



# Convergent Lifted Lasserre Hierarchy of SDPs for Minimizing Expectation of Piecewise Polynomial Loss over Wasserstein Balls

Neil De Vera Dizon<sup>1</sup> · Queenie Yingkun Huang<sup>1</sup> · Thai Doan Chuong<sup>2</sup> · Guoyin Li<sup>1</sup> · Vaithilingam Jeyakumar<sup>1</sup>

Received: 11 July 2025 / Accepted: 1 April 2026  
© The Author(s) 2026

## Abstract

This paper investigates the minimization of the expectation of piecewise polynomial loss functions over Wasserstein balls. This optimization problem often appears as a key sub-problem of distributionally robust optimization problems. We establish the asymptotic convergence of a hierarchy of semi-definite programming (SDP) relaxations, providing a framework for approximating the optimal values of these inherently infinite-dimensional optimization problems. A central foundational contribution is the development of a new lifted positivity certificate: we demonstrate that piecewise polynomials positive over Archimedean basic semi-algebraic sets admit a structured system of sum-of-squares (SOS) representations. Furthermore, we prove that the proposed hierarchy achieves finite convergence under suitable conditions when the

Communicated by Xiaoqi Yang.

Research of Dr. Neil Dizon was supported by a grant from the Australian Research Council. The research of Ms. Queenie Yingkun Huang was partially supported by a grant from UNSW Sydney. The research of Dr. Thai Doan Chuong was partially supported by his Brunel Research Initiative and Enterprise Fund (BRIEF) Award for 2025-2026. The research of Prof. Guoyin Li was supported by a grant from the Australian Research Council. The research of Prof. Vaithilingam Jeyakumar was supported by a grant from the Australian Research Council.

✉ Neil De Vera Dizon  
n.dizon@unsw.edu.au

Queenie Yingkun Huang  
yingkun.huang@unsw.edu.au

Thai Doan Chuong  
chuongthaidoan@yahoo.com; chuong.thaidoan@brunel.ac.uk

Guoyin Li  
g.li@unsw.edu.au

Vaithilingam Jeyakumar  
v.jeyakumar@unsw.edu.au

<sup>1</sup> Department of Applied Mathematics, University of New South Wales, Sydney 2052, Australia

<sup>2</sup> Department of Mathematics, Brunel University London, London, England

defining polynomials are convex. The practical utility and versatility of this approach are demonstrated via numerical experiments in revenue estimation and portfolio optimization.

**Keywords** Robust optimization · Distributionally robust optimization · Semi-definite linear programs · Sum-of-squares polynomials · Semi-infinite optimization

## 1 Introduction

Consider the expectation problem:

$$\min_{\mathbb{Q} \in \mathbb{B}_\varepsilon(\widehat{\mathbb{P}})} \mathbb{E}^{\mathbb{Q}} \left[ \min_{k=1, \dots, K} \max_{j=1, \dots, J} g_j^{(k)}(\boldsymbol{\xi}) \right] := \min_{\mathbb{Q} \in \mathbb{B}_\varepsilon(\widehat{\mathbb{P}})} \int_{\Xi} \min_{k=1, \dots, K} \max_{j=1, \dots, J} g_j^{(k)}(\boldsymbol{\xi}) \mathbb{Q}(d\boldsymbol{\xi}), \quad (\text{EP})$$

where each  $g_j^{(k)}$  is a polynomial and  $g(\boldsymbol{\xi}) := \min_{k=1, \dots, K} \max_{j=1, \dots, J} g_j^{(k)}(\boldsymbol{\xi})$  is termed as a *piecewise polynomial* loss. A piecewise polynomial is not necessarily a polynomial. Also, it is not necessarily a convex function even if  $g_j^{(k)}$ 's are all convex polynomials. The support set  $\Xi = \{\boldsymbol{\xi} \in \mathbb{R}^m : h_\ell(\boldsymbol{\xi}) \geq 0, \ell = 1, \dots, L\}$  is a basic semi-algebraic set. The set  $\mathbb{B}_\varepsilon(\widehat{\mathbb{P}}) = \{\mathbb{Q} \in \mathcal{P}(\Xi) : d_W(\mathbb{Q}, \widehat{\mathbb{P}}) \leq \varepsilon\}$  is the Wasserstein ball with  $\widehat{\mathbb{P}}$  the empirical distribution of past realizations  $\boldsymbol{\xi}_i \in \Xi, i = 1, \dots, N$  obtained from datasets.

Recent years have seen the emergence of the distributionally robust optimization (DRO) paradigm as a powerful framework for decision-making under uncertainty. At its core, the objective of DRO is to safeguard decisions against a range of potential probability distributions contained within a defined ambiguity set. *A fundamental component of this framework is the expectation problem, which serves as a key subproblem in many DRO formulations [30, 42].*

DRO offers distinct advantages over traditional methods. Unlike stochastic programming, which relies on a fixed distribution, DRO often demonstrates superior out-of-sample performance and enhanced computational tractability [30]. Simultaneously, it mitigates the inherent conservatism of deterministic robust optimization [5] by incorporating distributional information rather than focusing solely on worst-case scenarios.

**Motivation.** The piecewise polynomials appeared widely in DRO models as loss functions, although not all of the resulting DRO problems are numerically tractable. When  $J = 1$ , the case where  $g_1^{(k)}$ 's are linear was considered in [13], the case where  $g_1^{(k)}$ 's are quadratic functions was examined in [23], or more generally, the case where  $g_1^{(k)}$ 's are convex functions was studied in [17, 30, 44]. Moreover, when  $J \geq 1$ ,  $g_j^{(k)}$ 's can be SOS-convex polynomials [20], quadratic functions [21], or univariate polynomials [33].

Furthermore, decision-making in the face of uncertain conditions in a broad range of fields in finance, management, and economics can be modelled as an optimization

problem of the form (EP) with piecewise polynomials. For example, utility functions are frequently used in economics to quantify a person's risk preferences, and the piecewise power utility functions of order  $c = 2, 3, \dots$ , [15] defined as:

$$u(\xi) = \begin{cases} a(\xi - b)^c + d, & \text{if } 0 \leq \xi < b, \\ d, & \text{if } \xi \geq b, \end{cases}$$

where  $a \in \mathbb{R}$ ,  $b, d > 0$ , and  $\xi$  may represent wealth or consumption, can be expressed as a piecewise polynomial. Then, minimizing the negative of the power utility function can be cast into the form of (EP). The piecewise representation of the cubic utility function can be found in Section 5.1.

In portfolio optimization, the piecewise utility function  $\min_{k=1, \dots, K} \{a_k \xi^\top \mathbf{x} + b_k\}$ , where  $\xi$  and  $\mathbf{x}$  represent uncertain investment returns and fixed portfolio allocation respectively, is a special case of the DRO problem [13]. In addition, the risk measure component of conditional value-at-risk (CVaR) [36] and the expected shortfall [41] are both piecewise polynomials. Both risk measures have been applied in portfolio optimization [30, 41], and the CVaR is also applied in insurance management [38] and adversarial classification [19].

**Approach and methodology.** Optimization problems of the form (EP) are challenging infinite-dimensional optimization problems, and their numerical solvability depends on the structures of the loss function  $g$ , support set  $\Xi$ , and the ambiguity set  $\mathcal{P}$ . Various types of ambiguity sets have been studied in the DRO literature. The moment ambiguity sets are described by moment constraints using the mean and variance [13]. This approach is complementary to the probability distance-based ambiguity sets, which contain probability distributions that are “close enough” to a reference distribution [14].

A common approach for studying the problem (EP) is to employ duality techniques to convert it into a semi-infinite program [3, 16] that allows for the computation of solutions for problems with certain classes of loss functions of practical interest [30, 40, 44]. Consequently, computationally tractable optimization reformulations and approximation schemes have been proposed for DRO problems [13, 24, 26, 34, 40] in the sense that they can be expressed as efficiently solvable exact conic linear optimization problems or a convergent sequence of solvable conic linear problems.

**The Lasserre hierarchy of SDP relaxations.** Lasserre's semi-definite program (SDP) hierarchies have been applied in DRO and moment problems. Notably, [26] proposes a convergent hierarchy of SDPs to find the optimal value and optimal solution of a moment problem. Meanwhile, [31] presents a moment-sum-of-squares relaxation method for a class of moment-ambiguity DRO problems with a polynomial loss function. Also, [10] outlines an SDP hierarchy for a DRO problem where the ambiguity set consists of probability distributions with sum-of-squares density functions. For a survey of SDP approaches to moment problems, readers are referred to [12]. Further related results for min-max polynomial optimization can be found in [4, 25].

In this paper, we will show that asymptotic convergence of a lifted Lasserre hierarchy of SDPs holds for (EP), which can be used to approximate the optimal values of

the infinite-dimensional optimization problem (EP). This approach provides a computationally tractable method and algorithm for finding optimal values for various decision-making problems under uncertainty.

*Key contributions.* Our key contributions are itemized below:

- We present a new lifted sum-of-squares representation of positivity for piecewise polynomials, establishing that piecewise polynomials positive over an Archimedean basic semi-algebraic set admit a system of sum-of-squares representations.
- Building on this positivity representation, we propose a lifted hierarchy of semi-definite programs for the Wasserstein expectation problem of the form (EP) with piecewise polynomials. By establishing asymptotic convergence, we show that the hierarchy of semi-definite program relaxations can approximate the optimal values of these infinite-dimensional expectation problems.
- Exploiting the structure of piecewise convexity, we further show that the SDP hierarchy of (EP) achieves *finite convergence*, under suitable conditions, when all the defining polynomials are convex. This shows that finite convergence of hierarchies for convex polynomial optimization continues to hold for expectation problems with piecewise convex polynomials that are not necessarily convex. It is done by invoking the convex-concave minimax theorem and Lagrange duality of convex analysis [43] together with Scheiderer's sum-of-squares representation [37] of non-negative polynomials.
- Finally, to illustrate the versatility of our approach, we apply it to concrete models of expectation problems (EP) and distributionally robust decision-making models. We present results of a revenue estimation model with a piecewise polynomial utility function that is neither convex nor concave, exemplifying our SDP hierarchy. We also present results of numerical experiments, assessing the impact of increasing the relaxation level of our hierarchy on solution quality of a mean-CVaR portfolio optimization model.

*Originality.* This work exploits the piecewise structure of polynomials to represent a wide range of practical piecewise functions and extends the recent advances in moment-based DRO, from the piecewise sum-of-squares-convex functions [20] and piecewise quadratic functions [21] to Wasserstein DRO and general piecewise polynomials that are prevalent in a broad range of decision-making problems. A particular novelty of the present work lies in the derivation of a positivity certificate showing that piecewise polynomials that are positive over Archimedean basic semi-algebraic sets admit a lifted system of sum-of-squares representations. By combining techniques from the positivstellensatz of algebraic geometry [27] with the convex-concave minimax theorem of convex analysis [43], our approach enables the computation or estimation of the optimal values of otherwise infinite-dimensional expectation problems through SDPs, which can be efficiently solved using off-the-shelf software.

*Organization.* The organization of the paper is as follows. We recall preliminary materials on the Wasserstein ball and positive polynomials in Section 2. Section 3 presents the positivity representation for piecewise polynomials, from which we derive an asymptotically convergent SDP hierarchy for the expectation problem (EP). Section 4

discusses a finitely convergent SDP hierarchy for (EP). The versatility of our approach is illustrated in Section 5 through concrete models of expectation and distributionally robust decision-making problems. Finally, Section 6 concludes with future research directions.

## 2 Preliminaries

This section presents the notation and definitions used in the paper, and preliminary details on the Wasserstein ambiguity set and positive polynomials on a compact basic semi-algebraic set.

Denote  $\mathbb{R}^m$  the Euclidean space of dimension  $m$ ,  $\mathbb{R}_+^m$  the non-negative orthant of  $\mathbb{R}^m$ ,  $\mathbf{a}^\top \mathbf{b}$ ,  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$ , the standard inner product on  $\mathbb{R}^m$ , and  $\Delta = \{\boldsymbol{\delta} \in \mathbb{R}_+^m : \sum_{j=1}^m \delta_j = 1\}$  the simplex. Let  $\mathbf{e}_j$  be the  $j^{\text{th}}$  standard Euclidean basis vector,  $\mathbf{e}$  the vector of ones, and  $\mathbf{0}$  the vector of zeros in appropriate dimensions. For integers  $m, n$  with  $m \leq n$ , we define  $[m:n] := \{m, \dots, n\}$ .

Denote  $\mathbb{S}^m$  the space of  $(m \times m)$  real symmetric matrices. A matrix  $B \in \mathbb{S}^m$  is positive semi-definite, denoted as  $B \succeq 0$  (resp. positive definite, denoted as  $B \succ 0$ ), if  $\mathbf{x}^\top B \mathbf{x} \geq 0$  for all  $\mathbf{x} \in \mathbb{R}^m$  (resp.  $\mathbf{x}^\top B \mathbf{x} > 0$  for all  $\mathbf{x} \in \mathbb{R}^m, \mathbf{x} \neq \mathbf{0}$ ). Let  $\mathbb{S}_+^m$  be the cone of symmetric  $(m \times m)$  positive semi-definite matrices. We use lowercase letters for scalars, lowercase boldface letters for vectors, and capital letters for matrices.

Let  $X$  and  $X'$  be real topological vector spaces. The spaces  $X$  and  $X'$  are paired if a bilinear form  $\langle \cdot, \cdot \rangle : X' \times X \rightarrow \mathbb{R}$  is defined. The dual cone (positive polar cone) of a cone  $S \subseteq X$  is defined as  $S^* = \{x' \in X' : \langle x', x \rangle \geq 0, \forall x \in S\}$ .

Let  $\Xi$  be a compact set in  $\mathbb{R}^m$ . Denote by  $\mathcal{B}$  the Borel  $\sigma$ -algebra of  $\Xi$ . Let  $\mathcal{M}(\Xi)$  be the space of all (finite signed) regular Borel measures supported on  $\Xi$ . Let  $C(\Xi)$  be the space of continuous functions supported on  $\Xi$  equipped with the supremum norm.

We use  $\int_{\Xi} f(\boldsymbol{\xi}) \mathbb{P}(d\boldsymbol{\xi})$  to refer to the expectation of the random variable  $f(\boldsymbol{\xi})$  with respect to the measure  $\mathbb{P}$  supported on  $\Xi$ . We also denote by  $\mathcal{M}(\Xi)_+$  the convex cone of non-negative, finite, regular Borel measures supported on  $\Xi$ , that is,  $\mathcal{M}(\Xi)_+ = \{\mu \in \mathcal{M}(\Xi) : \mu \succeq_{\mathcal{B}} 0\}$ , where  $\mu \succeq_{\mathcal{B}} 0$  means that

*boldsymbolsymbol*  $\xi \in A = \mu(\xi \in A) \geq 0$  for all  $\mathcal{B}$ -measurable sets  $A$ . Let  $\mathcal{P}(\Xi)$  be the set of all probability measures supported on  $\Xi$ . The Dirac measure  $\mathbb{1}_{\boldsymbol{\xi}'}$  takes a mass of 1 at the point  $\boldsymbol{\xi}' \in \Xi$  and 0 otherwise, so  $\int_{\Xi} g(\boldsymbol{\xi}) \mathbb{1}_{\boldsymbol{\xi}'}(d\boldsymbol{\xi}) = g(\boldsymbol{\xi}')$ .

### 2.1 Wasserstein Ambiguity

Suppose that the empirical distribution is a discrete distribution given by  $\widehat{\mathbb{P}} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\widehat{\boldsymbol{\xi}}_i}$ , where  $\widehat{\boldsymbol{\xi}}_i \in \Xi, i \in [1:N]$ , are the past realizations obtained from datasets. Define the Wasserstein ambiguity set around the empirical distribution of radius  $\varepsilon \geq 0$  as:

$$\mathbb{B}_{\varepsilon}(\widehat{\mathbb{P}}) := \{\mathbb{Q} \in \mathcal{P}(\Xi) : d_W(\mathbb{Q}, \widehat{\mathbb{P}}) \leq \varepsilon\},$$

where  $d_W$  denotes the (order 2) Wasserstein distance given by

$$d_W(\mathbb{Q}, \mathbb{P}) = \min_{\Pi \in \mathcal{P}(\Xi \times \Xi)} \left\{ \left( \int_{\Xi \times \Xi} \|\xi - \xi'\|_2^2 \Pi(d\xi, d\xi') \right)^{1/2} : \begin{array}{l} \Pi \text{ is a joint distribution of } \xi \text{ and } \xi' \\ \text{with marginal distributions} \\ \mathbb{Q} \text{ and } \mathbb{P}, \text{ respectively} \end{array} \right\}$$

with  $\|\cdot\|_2$  the Euclidean norm in  $\mathbb{R}^m$ . This Wasserstein distance is also called the quadratic Wasserstein distance (cf. [32, Section 2.2]).

The Wasserstein distance of two distributions  $\mathbb{Q}$  and  $\mathbb{P}$  can be viewed as the minimum transportation cost for moving the probability mass from  $\mathbb{Q}$  to  $\mathbb{P}$ . The Wasserstein ambiguity set  $\mathbb{B}_\varepsilon(\widehat{\mathbb{P}})$  can be interpreted as the set of probability measures centred at the empirical distribution  $\widehat{\mathbb{P}}$  with radius  $\varepsilon \geq 0$ . The larger the radius, the richer the ambiguity set. On the other hand, if the radius  $\varepsilon$  shrinks to zero, the Wasserstein ball becomes the singleton set of the empirical distribution  $\{\widehat{\mathbb{P}}\}$ .

Consider the key expectation problem:

$$\min_{\mathbb{Q} \in \mathbb{B}_\varepsilon(\widehat{\mathbb{P}})} \mathbb{E}^{\mathbb{Q}}[g(\xi)], \tag{1}$$

where  $g : \mathbb{R}^m \rightarrow \mathbb{R}$  is a continuous function.

We associate with (1) the following semi-infinite program:

$$\max_{\lambda \geq 0, \alpha_i \in \mathbb{R}} \left\{ -\lambda \varepsilon^2 + \frac{1}{N} \sum_{i=1}^N \alpha_i : \inf_{\xi \in \Xi} \left( g(\xi) + \lambda \|\xi - \widehat{\xi}_i\|_2^2 - \alpha_i \right) \geq 0, \forall i \in [1:N] \right\}. \tag{ED}$$

Note that (ED) is a semi-infinite optimization problem because the decision variable space  $\mathbb{R} \times \mathbb{R}^N$  is finite-dimensional, but it can contain infinitely many non-negativity constraints  $g(\xi) + \lambda \|\xi - \widehat{\xi}_i\|_2^2 - \alpha_i \geq 0, \forall \xi \in \Xi$ , due to the (possibly) infinite set  $\Xi \subseteq \mathbb{R}^m$ .

The reduction from (1) over the infinite-dimensional space of probability measures to the semi-infinite problem (ED) is well-known in the literature. For example, [7, Theorem 1] provides a strong duality result when the ambiguity set is defined through an optimal transport cost more general than the  $\|\cdot\|_2^2$  in our setting. A strong duality result for a general Wasserstein distance is also noted in [14, Theorem 1]. We provide the detailed proof in the appendix for the sake of completeness.

**Lemma 2.1** (Reduction to finite-dimensional program [14]) *For the problem (1), let  $\Xi \subseteq \mathbb{R}^m$  be a compact set,  $g : \mathbb{R}^m \rightarrow \mathbb{R}$  be a continuous function, and  $\varepsilon > 0$ . Then,  $\min(1) = \max(ED)$ .*

*Proof* See Appendix A. □

### 2.2 Polynomials Positive on a Compact Basic Semi-algebraic Set

We recall some basic definitions and results of sum-of-squares polynomials. Let  $\mathbb{R}[\xi]$  be the space of polynomials with real coefficients over  $\xi \in \mathbb{R}^m$ . A polynomial  $f \in$

$\mathbb{R}[\xi]$  is called a Sum-of-Squares (SOS) polynomial if there exist polynomials  $f_i \in \mathbb{R}[\xi]$ ,  $i \in [1:n]$ , such that  $f = \sum_{i=1}^n f_i^2$ . We use  $\Sigma^2(\xi)$  to denote the set of all SOS polynomials  $f$  with respect to the variable  $\xi \in \mathbb{R}^m$ . We also use  $\Sigma_r^2(\xi)$  to denote the set of all SOS polynomials  $f$  of degree at most  $r$  with respect to the variable  $\xi \in \mathbb{R}^m$ . The SOS property is a computationally tractable relaxation of the non-negativity of polynomials, and certifying whether a polynomial is SOS amounts to solving a semi-definite program [28, Proposition 2.1].

In what follows, we assume that the support set  $\Xi \neq \emptyset$  is a compact basic semi-algebraic set, i.e., for  $\{h_\ell\}_{\ell=1}^L \subset \mathbb{R}[\xi]$ ,

$$\Xi = \{\xi \in \mathbb{R}^m : h_\ell(\xi) \geq 0, \forall \ell \in [1:L]\}. \tag{2}$$

From now on, we fix the basic semi-algebraic set  $\Xi$  as in (2), and the cone

$$\mathcal{Q}(h_1, \dots, h_L) := \left\{ f \in \mathbb{R}[\xi] : f = \sigma_0 + \sum_{\ell=1}^L \sigma_\ell h_\ell, \sigma_\ell \in \Sigma^2(\xi), \forall \ell \in [0:L] \right\} \tag{3}$$

is the quadratic module generated by  $\{h_\ell\}_{\ell=1}^L$ .

We also recall the following Archimedean condition [cf.29, Theorem2.16] and Putinar’s Positivstellensatz ([35], [cf.12, Theorem8]), which is a representation result for polynomials that are strictly positive on a compact basic semi-algebraic set.

**Assumption 2.2** (Archimedean condition) Let  $\{h_\ell\}_{\ell=1}^L \subset \mathbb{R}[\xi]$ ,  $\Xi$  defined in (2) be compact. The Archimedean condition states that there exists  $R \in \mathbb{N}$  such that  $R - \|\xi\|_2^2 \in \mathcal{Q}(h_1, \dots, h_L)$ .

**Proposition 2.3** (Putinar’s Positivstellensatz) Assume  $\Xi$  is as defined in (2) and Assumption 2.2 holds. If  $f \in \mathbb{R}[\xi]$  is strictly positive on  $\Xi$ , then  $f$  belongs to  $\mathcal{Q}(h_1, \dots, h_L)$ .

Another key result that we will need is the following non-negativity representation by Scheiderer [37, Example 3.18].

**Proposition 2.4** (Scheiderer’s non-negativity representation) Let  $h \in \mathbb{R}[\xi]$  be a polynomial for which the level set  $\Omega = \{\xi : h(\xi) \geq 0\}$  is compact. Let  $f(\xi)$  be non-negative on  $\Omega$ . Assume that the following conditions hold:

- (i)  $f$  has only finitely many zeros in  $\Omega$ , each lying in the interior of  $\Omega$ , and
- (ii) the Hessian of  $f$  is positive definite at each of the zeros in  $\Omega$ .

Then,  $f = \sigma_0 + \sigma_1 h$  for some  $\sigma_0, \sigma_1 \in \Sigma^2(\xi)$ .

### 3 Expectation Problems: Asymptotically Converging SDPs

This section presents a hierarchy of SDPs that approximates the expectation problem (EP), and its asymptotic convergence when the objective function involves a piecewise

polynomial (see Definition 3.1). This hierarchy is achieved by first establishing a sum-of-squares representation of a positive piecewise polynomial, which, in turn, guarantees the semi-infinite constraint in (ED). We begin by defining the notion of piecewise polynomials.

**Definition 3.1** (*Piecewise polynomial*) A function  $g : \mathbb{R}^m \rightarrow \mathbb{R}$  is called a piecewise polynomial if it can be expressed as

$$g(\xi) = \min_{k \in [1:K]} \max_{j \in [1:J]} g_j^{(k)}(\xi), \quad \xi \in \mathbb{R}^m,$$

for some polynomials  $g_j^{(k)} \in \mathbb{R}[\xi]$ .

This form of piecewise polynomial appears in a variety of real-world application settings, and it is, in general, not a polynomial. It extends the piecewise sum-of-squares-convex (SOS-convex) function in [20] where each  $g_j^{(k)}$  is a SOS-convex polynomial (see, e.g., [2, Definition 2.4] or [18, Section 1]), minimax quadratic in [21] where each  $g_j^{(k)}$  is a quadratic function, and piecewise affine functions.

### 3.1 Positivity Representation for Piecewise Polynomials

The goal of this subsection is to derive a system of sum-of-squares representations for the positivity of a piecewise polynomial over a compact semi-algebraic set.

**Theorem 3.2** (Lifted Putinar’s positivstellensatz for piecewise polynomials) *Let  $\Xi = \{\xi \in \mathbb{R}^m : h_\ell(\xi) \geq 0, \forall \ell \in [1:L]\}$  as in (2), and let the Archimedean condition in Assumption 2.2 hold for  $\{h_\ell\}_{\ell=1}^L$ . For each  $k \in [1:K]$ , let  $g_j^{(k)} \in \mathbb{R}[\xi]$ ,  $j \in [1:J]$ , and let  $\tau_1^{(k)}, \tau_2^{(k)} \in \mathbb{R}$  be such that*

$$\max_{\xi \in \Xi} \max_{j \in [1:J]} g_j^{(k)}(\xi) < \tau_1^{(k)} \quad \text{and} \quad \tau_2^{(k)} < \min_{\xi \in \Xi} \min_{j \in [1:J]} g_j^{(k)}(\xi). \tag{4}$$

Let  $f \in \mathbb{R}[\xi]$ . If

$$f(\xi) + \min_{k \in [1:K]} \max_{j \in [1:J]} g_j^{(k)}(\xi) > 0 \quad \text{for all } \xi \in \Xi,$$

then, for each  $k \in [1:K]$ , there exist  $\eta_j^{(k)} \in \Sigma^2(\xi, \tilde{\xi})$ ,  $j \in [1:(J+2)]$ ,  $\sigma_\ell^{(k)} \in \Sigma^2(\xi, \tilde{\xi})$ ,  $\ell \in [1:L]$ , such that

$$\psi + f - \sum_{\ell=1}^L \sigma_\ell^{(k)} h_\ell - \sum_{j=1}^{J+2} \eta_j^{(k)} \tilde{g}_j^{(k)} \in \Sigma^2(\xi, \tilde{\xi}), \tag{5}$$

where  $\psi(\xi, \tilde{\xi}) := \tilde{\xi}$  and

$$\tilde{g}_j^{(k)}(\xi, \tilde{\xi}) := \begin{cases} \tilde{\xi} - g_j^{(k)}(\xi), & \text{for } j \in [1:J], \\ \tau_1^{(k)} - \tilde{\xi} & \text{for } j = J + 1, \\ \tilde{\xi} - \tau_2^{(k)}, & \text{for } j = J + 2. \end{cases}$$

**Proof** This part follows from an application of Putinar’s Positivstellensatz to a compact lifted set obtained by augmenting  $\Xi$  with polynomial inequalities that encode upper and lower bounds on the functions  $g_j^{(k)}$  over  $\Xi$ .

Fix  $k \in [1:K]$ , then,

$$\min_{\xi \in \Xi} \left\{ f(\xi) + \max_{j \in [1:J]} g_j^{(k)}(\xi) \right\} > 0. \tag{6}$$

Under the Archimedean condition,  $\Xi$  is compact and so  $\tau_1^{(k)}, \tau_2^{(k)}$  in Eqn. (4) exist, and so,

$$\tau_1^{(k)} - g_j^{(k)}(\xi) > 0 \text{ and } g_j^{(k)}(\xi) - \tau_2^{(k)} > 0 \tag{7}$$

for all  $\xi \in \Xi$  and for all  $j \in [1:J]$ .

Now, consider the set

$$\tilde{\Xi}_k := \{(\xi, \tilde{\xi}) \in \mathbb{R}^m \times \mathbb{R} : \tilde{g}_j^{(k)}(\xi, \tilde{\xi}) \geq 0, \forall j \in [1:(J+2)], h_\ell(\xi) \geq 0, \forall \ell \in [1:L]\}, \tag{8}$$

where  $\tilde{g}_j^{(k)}$  are defined under Eqn. (5). Note that  $\tilde{\Xi}_k$  is non-empty as  $(\xi, \tau_1^{(k)}) \in \tilde{\Xi}_k$  for any  $\xi \in \Xi$ . Further, direct verification shows that

$$\begin{aligned} -\tilde{\xi}^2 &= \frac{[\tilde{\xi} - \tau_2^{(k)}]^2}{\tau_1^{(k)} - \tau_2^{(k)}} (\tau_1^{(k)} - \tilde{\xi}) + (\tau_1^{(k)} + \tau_2^{(k)})(\tau_1^{(k)} - \tilde{\xi}) + \frac{[\tau_1^{(k)} - \tilde{\xi}]^2}{\tau_1^{(k)} - \tau_2^{(k)}} (\tilde{\xi} - \tau_2^{(k)}) - (\tau_1^{(k)})^2 \\ &= \frac{[\tilde{g}_{J+2}^{(k)}(\xi, \tilde{\xi})]^2}{\tau_1^{(k)} - \tau_2^{(k)}} \tilde{g}_{J+1}^{(k)}(\xi, \tilde{\xi}) + (\tau_1^{(k)} + \tau_2^{(k)}) \tilde{g}_{J+1}^{(k)}(\xi, \tilde{\xi}) + \frac{[\tilde{g}_{J+1}^{(k)}(\xi, \tilde{\xi})]^2}{\tau_1^{(k)} - \tau_2^{(k)}} \tilde{g}_{J+2}^{(k)}(\xi, \tilde{\xi}) - (\tau_1^{(k)})^2, \end{aligned}$$

where the second equality follows by the definitions of  $\tilde{g}_{J+1}^{(k)}$  and  $\tilde{g}_{J+2}^{(k)}$ .

Since the Archimedean condition holds for  $\{h_\ell\}_{\ell=1}^L$ , there exist  $R_0 \in \mathbb{N}, \sigma_\ell \in \Sigma^2(\xi), \ell \in [0:L]$  such that  $R_0 - \|\xi\|_2^2 = \sigma_0(\xi) + \sum_{\ell=1}^L \sigma_\ell(\xi) h_\ell(\xi)$ . Choose  $R_1 \in \mathbb{N}$  such that  $R_1 \geq (\tau_1^{(k)})^2$ . This implies that

$$\begin{aligned} &R_0 + R_1 - \|(\xi, \tilde{\xi})\|_2^2 \\ &= R_0 - \|\xi\|_2^2 + R_1 - \tilde{\xi}^2 \\ &= \sigma_0(\xi) + \sum_{\ell=1}^L \sigma_\ell(\xi) h_\ell(\xi) + \frac{[\tilde{g}_{J+2}^{(k)}(\xi, \tilde{\xi})]^2}{\tau_1^{(k)} - \tau_2^{(k)}} \tilde{g}_{J+1}^{(k)}(\xi, \tilde{\xi}) + (\tau_1^{(k)} + \tau_2^{(k)}) \tilde{g}_{J+1}^{(k)}(\xi, \tilde{\xi}) \\ &\quad + \frac{[\tilde{g}_{J+1}^{(k)}(\xi, \tilde{\xi})]^2}{\tau_1^{(k)} - \tau_2^{(k)}} \tilde{g}_{J+2}^{(k)}(\xi, \tilde{\xi}) + (R_1 - (\tau_1^{(k)})^2), \end{aligned}$$

which means that the Archimedean condition also holds for  $(\{h_\ell\}_{\ell=1}^L \cup \{\tilde{g}_{J+1}^{(k)}, \tilde{g}_{J+2}^{(k)}\}) \subseteq \mathbb{R}[(\xi, \tilde{\xi})]$ , and thus for  $(\{h_\ell\}_{\ell=1}^L \cup \{\tilde{g}_j^{(k)}\}_{j=1}^{J+2}) \subseteq \mathbb{R}[(\xi, \tilde{\xi})]$ .

Consider the function  $\varphi(\xi, \tilde{\xi}) := \xi + f(\xi)$  for  $(\xi, \tilde{\xi}) \in \mathbb{R}^m \times \mathbb{R}$ . Under the validation of Eqn. (6), we can verify that  $\varphi(\xi, \tilde{\xi}) > 0$  for all  $(\xi, \tilde{\xi}) \in \tilde{\Xi}_k$ . Therefore, by Proposition 2.3, there exist  $\sigma_\ell^{(k)} \in \Sigma^2(\xi, \tilde{\xi})$ ,  $\ell \in [0:L]$ ,  $\eta_j^{(k)} \in \Sigma^2(\xi, \tilde{\xi})$ ,  $j \in [1:(J+2)]$  such that

$$\varphi = \sigma_0^{(k)} + \sum_{\ell=1}^L \sigma_\ell^{(k)} h_\ell + \sum_{j=1}^{J+2} \eta_j^{(k)} \tilde{g}_j^{(k)},$$

or in other words,

$$\sigma_0^{(k)} = \varphi - \sum_{\ell=1}^L \sigma_\ell^{(k)} h_\ell - \sum_{j=1}^{J+2} \eta_j^{(k)} \tilde{g}_j^{(k)} \in \Sigma^2(\xi, \tilde{\xi}),$$

arriving at the conclusion. □

**Remark 3.3** While the above positivity representation is established using the celebrated Putinar’s Positivstellensatz, it provides a flexible and convenient tool for deriving the asymptotic convergence of the SDP hierarchy used to compute the optimal value of the expectation problem with piecewise polynomial loss functions over Wasserstein balls.

In passing, we also note that by letting  $K = J = 1$ ,  $g_1^{(1)} \equiv 0$  in Theorem 3.2 and setting  $\tilde{\xi} = 0$ ,

$$\tau_1^{(1)} = 1 > \max_{\xi \in \Xi} g_1^{(1)}(\xi) = 0 \quad \text{and} \quad \tau_2^{(1)} = -1 < \min_{\xi \in \Xi} g_1^{(1)}(\xi) = 0,$$

the conclusion of Theorem 3.2 collapses to the one in the classical Putinar’s Positivstellensatz (see Proposition 2.3).

The next subsection moves to presenting SDP hierarchies for the expectation problems.

### 3.2 Converging Hierarchy of SDPs

We first recall the expectation problem:

$$\min_{\mathbb{Q} \in \mathbb{B}_\varepsilon(\hat{\mathbb{P}})} \mathbb{E}^{\mathbb{Q}} \left[ \min_{k \in [1:K]} \max_{j \in [1:J]} g_j^{(k)}(\xi) \right]. \tag{EP}$$

The support set  $\Xi = \{\xi \in \mathbb{R}^m : h_\ell(\xi) \geq 0, \forall \ell \in [1:L]\}$  is a basic semi-algebraic set as in equation (2). Recall that  $\mathbb{B}_\varepsilon(\hat{\mathbb{P}}) = \{\mathbb{Q} \in \mathcal{P}(\Xi) : d_W(\mathbb{Q}, \hat{\mathbb{P}}) \leq \varepsilon\}$  is the Wasserstein ambiguity set with  $\hat{\mathbb{P}} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\hat{\xi}_i}$  the empirical distribution of past realizations  $\hat{\xi}_i \in \Xi, i \in [1:N]$  obtained from datasets. The basic semi-algebraic support

set bound by inequalities has also appeared in [26, 31]. It covers some of the most practical support sets, such as the ellipsoid [13] and the box [30].

Throughout this subsection, we assume that the Archimedean condition (Assumption 2.2) holds for  $\{h_\ell\}_{\ell=1}^L$ . This implies that  $\Xi$  is compact, and so  $\tau_1^{(k)}$  and  $\tau_2^{(k)}$  as in (4) exist for each  $k \in [1:K]$ .

We associate (EP) with the following hierarchy of SDP programs, for a given  $r \in \mathbb{N}$ :

$$\begin{aligned} \max_{\substack{\lambda \geq 0, \alpha_i \in \mathbb{R} \\ \eta_{j,i}^{(k)}, \sigma_{\ell,i}^{(k)} \in \Sigma^2}} & -\lambda \varepsilon^2 + \frac{1}{N} \sum_{i=1}^N \alpha_i & (\text{AD}_r) \\ \text{s.t.} & \psi + \lambda \varphi_i - \alpha_i - \sum_{\ell=1}^L \sigma_{\ell,i}^{(k)} h_\ell - \sum_{j=1}^{J+2} \eta_{j,i}^{(k)} \tilde{g}_j^{(k)} \in \Sigma_{2r}^2(\xi, \tilde{\xi}), \quad \forall k \in [1:K], i \in [1:N], \\ & \deg(\sigma_{\ell,i}^{(k)} h_\ell) \leq 2r, \quad \deg(\eta_{j,i}^{(k)} \tilde{g}_j^{(k)}) \leq 2r, \quad \forall k \in [1:K], \ell \in [1:L], j \in [1:(J+2)], i \in [1:N], \end{aligned}$$

where  $\varepsilon > 0$ ,  $\psi(\xi, \tilde{\xi}) = \tilde{\xi}$ ,  $\varphi_i(\xi, \tilde{\xi}) = \|\xi - \widehat{\xi}_i\|_2^2$ ,  $\tilde{g}_j^{(k)}(\xi, \tilde{\xi}) = \tilde{\xi} - g_j^{(k)}(\xi)$  for each  $j \in [1:J]$ ,  $\tilde{g}_{J+1}^{(k)}(\xi, \tilde{\xi}) = \tau_1^{(k)} - \tilde{\xi}$  and  $\tilde{g}_{J+2}^{(k)}(\xi, \tilde{\xi}) = \tilde{\xi} - \tau_2^{(k)}$ . Here,  $\widehat{\xi}_i \in \Xi, i \in [1:N]$ , are the past realizations obtained from datasets as given in the definition of the Wasserstein ambiguity set.

We let  $\max(\text{AD}_r) = -\infty$  if its feasibility region is empty. We know that the feasibility region enlarges as the step  $r$  increases, and so the feasibility region will eventually become non-empty.

While the constraints in (AD<sub>r</sub>) are written as sum-of-squares constraints, they admit an equivalent representation as linear matrix inequalities [29, Proposition 2.1], and so, (AD<sub>r</sub>) can be written as a semi-definite program. Accordingly, we refer to our hierarchy as a hierarchy of SDP programs.

**Theorem 3.4** (Asymptotic convergence) *For the problem (EP) under the Wasserstein ambiguity set  $\mathbb{B}_\varepsilon(\mathbb{P})$  and support set  $\Xi$  in (2), let  $\Xi$  be a convex set; let the Archimedean condition (Assumption 2.2) hold; and let  $\varepsilon > 0$ . Then,*

$$\max(\text{AD}_r) \leq \max(\text{AD}_{r+1}) \leq \min(\text{EP})$$

for all  $r \in \mathbb{N}$ . Moreover,  $\lim_{r \rightarrow \infty} \max(\text{AD}_r) = \min(\text{EP})$ .

**Proof** Since the Archimedean condition (Assumption 2.2) holds for  $\{h_\ell\}_{\ell=1}^L$ ,  $\Xi$  is compact. We first observe that, by construction,  $\max(\text{AD}_r) \leq \max(\text{AD}_{r+1})$  for each given  $r \in \mathbb{N}$ .

Next, we show  $\max(\text{AD}_r) \leq \min(\text{EP})$  for any given  $r \in \mathbb{N}$ . To see this, note from Lemma 2.1 with  $g = \min_{k \in [1:K]} \max_{j \in [1:J]} g_j^{(k)}$  that  $\min(\text{EP}) = \max(9)$ , where (9) is given by

$$\max_{\lambda \geq 0, \alpha_i \in \mathbb{R}} -\lambda \varepsilon^2 + \frac{1}{N} \sum_{i=1}^N \alpha_i \tag{9}$$

$$\text{s.t. } \min_{\xi \in \Xi} \left\{ \lambda \|\xi - \widehat{\xi}_i\|_2^2 - \alpha_i + \min_{k \in [1:K]} \max_{j \in [1:J]} g_j^{(k)}(\xi) \right\} \geq 0, \quad \forall i \in [1:N].$$

Thus, the conclusion follows if one shows that  $\max(\text{AD}_r) \leq \max(9)$  for any given  $r \in \mathbb{N}$ . Now, take any feasible point  $\lambda \geq 0, \alpha_i \in \mathbb{R}, \eta_{j,i}^{(k)}, \sigma_{\ell,i}^{(k)} \in \Sigma_{2r}^2(\xi, \tilde{\xi})$  from the problem  $(\text{AD}_r)$ , then,

$$\tilde{\xi} + \lambda \|\xi - \widehat{\xi}_i\|_2^2 - \alpha_i - \sum_{\ell=1}^L \sigma_{\ell,i}^{(k)}(\xi, \tilde{\xi}) h_{\ell}(\xi) - \sum_{j=1}^{J+2} \eta_{j,i}^{(k)}(\xi, \tilde{\xi}) \tilde{g}_j^{(k)}(\xi, \tilde{\xi}) \geq 0,$$

for all  $(\xi, \tilde{\xi}) \in \tilde{\Xi}_k, k \in [1:K], i \in [1:N]$ , where  $\tilde{\Xi}_k$  is defined in Eqn. (8).

Fix an index  $k \in [1:K], i \in [1:N]$ . The non-negativity over  $\tilde{\Xi}_k$  means that  $\tilde{\xi} + \lambda \|\xi - \widehat{\xi}_i\|_2^2 - \alpha_i \geq 0$  for all  $(\xi, \tilde{\xi}) \in \tilde{\Xi}_k$ . Because  $(\xi, \max_{j \in [1:J]} g_j^{(k)}(\xi)) \in \tilde{\Xi}_k$  for any  $\xi \in \Xi$ , so

$$\max_{j \in [1:J]} g_j^{(k)}(\xi) + \lambda \|\xi - \widehat{\xi}_i\|_2^2 - \alpha_i \geq 0$$

for all  $\xi \in \Xi$ . Since this is true for any  $k \in [1:K]$ , so  $\lambda \|\xi - \widehat{\xi}_i\|_2^2 - \alpha_i + \min_{k \in [1:K]} \max_{j \in [1:J]} g_j^{(k)}(\xi) \geq 0$  for all  $\xi \in \Xi, i \in [1:N]$ . This implies that the chosen  $\lambda \geq 0, \alpha_i \in \mathbb{R}$  is feasible for the problem (9). As the problems (9) and  $(\text{AD}_r)$  share the same objective function, one has  $\max(\text{AD}_r) \leq \max(9)$ . Therefore, for any  $r \in \mathbb{N}$ ,

$$\max(\text{AD}_r) \leq \max(\text{AD}_{r+1}) \leq \max(9) = \min(\text{EP}). \tag{10}$$

To prove the asymptotic convergence, we first notice from Eqn. (10) that

$$\limsup_{r \rightarrow \infty} \max(\text{AD}_r) \leq \max(9) = \min(\text{EP}).$$

So, the asymptotic convergence will follow if one shows that  $\liminf_{r \rightarrow \infty} \max(\text{AD}_r) \geq \max(9)$ . To see this, take an arbitrary  $\gamma > 0$ . From the attainment of the problem (9), there exist  $\lambda^* \geq 0$  and  $\alpha_i^* \in \mathbb{R}, i \in [1:N]$ , such that

$$\max(9) - \frac{\gamma}{2} \leq -\lambda^* \varepsilon^2 + \frac{1}{N} \sum_{i=1}^N \alpha_i^*$$

and

$$\min_{\xi \in \Xi} \left\{ \lambda^* \|\xi - \widehat{\xi}_i\|_2^2 - \alpha_i^* + \min_{k \in [1:K]} \max_{j \in [1:J]} g_j^{(k)}(\xi) \right\} \geq 0, \quad \forall i \in [1:N].$$

This implies that  $\max(9) - \gamma \leq -\lambda^* \varepsilon^2 + \frac{1}{N} \sum_{i=1}^N (\alpha_i^* - \frac{\gamma}{2})$  and

$$\min_{\xi \in \Xi} \left\{ \lambda^* \|\xi - \widehat{\xi}_i\|_2^2 - (\alpha_i^* - \frac{\gamma}{2}) + \min_{k \in [1:K]} \max_{j \in [1:J]} g_j^{(k)}(\xi) \right\} > 0, \quad \forall i \in [1:N]. \tag{11}$$

For each  $i \in [1:N]$  and  $k \in [1:K]$ , by Theorem 3.2, Eqn. (11) implies that there exist

$$\eta_{j,i}^{(k)} \in \Sigma^2(\xi, \tilde{\xi}), \quad j \in [1:(J+2)], \quad \sigma_{\ell,i}^{(k)} \in \Sigma^2(\xi, \tilde{\xi}), \quad \ell \in [1:L],$$

such that

$$\psi + \lambda^* \varphi_i - \left(\alpha_i^* - \frac{\gamma}{2}\right) - \sum_{\ell=1}^L \sigma_{\ell,i}^{(k)} h_\ell - \sum_{j=1}^{J+2} \eta_{j,i}^{(k)} \tilde{g}_j^{(k)} \in \Sigma_{2r_0}^2(\xi, \tilde{\xi}),$$

for some  $r_0 \in \mathbb{N}$  with  $\deg(\sigma_{\ell,i}^{(k)} h_\ell) \leq 2r_0, \ell \in [1:L]$  and  $\deg(\eta_{j,i}^{(k)} \tilde{g}_j^{(k)}) \leq 2r_0, j \in [1:(J+2)]$ . Thus, for each  $\gamma > 0$ , there exists  $r_0 \in \mathbb{N}$  such that  $(\lambda^*, \alpha_i^* - \frac{\gamma}{2}, \eta_{j,i}^{(k)}, \sigma_{\ell,i}^{(k)})$  is feasible for  $(AD_{r_0})$ , which implies that  $\max(9) - \gamma \leq \max(AD_{r_0})$ . So,  $\liminf_{r \rightarrow \infty} \max(AD_r) \geq \max(9)$ , and hence  $\lim_{r \rightarrow \infty} \max(AD_r) = \max(9) = \min(EP)$ .  $\square$

When  $J = 1$ , the objective function takes the form  $g(\xi) = \min_{k \in [1:K]} g_1^{(k)}(\xi)$ , where  $g_1^{(k)}$ 's are polynomials. Using Theorem 3.2, the SOS constraints are simplified, and the hierarchy of SDP programs in  $(AD_r)$  becomes:

$$\begin{aligned} \max_{\lambda \geq 0, \alpha_i \in \mathbb{R}, \sigma_{\ell,i}^{(k)} \in \Sigma^2} & -\lambda \varepsilon^2 + \frac{1}{N} \sum_{i=1}^N \alpha_i & (\widetilde{AD}_r) \\ \text{s.t.} & g_1^{(k)} + \lambda \varphi_i - \alpha_i - \sum_{\ell=1}^L \sigma_{\ell,i}^{(k)} h_\ell \in \Sigma_{2r}^2(\xi), \quad \forall k \in [1:K], \quad i \in [1:N], \\ & \deg(\sigma_{\ell,i}^{(k)} h_\ell) \leq 2r, \quad \forall k \in [1:K], \quad \ell \in [1:L], \quad i \in [1:N], \end{aligned}$$

for a given  $r \in \mathbb{N}$  with  $\varepsilon > 0$  and  $\varphi_i(\xi) = \|\xi - \widehat{\xi}_i\|_2^2$ .

**Corollary 3.5** (Asymptotic convergence for min-polynomials) *For the problem (EP), let  $\Xi = \{\xi \in \mathbb{R}^m : h_\ell(\xi) \geq 0, \forall \ell \in [1:L]\}$  as in (2); let Assumption 2.2 hold; let  $J = 1$  and  $g_1^{(k)} : \mathbb{R}^m \rightarrow \mathbb{R}$  be a polynomial for each  $k \in [1:K]$ ; and let  $\varepsilon > 0$ . Then,*

$$\max(\widetilde{AD}_r) \leq \max(\widetilde{AD}_{r+1}) \leq \min(EP)$$

for all  $r \in \mathbb{N}$ . Moreover,  $\lim_{r \rightarrow \infty} \max(\widetilde{AD}_r) = \min(EP)$ .

**Proof** As in the proof of Theorem 3.4,  $\max(\widetilde{AD}_r) \leq \max(ED)$  for any given degree  $r \in \mathbb{N}$ . This follows from the fact that, for each  $i \in [1:N]$ ,

$$g_1^{(k)} + \lambda \varphi_i - \alpha_i - \sum_{\ell=1}^L \sigma_{\ell,i}^{(k)} h_\ell \in \Sigma_{2r}^2(\xi), \quad \forall k \in [1:K],$$

which, in turn, implies that

$$\lambda \|\xi - \widehat{\xi}_i\|_2^2 - \alpha_i + \min_{k \in [1:K]} g_1^{(k)}(\xi) \geq 0, \forall \xi \in \Xi,$$

for any  $\lambda \geq 0, \alpha_i \in \mathbb{R}$ .

Hence,  $\max(\widetilde{AD}_r) \leq \max(\widetilde{AD}_{r+1}) \leq \max(\mathbf{ED}) = \min(\mathbf{EP})$  holds, and the convergence result follows the same line of arguments as in the proof of asymptotic convergence of Theorem 3.4.  $\square$

### 4 Piecewise Convex Polynomials and Finite Convergence

Recall that we are considering the setting where the loss function is given by

$$g(\xi) = \min_{k \in [1:K]} \max_{j \in [1:J]} g_j^{(k)}(\xi),$$

and the support set takes the form  $\Xi = \{\xi \in \mathbb{R}^m : h_\ell(\xi) \geq 0, \forall \ell \in [1:L]\}$  as defined in Eqn. (2).

In this section, we assume that  $g_j^{(k)}, j \in [1:J], k \in [1:K]$ , and  $(-h_\ell), \ell \in [1:L]$ , are convex polynomials, and show that the hierarchy for the problem (EP) exhibits finite convergence. It is worth noting that a piecewise convex polynomial is not necessarily a convex function.

We associate with (EP) the following hierarchy of SDP programs:

$$\begin{aligned} & \max_{\substack{\lambda \geq 0, \alpha_i \in \mathbb{R} \\ \delta_i^{(k)} \in \Delta, \sigma_{\ell,i}^{(k)} \in \Sigma_{2r}^2(\xi)}} -\lambda \varepsilon^2 + \frac{1}{N} \sum_{i=1}^N \alpha_i & (\mathbf{FD}_r) \\ \text{s.t.} & \sum_{j=1}^J \delta_{j,i}^{(k)} g_j^{(k)} + \lambda \varphi_i - \alpha_i - \sum_{\ell=1}^L \sigma_{\ell,i}^{(k)} h_\ell \in \Sigma_{2r}^2(\xi), \quad \forall k \in [1:K], i \in [1:N], \\ & \deg(\sigma_{\ell,i}^{(k)} h_\ell) \leq 2r, \quad \forall k \in [1:K], \ell \in [1:L], i \in [1:N], \end{aligned}$$

where  $\varepsilon > 0, \varphi_i(\xi) = \|\xi - \widehat{\xi}_i\|_2^2$  and  $r \in \mathbb{N}$  is a given number with  $\deg(g_j^{(k)}) \leq 2r, \forall j \in [1:J], k \in [1:K]$ . We let  $\max(\mathbf{FD}_r) = -\infty$  if its feasibility region is empty.

We note that, in comparison to the SDP hierarchy (AD<sub>r</sub>) for general polynomials, the problem (FD<sub>r</sub>) contains fewer SOS polynomial variables. The SOS polynomials in the problem (AD<sub>r</sub>) are (m + 1)-variate compared to m-variate in (FD<sub>r</sub>). Moreover, the multipliers  $\eta_{j,i}^{(k)} \in \Sigma^2(\xi, \tilde{\xi}), j \in [1:J]$ , for the problem (AD<sub>r</sub>) are replaced with  $\delta_i^{(k)} \in \Delta$ , which makes (FD<sub>r</sub>) easier to solve.

We show that the hierarchy in (FD<sub>r</sub>) exhibits finite convergence.

**Theorem 4.1 (Finite convergence)** *For the problem (EP) under the Wasserstein ambiguity set  $\mathbb{B}_\varepsilon(\widehat{\mathbb{P}})$  and support set  $\Xi$  in (2), let  $(-h_\ell), \ell \in [1:L], g_j^{(k)}, j \in [1:J], k \in [1:K]$ ,*

be convex polynomials; let the Archimedean condition (Assumption 2.2) hold; and let  $\varepsilon > 0$ . Suppose further that:

- (Slater’s condition) There exists  $\bar{\xi} \in \mathbb{R}^m$  such that  $h_\ell(\bar{\xi}) > 0$  for all  $\ell \in [1:L]$ ; and
- (Hessian condition) For each  $k \in [1:K]$ , the Hessian of  $\sum_{j=1}^J a_j^{(k)} g_j^{(k)}$  is positive definite at every minimizer of  $\sum_{j=1}^J a_j^{(k)} g_j^{(k)}$  in  $\Xi$ , for all  $(a_1^{(k)}, \dots, a_J^{(k)}, \dots, a_L^{(k)}) \in \Delta$ .

Then, for all  $r \in \mathbb{N}$ ,

$$\max(\text{FD}_r) \leq \max(\text{FD}_{r+1}) \leq \min(\text{EP}).$$

Moreover, there exists  $r^* \in \mathbb{N}$  such that  $\max(\text{FD}_{r^*}) = \min(\text{EP})$ .

Before proceeding with the proof, we note that the Hessian condition is automatically satisfied when  $g_j^{(k)}$  is strongly convex for each  $k \in [1:K]$ ,  $j \in [1:J]$ .

**Proof** Since the Archimedean condition (Assumption 2.2) holds for  $\{h_\ell\}_{\ell=1}^L$ ,  $\Xi$  is compact. By construction,  $\max(\text{FD}_r) \leq \max(\text{FD}_{r+1})$  for any  $r \in \mathbb{N}$ .

We start by proving  $\max(\text{FD}_r) \leq \max(9)$  for any  $r \in \mathbb{N}$ . Fix an index  $k \in [1:K]$ ,  $i \in [1:N]$ . Take any feasible point  $\lambda \geq 0$ ,  $\alpha_i \in \mathbb{R}$ ,  $\delta_i^{(k)} \in \Delta$ ,  $\sigma_{\ell,i}^{(k)} \in \Sigma_{2r}^2(\xi)$  from the problem  $(\text{FD}_r)$ . The SOS constraint in  $(\text{FD}_r)$  gives us that

$$\sum_{j=1}^J \delta_{j,i}^{(k)} g_j^{(k)}(\xi) + \lambda \|\xi - \widehat{\xi}\|_2^2 - \alpha_i \geq 0,$$

for all  $\xi \in \Xi$ . Since  $\sum_{j=1}^J \delta_{j,i}^{(k)} g_j^{(k)}(\xi) \leq \max_{j \in [1:J]} g_j^{(k)}(\xi)$  for any  $\xi \in \mathbb{R}^m$ , it then follows that  $\lambda \|\xi - \widehat{\xi}\|_2^2 - \alpha_i + \max_{j \in [1:J]} g_j^{(k)}(\xi) \geq 0$  for all  $\xi \in \Xi$ . This holds for any  $k \in [1:K]$ , so

$$\lambda \|\xi - \widehat{\xi}\|_2^2 - \alpha_i + \min_{k \in [1:K]} \max_{j \in [1:J]} g_j^{(k)}(\xi) \geq 0, \quad \text{for all } \xi \in \Xi. \tag{12}$$

This implies that  $\lambda \geq 0$ ,  $\alpha_i \in \mathbb{R}$  is feasible for the problem (9) for  $i \in [1:N]$ , and we obtain  $\max(\text{FD}_r) \leq \max(9)$ . Consequently, we have

$$\max(\text{FD}_r) \leq \max(\text{FD}_{r+1}) \leq \max(9) = \min(\text{EP}),$$

where  $\min(\text{EP}) = \max(9)$  follows from Lemma 2.1.

To prove finite convergence, let  $(\lambda^*, \alpha_i^*)$  be the optimal solution for the semi-infinite problem (9) with attainment of  $\max(9)$  followed from Lemma 2.1 and let  $i \in [1:N]$  be fixed. The non-negativity constraint in Equation (12) is equivalent to

$$\min_{\xi \in \Xi} \left\{ \max_{j \in [1:J]} g_j^{(k)}(\xi) + \lambda^* \|\xi - \widehat{\xi}\|_2^2 - \alpha_i^* \right\}$$

$$= \min_{\xi \in \Xi} \max_{\delta \in \Delta} \underbrace{\left\{ \sum_{j=1}^J \delta_j g_j^{(k)}(\xi) + \lambda^* \|\xi - \widehat{\xi}_i\|_2^2 - \alpha_i^* \right\}}_{P_i^{(k)}(\xi, \delta)} \geq 0 \tag{13}$$

for each  $k \in [1:K]$ . Notice that  $P_i^{(k)}(\cdot, \delta)$  is convex for each fixed  $\delta \in \Delta \subset \mathbb{R}^J$ , and  $P_i^{(k)}(\xi, \cdot)$  is affine for each fixed  $\xi \in \Xi \subset \mathbb{R}^m$ . Moreover, both the simplex  $\Delta$  and  $\Xi$  are convex compact sets. Invoking the Convex-Concave Minimax Theorem [cf.43, Theorem 2.10.2] yields  $\max_{\delta \in \Delta} \min_{\xi \in \Xi} P_i^{(k)}(\xi, \delta) \geq 0$ , and so there exist  $\delta_i^{(k)} \in \Delta, k \in [1:K]$ , such that

$$\begin{aligned} & \min_{\xi \in \Xi} P_i^{(k)}(\xi, \delta_i^{(k)}) \geq 0 \\ \iff & \widetilde{P}_i^{(k)}(\xi) := \sum_{j=1}^J \delta_{j,i}^{(k)} g_j^{(k)}(\xi) + \lambda^* \|\xi - \widehat{\xi}_i\|_2^2 - \alpha_i^* \geq 0 \text{ for all } \xi \in \Xi, k \in [1:K]. \end{aligned} \tag{14}$$

Let  $k \in [1:K]$  be fixed and let  $v_{\min}^{i,k} := \min_{\xi \in \Xi} \widetilde{P}_i^{(k)}(\xi)$ , which is non-negative by Eqn. (14). By Lagrangian duality of convex optimization [cf.43, Theorem 2.9.2] under Slater’s condition, there exist  $\theta_{\ell,i}^{(k)} \in \mathbb{R}_+, \ell \in [1:L]$ , such that

$$\begin{aligned} v_{\min}^{i,k} &= \min_{\xi \in \mathbb{R}^m} \left\{ \widetilde{P}_i^{(k)}(\xi) - \sum_{\ell=1}^L \theta_{\ell,i}^{(k)} h_{\ell}(\xi) \right\} \\ \iff & v_{\min}^{i,k} \leq \widetilde{P}_i^{(k)}(\xi) - \sum_{\ell=1}^L \theta_{\ell,i}^{(k)} h_{\ell}(\xi) \text{ for all } \xi \in \mathbb{R}^m. \end{aligned} \tag{15}$$

This gives us that the following convex polynomial (since  $\widetilde{P}_i^{(k)}$  and  $(-h_{\ell}), \ell \in [1:L]$ , are convex, and  $\theta_{\ell,i}^{(k)} \geq 0, \ell \in [1:L]$ ):

$$\widetilde{f}_i^{(k)}(\xi) := \widetilde{P}_i^{(k)}(\xi) - \sum_{\ell=1}^L \theta_{\ell,i}^{(k)} h_{\ell}(\xi) - v_{\min}^{i,k}$$

is non-negative for all  $\xi \in \mathbb{R}^m$ .

If  $\xi^*$  is a minimizer of  $\widetilde{P}_i^{(k)}$  in  $\Xi$ , then  $v_{\min}^{i,k} = \widetilde{P}_i^{(k)}(\xi^*)$ . It follows from Eqn. (15) that  $\sum_{\ell=1}^L \theta_{\ell,i}^{(k)} h_{\ell}(\xi^*) = 0$ , and so,  $\xi^*$  is the minimizer of  $\widetilde{f}_i^{(k)}$  on  $\mathbb{R}^m$  with  $\widetilde{f}_i^{(k)}(\xi^*) = 0$ .

Meanwhile, by the Archimedean condition of  $\{h_{\ell}\}_{\ell=1}^L$ , there exists  $R \in \mathbb{N}$  such that

$$R - \|\xi\|_2^2 = \widehat{\sigma}_0(\xi) + \sum_{\ell=1}^L \widehat{\sigma}_{\ell}(\xi) h_{\ell}(\xi), \quad \widehat{\sigma}_{\ell} \in \Sigma^2(\xi), \quad \ell \in [0:L].$$

The set  $S := \{\xi : R - \|\xi\|_2^2 \geq 0\}$  contains  $\Xi$ . Without loss of generality, we assume that  $\Xi \subset \text{interior}(S)$  (otherwise, replace  $R$  in the definition of  $S$  with  $\widehat{R} \geq R$ , and still have  $\widehat{R} - \|\xi\|_2^2 \in \mathcal{Q}(h_1, \dots, h_L)$ ).

We now verify that  $\widetilde{f}_i^{(k)}$  satisfies the conditions (i) and (ii) of Proposition 2.4 on the set  $S$ , and we consider the cases  $[\lambda^* = 0]$  and  $[\lambda^* > 0]$  separately.

$[\lambda^* = 0]$ . In this case, note that  $\widetilde{P}_i^{(k)} = \sum_{j=1}^J \delta_{j,i}^{(k)} g_j^{(k)} - \alpha_i^*$  is a convex function, and by the assumption that the Hessian of  $\sum_{j=1}^J \delta_{j,i}^{(k)} g_j^{(k)}$  is positive definite at its minimizer, the global minimizer of  $\widetilde{P}_i^{(k)}$  on  $\Xi$  is unique. Moreover, if  $\xi^* \in \Xi$  is the unique global minimizer of  $\widetilde{P}_i^{(k)}$ , then  $\xi^*$  is the unique zero of  $\widetilde{f}_i^{(k)}$  on  $\mathbb{R}^m$ . Thus,  $\xi^*$  is the only zero of  $\widetilde{f}_i^{(k)}$  on  $\mathbb{R}^m$ , in particular it is the only zero in  $\text{interior}(S)$ . This verifies condition (i) of Proposition 2.4.

Additionally, we note that the Hessian

$$\nabla^2 \widetilde{f}_i^{(k)} = \nabla^2 \left[ \sum_{j=1}^J \delta_{j,i}^{(k)} g_j^{(k)} - \alpha_i^* - \sum_{\ell=1}^L \theta_{\ell,i}^{(k)} h_\ell - v_{\min}^{i,k} \right]$$

is also positive definite at the point  $\xi^*$ , which is the unique zero of  $\widetilde{f}_i^{(k)}$ . This verifies condition (ii) of Proposition 2.4.

$[\lambda^* > 0]$ . The Hessian

$$\nabla^2 \widetilde{f}_i^{(k)} = \nabla^2 \left[ \sum_{j=1}^J \delta_{j,i}^{(k)} g_j^{(k)} + \lambda^* \varphi_i - \alpha_i^* - \sum_{\ell=1}^L \theta_{\ell,i}^{(k)} h_\ell - v_{\min}^{i,k} \right], \quad \varphi_i(\xi) = \|\xi - \widehat{\xi}_i\|_2^2,$$

is positive definite at all  $\xi \in \mathbb{R}^m$  due to the strong convexity of  $\lambda^* \varphi_i$ . Moreover, similar to the  $[\lambda^* = 0]$  case,  $\xi^*$  is the unique zero of  $\widetilde{f}_i^{(k)}$  in  $\text{interior}(S)$ .

Since  $\widetilde{f}_i^{(k)}$  is non-negative on  $\mathbb{R}^m$ , hence on  $S$ , then  $\widetilde{f}_i^{(k)}$  satisfies the conditions of Proposition 2.4 on the set  $S$ . Applying Proposition 2.4 yields

$$\widetilde{f}_i^{(k)}(\xi) = \widetilde{\sigma}_{0,i}^{(k)}(\xi) + \widetilde{\sigma}_{1,i}^{(k)}(\xi)(R - \|\xi\|_2^2)$$

for some  $\widetilde{\sigma}_{0,i}^{(k)}, \widetilde{\sigma}_{1,i}^{(k)} \in \Sigma^2(\xi)$ .

Substituting the Archimedean representation of  $(R - \|\xi\|_2^2)$  and then rearranging for  $\widetilde{P}_i^{(k)}$ , we obtain

$$\widetilde{P}_i^{(k)}(\xi) = \underbrace{v_{\min}^{i,k} + \widetilde{\sigma}_{0,i}^{(k)}(\xi) + \widetilde{\sigma}_{1,i}^{(k)}(\xi) \widehat{\sigma}_0(\xi)}_{\sigma_{0,i}^{(k)}(\xi)} + \sum_{\ell=1}^L h_\ell(\xi) \underbrace{\left[ \theta_{\ell,i}^{(k)} + \widetilde{\sigma}_{1,i}^{(k)}(\xi) \widehat{\sigma}_\ell(\xi) \right]}_{\sigma_{\ell,i}^{(k)}(\xi)},$$

where  $\sigma_{\ell,i}^{(k)} \in \Sigma^2(\xi)$ ,  $\ell \in [0:L]$ , since  $v_{\min}^{i,k} \geq 0$ . Therefore, there exists  $r^* \in \mathbb{N}$  such that  $\deg(\sigma_{0,i}^{(k)}) \leq 2r^*$  and the equation above leads to

$$\begin{aligned} \Sigma_{2r^*}^2(\xi) \ni \sigma_{0,i}^{(k)}(\xi) &= \tilde{P}_i^{(k)}(\xi) - \sum_{\ell=1}^L \sigma_{\ell,i}^{(k)}(\xi) h_{\ell}(\xi) \\ &= \sum_{j=1}^J \delta_{j,i}^{(k)} g_j^{(k)}(\xi) + \lambda^* \varphi_i(\xi) - \alpha_i^* - \sum_{\ell=1}^L \sigma_{\ell,i}^{(k)}(\xi) h_{\ell}(\xi), \end{aligned}$$

or equivalently,

$$\sum_{j=1}^J \delta_{j,i}^{(k)} g_j^{(k)} + \lambda^* \varphi_i - \alpha_i^* - \sum_{\ell=1}^L \sigma_{\ell,i}^{(k)} h_{\ell} \in \Sigma_{2r^*}^2(\xi).$$

So  $(\lambda^*, \alpha_i^*, \delta_i^{(k)}, \sigma_{\ell,i}^{(k)})$  is feasible for  $(\text{FD}_{r^*})$ , and

$$\max(\text{FD}_{r^*}) \geq -\lambda^* \varepsilon^2 + \frac{1}{N} \sum_{i=1}^N \alpha_i^* = \max(9).$$

Since  $\max(\text{FD}_r) \leq \max(9)$  for all  $r$ , the equality holds at  $r^*$ .  $\square$

**Remark 4.2** (*Finite and one-step convergence*) In the special case where  $J = K = 1$ , our Hessian condition of Theorem 4.1 collapses to the condition that  $g_1^{(1)}$  is positive definite at its minimizers in  $\Xi$ . Related conditions have been used to study finite convergence of the Lasserre hierarchy for convex polynomial problems (see [11, Corollary 3.3] or [22, Theorem 3.3]).

By assuming that  $g_j^{(k)}$ 's in the objective function and  $(-h_{\ell})$ 's in the support set to be SOS-convex polynomials, the optimal value and solution of (EP) can be computed by a single SDP.

## 5 Numerical Experiments

This section is dedicated to numerical experiments on concrete decision-making problems of revenue estimation in Section 5.1 and portfolio optimization in Section 5.2 to examine the computational tractability of the resulting SDP hierarchies and test their scalability across a range of problem sizes. In addition, Section 5.3 discusses the strengths and limitations of the computational studies.

All numerical experiments are carried out in MATLAB, using YALMIP as the modeling framework and MOSEK as the optimization solver. Computations were performed on a mid-2024 Apple MacBook Air equipped with an Apple M3 chip (8-core CPU, 10-core GPU) and 16 GB of unified memory. In our implementation, the SOS constraints appearing in our formulations are provided directly to YALMIP using

its native SOS interface. YALMIP automatically converts these SOS constraints into equivalent semi-definite constraints, and the resulting SDPs are solved using MOSEK.

### 5.1 Data-Driven Revenue Estimation

This experiment exemplifies the hierarchy in  $(AD_r)$  of Section 3.2 to piecewise polynomials involving third-degree polynomials in the context of revenue estimation.

In revenue estimation, a merchant offers a random quantity  $\xi \in [0, R] \subseteq \mathbb{R}$ , where  $R > 0$ , of goods to  $K$  customers, each of whom quotes a price depending on the quantity. The merchant may sell to only one customer at a time and aims to maximize revenue by accepting the highest bid.

We model the *offer price*  $f_k$  from the  $k^{\text{th}}$  customer ( $k \in [1:K]$ ) using a third-order utility function:

$$f_k(\xi) = \begin{cases} a_k(\xi - b_k)^3 + d_k, & \text{if } 0 \leq \xi \leq b_k, \\ d_k, & \text{if } \xi > b_k, \end{cases}$$

where  $a_k > 0$ ,  $d_k, b_k \geq 0$ , and  $-a_k b_k^3 + d_k \geq 0$ .

Cubic utility functions have been used in mathematical models in economics and finance [6, 15]. They are concave and non-decreasing, and they incorporate information about the mean, variance, and skewness of the random quantity [6]. A related revenue estimation model with a quadratic utility function is studied in [17].

The merchant's *revenue* is thus the maximum offer price given by  $\max_{k \in [1:K]} f_k(\xi)$ . Estimating an upper bound for the expected revenue with Wasserstein ambiguity can be regarded as an expectation model:

$$\max_{\mathbb{Q} \in \mathbb{B}_\varepsilon(\hat{\mathbb{P}})} \mathbb{E}^{\mathbb{Q}} \left[ \max_{k \in [1:K]} f_k(\xi) \right], \tag{RP}$$

which can be written as

$$- \min_{\mathbb{Q} \in \mathbb{B}_\varepsilon(\hat{\mathbb{P}})} \mathbb{E}^{\mathbb{Q}} \left[ g(\xi) := \min_{k \in [1:K]} \max_{j \in [1:2]} g_j^{(k)}(\xi) \right],$$

where

$$g_1^{(k)}(\xi) = -a_k(\xi - b_k)^3 - d_k \text{ and } g_2^{(k)}(\xi) = -d_k, \text{ for } k \in [1:K]. \tag{16}$$

We assume that the random quantity is governed by an unknown true distribution  $\mathbb{P}$ , approximated by the empirical distribution  $\hat{\mathbb{P}} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\hat{\xi}_i}$  through a training dataset  $\hat{\Xi}_N = \{\hat{\xi}_i\}_{i=1}^N \subset \mathbb{R}$  consisting of  $N$  past observations of the random quantity  $\xi$ . Here, the support set  $\Xi = [0, R] = \{\xi \in \mathbb{R} : h_1(\xi) := \xi \geq 0, h_2(\xi) := R - \xi \geq 0\}$  corresponds to the range of the random supply quantity.

We now link (RP) to a hierarchy of SDP programs. Let  $\tau_1^{(k)}, \tau_2^{(k)} \in \mathbb{R}$  such that

$$\max_{\xi \in \Xi} \max_{j \in [1:2]} g_j^{(k)}(\xi) < \tau_1^{(k)} \quad \text{and} \quad \tau_2^{(k)} < \min_{\xi \in \Xi} \min_{j \in [1:2]} g_j^{(k)}(\xi).$$

Consider the hierarchy of SDP programs given by, for a given  $r \in \mathbb{N}$ :

$$\begin{aligned} \min_{\substack{\lambda \geq 0, \alpha_i \in \mathbb{R} \\ \eta_{j,i}^{(k)}, \sigma_{\ell,i}^{(k)} \in \Sigma^2}} \lambda \varepsilon^2 - \frac{1}{N} \sum_{i=1}^N \alpha_i & \tag{RD_r} \\ \text{s.t.} \quad \psi + \lambda \varphi_i - \alpha_i - \sigma_{1,i}^{(k)} h_1 - \sigma_{2,i}^{(k)} h_2 - \sum_{j=1}^4 \eta_{j,i}^{(k)} \tilde{g}_j^{(k)} \in \Sigma_{2r}^2(\xi, \tilde{\xi}), \quad \forall k \in [1:K], i \in [1:N], j \in [1:4], \\ \deg(\sigma_{\ell,i}^{(k)} h_\ell) \leq 2r, \quad \deg(\eta_{j,i}^{(k)} \tilde{g}_j^{(k)}) \leq 2r, \quad \forall k \in [1:K], j \in [1:4], i \in [1:N], \end{aligned}$$

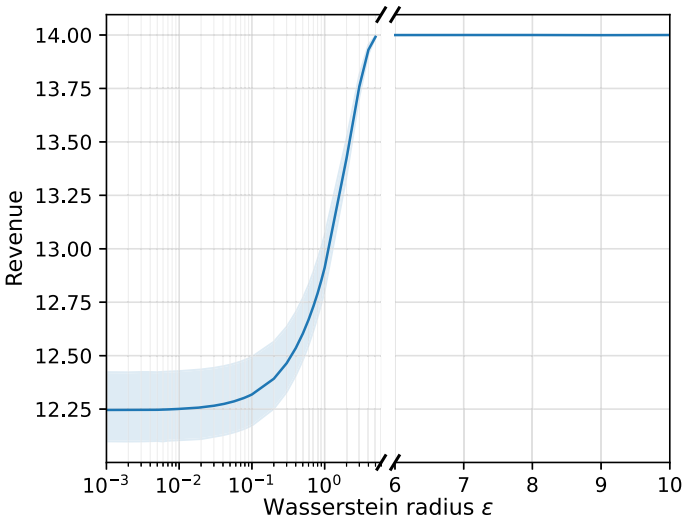
where  $\varepsilon > 0$ ,  $\psi(\xi, \tilde{\xi}) = \tilde{\xi}$ ,  $\varphi_i(\xi, \tilde{\xi}) = (\xi - \widehat{\xi}_i)^2$ ,  $\tilde{g}_j^{(k)}(\xi, \tilde{\xi}) = \tilde{\xi} - g_j^{(k)}(\xi)$ ,  $j \in [1:2]$ ,  $\tilde{g}_3^{(k)}(\xi, \tilde{\xi}) = \tau_1^{(k)} - \tilde{\xi}$ , and  $\tilde{g}_4^{(k)}(\xi, \tilde{\xi}) = \tilde{\xi} - \tau_2^{(k)}$ .

**Proposition 5.1** (Revenue estimation) *Consider the expectation problem (RP). Let  $\Xi = \{\xi \in \mathbb{R} : h_1(\xi) := \xi \geq 0, h_2(\xi) := R - \xi \geq 0\}$  be the support set, for some  $R > 0$ ; and let  $\varepsilon > 0$ . Let  $g_j^{(k)} \in \mathbb{R}[\xi]$ ,  $j \in [1:2], k \in [1:K]$  be as defined in (16). Then,  $\min(\text{RD}_r) \geq \min(\text{RD}_{r+1}) \geq \max(\text{RP})$ , for  $r = 2, 3, 4, \dots$ . Moreover,  $\lim_{r \rightarrow \infty} \min(\text{RD}_r) = \max(\text{RP})$ .*

**Proof** It is easy to check that the Archimedean condition for  $\{h_1, h_2\}$  is satisfied by choosing  $\bar{R} > R^2$ ,  $\sigma_0(\xi) = \bar{R} - R^2$ ,  $\sigma_1(\xi) = \frac{(R-\xi)^2}{R}$ ,  $\sigma_2(\xi) = \frac{\xi^2}{R} + R$ , so  $\bar{R} - \xi^2 = \sigma_0 + \sigma_1 h_1 + \sigma_2 h_2$ . Hence, the conclusion follows directly from Theorem 3.4. □

**Numerical setup.** In the following numerical experiment for revenue estimation, we consider  $K = 3$  customers. We fix the support set by choosing  $R = 12$ . We generate  $N = 30$  data points by sampling from a normal distribution with mean  $\frac{R}{2}$  and standard deviation  $\frac{R}{7}$ , followed by clipping the values to the interval  $[0, R]$ . The polynomials  $g_j^{(k)} \in \mathbb{R}[\xi]$ ,  $j \in [1:2], k \in [1:3]$  in (16) are chosen by fixing:  $[a_1, d_1, b_1] = [4, 9, \frac{3}{4}]$ ,  $[a_2, d_2, b_2] = [\frac{1}{4}, 11, \frac{7}{2}]$ , and  $[a_3, d_3, b_3] = [\frac{1}{110}, 14, \frac{23}{2}]$ .

**Results and Discussion.** Figure 1 displays the mean objective value (shown in solid line) obtained by solving the hierarchy in (RD<sub>r</sub>) with  $r = 2$ , as the Wasserstein radius  $\varepsilon$  varies over the interval  $[10^{-3}, 10]$ . The objective values correspond to the maximum revenues, averaged over ten independent simulation runs. The shaded region represents the inter-quantile range between the 20<sup>th</sup> and 80<sup>th</sup> quantiles. The results show that a larger Wasserstein radius, which enlarges the ambiguity set, yields a higher maximum expected revenue. Particularly, for Wasserstein radius  $\varepsilon \geq 6.0$ , the maximum revenue reaches \$14 (up to machine precision) which coincides with the (theoretical) maximum offer price of \$14, implying that the merchant successfully targets the right customer. On this interval  $[6, 10]$  of values for the Wasserstein radius, the theoretical upper



**Fig. 1** Average maximum revenue (in solid line) computed over ten independent simulation runs. The shaded region represents the inter-quantile range between the 20<sup>th</sup> and 80<sup>th</sup> quantiles.

bound on maximum revenue is already attained (up to machine precision), indicating that further increases in the relaxation level would not yield additional improvement.

We also observe that, constrained by computational resources,  $(RD_r)$  can be solved only for  $r = 2$  which is the minimum relaxation order possible.

**Scalability.** We investigate the scalability and efficiency of the SDP relaxation by varying the number of customers  $K$  and sample size  $N$  with fixed relaxation level  $r = 2$  and Wasserstein radius  $\epsilon = 10$ . The first three customers are fixed and coincide with those used in the numerical setup above. For  $k \geq 4$ , the remaining customer parameters are generated randomly with  $a_k > 0$ ,  $b_k \in [0, R]$ , and  $d_k \in [0, 14]$ .

Table 1 reports the CPU time in seconds together with corresponding maximum expected revenue. As  $K$  increases, the computational cost rises sharply, exceeding 300 seconds for  $K = 12$  in the largest instances, while remaining increasing only mildly with the sample size  $N$ . Table 2 highlights the growth in problem size as the number of customers  $K$  increases.

Across all tested configurations, the objective values remain essentially constant and saturate at the upper bound 14, up to negligible numerical error.

### 5.2 Data-driven Portfolio Optimization

We now explore the use of the hierarchy in  $(\widetilde{AD}_r)$  from Section 3.2 for min-polynomials, involving cubic loss functions in the context of portfolio optimization. Moreover, we report numerical results for several relaxation levels of the hierarchy.

Consider a capital market of  $m$  risky assets that can be chosen by an investor in the financial market. A portfolio is encoded by a vector of percentage weights  $y = (y_1, \dots, y_m)^T \in \Delta$ . Assume that the uncertainty in these assets is described by

**Table 1** CPU times (s) and objective values for varying sample size  $N$  and number of customers  $K$ .

		$K = 3$	$K = 6$	$K = 9$	$K = 12$
CPU Time (s)	$N = 30$	2.7699	16.8437	63.3640	202.7243
	$N = 60$	6.4975	32.2161	110.4003	24.6437
	$N = 90$	10.2844	51.6665	16.5743	81.4009
	$N = 120$	16.0445	10.3989	53.2011	182.8085
	$N = 150$	5.7485	31.3090	112.4932	316.2087
Objective Value	$N = 30$	14.0000	14.0000	14.0000	14.0000
	$N = 60$	13.9999	13.9999	14.0000	14.0000
	$N = 90$	14.0000	14.0000	13.9999	14.0000
	$N = 120$	14.0000	14.0000	14.0000	14.0000
	$N = 150$	14.0000	13.9999	13.9999	14.0000

**Table 2** Problem dimensions of the resulting SDP for  $(RD_r)$  at level  $r = 2$ , with fixed  $N = 150$  and increasing  $K$ , as reported by YALMIP.

$N$	$K$	Number of Constraints	Number of Scalar Variables	Number of Scalarized Matrix Variables
150	3	18,450	14,551	16,200
150	6	36,900	28,951	32,400
150	9	55,350	43,351	48,600
150	12	73,800	57,751	64,800

the random vector  $\xi \in \mathbb{R}^m$  governed by a true probability distribution  $\mathbb{P}$  supported on  $\Xi \subset \mathbb{R}^m$ . We further assume to have access to a training dataset  $\widehat{\Xi}_N = \{\widehat{\xi}_i\}_{i=1}^N$  of  $N$  past realizations of the random vector  $\xi$ , from which we form the empirical distribution  $\widehat{\mathbb{P}} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\widehat{\xi}_i}$ .

We consider the single-stage stochastic mean-risk portfolio optimization problem

$$\min_{y \in \Delta} \left\{ \mathbb{E}^{\mathbb{P}} \left[ \sum_{p=1}^m y_p c_p(\xi) \right] + \gamma \cdot \text{CVaR}_\eta \left[ \sum_{p=1}^m y_p c_p(\xi) \right] \right\}$$

for some continuous functions  $c_p(\xi)$ ,  $p = 1, \dots, m$ , representing the asset losses associated with the risk vector  $\xi$ . This formulation minimizes a weighted combination of the mean and the conditional value-at-risk (CVaR) of the portfolio loss modeled by  $\sum_{p=1}^m y_p c_p(\xi)$ , with  $\eta \in (0, 1]$  in the CVaR-term representing the confidence level, and  $\gamma \in \mathbb{R}_+$  reflects the investor’s level of risk aversion. The CVaR is a known measure for assessing tail risks, particularly in financial management [36]. For a random loss  $\text{Loss}(\xi)$ , the CVaR at a pre-specified confidence level  $\eta \in (0, 1]$  quantifies the expected loss in the worst  $\eta \times 100\%$  scenarios. It is defined by:

$$\text{CVaR}_\eta(\text{Loss}(\xi)) = \inf_{\tau \in \mathbb{R}} \left\{ \tau + \frac{1}{\eta} \mathbb{E}^{\mathbb{P}} \left[ \max \{ \text{Loss}(\xi) - \tau, 0 \} \right] \right\},$$

where  $\mathbb{P}$  denotes the true probability distribution of the underlying uncertainty  $\xi$ .

Inserting the definition of CVaR and robustifying against distributional uncertainty, the distributionally robust counterpart of this problem with respect to the Wasserstein ambiguity set  $\mathbb{B}_\varepsilon(\widehat{\mathbb{P}})$  is given by

$$\inf_{y \in \Delta, \tau \in \mathbb{R}} \max_{\mathbb{Q} \in \mathbb{B}_\varepsilon(\widehat{\mathbb{P}})} \mathbb{E}^{\mathbb{Q}} \left[ \max_{k \in [1:2]} g_1^{(k)}(y, \tau, \xi) \right], \tag{PP}$$

where

$$g_1^{(1)}(y, \tau, \xi) = \sum_{p=1}^m y_p c_p(\xi) + \gamma \tau \quad \text{and} \quad g_1^{(2)}(y, \tau, \xi) = \left(1 + \frac{\gamma}{\eta}\right) \sum_{p=1}^m y_p c_p(\xi) + \left(1 - \frac{1}{\eta}\right) \gamma \tau. \tag{17}$$

We consider the ball support set  $\Xi = \{\xi \in \mathbb{R}^m : h_1(\xi) := R^2 - \|\xi\|_2^2 \geq 0\}$  for some  $R > 0$ .

We associate (PP) with the following sequence of problems for  $r \in \mathbb{N}$  with  $2r \geq \max_{p \in [1:m]} \deg(c_p, 2)$ :

$$\begin{aligned} \min_{\substack{y \in \Delta, \tau \in \mathbb{R} \\ \lambda \geq 0, \alpha_i \in \mathbb{R} \\ \sigma_i^{(k)} \in \Sigma^2}} \lambda \varepsilon^2 - \frac{1}{N} \sum_{i=1}^N \alpha_i & \tag{PD}_r \\ \text{s.t.} \quad & - \sum_{p=1}^m y_p c_p - \gamma \tau + \lambda \varphi_i - \alpha_i - \sigma_i^{(1)} h_1 \in \Sigma_{2r}^2(\xi), \quad \forall i \in [1:N], \\ & - \left(1 + \frac{\gamma}{\eta}\right) \sum_{p=1}^m y_p c_p - \left(1 - \frac{1}{\eta}\right) \gamma \tau + \lambda \varphi_i - \alpha_i - \sigma_i^{(2)} h_1 \in \Sigma_{2r}^2(\xi), \quad \forall i \in [1:N], \\ & \deg(\sigma_i^{(1)}) \leq 2(r-1), \quad \deg(\sigma_i^{(2)}) \leq 2(r-1), \quad \forall i \in [1:N], \end{aligned}$$

where  $\varepsilon > 0$  and  $\varphi_i(\xi) = \|\xi - \widehat{\xi}_i\|_2^2$ .

For clarity, in what follows, we say that  $(y^*, \tau^*, \mathbb{Q}^*) \in (\Delta \times \mathbb{R} \times \mathbb{B}_\varepsilon(\widehat{\mathbb{P}}))$  is a solution for the min-max optimization problem (PP) if

$$\int_{\Xi} \max_{k \in [1:2]} g_1^{(k)}(y^*, \tau^*, \xi) \mathbb{Q}^*(d\xi) = \text{val}(\text{PP}),$$

where  $\text{val}(\text{PP})$  is the optimal value of (PP).

**Proposition 5.2 (Portfolio optimization)** *For the DRO problem (PP), let  $\Xi = \{\xi \in \mathbb{R}^m : h_1(\xi) := R^2 - \|\xi\|_2^2 \geq 0\}$  be the support set, and let  $\varepsilon > 0$ . Let  $g_1^{(1)}$  and  $g_1^{(2)}$  be as defined in (17). Assume that a solution  $(y^*, \tau^*, \mathbb{Q}^*)$  of (PP) and minimizers  $(y_r^*, \tau_r^*, \lambda_r^*, \alpha_{i,r}^*, \sigma_{i,r}^*)$  of  $(\text{PD}_r)$  exist for any  $r \in \mathbb{N}$ . Then  $\min(\text{PD}_r) \geq \min(\text{PD}_{r+1}) \geq \text{val}(\text{PP})$ , for any  $r \in \mathbb{N}$ . Moreover,  $\lim_{r \rightarrow \infty} \min(\text{PD}_r) = \text{val}(\text{PP})$ .*

**Proof** First, we note that the Archimedean condition is automatically satisfied for  $\Xi$ . Let  $\mathcal{X} := (\Delta \times \mathbb{R})$ . Fix  $\mathbf{y} \in \Delta$  and  $\tau \in \mathbb{R}$ . Consider the worst-case expectation subproblem of (PP):

$$\max_{\mathbb{Q} \in \widehat{\mathbb{B}}_\varepsilon(\mathbb{P})} \mathbb{E}^{\mathbb{Q}} \left[ \max_{k \in [1:2]} g_1^{(k)}((\mathbf{y}, \tau), \boldsymbol{\xi}) \right]. \tag{PP}^{(\mathbf{y}, \tau)}$$

By Theorem 3.5, max (PP)<sup>(y,τ)</sup> can be approximated by

$$\begin{aligned} \min_{\substack{\lambda \geq 0, \alpha_i \in \mathbb{R} \\ \sigma_i^{(k)} \in \Sigma^2}} \lambda \varepsilon^2 - \frac{1}{N} \sum_{i=1}^N \alpha_i & \tag{PD}_r^{(\mathbf{y}, \tau)} \\ \text{s.t. } - \sum_{p=1}^m y_p c_p - \gamma \tau + \lambda \varphi_i - \alpha_i - \sigma_i^{(1)} h_1 \in \Sigma_{2r}^2(\boldsymbol{\xi}), \quad \forall i \in [1:N], \\ & - \left(1 + \frac{\gamma}{\eta}\right) \sum_{p=1}^m y_p c_p - \left(1 - \frac{1}{\eta}\right) \gamma \tau + \lambda \varphi_i - \alpha_i - \sigma_i^{(2)} h_1 \in \Sigma_{2r}^2(\boldsymbol{\xi}), \quad \forall i \in [1:N], \\ & \deg(\sigma_i^{(1)}) \leq 2(r-1), \quad \deg(\sigma_i^{(2)}) \leq 2(r-1), \quad \forall i \in [1:N], \end{aligned}$$

in the sense that

$$\min(\text{PD}_r^{(\mathbf{y}, \tau)}) \geq \min(\text{PD}_{r+1}^{(\mathbf{y}, \tau)}) \geq \max(\text{PP}^{(\mathbf{y}, \tau)}) \quad \text{and} \quad \lim_{r \rightarrow \infty} \min(\text{PD}_r^{(\mathbf{y}, \tau)}) = \max(\text{PP}^{(\mathbf{y}, \tau)}). \tag{18}$$

To proceed, we define  $v_r : \mathcal{X} \rightarrow \mathbb{R}$  by  $v_r(\mathbf{y}, \tau) := \min(\text{PD}_r^{(\mathbf{y}, \tau)})$  for each  $(\mathbf{y}, \tau) \in \mathcal{X}$ , and  $v : \mathcal{X} \rightarrow \mathbb{R}$  by  $v(\mathbf{y}, \tau) := \max(\text{PP}^{(\mathbf{y}, \tau)})$  for each  $(\mathbf{y}, \tau) \in \mathcal{X}$ .

With these notation, (18) writes as:

$$v_r(\mathbf{y}, \tau) \geq v_{r+1}(\mathbf{y}, \tau) \geq v(\mathbf{y}, \tau) \quad \text{and} \quad \lim_{r \rightarrow \infty} v_r(\mathbf{y}, \tau) = v(\mathbf{y}, \tau). \tag{19}$$

For convenience, we denote

$$p_r := \min(\text{PD}_r) = \min_{(\mathbf{y}, \tau) \in \mathcal{X}} v_r(\mathbf{y}, \tau), \quad \text{and} \quad p^* := \text{val}(\text{PP}) = \min_{(\mathbf{y}, \tau) \in \mathcal{X}} v(\mathbf{y}, \tau). \tag{20}$$

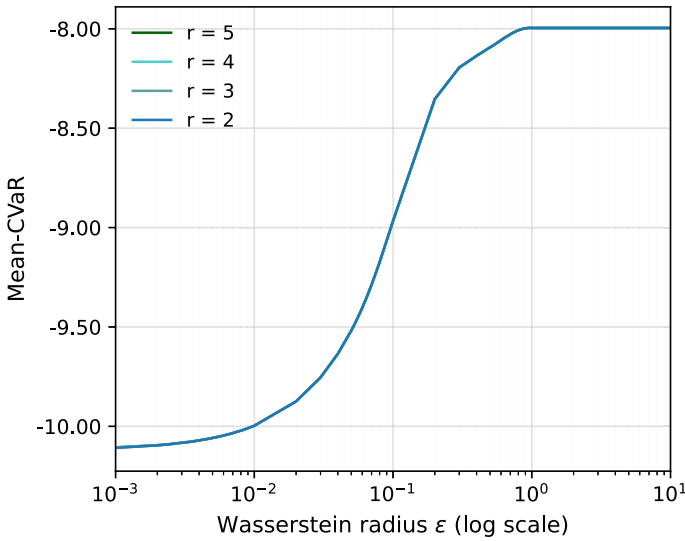
From our assumption of the existence of solutions, one sees that

$$p_r = v_r(\mathbf{y}_r^*, \tau_r^*) \quad \text{and} \quad p^* = \mathbb{E}^{\mathbb{Q}^*} \left[ \max_{k \in [1:2]} g_1^{(k)}((\mathbf{y}^*, \tau^*), \boldsymbol{\xi}) \right] = \max(\text{PP}^{(\mathbf{y}^*, \tau^*)}) = v(\mathbf{y}^*, \tau^*),$$

where  $(\mathbf{y}_r^*, \tau_r^*) \in \mathcal{X}$  and  $(\mathbf{y}^*, \tau^*) \in \mathcal{X}$  are given as in the assumptions.

Firstly, we claim that  $p_r \geq p_{r+1} \geq p^*$ . Indeed, by repeatedly using (19) and (20), we obtain

$$p_r = v_r(\mathbf{y}_r^*, \tau_r^*) \geq v_{r+1}(\mathbf{y}_r^*, \tau_r^*) \geq \min_{(\mathbf{y}, \tau) \in \mathcal{X}} v_{r+1}(\mathbf{y}, \tau) = v_{r+1}(\mathbf{y}_{r+1}^*, \tau_{r+1}^*) = p_{r+1},$$



**Fig. 2** Average in-sample mean-CVaR risk calculated over ten independent simulation runs, as the Wasserstein radius  $\epsilon$  varies within the interval  $[10^{-3}, 10]$ .

and  $p_{r+1} \geq v(\mathbf{y}_{r+1}^*, \tau_{r+1}^*) \geq \min_{(\mathbf{y}, \tau) \in \mathcal{X}} v(\mathbf{y}, \tau) = p^*$ .

So,  $p^*$  is a non-increasing sequence which is bounded below (by  $p^*$ ). In other words,  $\min(\text{PD}_r) \geq \min(\text{PD}_{r+1}) \geq \text{val}(\text{PP})$ .

Let  $\delta > 0$  be arbitrary. Since, by (19),  $\lim_{r \rightarrow \infty} v_r(\mathbf{y}^*, \tau^*) = v(\mathbf{y}^*, \tau^*)$ , then there exists  $\tilde{R} > 0$  such that for all  $r \geq \tilde{R}$ , we have

$$v_r(\mathbf{y}^*, \tau^*) < v(\mathbf{y}^*, \tau^*) + \delta \stackrel{(20)}{=} p^* + \delta \implies p_r \stackrel{(20)}{=} \min_{(\mathbf{y}, \tau) \in \mathcal{X}} v_r(\mathbf{y}, \tau) \leq v_r(\mathbf{y}^*, \tau^*) < p^* + \delta.$$

Noting that  $p_r \geq p^*$ , then for any  $r \geq \tilde{R}$ , we also have  $p_r - p^* = |p_r - p^*| < \delta$ . Since  $\delta$  is arbitrary, we conclude that  $\lim_{r \rightarrow \infty} p_r = p^*$ . Thus,  $\lim_{r \rightarrow \infty} \min(\text{PD}_r) = \text{val}(\text{PP})$ , and the asymptotic convergence holds.  $\square$

**Numerical setup.** In the following numerical experiments for portfolio optimization, we consider  $m = 3$  stocks. The support set is fixed by choosing  $R = 1$ . The  $N = 30$  data points are drawn uniformly from the unit sphere. The polynomials  $g_1^{(k)} \in \mathbb{R}[\xi]$ ,  $k \in [1:2]$ , in (17) are set with  $\gamma = 10$ ,  $\eta = 0.2$ ,  $c_1(\xi) = -1 + \xi_1 + \xi_1 \xi_2 - \xi_1 \xi_3 - 2\xi_1^3$ ,  $c_2(\xi) = -1 - \xi_1 \xi_2 + \xi_2^2 - \xi_2 \xi_3 + \xi_3^3$ , and  $c_3(\xi) = -1 + \xi_2 \xi_3 - \xi_3^2 - \xi_3^3$ , where  $c_1, c_2, c_3$  are consistent with [31, Example 5.5].

**Results and Discussion.** Figure 2 displays the mean objective values obtained by solving the hierarchy in  $(\text{PD}_r)$  over varying Wasserstein radius  $\epsilon \in [10^{-3}, 10]$  for different levels  $r \in \{2, 3, 4, 5\}$ . The objective values correspond to the mean-CVaR risks, and are averaged over ten independent simulation runs. The curves flatten once the Wasserstein radius reaches a certain threshold ( $\epsilon \approx 1$ ), indicating that higher uncertainty beyond a certain threshold does not lead to worse investment losses and

**Table 3** CPU times (seconds) for varying  $N$  and  $m$ .

	$m = 3$	$m = 6$	$m = 9$	$m = 12$
$N = 30$	0.7131	1.5564	5.7414	28.2044
$N = 60$	1.2574	3.0730	17.2813	65.2100
$N = 90$	1.3550	5.1252	22.2863	112.8436
$N = 120$	4.1532	8.8236	32.7673	192.6790
$N = 150$	2.9811	13.9251	57.2876	315.6149

**Table 4** Problem dimensions of the resulting SDP for  $(PD_r)$  at level  $r = 2$ , with fixed  $N = 150$  and increasing  $m$ , as reported by YALMIP.

$N$	$m$	Number of Constraints	Number of Scalar Variables	Number of Scalarized Matrix Variables
150	3	13,501	3,155	19,500
150	6	71,401	8,558	130,200
150	9	231,001	16,661	478,500
150	12	573,301	27,464	1,283,100

risks. The results show that the gains from increasing the relaxation level are negligible. We note that  $(PD_r)$  can still be solved for higher relaxation levels  $r > 5$ , although this comes at the expense of longer computation times.

**CPU time.** On average, a single instance of  $(PD_r)$  was solved in 0.4261 seconds for  $r = 2$ , 1.0180 seconds for  $r = 3$ , 2.2799 seconds for  $r = 4$ , and 6.4355 seconds for  $r = 5$ .

**Scalability.** Henceforth, we fix the relaxation level at  $r = 2$  and the Wasserstein radius at  $\varepsilon = 10.0$ , and vary both the number  $m$  of stocks and the number  $N$  of data points. In this manner, the dimension of the result optimization problem increases in two complementary directions: increasing  $m$  enlarges the decision space of the portfolio weights and the underlying polynomial model, while increasing  $N$  raises the number of sample-dependent constraints in the Wasserstein ambiguity set.

To extend the numerical setup for  $m = 3$  to cases where  $m > 3$ , we fix the cost functions for the additional stocks by setting  $c_p(\xi) = -\xi_p^3$  for  $p = 4, \dots, m$ , while retaining  $c_1, c_2$ , and  $c_3$  from the baseline setup. Specifically, we now consider  $m = 3, 6, 9, 12$  stocks. In varying  $N$ , we consider  $N = 30, 60, 90, 120, 150$ , corresponding to increasing the sample sizes in the empirical measure used to define the Wasserstein ambiguity set, which may be interpreted in the portfolio context as increasing the number of observed trading days.

The results in Table 3 show that the CPU time increases rapidly with the number of assets  $m$ , while the growth with respect to the number of samples  $N$  is comparatively moderate. This behavior is typical for polynomial optimization problems solved via SOS relaxations, as increasing  $m$  induces a combinatorial growth in the number of polynomial monomials and, consequently, in the size of the semi-definite matrix variables and the number of constraints. As illustrated in Table 4, the number of scalarized semi-definite variables grows by more than an order of magnitude as  $m$  increases,

which substantially raises the per-iteration cost of the interior-point method employed by MOSEK. In contrast, increasing  $N$  mainly introduces additional sample-dependent constraints and leads to a comparatively milder increase in computational effort.

### 5.3 Strengths and Limitations of the Computational Study

A key strength of our approach is that the SDP relaxations enable the computation of optimal values of these otherwise infinite-dimensional problems with piecewise polynomials that are prevalent in several decision-making models in diverse disciplines. The results confirm that the hierarchy can produce solutions within a reasonable time for problem instances of moderate dimensionality and relaxation degrees.

However, a notable limitation is computational scalability. Depending on the problem's dimensionality, available computational resources may restrict solving the SDP hierarchy beyond a certain level of relaxation (e.g., the revenue estimation in Section 5.1 can only be solved with relaxation level  $r = 2$ ), thereby limiting the approximation quality or the expressiveness of the model in higher dimensions. This is due to the size of the resulting SDPs, which grows rapidly with the degree and the number of variables.

## 6 Conclusion and Outlook

In this paper, we developed convergent semi-definite program hierarchies for a class of Wasserstein optimization problems with piecewise polynomials. These hierarchies enable the computation or estimation of the optimal values of otherwise infinite-dimensional DRO problems through SDPs, efficiently solvable by off-the-shelf software. We established conditions under which the hierarchies converge asymptotically and finitely.

The construction of these hierarchies relied on techniques from algebraic geometry, especially the sum-of-squares representations of positivity as well as non-negativity, and convex analysis tools such as convex duality and the minimax theorem. These tools allowed for the reformulation of the hard semi-infinite constraints into sum-of-squares constraints, expressible as linear matrix inequalities.

We demonstrated the computational feasibility of our approach through concrete examples in revenue estimation and portfolio optimization. Future research could explore:

- Moment ambiguity-based approach for hierarchies of piecewise polynomial functions, extending the piecewise SOS-convex case in [20].
- Bounded-degree SDP hierarchies for broad classes of DRO problems by using recent advances [9, 39].
- Other hierarchies based on scaled-diagonally-dominant-sum-of-squares [1], aimed at enhancing scalability.
- The recovery of optimal solutions of expectation problems (EP) from the solutions of associated SDPs.

## Appendix A Technical Details

This appendix provides technical proof.

**Proof of Lemma 2.1** Recall that  $\Xi$  is compact and  $\mathcal{M}(\Xi)$  refers to the space of (finite signed) regular Borel measures on  $\Xi$  and  $\mathcal{P}(\Xi) \subseteq \mathcal{M}(\Xi)$  is the set of probability measures. Let  $C(\Xi)$  be the space of continuous functions supported on  $\Xi$  equipped with the supremum norm. We now equip the  $\mathcal{M}(\Xi)$  with the weak\* topology.

By the law of total probability [see, e.g.,30, Theorem4.2],

$$\begin{aligned} \min (1) &= \inf_{\mathbb{Q}_i \in \mathcal{P}(\Xi)} \left\{ \frac{1}{N} \sum_{i=1}^N \mathbb{E}^{\mathbb{Q}_i} [g(\xi)] : \frac{1}{N} \sum_{i=1}^N \mathbb{E}^{\mathbb{Q}_i} [\|\xi - \widehat{\xi}_i\|_2^2] \leq \varepsilon^2 \right\} \\ &= \inf_{\mathbb{Q}_i \in \mathcal{M}(\Xi)} \left\{ \frac{1}{N} \sum_{i=1}^N \mathbb{E}^{\mathbb{Q}_i} [g(\xi)] + \sum_{i=1}^N \chi_{\mathcal{P}(\Xi)}(\mathbb{Q}_i) : \frac{1}{N} \sum_{i=1}^N \mathbb{E}^{\mathbb{Q}_i} [\|\xi - \widehat{\xi}_i\|_2^2] \leq \varepsilon^2 \right\}. \end{aligned} \tag{21}$$

Here,  $\chi_{\mathcal{P}(\Xi)}(\mathbb{Q})$  is the indicator function which takes value 0 if  $\mathbb{Q} \in \mathcal{P}(\Xi)$  and  $+\infty$  otherwise.

Since  $\mathcal{P}(\Xi)$  is a weakly\* compact and convex subset of  $\mathcal{M}(\Xi)$  [8, Example 2.122], and  $\mathcal{P}(\Xi)$  is non-empty due to  $\Xi \neq \emptyset$ , so the indicator function  $\chi_{\mathcal{P}(\Xi)}$  is a proper convex function. A direct verification shows that the problem of the right-hand side of (21) is a convex optimization problem on  $\prod_{i=1}^N \mathcal{M}(\Xi)$  with a proper convex objective function.

Note that Slater’s condition holds for the convex problem of the right-hand side of (21), i.e., the existence of  $\mathbb{Q}_i \in \mathcal{P}(\Xi), i \in [1:N]$ , such that  $\frac{1}{N} \sum_{i=1}^N \mathbb{E}^{\mathbb{Q}_i} [\|\xi - \widehat{\xi}_i\|_2^2] < \varepsilon^2$ . This can be seen by taking  $\mathbb{Q}_i = \mathbb{1}_{\widehat{\xi}_i}$  which gives  $\frac{1}{N} \sum_{i=1}^N \mathbb{E}^{\mathbb{1}_{\widehat{\xi}_i}} [\|\xi - \widehat{\xi}_i\|_2^2] = \frac{1}{N} \sum_{i=1}^N \|\widehat{\xi}_i - \widehat{\xi}_i\|_2^2 = 0 < \varepsilon^2$ .

Now, applying the convex duality theorem [43, Theorem 2.9.3], we have

$$\begin{aligned} \min (1) &= \max_{\lambda \geq 0} \inf_{\mathbb{Q}_i \in \mathcal{P}(\Xi)} \left\{ \frac{1}{N} \sum_{i=1}^N \mathbb{E}^{\mathbb{Q}_i} [g(\xi)] + \lambda \left( \frac{1}{N} \sum_{i=1}^N \mathbb{E}^{\mathbb{Q}_i} [\|\xi - \widehat{\xi}_i\|_2^2] - \varepsilon^2 \right) \right\} \\ &= \max_{\lambda \geq 0} \left\{ -\lambda \varepsilon^2 + \frac{1}{N} \sum_{i=1}^N \inf_{\mathbb{Q}_i \in \mathcal{P}(\Xi)} \mathbb{E}^{\mathbb{Q}_i} [g(\xi) + \lambda \|\xi - \widehat{\xi}_i\|_2^2] \right\}. \end{aligned}$$

Next, we claim that, for each  $i \in [1:N]$ ,

$$\inf_{\mathbb{Q}_i \in \mathcal{P}(\Xi)} \int_{\Xi} \left( g(\xi) + \lambda \|\xi - \widehat{\xi}_i\|_2^2 \right) \mathbb{Q}_i(d\xi) = \inf_{\xi \in \Xi} \left( g(\xi) + \lambda \|\xi - \widehat{\xi}_i\|_2^2 \right).$$

To see this, we note that the set of Dirac measures is a subset of  $\mathcal{P}(\Xi)$  since  $\mathbb{1}_{\xi} \in \mathcal{P}(\Xi)$  for any  $\xi \in \Xi$ , and so,  $\inf_{\mathbb{Q} \in \mathcal{P}(\Xi)} \int_{\Xi} f(\xi) \mathbb{Q}(d\xi) \leq \inf_{\xi \in \Xi} f(\xi)$  for any continuous function  $f$  on  $\mathbb{R}^m$ . The reverse inequality follows from the fact that  $\int_{\Xi} f(\xi) \mathbb{Q}(d\xi) \geq$

$\int_{\Xi} \left( \inf_{\xi' \in \Xi} f(\xi') \right) \mathbb{Q}(d\xi) = \inf_{\xi' \in \Xi} f(\xi')$  for  $\mathbb{Q} \in \mathcal{P}(\Xi)$  and for any continuous function  $f$ . Thus, the claim follows.

Therefore, by the claim, we see that

$$\begin{aligned} \min (1) &= \max_{\lambda \geq 0} \left\{ -\lambda \varepsilon^2 + \frac{1}{N} \sum_{i=1}^N \inf_{\xi \in \Xi} \left( g(\xi) + \lambda \|\xi - \widehat{\xi}_i\|_2^2 \right) \right\} \\ &= \max_{\lambda \geq 0, \alpha_i \in \mathbb{R}} \left\{ -\lambda \varepsilon^2 + \frac{1}{N} \sum_{i=1}^N \alpha_i : \inf_{\xi \in \Xi} \left( g(\xi) + \lambda \|\xi - \widehat{\xi}_i\|_2^2 - \alpha_i \right) \geq 0, \forall i \in [1:N] \right\}. \end{aligned}$$

Hence,  $\min (1) = \max (ED)$ .  $\square$

**Acknowledgements** The authors are grateful to the referees for their helpful comments and valuable suggestions which have contributed to the final preparation of the paper.

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions

**Data Availability** All data generated and analyzed in this study are provided within the article. The data used in the numerical experiments were generated randomly, and we have clearly described the procedure for reproducing them.

## Declarations

**Conflict of Interest** The last two authors declare that they are Associate Editors for the Journal of Optimization Theory and Applications. The other three authors have no conflict of interest to declare that are relevant to the content of this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ahmadi, A.A., Majumdar, A.: DSOS and SDSOS optimization: More tractable alternatives to sum of squares and semidefinite optimization. *SIAM Journal on Applied Algebra and Geometry* **3**(2), 193–230 (2019)
- Ahmadi, A.A., Parrilo, P.A.: A complete characterization of the gap between convexity and SOS-convexity. *SIAM J. Optim.* **23**(2), 811–833 (2013)
- Auslender, A., Goberna, M.A., López, M.A.: Penalty and smoothing methods for convex semi-infinite programming. *Math. Oper. Res.* **34**(2), 303–319 (2009)
- Bach, F.: Sum-of-squares relaxations for polynomial min-max problems over simple sets. *Math. Program.* **209**(1), 475–501 (2025)
- Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: *Robust Optimization*, vol. 28. Princeton University Press, Princeton (2009)

6. Benishay, H.: A fourth-degree polynomial utility function and its implications for investors' responses toward four moments of the wealth distribution. *Journal of Accounting, Auditing & Finance* **2**(3), 203–228 (1987)
7. Blanchet, J., Murthy, K.: Quantifying distributional model risk via optimal transport. *Math. Oper. Res.* **44**(2), 565–600 (2019)
8. Bonnans, J.F., Shapiro, A.: *Perturbation Analysis of Optimization Problems*. Springer Science & Business Media, New York (2013)
9. Chuong, T.D., Jeyakumar, V., Li, G.: A new bounded degree hierarchy with SOCP relaxations for global polynomial optimization and conic convex semi-algebraic programs. *J. Global Optim.* **75**(4), 885–919 (2019)
10. de Klerk, E., Kuhn, D., Postek, K.: Distributionally robust optimization with polynomial densities: Theory, models and algorithms. *Math. Program.* **181**, 265–296 (2020)
11. De Klerk, E., Laurent, M.: On the Lasserre hierarchy of semidefinite programming relaxations of convex polynomial optimization problems. *SIAM J. Optim.* **21**(3), 824–832 (2011)
12. de Klerk, E., Laurent, M.: A survey of semidefinite programming approaches to the generalized problem of moments and their error analysis. In *World Women in Mathematics 2018*, pp. 17–56, Switzerland, . Springer (2019)
13. Delage, E., Ye, Y.: Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* **58**(3), 595–612 (2010)
14. Gao, R., Kleywegt, A.: Distributionally robust stochastic optimization with Wasserstein distance. *Math. Oper. Res.* **48**(2), 603–655 (2023)
15. Gerber, H.U., Pafum, G.: Utility functions: From risk theory to finance. *North American Actuarial Journal* **2**(3), 74–91 (1998)
16. Goberna, M.A., Lopez, M.A.: Linear semi-infinite programming theory: An updated survey. *Eur. J. Oper. Res.* **143**(2), 390–405 (2002)
17. Han, S., Tao, M., Topcu, U., Owghadi, H., Murray, R.M.: Convex optimal uncertainty quantification. *SIAM J. Optim.* **25**(3), 1368–1387 (2015)
18. Helton, J.W., Nie, J.: Semidefinite representation of convex sets. *Math. Program.* **122**, 21–64 (2010)
19. Ho-Nguyen, N., Wright, S.J.: Adversarial classification via distributional robustness with Wasserstein ambiguity. *Math. Program.* **198**(2), 1411–1447 (2023)
20. Huang, Q., Jeyakumar, V., Li, G.: Piecewise SOS-convex moment optimization and applications via exact semi-definite programs. *EURO J. Comp. Optim.*, p. 100094, (2024)
21. Huang, Q., Jeyakumar, V., Li, G., Huyen, D.: Convexifiable quadratic inequality systems: New minimax S-lemma and exact SOCPs for classes of distributionally robust optimization problems. *J. Glob. Optim.*, pp. 1–34, (2025)
22. Jeyakumar, V., Pham, T., Li, G.: Convergence of the Lasserre hierarchy of SDP relaxations for convex polynomial programs without compactness. *Oper. Res. Lett.* **42**(1), 34–40 (2014)
23. Kuhn, D., Mohajerin Esfahani, P., Nguyen, V. A., Shafieezadeh-Abadeh, S.: Wasserstein distributionally robust optimization: Theory and applications in machine learning. In *Operations Research & Management Science in the Age of Analytics*, pp. 130–166. *Inform.*, (2019)
24. Kuhn, D., Shafiee, S., Wiesemann, W.: Distributionally Robust Optimization. *Acta Numer* **34**, 579–804 (2025)
25. Laraki, R., Lasserre, J.B.: Semidefinite programming for min-max problems and games. *Math. Program.* **131**, 305–332 (2012)
26. Lasserre, J.B.: A semidefinite programming approach to the generalized problem of moments. *Math. Program.* **112**, 65–92 (2008)
27. Lasserre, J.B.: Convexity in semialgebraic geometry and polynomial optimization. *SIAM J. Optim.* **19**(4), 1995–2014 (2009)
28. Lasserre, J.B.: *Moments, Positive Polynomials and Their Applications*, vol. 1. World Scientific, London (2009)
29. Lasserre, J.B.: *An Introduction to Polynomial and Semi-Algebraic Optimization*, vol. 52. Cambridge University Press, Cambridge (2015)
30. MohajerinEsfahani, P., Kuhn, D.: Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Math. Program.* **171**(1), 115–166 (2018)
31. Nie, J., Yang, L., Zhong, S., Zhou, G.: Distributionally robust optimization with moment ambiguity sets. *J. Sci. Comput.* **94**(1), 12 (2023)

32. Pflug, G. C., Pichler, A.: *Multistage Stochastic Optimization*, vol. 1104. Springer, (2014)
33. Popescu, I.: A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. *Math. Oper. Res.* **30**(3), 632–657 (2005)
34. Postek, K., den Hertog, D., Melenberg, B.: Computationally tractable counterparts of distributionally robust constraints on risk measures. *SIAM Rev.* **58**(4), 603–650 (2016)
35. Putinar, M.: Positive polynomials on compact semi-algebraic sets. *Indiana Univ. Math. J.* **42**(3), 969–984 (1993)
36. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. *J. Risk* **2**, 21–42 (2000)
37. Scheiderer, C.: Sums of squares on real algebraic curves. *Math. Z.* **245**, 725–760 (2003)
38. Van Parys, B.P., Goulart, P.J., Morari, M.: Distributionally robust expectation inequalities for structured distributions. *Math. Program.* **173**, 251–280 (2019)
39. Weisser, T., Lasserre, J.B., Toh, K.-C.: Sparse-BSOS: A bounded degree SOS hierarchy for large scale polynomial optimization with sparsity. *Math. Program. Comput.* **10**, 1–32 (2018)
40. Wiesemann, W., Kuhn, D., Sim, M.: Distributionally robust convex optimization. *Oper. Res.* **62**(6), 1358–1376 (2014)
41. Wozabal, D.: A framework for optimization under ambiguity. *Ann. Oper. Res.* **193**, 21–47 (2012)
42. Yue, M.-C., Kuhn, D., Wiesemann, W.: On linear optimization over Wasserstein balls. *Math. Program.* **195**(1), 1107–1122 (2022)
43. Zalinescu, C.: *Convex Analysis in General Vector Spaces*. World Scientific, Singapore (2002)
44. Zhen, J., Kuhn, D., Wiesemann, W.: A unified theory of robust and distributionally robust optimization via the primal-worst-equals-dual-best principle. *Operations Research*, (2023)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.