbb{R}_+^J \times \{0\}$, and $C = \{\mu \in X : \mu \succeq_{\mathcal{B}} 0\}$. Note that $g \in X'$, and C and K are convex cones that are closed in the respective topologies. Define a continuous linear map $A : X \to \mathbb{R}^{J+1}$ by $A\mu = (\langle h_1, \mu \rangle, \ldots, \langle h_J, \mu \rangle, \langle 1, \mu \rangle)$. Then,

$$\min (\mathrm{P}) = \min_{\mu \in C} \{\langle g, \mu \rangle : -A\mu + b \in K\}.$$

The adjoint mapping $A^* : \mathbb{R}^{J+1} \mapsto X'$ is given by $A^*\lambda = \sum_{j=1}^J \lambda_j h_j + \lambda_{J+1} \cdot 1$, and the (positive) polar cone of K is $K^+ = \mathbb{R}_+^J \times \mathbb{R}$. Recall that $C^+ = \{f \in X' : \langle f, \mu \rangle \geq 0, \forall \mu \in C\}$. Then, Assumption 3.2 can be equivalently rewritten as $b \in \mathrm{int}(A(C) + K)$. It follows from strong conic linear duality (Corollary B.2) that

$$\min (\mathrm{P}) = \max \left\{ -\sum_{j=1}^J \lambda_j \gamma_j - \lambda_{J+1} : g + \sum_{j=1}^J \lambda_j h_j + \lambda_{J+1} \cdot 1 \in C^+, \lambda \in K^+ \right\}, \qquad (5)$$

and the maximum of the problem in Eqn. (5) is attained.

[**Equivalent SOS representation of conic constraint**]. This means, there exists $\lambda \in \mathbb{R}_+^J \times \mathbb{R}$ with $g + \sum_{j=1}^J \lambda_j h_j + \lambda_{J+1} \cdot 1 \in C^+$ such that $\min (\mathrm{P}) = -\sum_{j=1}^J \lambda_j \gamma_j - \lambda_{J+1}$. By Lemma 3.1, $g + \sum_{j=1}^J \lambda_j h_j + \lambda_{J+1} \cdot 1 \in C^+$ if and only if

$$\min_{v \in \Omega} \left\{ \min_{k=1,\ldots,r} \max_{\ell=1,\ldots,L} g_\ell^k(v) + \sum_{j=1}^J \lambda_j h_j(v) + \lambda_{J+1} \right\} \geq 0. \qquad (6)$$

The polynomial $g_\ell^k + \sum_{j=1}^J \lambda_j h_j + \lambda_{J+1}$ is SOS-convex for each $\ell = 1, \ldots, L$, $k = 1, \ldots, r$. By Theorem 2.4, Eqn. (6) is equivalent to, for each $k = 1, \ldots, r$,

$$\begin{cases} \sum_{\ell=1}^L \delta_\ell^k g_\ell^k(v) + \sum_{j=1}^J \lambda_j h_j(v) + \lambda_{J+1} - \mathrm{tr}(Z_k F_0) - \sum_{i=1}^m v_i \mathrm{tr}(Z_k F_i) \in \Sigma_d^2(v), \\ \mathrm{tr}(Z_k M_t) = 0, \ t = 1, \ldots, N, \end{cases}$$

for some $Z_k \in \mathbb{S}_+^\nu$ and $\delta^k \in \Delta$. Therefore, $\min (\mathrm{P}) \leq \max (\mathrm{D})$.

[**Weak duality**]. Take any feasible point $\lambda \in \mathbb{R}_+^J \times \mathbb{R}$, $Z_k \in \mathbb{S}_+^\nu$, $\delta^k \in \Delta, k = 1, \ldots, r$, for (D).

Theorem 2.4 gives $\min\limits_{k=1,\ldots,r}\max\limits_{\ell=1,\ldots,L}\left\{g_\ell^k(v) + \sum\limits_{j=1}^{J}\lambda_j h_j(v) + \lambda_{J+1}\right\} \geq 0$ for all $v \in \Omega$, and thus,

$$\min_{k=1,\ldots,r}\max_{\ell=1,\ldots,L} g_\ell^k(v) \geq -\sum_{j=1}^{J}\lambda_j h_j(v) - \lambda_{J+1}, \text{ for all } v \in \Omega.$$

Take any feasible point μ for (P), then, $\mathbb{E}_\mu^\Omega\left[\min\limits_{k=1,\ldots,r}\max\limits_{\ell=1,\ldots,L} g_\ell^k(\omega)\right] \geq \mathbb{E}_\mu^\Omega\left[-\sum\limits_{j=1}^{J}\lambda_j h_j(\omega) - \lambda_{J+1}\right]$.

Further, $-\sum\limits_{j=1}^{J}\lambda_j\gamma_j - \lambda_{J+1} \leq \mathbb{E}_\mu^\Omega\left[-\sum\limits_{j=1}^{J}\lambda_j h_j(\omega) - \lambda_{J+1}\right]$. Thus, $\mathbb{E}_\mu^\Omega\left[\min\limits_{k=1,\ldots,r}\max\limits_{\ell=1,\ldots,L} g_\ell^k(\omega)\right] \geq$

$-\sum\limits_{j=1}^{J}\lambda_j\gamma_j - \lambda_{J+1}$, implying $\min\,(\mathrm{P}) \geq \max\,(\mathrm{D})$. Together, $\min\,(\mathrm{P}) = \max\,(\mathrm{D})$. $\qquad\square$

The SOS optimization problem (D) can be expressed equivalently as the following SDP (Proposition C.1),

$$\max_{\substack{\lambda\in\mathbb{R}_+^J\times\mathbb{R},\delta^k\in\Delta \\ Z_k\in\mathbb{S}_+^\nu, Q_k\in\mathbb{S}_+^{s(m,d/2)} \\ k=1,\ldots,r}} \quad -\sum_{j=1}^{J}\lambda_j\gamma_j - \lambda_{J+1} \tag{R}$$

$$\text{s.t.} \quad \sum_{\ell=1}^{L}\delta_\ell^k(g_\ell^k)_\alpha + \sum_{j=1}^{J}\lambda_j(h_j)_\alpha + \lambda_{J+1}(1)_\alpha - \mathrm{tr}(Z_k F_\alpha) = \mathrm{tr}(Q_k B_\alpha),$$

$$\text{for all } \alpha \in \mathcal{N}, \ k=1,\ldots,r,$$

$$\mathrm{tr}(Z_k M_t) = 0, \ t=1,\ldots,N, \ k=1,\ldots,r,$$

where $s(m,d) = \binom{m+d}{d}$, $\mathcal{N} = \{\alpha = (\alpha_1,\ldots,\alpha_m) \in (\{0\}\cup\mathbb{N})^m : \sum_{i=1}^{m}\alpha_i \leq d\}$, and $B_\alpha \in \mathbb{S}^{s(m,d/2)}$, $\alpha \in \mathcal{N}$, are matrices given in Appendix C. The notation $(f)_\alpha$ of any polynomial $f \in \mathbb{R}[v]$ of degree at most d refers to the α-th coefficient of f. In particular, $(1)_\alpha$ refers to the α-th coefficient of the constant function 1, that is, $(1)_\alpha = 1$ when $\alpha = 0$, and $(1)_\alpha = 0$ otherwise. The matrices are $F_0 := F_0 \in \mathbb{S}^\nu$, $F_{e_i} := F_i \in \mathbb{S}^\nu$, $i=1,\ldots,m$, and F_α, $\alpha \in \mathcal{N}\setminus\{0,e_1,\ldots,e_m\}$, is the zero matrix.

The dual problem of (R), which is also an SDP, is given by

$$\min_{\substack{y^k \in \mathbb{R}^{s(m,d)}, \xi^k \in \mathbb{R}^N \\ z^k \in \mathbb{R}, k=1,\dots,r}} \sum_{k=1}^{r} z^k \tag{S}$$

$$\text{s.t.} \quad \sum_{k=1}^{r} \sum_{\alpha \in \mathcal{N}} y_\alpha^k (h_j)_\alpha \leq \gamma_j, \ \ j = 1, \dots, J, \tag{7}$$

$$\sum_{\alpha \in \mathcal{N}} y_\alpha^k (g_\ell^k)_\alpha \leq z^k, \ \ \ell = 1, \dots, L, \ k = 1, \dots, r, \tag{8}$$

$$y_0^k F_0 + \sum_{i=1}^{m} y_{e_i}^k F_i + \sum_{t=1}^{N} \xi_t^k M_t \succeq 0, \ \ k = 1, \dots, r, \tag{9}$$

$$\sum_{k=1}^{r} y_0^k = 1, \ \ \sum_{\alpha \in \mathcal{N}} y_\alpha^k B_\alpha \succeq 0, \ \ k = 1, \dots, r, \tag{10}$$

Denote the minimum value of the problem in (S) as $\min(\text{S})$. Sufficient conditions for the strong duality between the conic programs (R) and (S) are standard [cf. 4, 35].

We will make use of (S) to recover an optimal probability measure for (P) under suitable sign assumptions. The relationship between these problems is depicted in Figure 1.
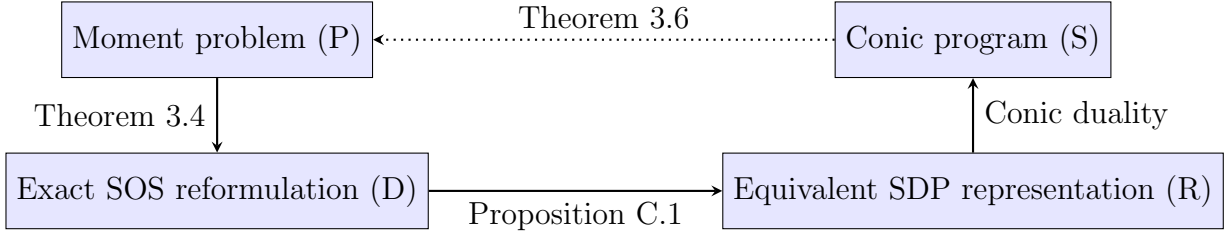

Figure 1: Recovering an optimal probability measure for (P) from (S).

We make use of the following generalized Jensen's inequality [23, Theorem 2.6].

Proposition 3.5. *Let $f \in \mathbb{R}[v]$ be an SOS-convex polynomial of an even degree d, $y = (y_\alpha)_{\alpha \in \mathcal{N}}$ satisfy $y_0 = 1$ and $\sum_{\alpha \in \mathcal{N}} y_\alpha B_\alpha \succeq 0$, and $L_y : \mathbb{R}[v] \to \mathbb{R}$ be the linear functional $L_y(f) = \sum_{\alpha \in \mathcal{N}} (f)_\alpha y_\alpha$. Then, $L_y(f) \geq f(L_y(\mathbf{v}))$, where $L_y(\mathbf{v}) = (L_y(\mathbf{v}_1), \dots, L_y(\mathbf{v}_m))$, and $\mathbf{v}_i$ denotes the polynomial that maps a vector $v \in \mathbb{R}^m$ to its i-th coordinate v_i, $i = 1, \dots, m$.*

Theorem 3.6 (Recovering an optimal solution). *Suppose that (P) admits a minimizer, Assumptions 3.2-3.3 are satisfied, and strong duality between (R) and (S) holds, i.e., $\max(\text{R}) = \min(\text{S})$. Let $(\bar{y}^k, \bar{\xi}^k, \bar{z}^k) \in \mathbb{R}^{s(m,d)} \times \mathbb{R}^N \times \mathbb{R}$, $\bar{y}^k = (\bar{y}_\alpha^k)_{\alpha \in \mathcal{N}}$, be a minimizer for (S) with $\bar{y}_0^k \geq 0$ for all $k = 1, \dots, r$. Denote $K := \{k \in \{1, \dots, r\} : \bar{y}_0^k > 0\}$ and $\hat{u}^k := \frac{1}{\bar{y}_0^k}(\bar{y}_{e_1}^k, \dots, \bar{y}_{e_m}^k) \in \mathbb{R}^m$ for $k \in K$. Suppose that, for all $k \notin K$,*

$$\bar{z}_k \geq 0 \ \text{ and } \ \sum_{\alpha \in \mathcal{N}} \bar{y}_\alpha^k (h_j)_\alpha \geq 0 \ \text{ for all } j = 1, \dots, J.$$

Then, $\widehat{\mu} := \sum_{k \in K} \bar{y}_0^k \mathbb{1}_{\widehat{u}^k}$, the linear combination of Dirac measures at points $\widehat{u}^k \in \mathbb{R}^m$, $k \in K$, is a minimizer for (P).

Proof. The equalities $\min(P) = \max(D)$, $\max(D) = \max(R)$, and $\max(R) = \min(S)$ follow from Theorem 3.4, Proposition C.1, and the strong duality assumption, respectively. Hence, $\min(P) = \min(S)$. Let $(\bar{y}^k, \bar{\xi}^k, \bar{z}^k) \in \mathbb{R}^{s(m,d)} \times \mathbb{R}^N \times \mathbb{R}$, $k = 1, \ldots, r$, be a solution for problem (S) with $\bar{y}_0^k \geq 0$ for all $k = 1, \ldots, r$, so it satisfies Eqns. (7)-(10).

For $k \in K$, define $\widehat{y}^k := \frac{1}{\bar{y}_0^k} \bar{y}^k \in \mathbb{R}^{s(m,d)}$, which satisfies $\widehat{y}_0^k = 1$ and by Eqn. (10),

$$\sum_{\alpha \in \mathcal{N}} \widehat{y}_\alpha^k B_\alpha = \frac{1}{\bar{y}_0^k} \sum_{\alpha \in \mathcal{N}} \bar{y}_\alpha^k B_\alpha \succeq 0.$$

Let $\widehat{u}^k = (\widehat{u}_1^k, \ldots, \widehat{u}_m^k) = (\widehat{y}_{e_1}^k, \ldots, \widehat{y}_{e_m}^k)$, $k \in K$. Clearly, $\widehat{u}^k \in \Omega$ for all $k \in K$ because we can take $\widehat{\xi}^k := \frac{1}{\bar{y}_0^k} \bar{\xi}^k \in \mathbb{R}^N$ such that, by Eqn. (9),

$$F_0 + \sum_{i=1}^m \widehat{u}_i^k F_i + \sum_{t=1}^N \widehat{\xi}_t^k M_t = \frac{1}{\bar{y}_0^k} \left(\bar{y}_0^k F_0 + \sum_{i=1}^m \bar{y}_{e_i}^k F_i + \sum_{t=1}^N \bar{\xi}_t^k M_t \right) \succeq 0, \ k \in K.$$

Define $\widehat{\mu} := \sum_{k \in K} \bar{y}_0^k \mathbb{1}_{\widehat{u}^k}$. We note that for any $f \in X'$,

$$\mathbb{E}_{\widehat{\mu}}^\Omega[f(\omega)] = \langle f, \widehat{\mu} \rangle = \langle f, \sum_{k \in K} \bar{y}_0^k \mathbb{1}_{\widehat{u}^k} \rangle = \sum_{k \in K} \bar{y}_0^k \langle f, \mathbb{1}_{\widehat{u}^k} \rangle = \sum_{k \in K} \bar{y}_0^k f(\widehat{u}^k).$$

By Eqn. (10), $\widehat{\mu}$ is a probability measure supported on Ω since $\mathbb{E}_{\widehat{\mu}}^\Omega[1] = \sum_{k \in K} \bar{y}_0^k = \sum_{k=1}^r \bar{y}_0^k = 1$. Now, $\widehat{u}^k = L_{\widehat{y}^k}(\mathbf{v})$, giving $h_j(\widehat{u}^k) = h_j(L_{\widehat{y}^k}(\mathbf{v}))$ for $j = 1, \ldots, J$, $k \in K$. Using Proposition 3.5 and the fact that h_j's are SOS-convex polynomials, we have

$$h_j(\widehat{u}^k) \leq L_{\widehat{y}^k}(h_j) = \sum_{\alpha \in \mathcal{N}} \widehat{y}_\alpha^k (h_j)_\alpha = \frac{1}{\bar{y}_0^k} \sum_{\alpha \in \mathcal{N}} \bar{y}_\alpha^k (h_j)_\alpha, \ j = 1, \ldots, J, \ k \in K.$$

By the assumption that $\sum_{k \notin K} \sum_{\alpha \in \mathcal{N}} \bar{y}_\alpha^k (h_j)_\alpha \geq 0$ and Eqn. (7), it follows that

$$\mathbb{E}_{\widehat{\mu}}^\Omega[h_j(\omega)] = \sum_{k \in K} \bar{y}_0^k h_j(\widehat{u}^k) \leq \sum_{k \in K} \sum_{\alpha \in \mathcal{N}} \bar{y}_\alpha^k (h_j)_\alpha \leq \sum_{k=1}^r \sum_{\alpha \in \mathcal{N}} \bar{y}_\alpha^k (h_j)_\alpha \leq \gamma_j, \ j = 1, \ldots, J,$$

so $\widehat{\mu}$ is feasible for (P).

Similarly, one can derive that $\bar{y}_0^k g_\ell^k(\widehat{u}^k) \leq \sum_{\alpha \in \mathcal{N}} \bar{y}_\alpha^k (g_\ell^k)_\alpha$, $\ell = 1, \ldots, L$, $k \in K$. From Eqn. (8), for each $k \in K$, $\sum_{\alpha \in \mathcal{N}} \bar{y}_\alpha^k (g_\ell^k)_\alpha \leq \bar{z}^k$. This shows that $\bar{y}_0^k g_\ell^k(\widehat{u}^k) \leq \bar{z}^k$ for all $\ell = 1, \ldots, L$, and so, $\max_{\ell=1,\ldots,L} \bar{y}_0^k g_\ell^k(\widehat{u}^k) \leq \bar{z}^k$. Let $g(v) = \min_{k=1,\ldots,r} \max_{\ell=1,\ldots,L} g_\ell^k(v)$. Recall that $g \in X'$ and, for each $k = 1, \ldots, r$,

$g(v) \leq \max_{\ell=1,\dots,L} g_\ell^k(v)$ for any $v \in \Omega$. Then, since $\bar{y}_0^k > 0$ for all $k \in K$, one has

$$\mathbb{E}_{\widehat{\mu}}^\Omega[g(\omega)] = \sum_{k \in K} \bar{y}_0^k\, g(\widehat{u}^k) \leq \sum_{k \in K} \bar{y}_0^k \left(\max_{\ell=1,\dots,L} g_\ell^k(\widehat{u}^k) \right) = \sum_{k \in K} \left(\max_{\ell=1,\dots,L} \bar{y}_0^k g_\ell^k(\widehat{u}^k) \right) \leq \sum_{k \in K} \bar{z}^k.$$

Using the assumption that $\sum_{k \notin K} \bar{z}^k \geq 0$, one has $\mathbb{E}_{\widehat{\mu}}^\Omega[g(\omega)] \leq \sum_{k \in K} \bar{z}^k \leq \sum_{k=1}^{r} \bar{z}^k = \min(\mathrm{S})$, giving $\min(\mathrm{P}) \leq \min(\mathrm{S})$.

Conversely, $\min(\mathrm{P}) = \min(\mathrm{S})$ ensures that $\min(\mathrm{P}) = \mathbb{E}_{\widehat{\mu}}^\Omega \left[\min_{k=1,\dots,r} \max_{\ell=1,\dots,L} g_\ell^k(\omega) \right]$. Therefore, $\widehat{\mu}$ is an optimal solution for the original moment problem (P). $\square$

The optimal probability measure $\widehat{\mu} := \sum_{k \in K} \bar{y}_0^k \mathbb{1}_{\widehat{u}^k}$ refers to the discrete distribution with probabilities $\mathbb{P}(\omega = \widehat{u}^k) = \bar{y}_0^k$, $k \in K$, i.e., the distribution with masses of $\bar{y}_0^k$ at points $\widehat{u}^k \in \Omega$, $k \in K$.

We present a recovery result under a stronger condition that is easy to verify.

Corollary 3.7 (**Recovery under a simple stronger condition**). *Suppose that* (P) *admits a minimizer, Assumptions 3.2-3.3 are satisfied, and* $\max(\mathrm{R}) = \min(\mathrm{S})$. *Let* $(\bar{y}^k, \bar{\xi}^k, \bar{z}^k) \in \mathbb{R}^{s(m,d)} \times \mathbb{R}^N \times \mathbb{R}$, $\bar{y}^k = (\bar{y}_\alpha^k)_{\alpha \in \mathcal{N}}$, *be a minimizer for* (S) *with* $\bar{y}_0^k > 0$ *for all* $k = 1, \dots, r$. *Denote* $\widehat{u}^k := \frac{1}{\bar{y}_0^k}(\bar{y}_{e_1}^k, \dots, \bar{y}_{e_m}^k) \in \mathbb{R}^m$, $k = 1, \dots, r$. *Then,* $\widehat{\mu} := \sum_{k=1}^{r} \bar{y}_0^k \mathbb{1}_{\widehat{u}^k}$ *is a minimizer for* (P).

Proof. This follows from Theorem 3.6 by setting $K = \{1, \dots, r\}$. $\square$

As an application to polynomial optimization, our recovery result reduces to the known Lasserre's result in [23]. More explicitly, consider

$$\min_{v \in \Omega} g(v), \tag{11}$$

where $\Omega \subset \mathbb{R}^m$ is given as in (LMI) and g is an SOS-convex polynomial. Notice that, in [23, Theorem 3.3], the feasible set Ω is a compact semi-algebraic set formed by SOS-convex polynomial inequalities, and thus can be represented as a projected spectrahedron [15].

We associate (11) with the following conic programs

$$\max_{\lambda \in \mathbb{R}, Z \in \mathbb{S}_+^\nu} \left\{ -\lambda : g(v) + \lambda - \mathrm{tr}(ZF_0) - \sum_{i=1}^{m} v_i \mathrm{tr}(ZF_i) \in \Sigma_d^2(v),\ \mathrm{tr}(ZM_t) = 0,\ t = 1, \dots, N \right\}, \tag{12}$$

and its dual problem

$$\min_{y \in \mathbb{R}^{s(m,d)}, \xi \in \mathbb{R}^N} \left\{ \sum_{\alpha \in \mathcal{N}} y_\alpha (g)_\alpha : F_0 + \sum_{i=1}^{m} y_{e_i} F_i + \sum_{t=1}^{N} \xi_t M_t \succeq 0,\ y_0 = 1,\ \sum_{\alpha \in \mathcal{N}} y_\alpha B_\alpha \succeq 0 \right\}. \tag{13}$$

The following Corollary shows how an optimal solution of (11) can be obtained by solving the SDP program (13).

14

Corollary 3.8. *For the problem* (11), *suppose that* Ω *is compact, there exist* $\bar{v} \in \mathbb{R}^m$ *and* $\bar{\xi} \in \mathbb{R}^N$ *such that* $F_0 + \sum_{i=1}^m \bar{v}_i F_i + \sum_{t=1}^N \bar{\xi}_t M_t \succ 0$, *and* $\max(12) = \min(13)$. *Let* $(\bar{y}, \bar{\xi}) \in \mathbb{R}^{s(m,d)} \times \mathbb{R}^N$ *be a minimizer for* (13). *Then,* $(\bar{y}_{e_1}, \ldots, \bar{y}_{e_m}) \in \mathbb{R}^m$ *is a minimizer for* (11).

Proof. Consider the moment problems

$$\min_{\mu \in \mathcal{P}_\Omega} \ \mathbb{E}_\mu^\Omega[g(\omega)] \tag{14}$$

and

$$\min_{\mu \in D} \ \mathbb{E}_\mu^\Omega[g(\omega)], \tag{15}$$

where $D = \{\mathbb{1}_v \in X : v \in \Omega\} \subset \mathcal{P}_\Omega$ is the set of all Dirac measures supported on Ω, giving $\min(14) \leq \min(15)$. Note that $\min(15) = \min(11)$ because $\mathbb{E}_{\mathbb{1}_v}^\Omega[g(\omega)] = g(v)$ for all $v \in \Omega$.

The interior point condition (Assumption 3.2) for the moment problem (14) requires that $1 \in \text{int}\{\mathbb{E}_\mu^\Omega[1] : \mu \in C\}$, where $C = \{\mu \in X : \mu \succeq_\mathcal{B} 0\}$. Since $C = \text{cone}(\mathcal{P}_\Omega)$ and $\mathbb{E}_\mathbb{P}^\Omega[1] = 1$ for any $\mathbb{P} \in \mathcal{P}_\Omega$, a direct verification shows that $\{\mathbb{E}_\mu^\Omega[1] : \mu \in C\} = \mathbb{R}_+$. So, Assumption 3.2 is satisfied automatically.

By Corollary 3.7 with $\bar{y}_0 = 1$, we have that $\widehat{\mu} := \mathbb{1}_{(\bar{y}_{e_1}, \ldots, \bar{y}_{e_m})}$ is a minimizer for (14) and $(\bar{y}_{e_1}, \ldots, \bar{y}_{e_m}) \in \Omega$. As $\mathbb{E}_{\widehat{\mu}}^\Omega[g(\omega)] = \min(14) \leq \min(15)$ and $\widehat{\mu} \in D$, so $\widehat{\mu}$ is also a minimizer for (15). Noting that $\mathbb{E}_{\widehat{\mu}}^\Omega[g(\omega)] = g(\bar{y}_{e_1}, \ldots, \bar{y}_{e_m})$, one conclude that $(\bar{y}_{e_1}, \ldots, \bar{y}_{e_m})$ is a minimizer for the polynomial optimization problem (11). $\square$

4 Applications to Newsvendor and Revenue Maximization

We devote this section to presenting SDP reformulations and numerical results for the Newsvendor and revenue maximization problems.

Generalized Newsvendor Problems. A company orders n goods, where $n \geq 1$. For each of the i-th goods, the unit cost of upfront ordering is \$$c_i$ with $0 < c_i < 1$. For a fixed order quantity $0 \leq x_i \leq R_i$, where R_i refers to the capacity, the company wishes to estimate a worst-case (upper) bound for the cost of ordering to meet the random demands. If the demand ω_i exceeds the upfront order x_i, the company incurs a unit back-ordering cost of \$1, or a total back-ordering cost of \$$(\omega_i - x_i)$, otherwise, the cost of back-ordering is \$0. This results in a total ordering cost of $c_i x_i + \max\{\omega_i - x_i, 0\}$.

Assume that the demand for each goods is independent and the support is $\prod_{i=1}^n \Omega_i$, where $\Omega_i = [\underline{\omega}_i, \overline{\omega}_i]$ with $0 \leq \underline{\omega}_i < \overline{\omega}_i$, $i = 1, \ldots, n$. Then, the multi-product Newsvendor model reduces to single-product Newsvendor models, and the worst-case ordering cost can be found by solving $\max_{\mu \in \mathscr{P}_i}\{c_i x_i + \mathbb{E}_\mu^{\Omega_i}[\max\{\omega_i - x_i, 0\}]\}$ for all $i = 1, \ldots, n$, where $\mathscr{P}_i$ consists of various moment constraints.

We focus on the single-product Newsvendor model and suppress the index $i \in \{1, \ldots, n\}$. For a fixed order quantity $0 \leq x \leq R$, the worst-case ordering cost can be found from the following

moment problem,

$$\max_{\mu \in \mathscr{P}} \ cx + \mathbb{E}_\mu^\Omega[\max\{\omega - x, 0\}], \tag{16}$$

where $\Omega = [\underline{\omega}, \overline{\omega}] = \{v \in \mathbb{R} : F_0 + vF_1 \succeq 0\}, \underline{\omega} < \overline{\omega}$, with $F_0 = \begin{pmatrix} -\underline{\omega} & 0 \\ 0 & \overline{\omega} \end{pmatrix}$ and $F_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Another interpretation of (16) at $c = 0$ comes from pricing a European call option [3, 22]. The random variable ω represents the price of the underlying stock, and the fixed value x is regarded as the strike price. This gives the expected price of $\mathbb{E}_\mu^\Omega[\max\{\omega - x, 0\}]$, and thus a sharp upper bound $\max_{\mu \in \mathscr{P}} \mathbb{E}_\mu^\Omega[\max\{\omega - x, 0\}]$ of the price.

Numerically tractable reformulations for (16) are available for a limited choice of $\mathscr{P}$. For example, [27] provides a closed-form formula when $\mathscr{P}$ specifies known values for the mean and variance; recently, [12] presents a closed-form formula when $\mathscr{P}$ specifies known values for mean plus one other moment of any order; [3] reformulates (16) as an SDP when $\mathscr{P}$ contains the first J moments; and [33] reformulates (16) as a conic program when the random demand variable follows a distribution defined by robust statistics. For other related models with tractable reformulations, see [8].

We study the problem (16) when $\mathscr{P}_1 = \{\mu \in \mathcal{P}_\Omega : \mathbb{E}_\mu^\Omega[\omega] \leq \gamma_1, \mathbb{E}_\mu^\Omega[\omega^2] \leq \gamma_2\}$ and $\mathscr{P}_2 = \mathscr{P}_1 \cap \{\mu \in \mathcal{P}_\Omega : \mathbb{E}_\mu^\Omega[\omega^4] \leq \gamma_3\}$, where $\mathcal{P}_\Omega$ is the set of probability measures on Ω. The set $\mathscr{P}_1$ contains upper bounds for the mean and the second-order moment which is related to the variance, while $\mathscr{P}_2$ contains an additional bound for the fourth-order moment which is related to the kurtosis of a distribution. The mean, variance, and kurtosis of the random demand distribution are usually accessible in real-world applications.

The standard minimization form for the Newsvendor problem is

$$\min(17) := \min_{\mu \in \mathscr{P}} \ \mathbb{E}_\mu^\Omega \Big[\min_{k=1,2} g_1^k(\omega) \Big], \ g_1^1(v) = x - v, \ g_1^2(v) = 0, \tag{17}$$

and the worst-case expected ordering cost is equal to $[cx - \min(17)]$.

The associated SOS optimization problem for (17) when $\mathscr{P} = \mathscr{P}_1$ is

$$\max(18) := \max_{\substack{\lambda \in \mathbb{R}_+^2 \times \mathbb{R} \\ Z, Z' \in \mathbb{S}_+^2}} \ -\lambda_1 \gamma_1 - \lambda_2 \gamma_2 - \lambda_3 \tag{18}$$

$$\text{s.t.} \quad [x + \lambda_3 - \mathrm{tr}(ZF_0)] + [-1 + \lambda_1 - \mathrm{tr}(ZF_1)]v + \lambda_2 v^2 \in \Sigma_2^2(v),$$
$$[\lambda_3 - \mathrm{tr}(Z'F_0)] + [\lambda_1 - \mathrm{tr}(Z'F_1)]v + \lambda_2 v^2 \in \Sigma_2^2(v).$$

Its equivalent SDP representation is

$$\max(19) := \max_{\substack{\lambda \in \mathbb{R}_+^2 \times \mathbb{R} \\ Z, Z' \in \mathbb{S}_+^2}} \ -\lambda_1 \gamma_1 - \lambda_2 \gamma_2 - \lambda_3 \tag{19}$$

$$\text{s.t.} \quad \begin{pmatrix} x + \lambda_3 - \mathrm{tr}(ZF_0) & \frac{1}{2}[-1 + \lambda_1 - \mathrm{tr}(ZF_1)] \\ \frac{1}{2}[-1 + \lambda_1 - \mathrm{tr}(ZF_1)] & \lambda_2 \end{pmatrix} \succeq 0,$$
$$\begin{pmatrix} \lambda_3 - \mathrm{tr}(Z'F_0) & \frac{1}{2}[\lambda_1 - \mathrm{tr}(Z'F_1)] \\ \frac{1}{2}[\lambda_1 - \mathrm{tr}(Z'F_1)] & \lambda_2 \end{pmatrix} \succeq 0.$$

Proposition 4.1. *Assume that* (17) *admits a minimizer for* $\mathscr{P} = \mathscr{P}_1$ *and* $(\gamma_1, \gamma_2, 1) \in$ $\mathrm{int}\{\{(\mathbb{E}^\Omega_\mu[\omega], \mathbb{E}^\Omega_\mu[\omega^2], \mathbb{E}^\Omega_\mu[1]) : \mu \in X, \mu \succeq_{\mathcal{B}} 0\} + \mathbb{R}^2_+ \times \{0\}\}$. *Then,* $\min(17) = \max(19)$.

Proof. By Theorem 3.4, $\min(17) = \max(18)$. Next, following the approach in Proposition C.1, the constraint $[x + \lambda_3 - \mathrm{tr}(ZF_0)] + v[-1 + \lambda_1 - \mathrm{tr}(ZF_1)] + \lambda_2 v^2 \in \Sigma^2_2(v)$ is equivalent to the existence of $Q \in \mathbb{S}^2_+$ such that $\mathrm{tr}(QB_0) = x + \lambda_3 - \mathrm{tr}(ZF_0)$, $\mathrm{tr}(QB_1) = -1 + \lambda_1 - \mathrm{tr}(ZF_1)$, and $\mathrm{tr}(QB_2) = \lambda_2$, where $B_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $B_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, and $B_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. This gives

$$Q = \begin{pmatrix} x + \lambda_3 - \mathrm{tr}(ZF_0) & \frac{1}{2}[-1 + \lambda_1 - \mathrm{tr}(ZF_1)] \\ \frac{1}{2}[-1 + \lambda_1 - \mathrm{tr}(ZF_1)] & \lambda_2 \end{pmatrix} \in \mathbb{S}^2_+.$$

Similar arguments apply for the other constraint, and hence $\max(18) = \max(19)$. $\qquad\square$

Now, we illustrate how one can recover an optimal measure for (17) via the following SDP

$$\min(20) := \min_{\substack{y^1, y^2 \in \mathbb{R}^3 \\ z^1, z^2 \in \mathbb{R}}} z^1 + z^2 \tag{20}$$

$$\text{s.t.} \quad y^1_1 + y^2_1 \le \gamma_1, \ y^1_2 + y^2_2 \le \gamma_2, \ y^1_0 + y^2_0 = 1,$$
$$xy^1_0 - y^1_1 \le z^1, \ 0 \le z^2,$$
$$y^k_0 F_0 + y^k_1 F_1 \succeq 0, \ \begin{pmatrix} y^k_0 & y^k_1 \\ y^k_1 & y^k_2 \end{pmatrix} \succeq 0, \ k = 1, 2.$$

Proposition 4.2. *Assume the same conditions as in Proposition 4.1 and* $\max(19) = \min(20)$. *Let* $(\bar{y}^k, \bar{z}^k) \in \mathbb{R}^3 \times \mathbb{R}$ *with* $\bar{y}^k = (\bar{y}^k_0, \bar{y}^k_1, \bar{y}^k_2)$ *be a minimizer for* (20) *with* $\bar{y}^k_0 > 0$, $k = 1, 2$. *Denote* $\widehat{u}^k := \frac{1}{\bar{y}^k_0} \bar{y}^k_1 \in \mathbb{R}$, $k = 1, 2$. *Then,* $\bar{y}^1_0 \mathbb{1}_{\widehat{u}^1} + \bar{y}^2_0 \mathbb{1}_{\widehat{u}^2}$ *is a minimizer for* (17) *for* $\mathscr{P} = \mathscr{P}_1$.

Proof. This follows from Corollary 3.7. $\qquad\square$

Note that the probability measure optimal for the standard minimization form (17) will also be optimal for the original Newsvendor maximization problem (16).

In a similar vein, we examine the Newsvendor model with $\mathscr{P} = \mathscr{P}_2$. Its associated SOS optimization problem is given by

$$\max(21) := \max_{\substack{\lambda \in \mathbb{R}^3_+ \times \mathbb{R} \\ Z, Z' \in \mathbb{S}^2_+}} -\lambda_1 \gamma_1 - \lambda_2 \gamma_2 - \lambda_3 \gamma_3 - \lambda_4 \tag{21}$$

$$\text{s.t.} \quad [x + \lambda_4 - \mathrm{tr}(ZF_0)] + [-1 + \lambda_1 - \mathrm{tr}(ZF_1)]v + \lambda_2 v^2 + \lambda_3 v^4 \in \Sigma^2_4(v),$$
$$[\lambda_4 - \mathrm{tr}(Z'F_0)] + [\lambda_1 - \mathrm{tr}(Z'F_1)]v + \lambda_2 v^2 + \lambda_3 v^4 \in \Sigma^2_4(v).$$

Its equivalent SDP representation is

$$\max\,(22) := \max_{\substack{\lambda \in \mathbb{R}^3_+ \times \mathbb{R} \\ Z, Z' \in \mathbb{S}^2_+, Q, Q' \in \mathbb{S}^3_+}} \quad -\lambda_1 \gamma_1 - \lambda_2 \gamma_2 - \lambda_3 \gamma_3 - \lambda_4 \tag{22}$$

$$\text{s.t.} \quad \text{tr}(QB_0) = x + \lambda_4 - \text{tr}(ZF_0), \ \text{tr}(Q'B_0) = \lambda_4 - \text{tr}(Z'F_0),$$
$$\text{tr}(QB_1) = -1 + \lambda_1 - \text{tr}(ZF_1), \ \text{tr}(Q'B_1) = \lambda_1 - \text{tr}(Z'F_1),$$
$$\text{tr}(QB_2) = \text{tr}(Q'B_2) = \lambda_2,$$
$$\text{tr}(QB_3) = \text{tr}(Q'B_3) = 0,$$
$$\text{tr}(QB_4) = \text{tr}(Q'B_4) = \lambda_3,$$

where

$$B_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, B_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, B_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, B_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, B_4 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{23}$$

Proposition 4.3. *Assume that* (17) *admits a minimizer for* $\mathscr{P} = \mathscr{P}_2$ *and* $(\gamma_1, \gamma_2, \gamma_3, 1) \in$ int$\{\{(\mathbb{E}^\Omega_\mu[\omega], \mathbb{E}^\Omega_\mu[\omega^2], \mathbb{E}^\Omega_\mu[\omega^4], \mathbb{E}^\Omega_\mu[1]) : \mu \in X, \mu \succeq_\mathcal{B} 0\} + \mathbb{R}^3_+ \times \{0\}\}$. *Then,* min (17) = max (22).

Proof. The proof is similar to Proposition 4.1. $\qquad\square$

We recover an optimal solution for (17) using the following SDP,

$$\min\,(24) := \min_{\substack{y^1, y^2 \in \mathbb{R}^5 \\ z^1, z^2 \in \mathbb{R}}} \quad z^1 + z^2 \tag{24}$$

$$\text{s.t.} \quad y^1_1 + y^2_1 \leq \gamma_1, \ y^1_2 + y^2_2 \leq \gamma_2, \ y^1_4 + y^2_4 \leq \gamma_3, \ y^1_0 + y^2_0 = 1,$$
$$xy^1_0 - y^1_1 \leq z^1, \ 0 \leq z^2,$$
$$y^k_0 F_0 + y^k_1 F_1 \succeq 0, \ \begin{pmatrix} y^k_0 & y^k_1 & y^k_2 \\ y^k_1 & y^k_2 & y^k_3 \\ y^k_2 & y^k_3 & y^k_4 \end{pmatrix} \succeq 0, \ k = 1, 2.$$

Proposition 4.4. *Assume the same conditions as in Proposition 4.3 and* max (22) = min (24). *Let* $(\bar{y}^k, \bar{z}^k) \in \mathbb{R}^5 \times \mathbb{R}$ *with* $\bar{y}^k = (\bar{y}^k_0, \bar{y}^k_1, \bar{y}^k_2, \bar{y}^k_3, \bar{y}^k_4)$ *be a minimizer for* (24) *with* $\bar{y}^k_0 > 0$, $k = 1, 2$. *Denote* $\hat{u}^k := \frac{1}{\bar{y}^k_0} \bar{y}^k_1 \in \mathbb{R}$, $k = 1, 2$. *Then,* $\bar{y}^1_0 \mathbb{1}_{\hat{u}^1} + \bar{y}^2_0 \mathbb{1}_{\hat{u}^2}$ *is a minimizer for* (17) *for* $\mathscr{P} = \mathscr{P}_2$.

Proof. This follows from Corollary 3.7. $\qquad\square$

We illustrate how the worst-case costs of ordering $[cx - \max\,(19)]$ and $[cx - \max\,(22)]$ change by varying the order quantity $x \in [0, R] = [0, 10]$ for $\mathscr{P}_1$ and $\mathscr{P}_2$. The SDPs are modelled by CVX MATLAB [11] and solved by Mosek [28]. Set $\Omega = [\underline{\omega}, \overline{\omega}] = [0, 100]$ for the support, $c = 0.1$ for the unit cost of upfront ordering, and $\gamma_1 = \gamma_2 = \gamma_3 = 1$ for the moment bounds. Figure 2 illustrates the cost of ordering (\$) for $\mathscr{P} = \mathscr{P}_1$ (blue) and $\mathscr{P} = \mathscr{P}_2$ (orange).
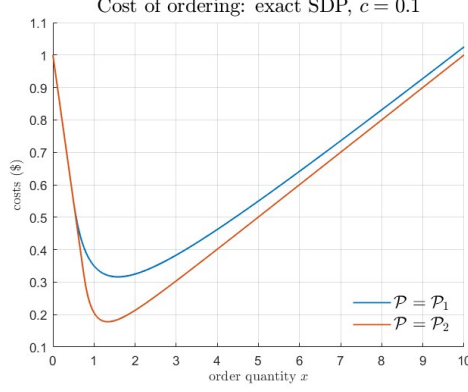
Figure 2: Costs of ordering (\$) for $\mathscr{P} = \mathscr{P}_1$ (blue) and $\mathscr{P} = \mathscr{P}_2$ (orange).

The curves in Figure 2 exhibit "U" shapes. Before the turning points ($x \in [0, 1.5811]$ blue, $x \in [0, 1.3337]$ orange), the more upfront orders at cost $c = \$0.1$ the company makes, the more it saves from back-ordering at cost \$1. For larger upfront ordering quantities ($x \in [1.5811, 10]$ blue, $x \in [1.3337, 10]$ orange) that can largely cover the demand, back-ordering is not needed. The slope of the curves in these regions is approximately $c = 0.1$.

Notice that $\mathscr{P}_2$ is a subset of $\mathscr{P}_1$, Figure 2 confirms that the costs of ordering for $\mathscr{P}_1$ (blue) are at least the costs for $\mathscr{P}_2$ (orange), and this leads to a higher minimum cost for $\mathscr{P}_1$. This could be justified that, by knowing more information about the demand, i.e., $\mathscr{P} = \mathscr{P}_2$, the future demand can be better predicted, and thus the ordering strategy could be better determined.

The minimum costs of ordering are achieved at the turning points of the curves. Specifically, the minimum costs are \$0.3162 and \$0.1778 if the company orders $x = 1.5811$ and $x = 1.3337$ units of goods under the ambiguity constraints $\mathscr{P} = \mathscr{P}_1$ (blue) and $\mathscr{P} = \mathscr{P}_2$ (orange), respectively.

For $\mathscr{P}_1$ with order quantity $x = 1.5811$, an optimal probability measure is $0.1\mathbb{1}_{3.1623} + 0.9\mathbb{1}_{0.0003}$. It refers to the discrete distribution with $\mathbb{P}(\omega = 3.1623) = 0.1$ and $\mathbb{P}(\omega = 0.0003) = 0.9$. For $\mathscr{P}_2$ with $x = 1.3337$, an optimal probability measure is $0.1\mathbb{1}_{1.7782} + 0.9\mathbb{1}_{0.0200}$.

Revenue Maximization. A merchant supplies a random quantity $\omega \in [0, R] \subseteq \mathbb{R}$ of goods and sells them to n customers. Each customer offers a different price based on the supply quantity. The goods can be sold to one customer exclusively at a time, and the merchant wishes to maximize revenue by selling the goods to the customer who offers the highest price.

This problem of revenue maximization can be formulated as

$$\max_{\mu \in \mathcal{P}_\Omega} \left\{ \mathbb{E}_\mu^\Omega \left[\max_{k=1,\dots,n} h_k(\omega) \right] : \mathbb{E}_\mu^\Omega[\omega] \leq \gamma_1, \mathbb{E}_\mu^\Omega[\omega^2] \leq \gamma_2 \right\}, \tag{25}$$

where h_k is the offer price from the k-th customer, $k = 1, \dots, n$, and $\Omega = [0, R]$, $R > 0$. The support can be expressed equivalently as a spectrahedron $\Omega = \{v \in \mathbb{R} : F_0 + vF_1 \succeq 0\}$ with $F_0 = \begin{pmatrix} 0 & 0 \\ 0 & R \end{pmatrix}$ and $F_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

The standard minimization form is

$$\min(26) := \min_{\mu \in \mathcal{P}_\Omega} \left\{ \mathbb{E}_\mu^\Omega[g(\omega)] : \mathbb{E}_\mu^\Omega[\omega] \leq \gamma_1, \mathbb{E}_\mu^\Omega[\omega^2] \leq \gamma_2 \right\}, \tag{26}$$

where g is given by $g(v) = -\max_{k=1,\dots,n} h_k(v)$. The maximum expected revenue is equal to $[-\min(26)]$.

Denoting $f_k(v) = -h_k(v)$, one has $g(v) = \min\limits_{k=1,\dots,n} f_k(v)$. Here, for each $k = 1, \dots, n$, we describe the offer price by

$$f_k(v) = \begin{cases} \alpha_k(v - b_k)^2 + \beta_k(v - b_k)^4 + c_k, & \text{if } 0 \leq v \leq b_k, \\ c_k, & \text{if } v > b_k, \end{cases} \tag{27}$$

where $\alpha_k, \beta_k, b_k \geq 0$ and $c_k \leq -\alpha_k b_k^2 - \beta_k b_k^4$. Note that g is a piecewise SOS-convex function with the following representation:

$$g(v) = \min\limits_{k=1,\dots,2n} \max\limits_{\ell=1,2} g_\ell^k(v), \tag{28}$$

where $g_1^k(v) = g_2^k(v) = \alpha_k(v - b_k)^2 + \beta_k(v - b_k)^4 + c_k$, $g_1^{n+k}(v) = -(\alpha_k b_k + \beta_k b_k^3)v + (\alpha_k b_k^2 + \beta_k b_k^4 + c_k)$, and $g_2^{n+k}(v) = c_k$, for $k = 1, \dots, n$. For details of this representation, see Appendix A.

In fact, $(-f_k)$ where f_k is given as in Eqn. (27), or h_k in problem (25), corresponds to a combined quadratic-quartic utility function that is concave, non-decreasing, and smooth [9]. The coefficients $(-\alpha_k)$ and $(-\beta_k)$ capture the rate at which the customer increases the offer price based on the supply; b_k is the maximum quantity of goods the k-th customer is willing to purchase, beyond which the customer will no longer be willing to increase the price; and c_k which excludes negative offer prices is the maximum price.

The function in Eqn. (27) reduces to the quadratic model in [13] by setting $\beta_k = 0$, $k = 1, \dots, n$, where an approximating upper bound for the maximum revenue is calculated through a convex program. Our model covers more diverse purchasing behaviours described by the quartic feature.

Suppose that there are three customers with parameters $\alpha_1 = 1$, $\alpha_2 = 1$, $\alpha_3 = \frac{1}{10}$, $\beta_1 = 1$, $\beta_2 = \frac{1}{16}$, $\beta_3 = \frac{1}{100}$, $b_1 = 1$, $b_2 = 2$, $b_3 = 4$, and $c_1 = -5$, $c_2 = -7$, $c_3 = -7.5$. The supply quantity ranges between $[0, R] = [0, 4]$. As shown in Figure 3, the piecewise SOS-convex function g is neither convex nor concave and is not smooth.
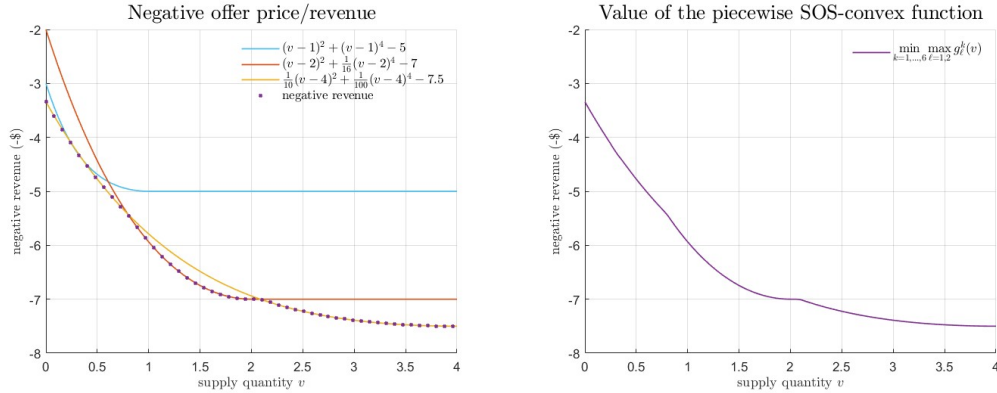

Figure 3: Left: negative offer prices (solid lines) and negative revenue (dotted lines). Right: piecewise SOS-convex function in Eqn. (28).

Associate (26) with the following SOS optimization problem with $n = 3$ customers,

$$\max_{\substack{\lambda \in \mathbb{R}_+^2 \times \mathbb{R}, \delta_k \in [0,1] \\ Z_k, Z_k' \in \mathbb{S}_+^2, k=1,\ldots,n}} - \lambda_1 \gamma_1 - \lambda_2 \gamma_2 - \lambda_3$$

$$\text{s.t.} \quad [\alpha_k b_k^2 + \beta_k b_k^4 + c_k + \lambda_3 - \text{tr}(Z_k F_0)] + [-2\alpha_k b_k - 4\beta_k b_k^3 + \lambda_1 - \text{tr}(Z_k F_1)]v +$$
$$[\alpha_k + 6\beta_k b_k^2 + \lambda_2]v^2 - 4\beta_k b_k v^3 + \beta_k v^4 \in \Sigma_4^2(v), \quad k = 1,\ldots,n,$$
$$[\delta_k \alpha_k b_k^2 + \delta_k \beta_k b_k^4 + c_k + \lambda_3 - \text{tr}(Z_k' F_0)] + [-\delta_k \alpha_k b_k - \delta_k \beta_k b_k^3 + \lambda_1 - \text{tr}(Z_k' F_1)]v +$$
$$\lambda_2 v^2 \in \Sigma_4^2(v), \quad k = 1,\ldots,n,$$

and its equivalent SDP representation

$$\max(29) := \max_{\substack{\lambda \in \mathbb{R}_+^2 \times \mathbb{R}, \delta_k \in [0,1] \\ Z_k, Z_k' \in \mathbb{S}_+^2, Q_k, Q_k' \in \mathbb{S}_+^3, k=1,\ldots,n}} - \lambda_1 \gamma_1 - \lambda_2 \gamma_2 - \lambda_3 \qquad (29)$$

$$\text{s.t.} \quad \text{for each } k = 1,\ldots,n,$$
$$\text{tr}(Q_k B_0) = \alpha_k b_k^2 + \beta_k b_k^4 + c_k + \lambda_3 - \text{tr}(Z_k F_0),$$
$$\text{tr}(Q_k B_1) = -2\alpha_k b_k - 4\beta_k b_k^3 + \lambda_1 - \text{tr}(Z_k F_1),$$
$$\text{tr}(Q_k B_2) = \alpha_k + 6\beta_k b_k^2 + \lambda_2,$$
$$\text{tr}(Q_k B_3) = -4\beta_k b_k, \ \text{tr}(Q_k B_4) = \beta_k,$$
$$\text{tr}(Q_k' B_0) = \delta_k \alpha_k b_k^2 + \delta_k \beta_k b_k^4 + c_k + \lambda_3 - \text{tr}(Z_k' F_0),$$
$$\text{tr}(Q_k' B_1) = -\delta_k \alpha_k b_k - \delta_k \beta_k b_k^3 + \lambda_1 - \text{tr}(Z_k' F_1),$$
$$\text{tr}(Q_k' B_2) = \lambda_2, \ \text{tr}(Q_k' B_3) = \text{tr}(Q_k' B_4) = 0,$$

where $B_0, \ldots, B_4$ are given in Eqn. (23).

Proposition 4.5. *Assume that* (26) *admits a minimizer and* $(\gamma_1, \gamma_2, 1) \in \text{int}\{\{(\mathbb{E}_\mu^\Omega[\omega], \mathbb{E}_\mu^\Omega[\omega^2], \mathbb{E}_\mu^\Omega[1]) : \mu \in X, \mu \succeq_\mathcal{B} 0\} + \mathbb{R}_+^2 \times \{0\}\}$. *Then,* $\min(26) = \max(29)$.

Proof. The proof is similar to Proposition 4.1. ☐

Similar to the Newsvendor application, higher-order moments can be incorporated, and the resulting SDP reformulation can be obtained.

The maximum revenue is equal to $[-\max(29)]$. When the first and second-order moments of the supply quantity are at most 2, i.e., $\gamma_1 = \gamma_2 = 2$, the maximum expected revenue is \$6.6495.

We recover an optimal measure via the SDP below,

$$\min_{\substack{y^k \in \mathbb{R}^5, z^k \in \mathbb{R} \\ k=1,\dots,2n}} \sum_{k=1}^{2n} z^k \tag{30}$$

$$\text{s.t.} \quad \sum_{k=1}^{2n} y_1^k \le \gamma_1, \ \sum_{k=1}^{2n} y_2^k \le \gamma_2, \ \sum_{k=1}^{2n} y_0^k = 1,$$

$$[\alpha_k b_k^2 + \beta_k b_k^4 + c_k]y_0^k - [2\alpha_k b_k + 4\beta_k b_k^3]y_1^k + [\alpha_k + 6\beta_k b_k^2]y_2^k -$$
$$4\beta_k b_k y_3^k + \beta_k y_4^k \le z^k, \ k = 1, \dots, n,$$

$$[\alpha_k b_k^2 + \beta_k b_k^4 + c_k]y_0^{n+k} - [\alpha_k b_k + \beta_k b_k^3]y_1^{n+k} \le z^{n+k}, \ k = 1, \dots, n,$$

$$c_k y_0^{n+k} \le z^{n+k}, \ k = 1, \dots, n,$$

$$y_0^k F_0 + y_1^k F_1 \succeq 0, \ \begin{pmatrix} y_0^k & y_1^k & y_2^k \\ y_1^k & y_2^k & y_3^k \\ y_2^k & y_3^k & y_4^k \end{pmatrix} \succeq 0, \ k = 1, \dots, 2n.$$

Solving (30) for $n = 3$ gives $\bar{y}^1 = (\bar{y}_0^1, \bar{y}_1^1, \bar{y}_2^1, \bar{y}_3^1, \bar{y}_4^1) = (0,0,0,0,0)$, $\bar{y}^2 = (1, 1.4142, 2, 2.8284, 4)$, $\bar{y}^3 = (0,0,0,0,0)$, $\bar{y}^4 = (0,0,0,0,1.5563)$, $\bar{y}^5 = (0,0,0,0,1.5871)$, $\bar{y}^6 = (0,0,0,0,1.4684)$, $\bar{z}^k = 0$, $k = 1, 3, 4, 5, 6$, and $\bar{z}^2 = -6.6495$. Clearly, $\bar{y}_0^k \ge 0$ and $K = \{k \in \{1, \dots, 6\} : \bar{y}_0^k > 0\} = \{2\}$. Moreover, $\bar{z}^k \ge 0$ for all $k \notin K$. For $h_1(v) = v$, it is satisfied that $\sum_{\alpha=0}^4 \bar{y}_\alpha^k (h_1)_\alpha = \bar{y}_1^k = 0$ for all $k \notin K$, and for $h_2(v) = v^2$, $\sum_{\alpha=0}^4 \bar{y}_\alpha^k (h_2)_\alpha = \bar{y}_2^k = 0$ for all $k \notin K$. By Theorem 3.6, an optimal probability measure for problem (26), and thus for (25) is $\mathbb{1}_{1.4142}$, which is the discrete distribution $\mathbb{P}(\omega = 1.4142) = 1$. This result can be interpreted by the fact that the maximum expected revenue \$6.6495 can be achieved when the merchant supplies 1.4142 units of goods.

We want to remark that the assumption $\bar{y}_0^k > 0$, $k = 1, \dots, r$, in Corollary 3.7 is sufficient but not necessary for an optimal solution to be $\sum_{k=1}^r \bar{y}_0^k \mathbb{1}_{\hat{u}^k}$. Indeed, with $r = 2n = 6$, the above revenue problem offers an example that $\bar{y}_0^k > 0$, $k = 1, \dots, r$, is not a necessary condition.

Figure 4 (left) shows how the maximum expected revenue changes with respect to the mean bound γ_1. The bound for the second-order moment is set to be $\gamma_2 = \gamma_1^2$. The curve exhibits an increasing trend. For a large mean bound γ_1, the company can target the customer offering a higher price, and thus the revenue is higher. Figure 4 (right), on the contrary, fixes $\gamma_1 = 2$ and increases γ_2. The curve exhibits an upward shape, which can be justified by the fact that the larger the second-order moment (and thus the variance), the wider the customers the company can sell products to, and thus the higher the revenue.
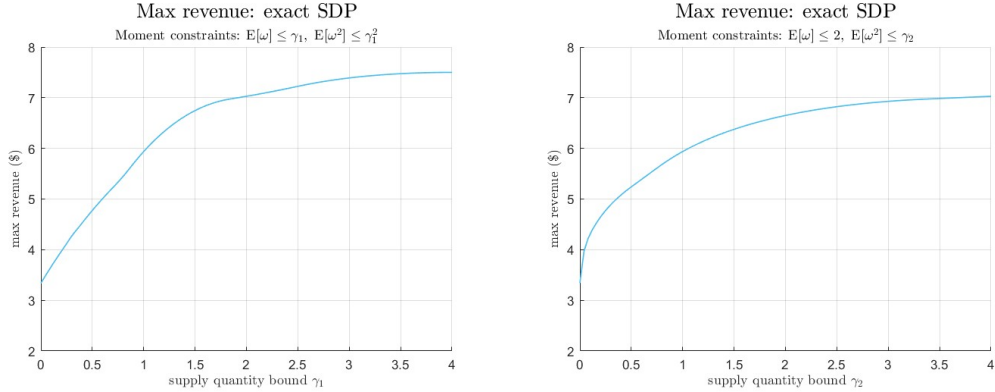
Figure 4: Maximum expected revenue by varying γ_1 (left) and γ_2 (right).

5 Conclusion and Outlook

We showed how to derive Sum-of-Squares (SOS) reformulations for important classes of infinite-dimensional moment optimization problems involving piecewise SOS-convex functions. We also showed how to recover an optimal probability measure from the associated Semi-Definite Program (SDP) reformulation.

It is worth noting that the class of piecewise SOS-convex functions as given in Definition 2.2 facilitates tractable representations of the corresponding moment problem (Theorem 3.4) and it is rich enough to cover a broad class of functions encountered in optimization. Our definition 2.2 of piecewise functions may be extended to more general settings where one piece of SOS-convex polynomial is given at one partition of $\mathbb{R}^m$, in line with the piecewise linear-quadratic function studied in [31, Definition 10.20]. It would be intriguing to explore the applicability and tractability of these generalized classes of piecewise functions.

Further, our method in Theorem 2.4 suggests that the numerically tractable SOS representation for non-negative piecewise SOS-convex functions over a projected spectrahedron can be extended to more general settings. For instance, it would be of interest to examine similar representations for piecewise SOS-convex functions over any non-convex sets whose convex hulls are semi-definite-representable, leading to SDP reformulations or relaxations for broad classes of optimization problems.

As applications, we presented numerical results for a class of generalized Newsvendor and revenue maximization problems with higher-order moments by solving their equivalent SDP reformulations. Our approach opens new avenues for further research, such as conic program reformulations for distributionally robust optimization problems [33, 36] involving piecewise SOS-convex functions, and applications to practical models such as the lot-sizing and product management problems in the face of uncertain conditions [2]. These problems will be examined in our forthcoming studies.

Acknowledgment

Declarations

Data availability statement. No data was used for the research described in the article.

Conflict of interest statement. The authors have no conflict of interest to declare that are relevant to the content of this article.

Appendix A   Piecewise SOS-Convex Representation

In this appendix, we present explicit piecewise SOS-convex representations for the examples mentioned in the Introduction (Section 1). We note that the representation is, in general, not unique.

- The truncated ℓ_1-norm, $p_\varepsilon(v) = \min\{1, \varepsilon|v|\}$, $\varepsilon > 0$, can be expressed as $p_\varepsilon(v) = \min_{k=1,2} \max_{\ell=1,2} g_\ell^k(v)$ where $g_1^1(v) = g_2^1(v) = 1, g_1^2(v) = \varepsilon v, g_2^2(v) = -\varepsilon v$. Note that the truncated ℓ_1-norm is non-convex and non-smooth. We plot $\max_{\ell=1,2} g_\ell^1(v)$ versus $\max_{\ell=1,2} g_\ell^2(v)$ in Figure 5 (left) and the truncated ℓ_1-norm in Figure 5 (right).
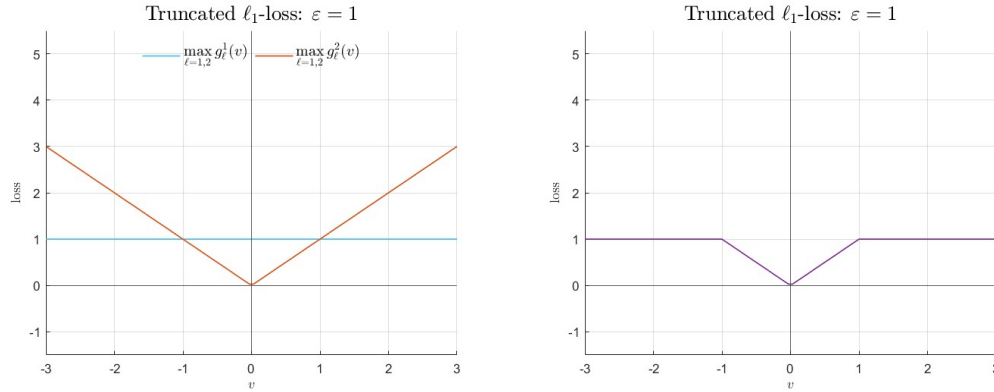

Figure 5: Piecewise representation (left) and truncated ℓ_1-norm (right).

- The piecewise quadratic function, $a, b \geq 0$, $c \in \mathbb{R}$,

$$f(v) = \begin{cases} a(v - b)^2 + c, & \text{if } 0 \leq v \leq b, \\ c, & \text{if } v > b, \end{cases}$$

can be written as $f(v) = \min_{k=1,2} \max_{\ell=1,2} g_\ell^k(v)$ where $g_1^1(v) = g_2^1(v) = a(v - b)^2 + c$, $g_1^2(v) = -abv + ab^2 + c$, and $g_2^2(v) = c$. We plot $\max_{\ell=1,2} g_\ell^1(v)$ versus $\max_{\ell=1,2} g_\ell^2(v)$ in Figure 6 (left) and the full piecewise function in Figure 6 (right).
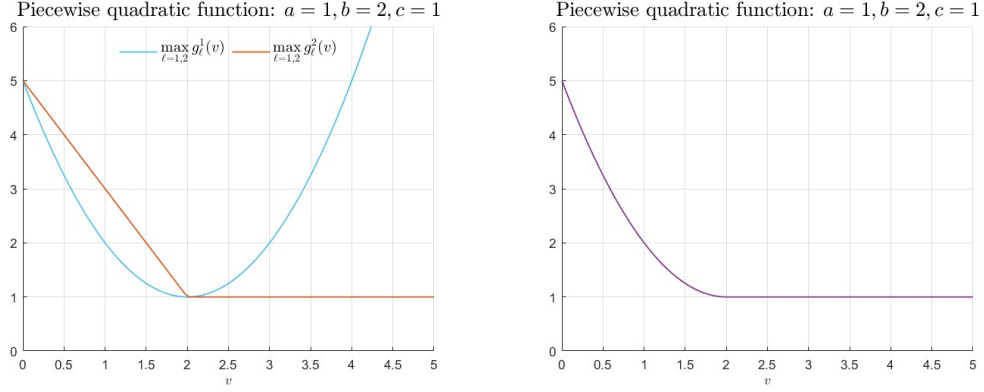
Figure 6: Piecewise representation (left) and the piecewise quadratic function (right).

- The piecewise quadratic-quartic function from Section 4,

$$f(v) = \begin{cases} \alpha(v - b)^2 + \beta(v - b)^4 + c, & \text{if } 0 \le v \le b, \\ c, & \text{if } v > b, \end{cases}$$

with $\alpha, \beta, b \ge 0$, $c \in \mathbb{R}$, can be expressed as $f(v) = \min_{k=1,2} \max_{\ell=1,2} g_\ell^k(v)$, where $g_1^1(v) = g_2^1(v) = \alpha(v - b)^2 + \beta(v - b)^4 + c$, $g_1^2(v) = -(\alpha b + \beta b^3)v + (\alpha b^2 + \beta b^4 + c)$, and $g_2^2(v) = c$. See Figure 7 for illustrations.
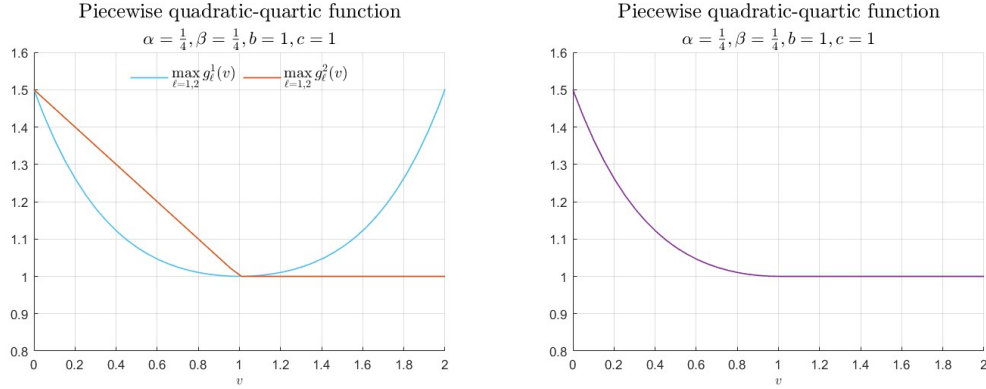

Figure 7: Piecewise representation (left) and the piecewise quadratic-quartic function (right).

- The Huber loss function [33] with parameter $\varepsilon > 0$,

$$H_\varepsilon(v) = \begin{cases} \frac{1}{2}v^2, & \text{if } |v| \le \varepsilon, \\ \varepsilon|v| - \frac{1}{2}\varepsilon^2, & \text{otherwise}, \end{cases}$$

can be expressed equivalently as $H_\varepsilon(v) = \min_{k=1,2} \max_{\ell=1,2,3,4} g_\ell^k(v)$, where $g_\ell^1(v) = \frac{1}{2}v^2$, $\ell = 1, \ldots, 4$, $g_1^2(v) = \varepsilon v - \frac{1}{2}\varepsilon^2$, $g_2^2(v) = -\varepsilon v - \frac{1}{2}\varepsilon^2$, $g_3^2(v) = \frac{1}{2}\varepsilon v$, and $g_4^2(v) = -\frac{1}{2}\varepsilon v$. See Figure 8 for illustrations.
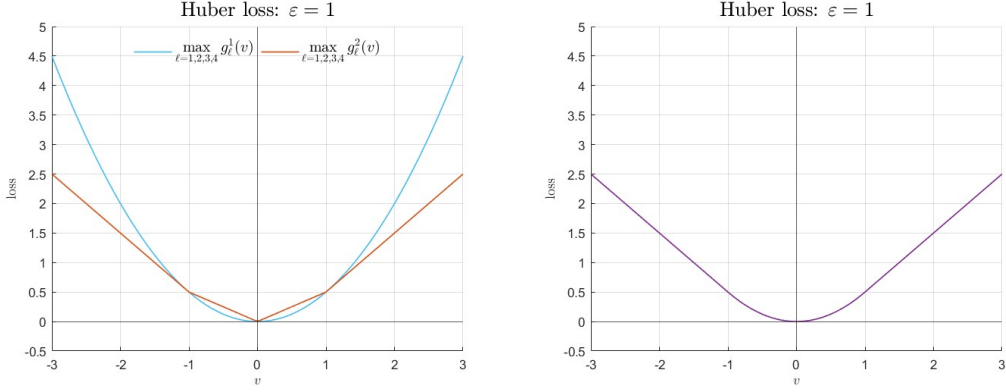
Figure 8: Piecewise representation (left) and Huber loss function (right).

Appendix B   Conic Duality in Topological Vector Spaces

This section presents duality results for infinite-dimensional conic linear programs that were used in Section 3.

Let X and X' be real topological vector spaces. They are paired if a bilinear form $\langle \cdot, \cdot \rangle : X' \times X \to \mathbb{R}$ is defined. For the linear map $A : X \to \mathbb{R}^J$, assume that for any $\lambda \in \mathbb{R}^J$, there is a unique $x' \in X'$ satisfying $\lambda^\top A x = \langle x', x \rangle$ for all $x \in X$. See [32, Assumption A1] for a discussion of this assumption. For the continuous linear map $A : X \to \mathbb{R}^J$, the adjoint mapping $A^* : \mathbb{R}^J \to X'$ is defined as $\langle A^* \lambda, x \rangle = \lambda^\top A x$ for any $x \in X, \lambda \in \mathbb{R}^J$. See [4, 35] for more details on the convexity of sets and functions.

We consider the problem

$$\min_{x \in X} \ \langle v, x \rangle \tag{CP}$$
$$\text{s.t. } x \in C, -Ax + b \in K,$$

where $C \subset X$ and $K \subset \mathbb{R}^J$ are convex cones that are closed in the respective topologies, $A : X \to \mathbb{R}^J$ is a linear map, and $v \in X', b \in \mathbb{R}^J$.

Associate (CP) with the set $D = \{x \in X : -Ax + b \in K\}$ and the dual problem

$$\max_{\lambda \in \mathbb{R}^J} \ -\lambda^\top b \tag{CD}$$
$$\text{s.t. } v + A^* \lambda \in C^+, \lambda \in K^+,$$

where $C^+ = \{x' \in X' : \langle x', x \rangle \geq 0, \forall x \in C\}$ and $K^+ = \{w \in \mathbb{R}^J : w^\top x \geq 0, \forall x \in K\}$.

We present a strong duality theorem under a more general constraint qualification known as the Generalized-Sharpened strong Conical Hull Intersection Property (G-S strong CHIP [5]). A pair of sets $\{C, D\}$ is said to satisfy the G-S strong CHIP at $x^* \in (C \cap D)$ whenever $(C \cap D - x^*)^+ = (C - x^*)^+ + \cup_{\lambda \in K^+} \{-A^* \lambda : \lambda^\top (-Ax^* + b) = 0\}$.

Theorem B.1 (Strong duality under G-S strong CHIP). *Suppose that x^* is a minimizer for the problem* (CP), *the linear form $\langle u, \cdot \rangle$ is continuous at x^* for each $u \in X'$, the linear map $A : X \to \mathbb{R}^J$ is continuous, the convex cones C and K are closed in the respective topologies, and that the pair $\{C, D\}$ satisfies the G-S strong CHIP at x^*. Then, $\min$ (CP) $= \max$ (CD), and the*

maximum of (CD) *is attained.*

Proof. Firstly, we note that, by construction, weak duality holds, that is, $\min(\mathrm{CP}) \geq \max(\mathrm{CD})$. To see $\min(\mathrm{CP}) \leq \max(\mathrm{CD})$, let x^* be a minimizer of (CP). By the necessary optimality conditions [35], $0 \in \partial(\langle v, \cdot \rangle)(x^*) - (C \cap D - x^*)^+ = \{v\} - (C \cap D - x^*)^+$. This means $v \in (C \cap D - x^*)^+$, and thus $v \in (C - x^*)^+ + \cup_{\lambda \in K^+}\{-A^*\lambda : \lambda^\top(-Ax^* + b) = 0\}$ by the G-S strong CHIP assumption. Therefore, there exist $u \in (C - x^*)^+$ and $\bar\lambda \in K^+$ satisfying $u = v + A^*\bar\lambda \in (C - x^*)^+$, $\bar\lambda^\top(-Ax^* + b) = 0$, and

$$\langle v + A^*\bar\lambda, x - x^* \rangle \geq 0, \quad \text{for all} \quad x \in C. \tag{31}$$

Since C is a convex cone, we have $\langle v + A^*\bar\lambda, x \rangle \geq 0$ for all $x \in C$, and so $v + A^*\bar\lambda \in C^+$. In particular, $\bar\lambda$ is feasible for (CD). Moreover, letting $x = 0 \in C$ in Eqn. (31) gives $\langle v + A^*\bar\lambda, x^* \rangle = \langle v, x^* \rangle + \bar\lambda^\top Ax^* \leq 0$. But $\bar\lambda^\top Ax^* = \bar\lambda^\top b$, which implies $\langle v, x^* \rangle + \bar\lambda^\top b \leq 0$. Hence, $-\bar\lambda^\top b \geq \langle v, x^* \rangle$, and so,

$$\max(\mathrm{CD}) = \max\{-\lambda^\top b : \lambda \in K^+, \ v + A^*\lambda \in C^+\} \geq -\bar\lambda^\top b \geq \langle v, x^* \rangle = \min(\mathrm{CP}).$$

Therefore, we see that $\min(\mathrm{CP}) = \max(\mathrm{CD})$ and the maximum of (CD) is attained at $\bar\lambda$. □

The G-S strong CHIP is the weakest constraint qualification guaranteeing strong (Lagrangian) duality for convex optimization. For details, see [5] and other references therein. As we see below, strong duality under the interior point condition [32] is a consequence of Theorem B.1.

Corollary B.2 (Strong duality under interior point condition). *Suppose that x^* is a minimizer for the problem* (CP)*, the linear form $\langle u, \cdot \rangle$ is continuous at x^* for each $u \in X'$, the linear map $A : X \to \mathbb{R}^J$ is continuous, the convex cones C and K are closed in the respective topologies. If $b \in \mathrm{int}(A(C) + K)$, then, $\min(\mathrm{CP}) = \max(\mathrm{CD})$, and the maximum of* (CD) *is attained.*

Proof. The conclusion will follow from Theorem B.1 if we show that $b \in \mathrm{int}(A(C) + K)$ implies

$$(C \cap D - x^*)^+ = (C - x^*)^+ + \cup_{\lambda \in K^+}\{-A^*\lambda : \lambda^\top(-Ax^* + b) = 0\}.$$

Note that the inclusion $(C - x^*)^+ + \cup_{\lambda \in K^+}\{-A^*\lambda : \lambda^\top(-Ax^* + b) = 0\} \subseteq (C \cap D - x^*)^+$ holds by construction. Conversely, let $u \in (C \cap D - x^*)^+$. By definition, this means $\langle u, x - x^* \rangle \geq 0$ for all $x \in (C \cap D)$. Hence, $\langle u, x \rangle \geq \langle u, x^* \rangle$ for all $x \in (C \cap D)$, and x^* is a minimizer of the convex optimization problem $\min\{\langle u, x \rangle : x \in C, -Ax + b \in K\}$. By [32], there exists $\lambda \in K^+$ with $\langle u, x^* \rangle = -\lambda^\top b$ and $A^*\lambda + u \in C^+$. Since $-Ax^* + b \in K$ and $\lambda \in K^+$, we have $\lambda^\top(-Ax^* + b) \geq 0$.

On the other hand, $\lambda^\top(-Ax^* + b) = \langle -A^*\lambda - u, x^* \rangle \leq 0$ as $x^* \in C$ and $A^*\lambda + u \in C^+$. These together force $\lambda^\top(-Ax^* + b) = 0$. In addition, $\langle A^*\lambda + u, x - x^* \rangle = \langle A^*\lambda + u, x \rangle \geq 0$ for all $x \in C$. This is equivalent to $\langle A^*\lambda + u, x \rangle \geq 0$ for all $x \in (C - x^*)$, which further implies $A^*\lambda + u \in (C - x^*)^+$. Hence, $u \in (C - x^*)^+ + \cup_{\lambda \in K^+}\{-A^*\lambda : \lambda^\top(-Ax^* + b) = 0\}$, and the proof is complete. □

Appendix C SDP Representation of SOS Problems

It is known that the SOS optimization problem (D) can be expressed equivalently as an SDP. This appendix provides technical results for Section 3.

A monomial over $v \in \mathbb{R}^m$ of degree $\bar{d}$ is $v^\alpha = v_1^{\alpha_1} v_2^{\alpha_2} \ldots v_m^{\alpha_m}$ with $\bar{d} = \sum_{i=1}^m \alpha_i$, and $\alpha = (\alpha_1, \ldots, \alpha_m) \in (\{0\} \cup \mathbb{N})^m$ is a multi-index. The canonical basis is

$$y(v) := (1, v_1, \ldots, v_m, v_1^2, v_1 v_2, \ldots, v_m^2, \ldots, v_1^{\bar{d}}, \ldots, v_m^{\bar{d}})^\top,$$

which is of dimension $s(m, \bar{d}) := \binom{m+\bar{d}}{\bar{d}}$.

Let f be a real polynomial with an even degree $d = 2\bar{d}$ written as $f(v) = \sum_{\alpha \in \mathcal{N}} (f)_\alpha v^\alpha$, where $(f)_\alpha$ is the α-th coefficient of f and $\mathcal{N} = \{\alpha \in (\{0\} \cup \mathbb{N})^m : \sum_{i=1}^m \alpha_i \le d\}$. By [24, Proposition 2.1], f is SOS if and only if there exists $Q \in \mathbb{S}_+^{s(m,d/2)}$ such that $f(v) = y(v)^\top Q y(v)$. This expresses the coefficients of $f(v)$ as linear equations of the entries in Q. If we write $y(v)y(v)^\top = \sum_{\alpha \in \mathcal{N}} B_\alpha v^\alpha$ for appropriate matrices $B_\alpha \in \mathbb{S}^{s(m,d/2)}, \alpha \in \mathcal{N}$, checking whether f is SOS amounts to finding $Q \in \mathbb{S}_+^{s(m,d/2)}$ such that $\mathrm{tr}(Q B_\alpha) = (f)_\alpha$ for all $\alpha \in \mathcal{N}$.

The following proposition shows that the SOS problem (D) and the SDP program (R) share the same optimal values.

Proposition C.1. Let g_ℓ^k, $\ell = 1, \ldots, L$, $k = 1, \ldots, r$, h_j, $j = 1, \ldots, J$, be defined as in Theorem 3.4. Then, $\max(\mathrm{D}) = \max(\mathrm{R})$.

Proof. The SOS constraints

$$\sum_{\ell=1}^L \delta_\ell^k g_\ell^k(v) + \sum_{j=1}^J \lambda_j h_j(v) + \lambda_{J+1} - \mathrm{tr}(Z_k F_0) - \sum_{i=1}^m v_i \mathrm{tr}(Z_k F_i) \in \Sigma_d^2(v), \ k = 1, \ldots, r,$$

of (D) are equivalent to the existence of $Q_k \in \mathbb{S}_+^{s(m,d/2)}$, $k = 1, \ldots, r$, such that

$$\sum_{\ell=1}^L \delta_\ell^k (g_\ell^k)_\alpha + \sum_{j=1}^J \lambda_j (h_j)_\alpha + \lambda_{J+1}(1)_\alpha - \mathrm{tr}(Z_k F_\alpha) = \mathrm{tr}(Q_k B_\alpha), \text{ for all } \alpha \in \mathcal{N}, \ k = 1, \ldots, r,$$

where $F_0 := F_0 \in \mathbb{S}^\nu$, $F_{e_i} := F_i \in \mathbb{S}^\nu$, $i = 1, \ldots, m$, and F_α, $\alpha \in \mathcal{N} \setminus \{0, e_1, \ldots, e_m\}$, is the zero matrix. Thus, the conclusion follows. $\qquad \square$

References

[1] A. A. Ahmadi and P. A. Parrilo. A complete characterization of the gap between convexity and SOS-convexity. *SIAM Journal on Optimization*, 23(2):811–833, 2013. https://arxiv.org/pdf/1111.4587.pdf.

[2] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*, volume 28. Princeton University Press, Princeton, 2009.

[3] D. Bertsimas and I. Popescu. On the relation between option and stock prices: a convex optimization approach. *Operations Research*, 50(2):358–374, 2002.

[4] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer Science & Business Media, New York, 2013.

[5] N. Chieu, V. Jeyakumar, G. Li, and H. Mohebi. Constraint qualifications for convex optimization without convexity of constraints: new connections and applications to best approximation. *European Journal of Operational Research*, 265(1):19–25, 2018.

[6] E. de Klerk, D. Kuhn, and K. Postek. Distributionally robust optimization with polynomial densities: theory, models and algorithms. *Mathematical Programming*, 181:265–296, 2020.

[7] E. de Klerk and M. Laurent. A survey of semidefinite programming approaches to the generalized problem of moments and their error analysis. In *World Women in Mathematics 2018*, pages 17–56, Switzerland, 2019. Springer.

[8] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

[9] H. U. Gerber and G. Pafum. Utility functions: from risk theory to finance. *North American Actuarial Journal*, 2(3):74–91, 1998.

[10] J. Goh and M. Sim. Distributionally robust optimization and its tractable approximations. *Operations Research*, 58(4):902–917, 2010.

[11] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.0 beta. `http://cvxr.com/cvx`, 2020. Accessed: 11 Aug 2023.

[12] J. Guo, S. He, B. Jiang, and Z. Wang. A unified framework for generalized moment problems: a novel primal-dual approach. *https://arxiv.org/pdf/2201.01445.pdf*, 2022.

[13] S. Han, M. Tao, U. Topcu, H. Owhadi, and R. M. Murray. Convex optimal uncertainty quantification. *SIAM Journal on Optimization*, 25(3):1368–1387, 2015.

[14] G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. A distributionally robust perspective on uncertainty quantification and chance-constrained programming. *Mathematical Programming*, 151:35–62, 2015.

[15] J. W. Helton and J. Nie. Structured semidefinite representation of some convex sets. In *2008 47th IEEE Conference on Decision and Control*, pages 4797–4800. IEEE, 2008.

[16] J. W. Helton and J. Nie. Semidefinite representation of convex sets. *Mathematical Programming*, 122:21–64, 2010.

[17] Q. Y. Huang and V. Jeyakumar. A distributional Farkas' lemma and moment optimization problems with no-gap dual semi-definite programs. *Optimization Letters*, pages 1–16, 2024.

[18] V. Jeyakumar, G. M. Lee, J. H. Lee, and Q. Y. Huang. Sum-of-squares relaxations in robust DC optimization and feature selection. *Journal of Optimization Theory and Applications*, 200(1):308–343, 2024.

[19] V. Jeyakumar and G. Li. Exact SDP relaxations for classes of nonlinear semidefinite programming problems. *Operations Research Letters*, 40(6):529–536, 2012.

[20] V. Jeyakumar, G. Li, and J. Vicente-Pérez. Robust SOS-convex polynomial optimization problems: exact SDP relaxations. *Optimization Letters*, 9:1–18, 2015.

[21] V. Jeyakumar and J. Vicente-Pérez. Dual semidefinite programs without duality gaps for a class of convex minimax programs. *Journal of Optimization Theory and Applications*, 162:735–753, 2014.

[22] J. B. Lasserre. A semidefinite programming approach to the generalized problem of moments. *Mathematical Programming*, 112:65–92, 2008.

[23] J. B. Lasserre. Convexity in semialgebraic geometry and polynomial optimization. *SIAM Journal on Optimization*, 19(4):1995–2014, 2009.

[24] J. B. Lasserre. *Moments, Positive Polynomials and Their Applications*, volume 1. World Scientific, London, 2009.

[25] J. B. Lasserre. *An Introduction to Polynomial and Semi-Algebraic Optimization*, volume 52. Cambridge University Press, Cambridge, 2015.

[26] H. A. Le Thi, X. T. Vo, and T. Pham Dinh. Feature selection for linear SVMs under uncertain data: robust optimization based on difference of convex functions algorithms. *Neural Networks*, 59:36–50, 2014.

[27] A. W. Lo. Semi-parametric upper bounds for option prices and expected payoffs. *Journal of Financial Economics*, 19(2):373–387, 1987.

[28] MOSEK ApS. The mosek optimization toolbox for matlab manual. version 10.1. `https://docs.mosek.com/10.0/toolbox/index.html#`, 2024. Accessed: 23 Apr 2024.

[29] T. Netzer and D. Plaumann. *Geometry of Linear Matrix Inequalities*. Springer, 2023.

[30] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

[31] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317. Springer Science & Business Media, 2009.

[32] A. Shapiro. On duality theory of conic linear problems. *Semi-Infinite Programming: Recent Advances*, pages 135–165, 2001.

[33] W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

[34] H. Xu, Y. Liu, and H. Sun. Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods. *Mathematical Programming*, 169:489–529, 2018.

[35] C. Zalinescu. *Convex Analysis in General Vector Spaces*. World Scientific, Singapore, 2002.

[36] J. Zhen, D. Kuhn, and W. Wiesemann. A unified theory of robust and distributionally robust optimization via the primal-worst-equals-dual-best principle. *Operations Research*, 2023.