



# Distributionally Robust Optimization with Moment Ambiguity Sets

Jiawang Nie<sup>1</sup> · Liu Yang<sup>2</sup> · Suhan Zhong<sup>3</sup> · Guangming Zhou<sup>2</sup>

Received: 31 October 2021 / Revised: 14 October 2022 / Accepted: 7 November 2022 /  
Published online: 1 December 2022  
© The Author(s) 2022

## Abstract

This paper studies distributionally robust optimization (DRO) when the ambiguity set is given by moments for the distributions. The objective and constraints are given by polynomials in decision variables. We reformulate the DRO with equivalent moment conic constraints. Under some general assumptions, we prove the DRO is equivalent to a linear optimization problem with moment and psd polynomial cones. A Moment-SOS relaxation method is proposed to solve it. Its asymptotic and finite convergence are shown under certain assumptions. Numerical examples are presented to show how to solve DRO problems.

**Keywords** Distributionally robust optimization · Ambiguity set · Moment · Polynomial · Moment-SOS relaxation

**Mathematics Subject Classification** 90C23 · 90C15 · 90C22

## 1 Introduction

Many decision problems are involved with uncertainties. People often like to make a decision that works well with uncertain data. The distributionally robust optimization (DRO) is a

---

✉ Jiawang Nie  
njw@math.ucsd.edu

Liu Yang  
yangl410@xtu.edu.cn

Suhan Zhong  
suzhong@tamu.edu

Guangming Zhou  
zhougm@xtu.edu.cn

<sup>1</sup> Department of Mathematics, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

<sup>2</sup> School of Mathematics and Computational Sciences, Xiangtan University, Xiangtan 411105, Hunan, China

<sup>3</sup> Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, USA

frequently used model for this kind of decision problems. A typical DRO problem is

$$\min_{x \in X} f(x) \quad s.t. \quad \inf_{\mu \in \mathcal{M}} \mathbb{E}_\mu[h(x, \xi)] \geq 0, \tag{1.1}$$

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}, h : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}, x := (x_1, \dots, x_n)$  is the decision variable constrained in a set  $X \subseteq \mathbb{R}^n$  and  $\xi := (\xi_1, \dots, \xi_p) \in \mathbb{R}^p$  is the random variable obeying the distribution of a measure  $\mu \in \mathcal{M}$ . The notation  $\mathbb{E}_\mu[h(x, \xi)]$  stands for the expectation of the random function  $h(x, \xi)$  with respect to the distribution of  $\xi$ . The set  $\mathcal{M}$  is called the *ambiguity set*, which is used to describe the uncertainty of the measure  $\mu$ .

The ambiguity set  $\mathcal{M}$  is often moment-based or discrepancy-based. For the moment-based ambiguity, the set  $\mathcal{M}$  is usually specified by the first, second moments [11, 17, 50]. Recently, higher order moments are also often used [8, 15, 28], especially in relevant applications with machine learning. For discrepancy-based ambiguity sets, popular examples are the  $\phi$ -divergence ambiguity sets [2, 31] and the Wasserstein ambiguity sets [40]. There are also some other types of ambiguity sets. For instance, [22] assumes  $\mathcal{M}$  is given by distributions with sum-of-squares (SOS) polynomial density functions of known degrees.

We are mostly interested in Borel measures whose supports and moments, up to a given degree  $d$ , are respectively contained in given sets  $S \subseteq \mathbb{R}^p$  and  $Y \subseteq \mathbb{R}^{\binom{p+d}{d}}$ . Let  $\mathcal{B}(S)$  denote the set of Borel measures supported in  $S$ . We assume the ambiguity set is given as

$$\mathcal{M} := \left\{ \mu \in \mathcal{B}(S) : \mathbb{E}_\mu([\xi]_d) \in Y \right\}, \tag{1.2}$$

where  $[\xi]_d$  is the monomial vector

$$[\xi]_d := [1 \ \xi_1 \ \dots \ \xi_p \ (\xi_1)^2 \ \xi_1 \xi_2 \ \dots \ \xi_p^d]^T.$$

The problem (1.1) equipped with the above ambiguity set is called the distributionally robust optimization of moment (DROM). When all the defining functions are polynomials, the DROM is an important class of distributionally robust optimization. It has broad applications. Polynomial and moment optimization are studied extensively [25, 29, 34, 38]. This paper studies how to solve DROM in the form (1.1) by using Moment and SOS relaxations (see the preliminary section for a brief review of them). Currently, there exists relatively few work on this topic.

Solving DROM is of broad interests recently. It is studied in [22, 31] when the density functions are given by polynomials. Polynomial and moment optimization are studied extensively [25, 29, 34, 38]. In this paper, we study how to solve DROM in the form (1.1) by using Moment-SOS relaxations. Currently, there exists relatively less work on this topic.

We remark that the distributionally robust min-max optimization

$$\min_{x \in X} \max_{\mu \in \mathcal{M}} \mathbb{E}_\mu[F(x, \xi)] \tag{1.3}$$

is a special case of the distributionally robust optimization in the form (1.1). Assume each  $\mu \in \mathcal{M}$  is a probability measure (i.e.,  $\mathbb{E}_\mu[1] = 1$ ), then the min-max optimization (1.3) is equivalent to

$$\min_{(x, x_0) \in X \times \mathbb{R}} x_0 \quad s.t. \quad \inf_{\mu \in \mathcal{M}} \mathbb{E}_\mu[x_0 - F(x, \xi)] \geq 0. \tag{1.4}$$

This is a distributionally robust optimization problem in the form (1.1).

The distributionally robust optimization is frequently used to model uncertainties in various applications. It is closely related to stochastic optimization and robust optimization. Under certain conditions, the DRO can be transformed into other two kinds of problems. In stochastic optimization (see [6, 14, 24, 45, 47]), people often need to solve decision problems

such that the true distributions can be well approximated by sampling. The performance of computed solutions heavily relies on the quality of sampling. In order to get more stable solutions, regularization terms can be added to the optimization (see [39, 43, 49]). In robust optimization (see [1, 4]), the uncertainty is often assumed to be freely distributed in some sets. This approach is often computationally tractable and suitable for large-scale data. However, it may produce too pessimistic decisions for certain applications. Combining these two approaches may give more reasonable decisions sometimes. Some information of random variables may be well estimated, or even exactly generated from the sampling or historic data. For instance, people may know the support of the measure, discrepancy from a reference distribution, or its descriptive statistics. The ambiguity set can be given as a collection of measures satisfying such properties. It contains some exact information of distributions, as well as some uncertainties. For decision problems with ambiguity sets, it is naturally to find optimal decisions that work well under uncertainties. This gives rise to distributionally robust optimization like (1.1).

We refer to [7, 22, 41, 51, 53, 54, 56] for recent work on distributionally robust optimization. For the min-max robust optimization (1.3), we refer to [11, 46, 50]. The distributionally robust optimization has broad applications, e.g., portfolio management [11, 13, 55], network design [31, 52], inventory problems [5, 50] and machine learning [12, 16, 32]. For more general work on distributionally robust optimization, we refer to the survey [44] and the references therein.

### Contributions

This article studies the distributionally robust optimization (1.1) with a moment ambiguity set  $\mathcal{M}$  as in (1.2). Assume the measure support set  $S$  is a semi-algebraic set given by a tuple  $g := (g_1, \dots, g_{m_1})$  of polynomials in  $\xi$ . Similarly, assume the feasible set  $X$  is given by a polynomial tuple  $c := (c_1, \dots, c_{m_2})$  in  $x$ . We consider the case that the objective  $f(x)$  is a polynomial in  $x$ , constrained in a set  $X \subseteq \mathbb{R}^n$ , and that the function  $h(x, \xi)$  is polynomial in the random variable  $\xi$  and is linear in  $x$ . The function  $h(x, \xi)$  can be written as

$$h(x, \xi) := \sum_{\substack{\alpha := (\alpha_1, \dots, \alpha_p) \\ \alpha_1 + \dots + \alpha_p \leq d}} h_\alpha(x) \cdot \xi_1^{\alpha_1} \dots \xi_p^{\alpha_p}, \tag{1.5}$$

where each coefficient  $h_\alpha(x)$  is a linear function in  $x$ . The total degree in  $\xi := (\xi_1, \dots, \xi_p)$  is at most  $d$ . For neatness, we also write that

$$h(x, \xi) = (Ax + b)^T [\xi]_d, \tag{1.6}$$

for a given matrix  $A$  and vector  $b$ . Recall that  $\mathcal{M}$  has the expression (1.2). It is clear that the set  $\mathcal{M}$  consists of truncated moment sequences (tms)

$$y := (y_\alpha), \quad \text{where } \alpha := (\alpha_1, \dots, \alpha_p), \quad |\alpha| := \alpha_1 + \dots + \alpha_p \leq d,$$

such that the moment vector  $y = \int [\xi]_d d\mu$  is contained in a given set  $Y$ . In this paper, we focus on the case that  $S$  is compact and that  $Y$  is a set whose conic hull  $\text{cone}(Y)$  can be represented by linear, second order or semidefinite conic inequalities. For convenience, define the conic hull of moments

$$K := \text{cone}(\{\mathbb{E}_\mu([\xi]_d) : \mu \in \mathcal{M}\}). \tag{1.7}$$

Note that  $K$  can also be expressed with  $\text{cone}(Y)$ ; see (3.6). The constraint in (1.1) is the same as

$$(Ax + b)^T y \geq 0 \quad \forall y \in K.$$

Let  $K^*$  denote the dual cone of  $K$ , then the above is equivalent to  $Ax + b \in K^*$ . Therefore, the problem (1.1) can be equivalently reformulated as

$$\min_{x \in X} f(x) \quad \text{s.t.} \quad Ax + b \in K^*. \tag{1.8}$$

The moment constraining cone  $K$  and its dual cone  $K^*$  are typically difficult to describe computationally. However, they can be successfully solved by Moment-SOS relaxations (see [35, 38]).

A particularly interesting case is that  $\xi$  is a univariate random variable, i.e.,  $p = 1$ . For this case, the dual cone  $K^*$  can be exactly represented by semidefinite programming constraints. For instance, if  $d = 4$ ,  $Y$  is the hypercube  $[0, 1]^5$  and  $S = [a_1, a_2]$ , then  $\text{cone}(Y)$  is the nonnegative orthant and the cone  $K$  can be expressed by the constraints

$$\begin{bmatrix} y_0 & y_1 & y_2 \\ y_1 & y_2 & y_3 \\ y_2 & y_3 & y_4 \end{bmatrix} \geq 0, \quad (a_1 + a_2) \begin{bmatrix} y_1 & y_2 \\ y_2 & y_3 \end{bmatrix} \geq a_1 a_2 \begin{bmatrix} y_0 & y_1 \\ y_1 & y_2 \end{bmatrix} + \begin{bmatrix} y_2 & y_3 \\ y_3 & y_4 \end{bmatrix},$$

$$(y_0, y_1, y_2, y_3, y_4) \geq 0.$$

In the above,  $X_1 \geq X_2$  means that  $X_1 - X_2$  is a positive semidefinite (psd) matrix. The dual cone  $K^*$  can be given by semidefinite programming constraints dual to the above. The proof for such expression is shown in Theorem 4.6.

For the case that  $\xi$  is multi-variate, i.e.,  $p > 1$ , there typically do not exist explicit semidefinite programming representations for the cone  $K$  and its dual cone  $K^*$ . However, they can be approximated efficiently by Moment-SOS relaxations (see [35, 38]).

This paper studies how to solve the equivalent optimization problem (1.8) by Moment-SOS relaxations. In computation, the cone of  $Y$  is usually expressed as a Cartesian product of linear, second order, or semidefinite conic constraints. A hierarchy of Moment-SOS relaxations is proposed to solve (1.8) globally, which is equivalent to the distributionally robust optimization (1.1). It is worthy to note that our convex relaxations use both ‘‘moment’’ and ‘‘SOS’’ relaxation techniques, which are different from the classic work of polynomial optimization and DROM problems. In most prior work, usually one of moment and SOS relaxation is used, but rarely two are used simultaneously. Under some general assumptions (e.g., the compactness or archimedeaness), we prove the asymptotic and finite convergence of the proposed Moment-SOS method. The property of finite convergence makes our method very attractive for solving DROM. To check whether a Moment-SOS relaxation is tight or not, one can solve an  $\mathcal{A}$ -truncated moment problem with the method in [35]. By doing so, we not only compute the optimal values and optimizers of (1.8), but also obtain a measure  $\mu$  that achieves the worst case expectation constraint. This is a major advantage that most other methods do not own. In summary, our main contributions are:

- We consider the new class of distributionally robust optimization problems in the form (1.1), which are given by polynomial functions and moment ambiguity sets. The Moment-SOS relaxation method is proposed to solve them globally. It has more attractive properties than prior existing methods. Numerical examples are given to show the efficiency.
- When the objective  $f(x)$  and the constraining set  $X$  are given by SOS-convex polynomials, we prove the DROM is equivalent to a linear conic optimization problem.

- Under some general assumptions, we prove the asymptotic and finite convergence of the proposed method. There is little prior work on finite convergence for solving DROM. In particular, when the random variable  $\xi$  is univariate, we show that the lowest order Moment-SOS relaxation is sufficient for solving (1.8) exactly.
- We also show how to obtain the measure  $\mu^*$  that achieves the worst case expectation constraint.

The rest of the paper is organized as follows. Section 2 reviews some preliminary results about moment and polynomial optimization. In Sect. 3, we give an equivalent reformulation of the distributionally robust optimization, expressing it as a linear conic optimization problem. In Sect. 4, we give an algorithm of Moment-SOS relaxations to solve (1.8). Some numerical experiments and applications are given in Sect. 5. Finally, we make some conclusions and discussions in Sect. 6.

## 2 Preliminaries

### Notation

The symbol  $\mathbb{R}$  (resp.,  $\mathbb{R}_+, \mathbb{N}$ ) denotes the set of real numbers (resp., nonnegative real numbers, nonnegative integers). For  $t \in \mathbb{R}$ ,  $\lceil t \rceil$  denotes the smallest integer that is greater or equal to  $t$ . For an integer  $k > 0$ ,  $[k] := \{1, \dots, k\}$ . The symbol  $\mathbb{N}^n$  (resp.,  $\mathbb{R}^n$ ) stands for the set of  $n$ -dimensional vectors with entries in  $\mathbb{N}$  (resp.,  $\mathbb{R}$ ). For a vector  $v$ , we use  $\|v\|$  to denote its Euclidean norm. The superscript  $T$  denotes the transpose of a matrix or vector. For a set  $S$ , the notation  $\mathcal{B}(S)$  denotes the set of Borel measures whose supports are contained in  $S$ . For two sets  $S_1, S_2$ , the operation

$$S_1 + S_2 := \{s_1 + s_2 : s_1 \in S_1, s_2 \in S_2\}$$

is the Minkowski sum. The symbol  $e$  stands for the vector of all ones and  $e_i$  stands for the  $i$ th standard unit vector, i.e., its  $i$ th entry is 1 and all other entries are zeros. We use  $I_n$  to denote the  $n$ -by- $n$  identity matrix. A symmetric matrix  $W$  is positive semidefinite (psd) if  $v^T W v \geq 0$  for all  $v \in \mathbb{R}^n$ . We write  $W \geq 0$  to mean that  $W$  is psd. The strict inequality  $W > 0$  means that  $W$  is positive definite.

The symbol  $\mathbb{R}[x] := \mathbb{R}[x_1, \dots, x_n]$  denotes the ring of polynomials in  $x$  with real coefficients, and  $\mathbb{R}[x]_d$  is the subset of  $\mathbb{R}[x]$  with polynomials of degrees at most  $d$ . For a polynomial  $f \in \mathbb{R}[x]$ , we use  $\deg(f)$  to denote its degree. For a tuple  $f = (f_1, \dots, f_r)$  of polynomials, the  $\deg(f)$  denotes the highest degree of  $f_i$ . For a polynomial  $p(x)$ ,  $\text{vec}(p)$  is the coefficient vector of  $p$ . For  $\alpha := (\alpha_1, \dots, \alpha_n)$  and  $x := (x_1, \dots, x_n)$ , we denote that

$$x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}, \quad |\alpha| := \alpha_1 + \cdots + \alpha_n.$$

For a degree  $d$ , denote the power set

$$\mathbb{N}_d^n := \{\alpha \in \mathbb{N}^n : |\alpha| \leq d\}.$$

Let  $[x]_d$  denote the vector of all monomials in  $x$  that have degrees at most  $d$ , i.e.,

$$[x]_d := [1 \ x_1 \ \cdots \ x_n \ x_1^2 \ x_1 x_2 \ \cdots \ x_n^d]^T.$$

The notation  $\xi^\alpha$  and  $[\xi]_d$  are similarly defined for  $\xi := (\xi_1, \dots, \xi_p)$ . The notation  $\mathbb{E}_\mu[h(\xi)]$  denotes the expectation of the random function  $h(\xi)$  with respect to  $\mu$  for the random variable  $\xi$ . The Dirac measure, which is supported at a point  $u$ , is denoted as  $\delta_u$ .

Let  $V$  be a vector space over the real field  $\mathbb{R}$ . A set  $C \subseteq V$  is a cone if  $ax \in C$  for all  $x \in C$  and  $a > 0$ . For a set  $X \subset V$ , we denote its closure by  $\overline{X}$  in the Euclidean topology. Its conic hull, which is the minimum convex cone containing  $X$ , is denoted as  $\text{cone}(X)$ . The dual cone of the set  $X$  is

$$X^* := \{\ell \in V^* \mid \ell(x) \geq 0, \forall x \in X\}, \tag{2.1}$$

where  $V^*$  is the dual space of  $V$  (i.e., the space of linear functionals on  $V$ ). Note that  $X^*$  is a closed convex cone for all  $X$ . For two nonempty sets  $X_1, X_2 \in V$ , we have  $(X_1 + X_2)^* = X_1^* \cap X_2^*$ . When  $X_1 + X_2$  is a closed convex cone, we also have  $(X_1^* \cap X_2^*)^* = X_1 + X_2$ .

In the following, we review some basics in optimization about polynomials and moments. We refer to [21, 25, 27, 29, 36, 37] for more details about this topic.

### 2.1 SOS and Nonnegative Polynomials

A polynomial  $f \in \mathbb{R}[x]$  is said to be SOS if  $f = f_1^2 + \dots + f_k^2$  for some real polynomials  $f_i \in \mathbb{R}[x]$ . We use  $\Sigma[x]$  to denote the cone of all SOS polynomials in  $x$ . The  $d$ th degree truncation of the SOS cone  $\Sigma[x]$  is

$$\Sigma[x]_d := \Sigma[x] \cap \mathbb{R}[x]_d.$$

It is a closed convex cone for each degree  $d$ . For a polynomial  $f \in \mathbb{R}[x]$ , the membership  $f \in \Sigma[x]$  can be checked by solving semidefinite programs [25, 29]. In particular,  $f$  is said to be SOS-convex [18] if its Hessian matrix  $\nabla^2 f(x)$  is SOS, i.e.,  $\nabla^2 f = V(x)^T V(x)$  for a matrix polynomial  $V(x)$ .

In this paper, we also need to work with polynomials in  $\xi := (\xi_1, \dots, \xi_p)$ . For a tuple  $g := (g_1, \dots, g_{m_1})$  of polynomials in  $\xi$ , its quadratic module is the set

$$\text{QM}[g] := \Sigma[\xi] + g_1 \cdot \Sigma[\xi] + \dots + g_{m_1} \cdot \Sigma[\xi].$$

The  $d$ th degree truncation of  $\text{QM}[g]$  is

$$\text{QM}[g]_d := \Sigma[\xi]_d + g_1 \cdot \Sigma[\xi]_{d-\text{deg}(g_1)} + \dots + g_{m_1} \cdot \Sigma[\xi]_{d-\text{deg}(g_{m_1})}.$$

Let  $S = \{\xi \in \mathbb{R}^p : g(\xi) \geq 0\}$  be the set determined by  $g$  and let  $\mathcal{P}(S)$  denote the set of polynomials that are nonnegative on  $S$ . We also frequently use the  $d$ th degree truncation

$$\mathcal{P}_d(S) := \mathcal{P}(S) \cap \mathbb{R}[\xi]_d.$$

Then it holds that for all degree  $d$

$$\text{QM}[g]_d \subseteq \mathcal{P}_d(S).$$

The quadratic module  $\text{QM}[g]$  is said to be *archimedean* if there exists a polynomial  $\phi \in \text{QM}[g]$  such that  $\{\xi \in \mathbb{R}^p : \phi(\xi) \geq 0\}$  is compact. If  $\text{QM}[g]$  is archimedean, then  $S$  must be a compact set. The converse is not necessarily true. However, for compact  $S$ , the quadratic module  $\text{QM}[\tilde{g}]$  is archimedean if  $g$  is replaced by  $\tilde{g} := (g, N - \|\xi\|^2)$  for  $N$  sufficiently large. When  $\text{QM}[g]$  is archimedean, if a polynomial  $h > 0$  on  $S$ , then we have  $h \in \text{QM}[g]$  (see [42]). Furthermore, under some classical optimality conditions, we have  $h \in \text{QM}[g]$  if  $h \geq 0$  on  $S$  (see [36]).

### 2.2 Truncated Moment Problems

For the variable  $\xi \in \mathbb{R}^p$ , the space of truncated multi-sequences (tms) of degree  $d$  is

$$\mathbb{R}^{\mathbb{N}_d^p} := \{z = (z_\alpha)_{\alpha \in \mathbb{N}_d^p} : z_\alpha \in \mathbb{R}\}.$$

Each  $z \in \mathbb{R}^{\mathbb{N}_d^p}$  determines the linear Riesz functional  $\mathcal{L}_z$  on  $\mathbb{R}[\xi]_d$  such that

$$\mathcal{L}_z\left(\sum_{\alpha \in \mathbb{N}_d^p} h_\alpha \xi^\alpha\right) := \sum_{\alpha \in \mathbb{N}_d^p} h_\alpha z_\alpha. \tag{2.2}$$

For convenience of notation, we also write that

$$\langle q, z \rangle := \mathcal{L}_z(q), \quad q \in \mathbb{R}[\xi]_d. \tag{2.3}$$

For a polynomial  $q \in \mathbb{R}[\xi]_{2d}$  and a tms  $z \in \mathbb{R}^{\mathbb{N}_{2k}^p}$ , with  $k \geq d$ , the  $k$ th order *localizing matrix*  $L_q^{(d)}[z]$  is such that

$$\text{vec}(a)^T \left(L_q^{(k)}[z]\right) \text{vec}(b) = \mathcal{L}_z(qab) \tag{2.4}$$

for all  $a, b \in \mathbb{R}[\xi]_s$ , where  $s = k - \lceil \text{deg}(q)/2 \rceil$ . In particular, for  $q = 1$  (the constant one polynomial), the  $L_1^{(k)}[z]$  becomes the so-called *moment matrix*

$$M_k[z] := L_1^{(k)}[z]. \tag{2.5}$$

We can use the moment matrix and localizing matrices to describe dual cones of quadratic modules. For a polynomial tuple  $g = (g_1, \dots, g_{m_1})$  with  $\text{deg}(g) \leq 2k$ , define the tms cone

$$\mathcal{S}[g]_{2k} := \left\{z \in \mathbb{R}^{\mathbb{N}_{2k}^p} : M_k[z] \geq 0, L_{g_1}^{(k)}[z] \geq 0, \dots, L_{g_{m_1}}^{(k)}[z] \geq 0\right\}. \tag{2.6}$$

It can be verified that (see [38])

$$(\text{QM}[g]_{2k})^* = \mathcal{S}[g]_{2k}. \tag{2.7}$$

A tms  $z = (z_\alpha) \in \mathbb{R}^{\mathbb{N}_d^p}$  is said to admit a representing measure  $\mu$  supported in a set  $S \subseteq \mathbb{R}^p$  if  $z_\alpha = \int \xi^\alpha d\mu$  for all  $\alpha \in \mathbb{N}_d^p$ . Such a measure  $\mu$  is called an  $S$ -representing measure for  $z$ . In particular, if  $z = 0$  is the zero tms, then it admits the identically zero measure. Denote by  $\text{meas}(z, S)$  the set of  $S$ -measures admitted by  $z$ . This gives the moment cone

$$\mathcal{R}_d(S) := \{z \in \mathbb{R}^{\mathbb{N}_d^p} \mid \text{meas}(z, S) \neq \emptyset\}. \tag{2.8}$$

It is interesting to note that  $\mathcal{R}_d(S)$  can also be written as the conic hull

$$\mathcal{R}_d(S) = \text{cone}(\{[\xi]_d : \xi \in S\}). \tag{2.9}$$

Recall that  $\mathcal{P}_d(S)$  denotes the cone of polynomials in  $\mathbb{R}[\xi]_d$  that are nonnegative on  $S$ . It is a closed and convex cone. For all  $h \in \mathcal{P}_d(S)$  and  $z \in \mathcal{R}_d(S)$ , it holds that for every  $\mu \in \text{meas}(z, S)$ ,

$$\langle h, z \rangle = \sum_{\alpha \in \mathbb{N}_d^p} h_\alpha z_\alpha = \int h(\xi) d\mu \geq 0.$$

This implies that  $\mathcal{R}_d(S)^* = \mathcal{P}_d(S)$ . When  $S$  is compact, we also have  $\mathcal{P}_d(S)^* = \mathcal{R}_d(S)$ . If  $S$  is not compact, then

$$\mathcal{P}_d(S)^* = \overline{\mathcal{R}_d(S)}. \tag{2.10}$$

We refer to [29, Sect. 5.2] and [38] for this fact.

A frequent case is that  $S = \{\xi : g(\xi) \geq 0\}$  is determined by a polynomial tuple  $g = (g_1, \dots, g_{m_1})$ . For an integer  $k \geq \deg(g)/2$ , a tms  $z \in \mathbb{R}^{\mathbb{N}_{2k}^p}$  admits an  $S$ -representing measure  $\mu$  if  $z \in \mathcal{S}[g]_{2k}$  and

$$\text{rank } M_{k-d_0}[z] = \text{rank } M_k[z], \tag{2.11}$$

where  $d_0 = \lceil \deg(g)/2 \rceil$ . Moreover, the measure  $\mu$  is unique and is  $r$ -atomic, i.e.,  $|\text{supp}(\mu)| = r$ , where  $r = \text{rank } M_k[z]$ . The above rank condition is called *flat extension* or *flat truncation* [9, 34]. When it holds, the tms  $z$  is said to be a *flat* tms. When  $z$  is flat, one can obtain the unique representing measure  $\mu$  for  $z$  by computing Schur decompositions and eigenvalues (see [19]).

To obtain a representing measure for a tms  $y \in \mathbb{R}_d^{\mathbb{N}_d^p}$  that is not flat, a semidefinite relaxation method is proposed in [35]. Suppose  $S$  is compact and the quadratic module  $\text{QM}[g]$  is archimedean. Select a generic polynomial  $R \in \Sigma[\xi]_{2k}$ , with  $2k > \deg(g)$ , and then solve the moment optimization

$$\begin{cases} \min_{\omega} \langle R, \omega \rangle \\ \text{s.t. } \omega|_d = y, \omega \in \mathcal{S}[g]_{2k}. \end{cases} \tag{2.12}$$

In the above  $\omega|_d$  denotes the  $d$ th degree truncation of  $\omega$ , i.e.,

$$\omega|_d := (\omega_\alpha)_{|\alpha| \leq d}. \tag{2.13}$$

As  $k$  increases, by solving (2.12), one can either get a flat extension of  $y$ , or a certificate that  $y$  does not have any representing measure. We refer to [35] for more details about solving truncated moment problems.

### 3 Moment Optimization Reformulation

In this section, we reformulate the distributionally robust optimization equivalently as a linear conic optimization problem with moment constraints. We consider the DROM problem

$$\begin{cases} \min_{x \in \mathbb{R}^n} f(x) \\ \text{s.t. } \inf_{\mu \in \mathcal{M}} \mathbb{E}_\mu[h(x, \xi)] \geq 0, \\ x \in X, \end{cases} \tag{3.1}$$

where  $x$  is the decision variable constrained in a set  $X \subseteq \mathbb{R}^n$  and  $\xi \in \mathbb{R}^p$  is the random variable obeying the distribution of the measure  $\mu$  that belongs to the moment ambiguity set  $\mathcal{M}$ . We assume that the objective  $f(x)$  is a polynomial in  $x$  and  $h(x, \xi)$  is a polynomial in  $\xi$  whose coefficients are linear in  $x$ . Equivalently, one can write that

$$h(x, \xi) = (Ax + b)^T [\xi]_d, \quad A \in \mathbb{R}^{\binom{p+d}{d} \times n}, \quad b \in \mathbb{R}^{\binom{p+d}{d}}. \tag{3.2}$$

Suppose measures in the ambiguity set  $\mathcal{M}$  have supports contained in the set

$$S = \{\xi \in \mathbb{R}^p : g_1(\xi) \geq 0, \dots, g_{m_1}(\xi) \geq 0\}, \tag{3.3}$$

for a given tuple  $g := (g_1, \dots, g_{m_1})$  of polynomials in  $\xi$ . The ambiguity set  $\mathcal{M}$  can be expressed as

$$\mathcal{M} := \{\mu \in \mathcal{B}(S) \mid \mathbb{E}_\mu([\xi]_d) \in Y\}, \tag{3.4}$$



where  $Y$  is the constraining set for moments of  $\mu$ . The set  $Y$  is not necessarily closed or convex. The closure of its conic hull is denoted as  $\overline{\text{cone}(Y)}$ . In computation, it is often a Cartesian product of linear, second order or semidefinite cones. The constraining set  $X$  for  $x$  is assumed to be the set

$$X := \{x \in \mathbb{R}^n \mid c_1(x) \geq 0, \dots, c_{m_2}(x) \geq 0\}, \tag{3.5}$$

for a tuple  $c = (c_1, \dots, c_{m_2})$  of polynomials in  $x$ .

The DROM (3.1) can be equivalently reformulated as polynomial optimization with moment conic conditions. Observe that

$$\inf_{\mu \in \mathcal{M}} \mathbb{E}_\mu[h(x, \xi)] \geq 0 \iff (Ax + b)^T y \geq 0, \forall y \in \mathcal{R}_d(S) \cap \text{cone}(Y).$$

The set  $\mathcal{R}_d(S)$  is the moment cone defined as in (2.8). It consists of degree- $d$  tms' admitting  $S$ -measures. For convenience, we denote the intersection

$$K = \mathcal{R}_d(S) \cap \text{cone}(Y). \tag{3.6}$$

Therefore, we get that

$$\inf_{\mu \in \mathcal{M}} \mathbb{E}_\mu[h(x, \xi)] \geq 0 \iff Ax + b \in K^*, \tag{3.7}$$

where  $K^*$  denotes the dual cone of  $K$ . In view of (2.1) and (2.3), the dual cone  $Y^*$  is the following polynomial cone

$$Y^* = \{\phi \in \mathbb{R}[\xi]_d : \langle \phi, z \rangle \geq 0, \forall z \in Y\}. \tag{3.8}$$

Observe the dual cone relations

$$\begin{aligned} \mathcal{R}_d(S)^* &= \mathcal{P}_d(S), & \mathcal{P}_d(S)^* &= \overline{\mathcal{R}_d(S)}, \\ (\mathcal{P}_d(S) + Y^*)^* &= \overline{\mathcal{R}_d(S)} \cap \overline{\text{cone}(Y)}. \end{aligned}$$

When both  $\mathcal{R}_d(S)$  and  $\text{cone}(Y)$  are closed, we have

$$\overline{\mathcal{R}_d(S) \cap \text{cone}(Y)} = \overline{\mathcal{R}_d(S)} \cap \overline{\text{cone}(Y)}. \tag{3.9}$$

If one of them is not closed, the above may or may not be true. Note that  $C^{**} = C$  if  $C$  is a closed convex cone. When (3.9) holds and the sum  $\mathcal{P}_d(S) + Y^*$  is a closed cone, we can express the dual cone  $K^*$  as

$$K^* = \mathcal{P}_d(S) + Y^*. \tag{3.10}$$

As shown in [3, Proposition B.2.7], the above equality holds if  $\mathcal{R}_d(S)$ ,  $\text{cone}(Y)$  are closed and their interiors have non-empty intersection. Such conditions are often satisfied for most applications. Recall that  $h(x, \xi) = (Ax + b)^T [\xi]_d$ . The membership  $Ax + b \in K^*$  means that  $h(x, \xi) \in K^*$ . Therefore, we get the following result.

**Theorem 3.1** *Assume the set  $X$  is given as in (3.5). If the equality (3.10) holds, then (3.1) is equivalent to the following optimization*

$$\begin{cases} \min_{x \in \mathbb{R}^n} f(x) \\ \text{s.t. } c_1(x) \geq 0, \dots, c_{m_2}(x) \geq 0, \\ h(x, \xi) \in \mathcal{P}_d(S) + Y^*. \end{cases} \tag{3.11}$$

The membership constraint in (3.11) means that  $h(x, \xi)$ , as a polynomial in  $\xi$ , is the sum of a polynomial in  $\mathcal{P}_d(S)$  and a polynomial in  $Y^*$ . When  $f, c_1, \dots, c_{m_2}$  are all linear functions, (3.11) is a linear conic optimization problem. When  $f$  and every  $c_i$  are polynomials, we can apply Moment-SOS relaxations to solve it.

Recall that  $X$  is the set given as in (3.5). Denote the degree

$$d_1 := \max\{\deg(f)/2, \lceil \deg(c)/2 \rceil\}.$$

Observe that for all  $x \in X$  and  $w = [x]_{2d_1}$ , it holds that

$$\begin{aligned} \langle f, w \rangle &= f(x), \quad M_{d_1}[w] \geq 0, \\ L_{c_i}^{(d_1)}[w] &\geq 0, \quad i = 1, \dots, m_2. \end{aligned}$$

We refer to the Subsection 2.2 for the above notation. For convenience, define the projection map  $\pi : \mathbb{R}^{\mathbb{N}_{2d_1}^n} \rightarrow \mathbb{R}^n$  such that

$$\pi(w) := (w_{e_1}, \dots, w_{e_n}), \quad w \in \mathbb{R}^{\mathbb{N}_{2d_1}^n}. \tag{3.12}$$

So, the optimization (3.11) can be relaxed to

$$\begin{cases} \min_{(x,w)} \langle f, w \rangle \\ \text{s.t. } M_{d_1}[w] \geq 0, \quad L_{c_i}^{(d_1)}[w] \geq 0 \quad (i \in [m_2]), \\ \quad h(x, \xi) \in \mathcal{P}_d(S) + Y^*, \\ \quad w_0 = 1, \quad x = \pi(w), \quad w \in \mathbb{R}^{\mathbb{N}_{2d_1}^n}. \end{cases} \tag{3.13}$$

The relaxation (3.13) is said to be *tight* if it has the same optimal value as (3.11) does. Under the SOS-convexity assumption, the relaxation (3.13) is equivalent to (3.11). This is the following result.

**Theorem 3.2** *Suppose the ambiguity set  $\mathcal{M}$  is given as in (3.4) and the set  $X$  is given as in (3.5). Assume the polynomials  $f, -c_1, \dots, -c_{m_2}$  are SOS-convex. Then, the optimization problems (3.13) and (3.11) are equivalent in the following sense: they have the same optimal value, and  $w^*$  is a minimizer of (3.13) if and only if  $x^* := \pi(w^*)$  is a minimizer of (3.11).*

**Proof** Let  $w$  be a feasible point for (3.13) and  $x = \pi(w)$ , then  $Ax + b \in K^*$ . Since  $f, -c_1, \dots, -c_{m_2}$  are SOS-convex, by the Jensen’s inequality (see [26]), we have the inequalities

$$\begin{aligned} f(x) &= f(\pi(w)) \leq \langle f, w \rangle, \\ c_i(x) &= c_i(\pi(w)) \geq \langle c_i, w \rangle, \quad i = 1, \dots, m_2. \end{aligned}$$

The (1, 1)-entry of  $L_{c_i}^{(d_1)}[w]$  is  $\langle c_i, w \rangle$ , so  $L_{c_i}^{(d_1)}[w] \geq 0$  implies that  $\langle c_i, w \rangle \geq 0$ . This means that  $x = \pi(w) \in X$  for every  $w$  that is feasible for (3.13). Let  $f_0, f_1$  denote the optimal values of (3.11), (3.13) respectively. Since the latter is a relaxation of the former, it is clearly that  $f_0 \geq f_1$ . For every  $\epsilon > 0$ , there exists a feasible  $w$  such that  $\langle f, w \rangle \leq f_1 + \epsilon$ , which implies that  $f(\pi(w)) \leq f_1 + \epsilon$ . Hence  $f_0 \leq f_1 + \epsilon$  for every  $\epsilon > 0$ . Therefore,  $f_0 = f_1$ , i.e., (3.13) and (3.11) have the same optimal value.

If  $w^*$  is a minimizer of (3.13), we also have  $x^* = \pi(w^*) \in X$  and

$$f(x^*) = f(\pi(w^*)) \leq \langle f, w^* \rangle.$$

Since (3.13) is a relaxation of (3.11), they must have the same optimal value and  $x^*$  is a minimizer of (3.11). For the converse, if  $x^*$  is a minimizer of (3.11), then  $w^* := [x^*]_{2d_1}$  is

feasible for (3.13) and  $f(x^*) = \langle f, w^* \rangle$ . So  $w^*$  must also be a minimizer of (3.13), since (3.13) and (3.11) have the same optimal value.  $\square$

In the following, we derive the dual optimization of (3.13). As in Subsection 2.2, we have seen that

$$M_{d_1}[w] \geq 0, L_{c_i}^{(d_1)}[w] \geq 0 (i \in [m_2]) \iff w \in \mathcal{S}[c]_{2d_1},$$

where  $\mathcal{S}[c]_{2d_1}$  is given similarly as in (2.6). Recall the dual relationship

$$(\text{QM}[c]_{2d_1})^* = \mathcal{S}[c]_{2d_1},$$

as shown in (2.7). The Lagrange function for (3.13) is

$$\begin{aligned} \mathcal{L}(w; \gamma, q, y, z) &= \langle f, w \rangle - \gamma(w_0 - 1) - \langle q, w \rangle - \langle y, A\pi(w) + b \rangle \\ &= \langle f - q - y^T Ax - \gamma \cdot 1, w \rangle + \gamma - \langle b, y \rangle, \end{aligned}$$

for  $\gamma \in \mathbb{R}, q \in \text{QM}[c]_{2d_1}, y \in \overline{K}$ . (Note that the cone  $K$  is not necessarily closed.) To make  $\mathcal{L}(w; \gamma, q, y, z)$  have a finite infimum for  $w \in \mathbb{R}^{\mathbb{N}_{2d_1}^n}$ , we need the constraint

$$f - y^T Ax - \gamma = q.$$

Therefore, the dual optimization of (3.13) is

$$\begin{cases} \max_{(\gamma, y)} \gamma - \langle b, y \rangle \\ \text{s.t. } f(x) - y^T Ax - \gamma \in \text{QM}[c]_{2d_1}, \\ \gamma \in \mathbb{R}, y \in \overline{K}. \end{cases} \tag{3.14}$$

The first membership in (3.14) means that  $f(x) - y^T Ax - \gamma$ , as a polynomial in  $x$ , belongs to the truncated quadratic module  $\text{QM}[c]_{2d_1}$ . So it gives a constraint for both  $\gamma$  and  $y$ .

### 4 The Moment-SOS Relaxation Method

In this section, we give a Moment-SOS relaxation method for solving the distributionally robust optimization and prove its convergence.

In Sect. 3, we have seen that the DROM (3.1) is equivalent to the linear conic optimization (3.13) under certain assumptions. It is still hard to solve (3.13) directly, due to the membership constraint  $h(x, \xi) \in \mathcal{P}_d(S) + Y^*$ . This is because the nonnegative polynomial cone  $\mathcal{P}_d(S)$  typically does not have an explicit computational representation. For its dual problem (3.14), it is similarly difficult to deal with the conic membership  $y \in \overline{K}$ . However, both (3.13) and (3.14) can be solved efficiently by Moment-SOS relaxations.

Recall that  $S$  is a semi-algebraic set given as in (3.3). For every integer  $k \geq d/2$ , it holds the nesting containment

$$\text{QM}[g]_{2k} \cap \mathbb{R}[\xi]_d \subseteq \text{QM}[g]_{2k+2} \cap \mathbb{R}[\xi]_d \subseteq \dots \subseteq \mathcal{P}_d(S).$$

We thus consider the following restriction of (3.13):

$$\begin{cases} \min_{(x, w)} \langle f, w \rangle \\ \text{s.t. } M_{d_1}[w] \geq 0, L_{c_i}^{(d_1)}[w] \geq 0 (i \in [m_2]), \\ h(x, \xi) \in \text{QM}[g]_{2k} + Y^*, \\ w_0 = 1, x = \pi(w), w \in \mathbb{R}^{\mathbb{N}_{2d_1}^n}. \end{cases} \tag{4.1}$$

The integer  $k$  is called the relaxation order. Since  $(\text{QM}[g]_{2k})^* = \mathcal{S}[g]_{2k}$ , the dual optimization of (4.1) is

$$\begin{cases} \max_{(\gamma, y, z)} \gamma - \langle b, y \rangle \\ \text{s.t.} \quad f(x) - y^T Ax - \gamma \in \text{QM}[c]_{2d_1}, \\ \gamma \in \mathbb{R}, z \in \mathcal{S}[g]_{2k}, y \in \overline{\text{cone}(Y)}, y = z|_d. \end{cases} \tag{4.2}$$

We would like to remark that  $\text{QM}[g]$  is a quadratic module in the polynomial ring  $\mathbb{R}[\xi]$ , while  $\text{QM}[c]$  is a quadratic module in the polynomial  $\mathbb{R}[x]$ . The notation  $z|_d$  denotes the degree- $d$  truncation of  $z$ ; see (2.13) for its meaning. The optimization (4.2) is a relaxation of (3.14), since it has a bigger feasible set. There exist both quadratic module and moment constraints in (4.2). The primal-dual pair (4.1)-(4.2) can be solved as semidefinite programs. The following is a basic property about the above optimization.

**Theorem 4.1** *Assume (3.9) holds. Suppose  $(\gamma^*, y^*, z^*)$  is an optimizer of (4.2) for the relaxation order  $k$ . Then  $(\gamma^*, y^*)$  is a maximizer of (3.14) if and only if it holds that  $y^* \in \overline{\mathcal{R}_d(S)}$ .*

**Proof** If  $(\gamma^*, y^*)$  is a maximizer of (3.14), then it is clear that  $y^* \in \overline{\mathcal{R}_d(S)}$ . Conversely, if  $y^* \in \overline{\mathcal{R}_d(S)}$ , then  $(\gamma^*, y^*)$  is feasible for (3.14), since (3.9) holds. Since (4.2) is a relaxation of (3.14), we know  $(\gamma^*, y^*)$  must also be a maximizer of (3.14).  $\square$

If  $\mathcal{R}_d(S)$  is a closed cone, then we only need to check  $y^* \in \mathcal{R}_d(S)$  in the above. Interestingly, when  $S$  is compact, the moment cone  $\mathcal{R}_d(S)$  is closed [27, 29, 38]. As introduced in the Subsection 2.2, the membership  $y^* \in \mathcal{R}_d(S)$  can be checked by solving a truncated moment problem. This can be done by solving the optimization (2.12) for a generically selected objective. Once  $(\gamma^*, y^*)$  is confirmed to be a maximizer of (3.14), we show how to get a minimizer for (3.1). This is shown as follows.

**Theorem 4.2** *Assume (3.9) holds. For a relaxation order  $k$ , suppose  $(x^*, w^*)$  is a minimizer of (4.1) and  $(\gamma^*, y^*, z^*)$  is a maximizer of (4.2) such that  $y^* \in \overline{\mathcal{R}_d(S)}$ . Assume there is no duality gap between (4.1) and (4.2), i.e., they have the same optimal value. If the point  $x^*$  belongs to the set  $X$  and  $f(x^*) = \langle f, w^* \rangle$ , then  $x^*$  is a minimizer of (3.11). Moreover, if in addition the dual cone  $K^*$  can be expressed as in (3.10), then  $x^*$  is also a minimizer of (3.1).*

**Proof** Let  $f_1, f_2$  be optimal values of the optimization problems (3.13) and (3.14) respectively. Then, by the weak duality, it holds that

$$f_1 \geq f_2.$$

The membership  $y^* \in \overline{\mathcal{R}_d(S)}$  implies that  $(\gamma^*, y^*)$  is a maximizer of (3.14), by Theorem 4.1. So  $f_2 = \gamma^* - b^T y^*$ . By the assumption, the primal-dual pair (4.1)-(4.2) have the same optimal value, so

$$\langle f, w^* \rangle = \gamma^* - b^T y^* = f_2.$$

The constraint  $h(x^*, \xi) \in \text{QM}[g]_{2k} + Y^*$  implies that  $h(x^*, \xi) \in \mathcal{P}_d(S) + Y^*$ . Since  $x^* \in X$ , we know  $x^*$  is a feasible point of (3.11). The optimal value of (3.11) is greater than or equal to that of (3.13), hence

$$f_1 \geq f_2 = \langle f, w^* \rangle = f(x^*) \geq f_1.$$

So  $f(x^*) = f_1$ . This implies that  $x^*$  is a minimizer of (3.11). Moreover, if in addition  $K^*$  can be expressed as in (3.10), the optimization (3.1) is equivalent to (3.11), by Theorem 3.1. So  $x^*$  is also a minimizer of (3.1).  $\square$

In the above theorem, the assumptions that  $x^* \in X$  and  $f(x^*) = \langle f, w^* \rangle$  must hold if  $f, -c_1, \dots, -c_{m_2}$  are SOS-convex polynomials. We have the following theorem.

**Theorem 4.3** *Assume (3.9) holds. For a relaxation order  $k$ , suppose  $(x^*, w^*)$  is a minimizer of (4.1) and  $(\gamma^*, y^*, z^*)$  is a maximizer of (4.2) such that  $y^* \in \overline{\mathcal{B}_d(S)}$ . Assume there is no duality gap between (4.1) and (4.2), i.e., they have the same optimal value. If  $f, -c_1, \dots, -c_{m_2}$  are SOS-convex polynomials, then  $x^* := \pi(w^*)$  is a minimizer of (3.11). Moreover, if in addition  $K^*$  can be expressed as in (3.10), then  $x^*$  is also a minimizer of (3.1).*

**Proof** Since  $f$  and  $-c_1, \dots, -c_{m_2}$  are SOS-convex polynomials, by the Jensen’s inequality (see [26]), it holds that

$$\begin{aligned} f(x^*) &= f(\pi(w^*)) \leq \langle f, w^* \rangle, \\ c_i(x^*) &= c_i(\pi(w^*)) \geq \langle c_i, w^* \rangle, \quad i = 1, \dots, m_2. \end{aligned}$$

Similarly, the constraint  $L_{c_i}^{(d_i)}[w^*] \geq 0$  implies that  $\langle c_i, w^* \rangle \geq 0$ . So  $x^* \in X$  is a feasible point of (3.11). As in the proof of Theorem 4.2, we can similarly show that

$$f_1 \geq f_2 = \langle f, w^* \rangle \geq f(x^*) \geq f_1,$$

so  $f(x^*) = \langle f, w^* \rangle$ . The conclusions follow from Theorem 4.2. □

### 4.1 An Algorithm for Solving the DROM

Based on the above discussions, we now give the algorithm for solving the optimization problem (3.13) and its dual (3.14), as well as the DROM (3.1).

**Algorithm 4.4** For given  $f, h, \mathcal{M}, S, X, Y$  and the defining polynomial tuples  $g$  and  $c$ , do the following:

Step 0 Get a computational representation for  $\overline{\text{cone}(Y)}$  and the dual cone  $Y^*$ . Initialize

$$d_0 := \lceil \text{deg}(g)/2 \rceil, \quad t_0 := \lceil d/2 \rceil, \quad k := \lceil d/2 \rceil, \quad l := t_0 + 1.$$

Choose a generic polynomial  $R \in \Sigma[\xi]_{2t_0+2}$ .

Step 1 Solve (4.1) for a minimizer  $(x^*, w^*)$  and solve (4.2) for a maximizer  $(\gamma^*, y^*, z^*)$ .

Step 2 Solve the moment optimization

$$\begin{cases} \min_{\omega} \langle R, \omega \rangle \\ \text{s.t. } \omega|_d = y^*, \omega \in \mathcal{S}[g]_{2\ell}, \omega \in \mathbb{R}^{N_{2\ell}^p}. \end{cases} \tag{4.3}$$

If (4.3) is infeasible, then  $y^*$  admits no  $S$ -measure, update  $k := k + 1$  and go back to Step 1. Otherwise, solve (4.3) for a minimizer  $\omega^*$  and go to Step 3.

Step 3 Check whether or not there exists an integer  $s \in [\max(d_0, t_0), \ell]$  such that

$$\text{rank } M_{s-d_0}[\omega^*] = \text{rank } M_s[\omega^*].$$

If such  $s$  does not exist, update  $\ell := \ell + 1$  and go to Step 2. If such  $s$  exists, then  $y^* = \int[\xi]_d \bar{\alpha} \mu$  for the measure

$$\mu = \theta_1 \delta_{u_1} + \dots + \theta_r \delta_{u_r}.$$

In the above, the scalars  $\theta_1, \dots, \theta_r > 0$ ,  $u_1, \dots, u_r \in S$  are distinct points,  $r = \text{rank } M_s[\omega^*]$ , and  $\delta_{u_i}$  denotes the Dirac measure supported at  $u_i$ . Up to scaling, a measure  $\mu^* \in \mathcal{M}$  that achieves the worst case expectation constraint can be recovered as a multiple of  $\mu$ .

**Remark 4.5** All optimization problems in Algorithm 4.4 can be solved numerically by the software `GloptiPoly3` [20], `YALMIP` [30] and `SeDuMi` [48]. In Step 0, we assume  $\overline{\text{cone}(Y)}$  can be expressed by linear, second order or semidefinite cones. See Sect. 5 for more details. In Step 1, if (4.1) is unbounded from below, then (3.11) must also be unbounded from below. If (4.2) is unbounded from above, then (3.14) may be unbounded from above (and hence (3.11) is infeasible), or it may be because the relaxation order  $k$  is not large enough. We refer to [38] for how to verify unboundedness of (3.14). Generally, one can assume (4.1) and (4.2) have optimizers. In Step 3, the finitely atomic measure  $\mu$  can be obtained by computing Schur decompositions and eigenvalues. We refer to [19] for the method. It is also implemented in the software `GloptiPoly3`. Note that the measure  $\mu$  associated with  $y^*$  may not belong to  $\mathcal{M}$ . This is because (4.2) has the conic constraint  $y \in \overline{\text{cone}(Y)}$  instead of  $y \in Y$ . Once the atomic measure  $\mu$  is extracted, we can choose a scalar  $\beta > 0$  such that  $\beta\mu \in \mathcal{M}$ .

### 4.2 Convergence of Algorithm 4.4

In this subsection, we prove the convergence of Algorithm 4.4. The main results here are based on the work [35, 38].

First, we consider the relatively simple but still interesting case that  $\xi$  is a univariate random variable (i.e.,  $p = 1$ ) and the support set  $S = [a_1, a_2]$  is an interval. For this case, Algorithm 4.4 must terminate in the initial loop  $k := \lceil d/2 \rceil$  with  $y^* \in \mathcal{R}_d(S)$ , if  $(\gamma^*, y^*, z^*)$  is a maximizer of (4.2).

**Theorem 4.6** *Suppose the random variable  $\xi$  is univariate and the set  $S = [a_1, a_2]$ , for scalars  $a_1 < a_2$ , is an interval with the constraint  $g(\xi) := (\xi - a_1)(a_2 - \xi) \geq 0$ . If  $(\gamma^*, y^*, z^*)$  is a maximizer of (4.2) for  $k = \lceil d/2 \rceil$ , then we must have  $z^* \in \mathcal{R}_{2k}(S)$  and hence  $y^* \in \mathcal{R}_d(S)$ .*

**Proof** In the relaxation (4.2), the tms  $z$  has the even degree  $2k$ . We label the entries of  $z$  as  $z = (z_0, z_1, \dots, z_{2k})$ . The condition  $z \in \mathcal{S}[g]_{2k}$  implies that

$$M_k[z] \geq 0, \quad L_g^{(k)}[z] \geq 0. \tag{4.4}$$

Since  $g = (\xi - a_1)(a_2 - \xi)$ , one can verify that  $L_g^{(k)}[z] \geq 0$  is equivalent to

$$(a_1 + a_2) \begin{bmatrix} z_1 & z_2 & \cdots & z_k \\ z_2 & z_3 & \cdots & z_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ z_k & z_{k+1} & \cdots & z_{2k-1} \end{bmatrix} \succeq a_1 a_2 \begin{bmatrix} z_0 & z_1 & \cdots & z_{k-1} \\ z_1 & z_2 & \cdots & z_k \\ \vdots & \vdots & \ddots & \vdots \\ z_{k-1} & z_k & \cdots & z_{2k-2} \end{bmatrix} + \begin{bmatrix} z_2 & z_3 & \cdots & z_k \\ z_3 & z_4 & \cdots & z_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ z_k & z_{k+1} & \cdots & z_{2k} \end{bmatrix}.$$

As shown in [9, 23], the (4.4) are sufficient and necessary conditions for  $z \in \mathcal{R}_{2k}(S)$ . So, if  $(\gamma^*, y^*, z^*)$  is a maximizer of (4.2), then  $M_k[z^*] \geq 0$  and  $L_g^{(k)}[z^*] \geq 0$ . Hence, we have  $z^* \in \mathcal{R}_{2k}(S)$  and hence  $y^* = z^*|_d \in \mathcal{R}_d(S)$ . □

Second, we prove the asymptotic convergence of Algorithm 4.4 when the random variable  $\xi$  is multi-variate. It requires that the quadratic module  $\text{QM}[g]$  is archimedean and (3.13) has interior points.

**Theorem 4.7** *Assume that  $\text{QM}[g]$  is archimedean and there exists a point  $\hat{x} \in X$  such that  $h(\hat{x}, \xi) = a_1(\xi) + a_2(\xi)$  with  $a_1 > 0$  on  $S$  and  $a_2 \in Y^*$ . Suppose  $(\gamma^{(k)}, y^{(k)}, z^{(k)})$  is an optimal triple of (4.2) when its relaxation order is  $k$ . Then, the sequence  $\{y^{(k)}\}_{k=1}^\infty$  is bounded and every accumulation point of  $\{y^{(k)}\}_{k=1}^\infty$  belongs to the cone  $\mathcal{R}_d(S)$ . Therefore, every accumulation point of  $\{(\gamma^{(k)}, y^{(k)})\}_{k=1}^\infty$  is a maximizer of (3.14).*

**Proof** For every  $(\gamma, y, z)$  that is feasible for (4.2) and for  $\hat{w} := [\hat{x}]_{2d_1}$ , it holds that

$$\begin{aligned} \langle f, \hat{w} \rangle - (\gamma - \langle b, y \rangle) &= \langle f - y^T A x - \gamma, \hat{w} \rangle + (A \hat{x} + b)^T y \\ &\geq (A \hat{x} + b)^T y. \end{aligned} \tag{4.5}$$

There exists  $\epsilon > 0$  such that  $a_1(\xi) - \epsilon \in \text{QM}[g]_{2k_0}$ , for some  $k_0 \in \mathbb{N}$ , since  $\text{QM}[g]$  is archimedean. Noting  $a_2 \in Y^*$ , one can see that

$$(A \hat{x} + b)^T y = \langle h(\hat{x}, \xi), y \rangle = \langle a_1(\xi), y \rangle + \langle a_2(\xi), y \rangle \geq \langle a_1(\xi), y \rangle.$$

For all  $k \geq k_0$ , it holds that

$$\langle a_1(\xi), y \rangle = \langle a_1(\xi) - \epsilon, y \rangle + \epsilon \langle 1, y \rangle \geq \epsilon \langle 1, y \rangle = \epsilon y_0.$$

(Note  $\langle 1, y \rangle = y_0$ .) Let  $f_2$  be the optimal value of (3.14), then

$$\gamma^{(k)} - \langle b, y^{(k)} \rangle \geq f_2,$$

because  $(\gamma^{(k)}, y^{(k)}, z^{(k)})$  is an optimizer of (4.2), and (4.2) is a relaxation of the maximization (3.14). So (4.5) implies that

$$(A \hat{x} + b)^T y^{(k)} \leq \langle f, \hat{w} \rangle - f_2.$$

Hence, we can get that

$$(y^{(k)})_0 \leq \frac{1}{\epsilon} (\langle f, \hat{w} \rangle - f_2).$$

The sequence  $\{(y^{(k)})_0\}_{k=1}^\infty$  is bounded.

Since  $\text{QM}[g]$  is archimedean, there exists  $N > 0$  such that  $N - \|\xi\|^2 \in \text{QM}[g]_{2k_1}$  for some  $k_1 \geq k_0$ . For all  $k \geq k_1$ , the membership  $z^{(k)} \in \mathcal{S}[g]_{2k}$  implies that

$$N \cdot (z^{(k)})_0 - ((z^{(k)})_{2e_1} + \dots + (z^{(k)})_{2e_p}) \geq 0.$$

Note that  $y^{(k)} = z^{(k)}|_d$ , hence  $(y^{(k)})_0 = (z^{(k)})_0$ . Since  $z^{(k)} \in \mathcal{S}[g]_{2k}$  and the sequence  $\{(z^{(k)})_0\}_{k=1}^\infty$  is bounded, one can further show that the set

$$\{z^{(k)}|_d : z \in \mathcal{S}[g]_{2k}\}_{k=1}^\infty$$

is bounded. We refer to [38, Theorem 4.3] for more details about the proof. Therefore, the sequence  $\{y^{(k)}\}_{k=1}^\infty$  is bounded. Since  $\text{QM}[g]$  is archimedean, we also have

$$\mathcal{R}_d(S) = \bigcap_{k=1}^\infty S_k, \quad \text{where } S_k := \{z|_d : z \in \mathcal{S}[g]_{2k}\}.$$

This is shown in Proposition 3.3 of [38]. So, if  $\hat{y}$  is an accumulation point of  $\{y^{(k)}\}_{k=1}^\infty$ , then we must have  $\hat{y} \in \mathcal{R}_d(S)$ . Similarly, if  $(\hat{\gamma}, \hat{y}, \hat{z})$  is an accumulation point of  $\{(\gamma^{(k)}, y^{(k)}, z^{(k)})\}_{k=1}^\infty$ , then  $\hat{y} \in \mathcal{R}_d(S)$ . As in the proof of Theorem 4.1, one can similarly show that  $(\hat{\gamma}, \hat{y})$  is a maximizer of (3.14).  $\square$

Last, we prove that Algorithm 4.4 will terminate within finitely many steps under certain assumptions. Like Theorem 4.7, we also assume the archimedeaness of  $\text{QM}[g]$ . When  $\text{QM}[g]$  is not archimedean, if the set  $S = \{\xi \in \mathbb{R}^p : g(\xi) \geq 0\}$  is bounded, we can replace  $g$  by  $\tilde{g} = (g, N - \|\xi\|^2)$  where  $N$  is such that  $S \subseteq \{\|\xi\|^2 \leq N\}$ . Then  $\text{QM}[\tilde{g}]$  is archimedean. Moreover, we also need to assume the strong duality between (3.13) and (3.14), which is guaranteed under the Slater’s condition for (3.14). These assumptions typically hold for polynomial optimization.

**Theorem 4.8** *Assume  $\text{QM}[g]$  is archimedean and there is no duality gap between (3.13) and (3.14). Suppose  $(x^*, w^*)$  is a minimizer of (3.13) and  $(\gamma^*, y^*)$  is a maximizer of (3.14) satisfying:*

- (i) *There exists  $k_1 \in \mathbb{N}$  such that  $h(x^*, \xi) = h_1(\xi) + h_2(\xi)$ , with  $h_1 \in \text{QM}[g]_{2k_1}$  and  $h_2 \in Y^*$ .*
- (ii) *The polynomial optimization problem in  $\xi$*

$$\begin{cases} \min_{\xi \in \mathbb{R}^p} h_1(\xi) \\ \text{s.t. } g_1(\xi) \geq 0, \dots, g_{m_1}(\xi) \geq 0 \end{cases} \tag{4.6}$$

*has finitely many critical points  $u$  such that  $h_1(u) = 0$ .*

*Then, when  $k$  is large enough, for every optimizer  $(\gamma^{(k)}, y^{(k)}, z^{(k)})$  of (4.2), we must have  $y^{(k)} \in \mathcal{R}_d(S)$ .*

**Proof** Since there is no duality gap between (3.13) and (3.14),

$$0 = \langle f, w^* \rangle - \langle \gamma^* - \langle b, y^* \rangle \rangle = \langle f - (y^*)^T Ax - \gamma^*, w^* \rangle + \langle Ax^* + b \rangle^T y^*.$$

Due to the feasibility constraints, we further have

$$\langle f(x) - (y^*)^T Ax - \gamma^*, w^* \rangle = 0, \quad \langle Ax^* + b \rangle^T y^* = 0.$$

Therefore, it holds that

$$\langle Ax^* + b \rangle^T y^* = \langle h(x^*, \xi), y^* \rangle = \langle h_1(\xi), y^* \rangle + \langle h_2(\xi), y^* \rangle = 0.$$

The conic membership  $y^* \in \overline{K}$  implies that

$$\langle h_1(\xi), y^* \rangle = \langle h_2(\xi), y^* \rangle = 0.$$

We consider the polynomial optimization problem (4.6) in the variable  $\xi$ . For each order  $k \geq k_1$ , the  $k$ th order Moment-SOS relaxation pair for solving (4.6) is

$$\min \langle h_1(\xi), z \rangle \quad \text{s.t. } z \in \mathcal{S}[g]_{2k}, z_0 = 1, \tag{4.7}$$

$$v_k := \max \gamma \quad \text{s.t. } h_1(\xi) - \gamma \in \text{QM}[g]_{2k}. \tag{4.8}$$

The archimedeaness of  $\text{QM}[g]$  implies that  $S$  is compact, so

$$\overline{\mathcal{R}_d(S)} = \mathcal{R}_d(S).$$



The membership  $y^* \in \overline{K}$  implies that  $y^* \in \mathcal{R}_d(S)$ . Since

$$\langle h_1(\xi), y^* \rangle = 0,$$

the polynomial  $h_1(\xi)$  vanishes on the support of each  $S$ -representing measure for  $y^*$ , so the optimal value of (4.6) is zero. By the given assumption, the sequence  $\{v_k\}$  has finite convergence to the optimal value 0 and the relaxation (4.8) achieves its optimal value for all  $k \geq k_1$ . The optimization (4.6) has only finitely many critical points that are global optimizers. So, Assumption 2.1 of [34] for the optimization (4.6) is satisfied. Moreover, the given assumption also implies that  $(x^*, w^*)$  is an optimizer of (4.1) and  $(\gamma^*, y^*, z^*)$  is an optimizer of (4.2) for all  $k \geq k_1$ . Suppose  $(x^{(k)}, w^{(k)})$  is an arbitrary optimizer of (4.1) and  $(\gamma^{(k)}, y^{(k)}, z^{(k)})$  is an arbitrary optimizer of (4.2), for the relaxation order  $k$ .

When  $(z^{(k)})_0 = 0$ , we have  $vec(1)^T M_k[z^{(k)}]vec(1) = 0$ . Since  $M_k[z^{(k)}] \succeq 0$ ,

$$M_k[z^{(k)}]vec(1) = 0.$$

Consequently, we further have  $M_k[z^{(k)}]vec(\xi^\alpha) = 0$  for all  $|\alpha| \leq k - 1$  (see Lemma 5.7 of [29]). Then, for each power  $\alpha = \beta + \eta$  with  $|\beta|, |\eta| \leq k - 1$ , one can get

$$(z^{(k)})_\alpha = vec(\xi^\beta)^T M_k[z^{(k)}]vec(\xi^\eta) = 0.$$

This means that  $z^{(k)}|_{2k-2}$  is the zero vector and hence  $y^{(k)} \in \mathcal{R}_d(S)$ .

For the case  $(z^{(k)})_0 > 0$ , let  $\hat{z} := z^{(k)}/(z^{(k)})_0$ . The given assumption implies that  $(x^*, w^*)$  is also a minimizer of (4.1) and  $(\gamma^*, y^*, z^*)$  is optimal for (4.2), for all  $k \geq k_1$ . So there is no duality gap between (4.1) and (4.2). Since  $(\gamma^{(k)}, y^{(k)}, z^{(k)})$  is optimal for (4.2), so  $\langle h_1(\xi), z^{(k)} \rangle = 0$  and hence  $\hat{z}$  is a minimizer of (4.7) for all  $k \geq k_1$ . By Theorem 2.2 of [35], the minimizer  $z^{(k)}$  must have a flat truncation  $z^{(k)}|_{2t}$  for some  $t$ , when  $k$  is sufficiently big. This means that the truncation  $z^{(k)}|_{2t}$ , as well as  $y^{(k)}$ , has a representing measure supported in  $S$ . Therefore, we have  $y^{(k)} \in \mathcal{R}_d(S)$ .  $\square$

The conclusion of Theorem 4.8 is guaranteed to hold under conditions (i) and (ii), which depend on the constraints  $g$  and the set  $Y$ . These two conditions are not convenient to verify computationally. However, in computational practice of Algorithm 4.4, there is no need to check or verify them. The correctness of computational results by Algorithm 4.4 does not depend on conditions (i) and (ii). In other words, the conditions (i) and (ii) are sufficient for Algorithm 4.4 to have finite convergence, but they may not be necessary. It is possible that the finite convergence occurs even if some of them fail to hold. In our numerical experiments, the finite convergence is always observed. We also like to remark that the conditions (i) and (ii) generally hold, which is a main topic of the work [36]. In particular, when  $h_1$  has generic coefficients, the optimization (4.6) has finitely many critical points and so the condition (ii) holds. This is shown in [34].

## 5 Numerical Experiments

In this section, we give numerical experiments for Algorithm 4.4 to solve distributionally robust optimization problems. The computation is implemented in MATLAB R2018a, in a Laptop with CPU 8th Generation Intel® Core™ i5-8250U and RAM 16 GB. The software GloptiPoly3 [20], YALMIP [30] and SeDuMi [48] are used for the implementation. For neatness of presentation, we only display four decimal digits.

To apply implement Algorithm 4.4, we need a computational representation for the cone  $cone(Y)$ . For a given set  $Y$ , it may be mathematically hard to get a computationally efficient

description for the closure of its conic hull. However, in most applications, the set  $Y$  is often convex and there usually exist convenient representations for  $\overline{\text{cone}(Y)}$ . For instance, the  $\overline{\text{cone}(Y)}$  is often a polyhedra, second order, or semidefinite cone, or a Cartesian product of them. The following are some frequently appearing cases.

- If  $Y = \{y : Ty + u \geq 0\}$  is a nonempty polyhedron, given by some matrix  $T$  and vector  $u$ , then

$$\overline{\text{cone}(Y)} = \{y : Ty + su \geq 0, s \in \mathbb{R}_+\}. \tag{5.1}$$

It is also a polyhedron and is closed.

- Consider that  $Y = \{y : \mathcal{A}(y) + B \geq 0\}$  is given by a linear matrix inequality, for a homogeneous linear symmetric matrix valued function  $\mathcal{A}$  and a symmetric matrix  $B$ . If  $Y$  is nonempty and bounded, then

$$\overline{\text{cone}(Y)} = \{y : \mathcal{A}(y) + sB \geq 0, s \in \mathbb{R}_+\}. \tag{5.2}$$

When  $Y$  is unbounded, the  $\text{cone}(Y)$  may not be closed and its closure  $\overline{\text{cone}(Y)}$  may be tricky. We refer to the work [33] for such cases. When  $Y$  is given by second order conic conditions, we can do similar things for obtaining  $\overline{\text{cone}(Y)}$ .

**Example 5.1** Consider the DROM problem

$$\begin{cases} \min_{x \in \mathbb{R}^4} & f(x) = -x_1 - 2x_2 - x_3 + 2x_4 \\ \text{s.t.} & \inf_{\mu \in \mathcal{M}} \mathbb{E}_\mu[h(x, \xi)] \geq 0, \\ & x \geq 0, 1 - e^T x \geq 0, \end{cases} \tag{5.3}$$

where (the random variable  $\xi$  is univariate, i.e,  $p = 1$ )

$$\begin{aligned} h(x, \xi) &= (x_4 - x_1 - 2)\xi^5 + (x_4 - 1)\xi^4 + (2x_1 + x_2 + x_4 + 1)\xi^3 \\ &\quad + (2x_1 - x_2 + x_4 - 1)\xi^2 + (2 - x_2 - x_3)\xi, \\ S &= [0, 3], \quad g = 3\xi - \xi^2, \end{aligned}$$

$$Y = \left\{ y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_5 \end{bmatrix} \in \mathbb{R}^6 \mid \begin{array}{l} 1 \leq y_0 \leq y_1 \leq y_2 \leq \\ y_3 \leq y_4 \leq y_5 \leq 2 \end{array} \right\}.$$

The  $\overline{\text{cone}(Y)}$  is given as in (5.1). The objective  $f$  and constraints  $c_1, c_2$  are all linear. We start with  $k = 3$ , and the Algorithm 4.4 terminates in the initial loop. The optimal value  $F^*$  and the optimizer  $x^*$  for (3.11) are respectively

$$F^* \approx -0.0326, \quad x^* \approx (0.6775, 0.0000, 0.0000, 0.3225).$$

The optimizer for (4.2) is

$$y^* \approx (0.9355, 0.9355, 0.9517, 1.0163, 1.2260, 1.8710).$$

The measure  $\mu$  for achieving  $y^* = \int[\xi]_5 d\mu$  is supported at the points

$$u_1 \approx 0.9913, \quad u_2 \approx 3.0000.$$

By a proper scaling, we get the measure  $\mu^* = 0.9957\delta_{u_1} + 0.0043\delta_{u_2}$  that achieves the worst case expectation constraint.

**Example 5.2** Consider the DROM problem

$$\left\{ \begin{array}{l} \min_{x \in \mathbb{R}^3} f(x) = (x_1 - x_3 + x_1x_3)^2 + (2x_2 + 2x_1x_2 - x_3^2)^2 \\ \text{s.t.} \quad \inf_{\mu \in \mathcal{M}} \mathbb{E}_\mu[h(x, \xi)] \geq 0, \\ c_1(x) = 1 - x_1^2 - x_2^2 - x_3^2 \geq 0, \\ c_2(x) = 3x_3 - x_1^2 - 2x_2^4 \geq 0, \end{array} \right. \tag{5.4}$$

where (the random variable  $\xi$  is bivariate, i.e.  $p = 2$ )

$$\begin{aligned} h(x, \xi) &= (1 - x_3)\xi_1^2\xi_2^2 + (x_1 - x_2 + x_3 - 1)\xi_1\xi_2^2 \\ &\quad + (x_1 + x_2 + x_3 + 1)\xi_2^2 + (x_1 - x_3)\xi_1^2 - \xi_2, \\ S &= \{\xi \in \mathbb{R}^2 : 1 - \xi^T\xi \geq 0\}, \quad g := 1 - \xi^T\xi, \\ Y &= \left\{ y \in \mathbb{R}^{\mathbb{N}_4^2} \left| \begin{array}{l} y_{00} = 1, \quad 0.1 \leq y_\alpha \leq 1 \quad (0 < |\alpha| \leq 4) \\ \begin{pmatrix} y_{20} & y_{11} & y_{30} & y_{12} \\ y_{11} & y_{02} & y_{21} & y_{03} \\ y_{30} & y_{21} & y_{40} & y_{22} \\ y_{12} & y_{03} & y_{22} & y_{04} \end{pmatrix} \preceq 2I_4 \end{array} \right. \right\}. \end{aligned}$$

The  $\overline{\text{cone}}(Y)$  is given as in (5.2). One can verify that  $f$  and all  $-c_i$  are SOS-convex. We start with  $k = 2$ , and Algorithm 4.4 terminates in the initial loop. The optimal value  $F^*$  and optimizer  $x^*$  of (3.11) are respectively

$$F^* \approx 0.0160, \quad x^* \approx (0.4060, 0.0800, 0.4706).$$

The optimizer for (4.2) is

$$y^* \approx (0.3180, 0.2750, 0.1411, 0.2436, 0.1137, 0.0744, 0.2199, 0.0950, 0.0552, 0.0460, 0.2011, 0.0819, 0.0426, 0.0318, 0.0318).$$

The measure  $\mu$  for achieving  $y^* = \int[\xi]_4 d\mu$  is supported at the points

$$u_1 \approx (0.6325, 0.7745), \quad u_2 \approx (0.9434, 0.3317).$$

By a proper scaling, we get the measure  $\mu^* = 0.2527\delta_{u_1} + 0.7473\delta_{u_2}$  that achieves the worst case expectation constraint.

**Example 5.3** Consider the DROM problem

$$\left\{ \begin{array}{l} \min_{x \in \mathbb{R}^3} f(x) = x_1^4 - 2x_1^2 + 2x_2^3 + x_3^4 \\ \text{s.t.} \quad \inf_{\mu \in \mathcal{M}} \mathbb{E}_\mu[h(x, \xi)] \geq 0, \\ c_1(x) = x_1^2 + x_2^2 + x_3^2 - 1 \geq 0, \\ c_2(x) = 4 - x_1^2 - 2x_2^2 - x_3 \geq 0, \end{array} \right. \tag{5.5}$$

where (the random variable  $\xi$  is bivariate, i.e,  $p = 2$ )

$$\begin{aligned} h(x, \xi) &= (x_1 + x_2 + 1)\xi_2^4 + (3x_1 + x_2)\xi_1^2\xi_2 + (x_1 + 2x_2 + x_3 + 1)\xi_1^3 \\ &\quad + 2x_1 + x_2 - 2x_3, \\ S &= \{\xi \in \mathbb{R}^2 : g := (\xi_1, \xi_2, 1 - e^T \xi) \geq 0\}, \\ Y &= \left\{ y \in \mathbb{R}^{\mathbb{N}_4^2} \mid \begin{array}{l} y_{00} = 1, 0.2^i \leq y_{i0} \leq 0.6^i, \\ y_{i0} \geq 1.2y_{0i}, i = 1, 2, 3, 4 \end{array} \right\}. \end{aligned}$$

In the above,  $\overline{\text{cone}(Y)}$  is given as in (5.1). The objective  $f$  and  $-c_1$  are not convex. We start with  $k = 2$ , while the algorithm terminates at  $k = 3$ . In the last loop, the optimizers for (4.1) and (4.2) are

$$\begin{aligned} w^* &\approx (1.0000, 0.2692, -1.5454, -0.8493, 0.0725, -0.4161, -0.2287, 2.3884, \\ &\quad 1.3125, 0.7213, 0.0195, -0.1120, -0.0616, 0.6430, 0.3534, 0.1942, \\ &\quad -3.6911, -2.0284, -1.1147, -0.6126, 0.0053, -0.0302, -0.0166, \\ &\quad 0.1731, 0.0951, 0.0523, -0.9938, -0.5461, -0.3001, -0.1649, \\ &\quad 5.7044, 3.1348, 1.7227, 0.9467, 0.5202), \\ y^* &\approx (0.0871, 0.0488, 0.0383, 0.0300, 0.0188, 0.0195, 0.0184, 0.0116, \\ &\quad 0.0073, 0.0122, 0.0113, 0.0071, 0.0045, 0.0028, 0.0094). \end{aligned}$$

The optimal value  $F^* \approx -7.0017$  for both of them. The measure for achieving  $y^* = \int[\xi]_4 d\mu$  is supported at the points

$$u_1 \approx (0.0000, 1.0000), \quad u_2 \approx (0.6139, 0.3861).$$

By a proper scaling, we get the measure  $\mu^* = 0.0877\delta_{u_1} + 0.9123\delta_{u_2}$  that achieves the worst case expectation constraint. The point

$$x^* = \pi(w^*) \approx (0.2692, -1.5454, -0.8493),$$

is feasible for (5.5) as  $c_1(x^*) \approx 2.1822$  and  $c_2(x^*) \approx 3.9919 \cdot 10^{-8}$ . Moreover,  $F^* - f(x^*) \approx 1.2204 \cdot 10^{-7}$ . By Theorem 4.2, we know  $F^*$  is the optimal value and  $x^*$  is an optimizer for (5.5).

**Example 5.4** Consider the DROM problem

$$\left\{ \begin{array}{l} \min_{x \in \mathbb{R}^3} f(x) = x_1^4 - x_1x_2x_3 + x_3^3 + 3x_1x_3 + x_2^2 \\ \text{s.t.} \quad \inf_{\mu \in \mathcal{M}} \mathbb{E}_\mu[h(x, \xi)] \geq 0, \\ c_1(x) = x_1x_2 - 0.25 \geq 0, \\ c_2(x) = 6 - x_1^2 - 4x_1x_2 - x_2^2 - x_3^2 \geq 0, \end{array} \right. \tag{5.6}$$

where (the random variable  $\xi$  is bivariate, i.e,  $p = 2$ )

$$\begin{aligned}
 h(x, \xi) &= (2 - x_1 + x_2)\xi_2^4 + (x_1 + x_3 + 1)\xi_1\xi_2^2 + (2 - x_1 + 2x_2)\xi_2^3 \\
 &\quad + (x_1 + 2x_2 + x_3 + 2)\xi_1^2 + (3x_2 - x_1)\xi_2^2, \\
 S &= \{\xi \in \mathbb{R}^2 \mid 1 \leq \xi^T \xi \leq 4\}, \quad g = (\xi^T \xi - 1, 4 - \xi^T \xi), \\
 Y &= \left\{ y \in \mathbb{R}^{\mathbb{N}_4^2} \mid y_{00} = 1, \sum_{|\alpha| \geq 1} y_\alpha^2 = 36 \right\}.
 \end{aligned}$$

The set  $Y$  is not convex. Its convex hull is  $\|y\| \leq \sqrt{37}$  with  $y_{00} = 1$ . Hence,

$$\overline{\text{cone}(Y)} = \left\{ y \in \mathbb{R}^{\mathbb{N}_4^2} \mid \|y\|_2 \leq \sqrt{37}y_{00} \right\}.$$

The functions  $f$  and  $-c_1, -c_2$  are not convex. We begin with  $k = 2$ . The optimizers for (4.1) and (4.2) are respectively

$$\begin{aligned}
 w^* &\approx (1.0000, 0.6790, 0.3682, -2.0984, 0.4611, 0.2500, -1.4249, 0.1356, -0.7726, \\
 &\quad 4.4034, 0.3131, 0.1698, -0.9675, 0.0920, -0.5246, 2.9900, 0.0499, -0.2845, \\
 &\quad 1.6212, -9.2402, 0.2126, 0.1153, -0.6569, 0.0625, -0.3562, 2.0302, 0.0339, \\
 &\quad -0.1932, 1.1008, -6.2742, 0.0184, -0.1047, 0.5969, -3.4021, 19.3898), \\
 y^* &\approx (1.2272, 0.2992, -1.1902, 0.0730, -0.2902, 1.1543, 0.0178, -0.0708, 0.2814, \\
 &\quad -1.1194, 0.0043, -0.0173, 0.0686, -0.2729, 1.0857).
 \end{aligned}$$

The optimal value is  $F^* \approx -12.6420$  for both of them. The measure for achieving  $y^* = \int [\xi]_4 \bar{\alpha} \mu$  is  $\mu = 1.2272\delta_u$ , with  $u \approx (0.2438, -0.9698) \in S$ . So  $\mu^* = \delta_u$ . For the point

$$x^* = \pi(w^*) \approx (0.6790, 0.3682, -2.0984),$$

one can verify that  $x^*$  is feasible for (5.6), since

$$c_1(x^*) \approx -1.6654 \cdot 10^{-9}, \quad c_2(x^*) \approx 5.6235 \cdot 10^{-8}, \quad F^* - f(x^*) \approx -7.7271 \cdot 10^{-8}.$$

By Theorem 4.2, we know  $x^*$  is the optimizer for (5.6).

**Example 5.5 (Portfolio selection [11, 22])** Consider that there exist  $n$  risky assets that can be chosen by the investor in the financial market. The uncertain loss  $r_i$  of each asset can be described by the random risk variable  $\xi$  which admits a probability measure supported in  $S = [0, 1]^p$ . Assume the moments of  $\mu \in \mathcal{M}$  are constrained in the set

$$Y = \left\{ y \in \mathbb{R}^{\mathbb{N}_3^3} \mid y_{000} = 1, 0.1 \leq y_\alpha \leq 1, |\alpha| \geq 1 \right\}.$$

The cone  $\overline{\text{cone}(Y)}$  can be given as in (5.1). Minimizing the portfolio loss over the ambiguity set  $\mathcal{M}$  is equivalent to solving the following min-max optimization problem

$$\min_{x \in \Delta_3} \max_{\mu \in \mathcal{M}} \mathbb{E}_\mu [x_1 r_1(\xi) + x_2 r_2(\xi) + x_3 r_3(\xi)], \tag{5.7}$$

for the simplex  $\Delta_n := \{x \in \mathbb{R}^3 \mid e^T x = 1, x \geq 0\}$ . The functions  $r_i(\xi)$  are

$$\begin{cases} r_1(\xi) = -1 + \xi_1 + \xi_1 \xi_2 - \xi_1 \xi_3 - 2\xi_1^3, \\ r_2(\xi) = -1 - \xi_1 \xi_2 + \xi_2^2 - \xi_2 \xi_3 + \xi_2^3, \\ r_3(\xi) = -1 + \xi_2 \xi_3 - \xi_3^2 - \xi_3^3. \end{cases} \tag{5.8}$$

Then (5.7) can be equivalently reformulated as

$$\begin{cases} \min_{(x_0, x) \in \mathbb{R} \times \mathbb{R}^3} & x_0 \\ \text{s.t.} & \inf_{\mu \in \mathcal{M}} \mathbb{E}_\mu [x_0 - (x_1 r_1(\xi) + x_2 r_2(\xi) + x_3 r_3(\xi))] \geq 0, \\ & x \geq 0, e^T x = 1. \end{cases} \quad (5.9)$$

Applying Algorithm 4.4 to solve (5.9), we get the optimal value  $F^*$  and the optimizer  $(x_0^*, x^*)$  in the initial loop  $k = 2$ :

$$F^* \approx -1.0136, \quad (x_0^*, x^*) \approx (-1.0136, 0.1492, 0.3501, 0.5007).$$

The optimizer for (4.2) is

$$y^* \approx (1.0000, 0.6077, 0.4440, 0.3725, 0.3864, 0.3347, 0.2530, \\ 0.4440, 0.2666, 0.1803, 0.2560, 0.2523, 0.1771, 0.3347, \\ 0.2010, 0.1306, 0.4440, 0.2666, 0.1601, 0.1000).$$

The measure for achieving  $y^* = \int [\xi]_4 d\mu$  is

$$\mu = 0.5560\delta_{u_1} + 0.4440\delta_{u_2},$$

with the following two points in  $S$ :

$$u_1 \approx (0.4911, -0.0000, 0.1905), \quad u_2 \approx (0.7538, 1.0000, 0.6005).$$

Since  $\mu$  belongs to  $\mathcal{M}$ , it is also the measure that achieves the worst case expectation constraint. Therefore, the optimizer for (5.7) is  $x^*$  and the optimal value is  $-1.0136$ .

**Example 5.6 (Newsvendor problem [50])** Consider that there is a newsvendor trade product with an uncertain daily demand. Assume the demand quantity  $D(\xi)$  is affected by a random variable  $\xi \in \mathbb{R}^2$  such that

$$D(\xi) = 2 - \xi_1 + \xi_2 - \xi_1^2 + 2\xi_2^2 + \xi_1^4.$$

In each day, the newsvendor orders  $x$  units of the product at the wholesale price  $P_1$ , sells the product with quantity  $\min\{x, D(\xi)\}$  at the retail price  $P_2$  and clears the unsold stock at the salvage price  $P_0$ . Assume that  $P_0 < P_1 < P_2$ , then the newsvendor’s daily loss is given as

$$l(x, \xi) := (P_1 - P_2)x + (P_2 - P_0) \cdot \max\{x - D(\xi), 0\}.$$

Clearly, the newsboy will earn the most if he can buy the greatest order quantity that is guaranteed to be sold out. Suppose  $\xi$  admits a probability measure supported in  $S$  and has its true distribution contained in the ambiguity set  $\mathcal{M}$ . Then the best order decision for the newsvendor product can be obtained from the following DROM problem

$$\begin{cases} \min_{x \in \mathbb{R}} & (P_1 - P_2)x \\ \text{s.t.} & \inf_{\mu \in \mathcal{M}} \mathbb{E}_\mu [D(\xi) - x] \geq 0, \\ & x \geq 0. \end{cases} \quad (5.10)$$

Suppose  $P_0 = 0.25, P_1 = 0.5, P_2 = 1$ , and

$$S = [0, 5]^2, \quad Y = \left\{ y \in \mathbb{R}^{\mathbb{N}_4^2} \mid \begin{array}{l} y_{00} = 1, 1 \leq y_{01} \leq y_{02} \leq 4 \\ 2^i \leq y_{i0} \leq 4^i, i = 1, 2, 3, 4 \end{array} \right\}.$$

The cone  $\overline{\text{cone}(Y)}$  can be given as in (5.1). Applying Algorithm 4.4 to solve (5.10), we get the optimal value  $F$  and the optimizer  $x^*$  respectively

$$F^* \approx -7.5000, \quad x^* \approx 15.0000.$$

The optimizer of (4.2) is

$$y^* \approx (0.5000, 1.0000, 0.5000, 2.0000, 1.0000, 0.5000, 4.0000, 2.0000, \\ 1.0000, 0.5000, 8.0000, 4.0000, 2.0000, 1.0000, 0.5000).$$

The measure for achieving  $y^* = \int [\xi]_4 d\mu$  is  $\mu = 0.5\delta_u$  with

$$u = (2.0000, 1.0000) \in S.$$

So  $\mu^* = \delta_u$  achieves the worst case expectation constraint.

We would like to remark that the ambiguity set  $\mathcal{M}$  can be constructed by samples or historic data. It can also be updated as the sampling size increases. Assume the support set  $S$  is given and each  $\mu \in \mathcal{M}$  is a probability measure. The moment ambiguity set  $Y$  can be estimated by statistical samplings. Suppose  $T = \{\xi^{(1)}, \dots, \xi^{(N)}\}$  is a given sample set for  $\xi$ . One can randomly choose  $T_1, \dots, T_s \subseteq T$  such that each  $T_i$  contains  $\lceil N/2 \rceil$  samples. Choose a smaller sample size  $s$ , say,  $s = 5$ . For a given degree  $d$ , choose the moment vectors  $l, u \in \mathbb{R}^{\mathbb{N}_d^n}$  such that

$$l_\alpha = \min_{j=1, \dots, s} \left\{ \frac{1}{|T_j|} \sum_{i \in T_j} (\xi^{(i)})^\alpha, \frac{1}{|T \setminus T_j|} \sum_{i \in T \setminus T_j} (\xi^{(i)})^\alpha \right\}, \\ u_\alpha = \max_{j=1, \dots, s} \left\{ \frac{1}{|T_j|} \sum_{i \in T_j} (\xi^{(i)})^\alpha, \frac{1}{|T \setminus T_j|} \sum_{i \in T \setminus T_j} (\xi^{(i)})^\alpha \right\}$$

for every power  $\alpha \in \mathbb{N}_d^n$ . The moment constraining set  $Y$ , e.g., as in Example 5.5, can be estimated as

$$Y = \{y \in \mathbb{R}^{\mathbb{N}_d^n} : l \leq y \leq u\}. \tag{5.11}$$

Other types of moment constraining set  $Y$  can be estimated similarly. Suppose each  $\xi^{(i)}$  independently follows the distribution of  $\xi$ . As the sample size  $N$  increases, the moment ambiguity set  $\mathcal{M}$  with  $Y$  in (5.11) is expected to give a better approximation of the true distribution of  $\xi$ . This is indicated by the Law of Large Numbers and the convergence results of sample average approximations. The following is an example for how to do this.

**Example 5.7** Consider the portfolio selection optimization problem as in Example 5.5. The DROM is (5.7), or equivalently (5.9). Assume each  $r_i(\xi)$  is given as in (5.8). Suppose  $\xi = (\xi_1, \xi_2, \xi_3)$  is the random variable, where each  $\xi_i$  is independently distributed. Assume  $\xi_1$  follows the uniform distribution on  $[0, 1]$ ,  $\xi_2$  follows the truncated standard normal distribution on  $[0, 1]$  and  $\xi_3$  follows the truncated exponential distribution with the mean value 0.5 on  $[0, 1]$ . We use the MATLAB commands `makedist` and `truncate` to generate samples of  $\xi$  with the sample size  $N \in \{50, 100, 200\}$ , and then construct  $Y$  as in (5.11) with  $s = 5$  and  $d = n = 3$ .

(i). When  $N = 50$ , we get that

$$l = (1.0000, 0.4354, 0.3779, 0.3873, 0.2757, 0.1916, 0.1872, \\ 0.1975, 0.1549, 0.2018, 0.2027, 0.1299, 0.1161, 0.1111, \\ 0.0848, 0.1025, 0.1193, 0.0801, 0.0866, 0.1207),$$

**Table 1** Computational results for Example 5.7

Case	$F^*$	$x^*$	$\theta$	$u_1, u_2, u_3$
(i)	-0.9711	$\begin{bmatrix} 0.0159 \\ 0.2824 \\ 0.7017 \end{bmatrix}$	$\begin{bmatrix} 0.0746 \\ 0.7297 \\ 0.1957 \end{bmatrix}$	$\begin{bmatrix} 0.0000 \\ 0.4555 \\ 0.1266 \end{bmatrix}, \begin{bmatrix} 0.5675 \\ 0.0000 \\ 0.1074 \end{bmatrix}, \begin{bmatrix} 0.8492 \\ 1.0000 \\ 0.3832 \end{bmatrix}$
(ii)	-0.9743	$\begin{bmatrix} 0.0163 \\ 0.2711 \\ 0.7125 \end{bmatrix}$	$\begin{bmatrix} 0.0749 \\ 0.7320 \\ 0.1931 \end{bmatrix}$	$\begin{bmatrix} 0.0000 \\ 0.5170 \\ 0.1563 \end{bmatrix}, \begin{bmatrix} 0.5898 \\ 0.0000 \\ 0.1103 \end{bmatrix}, \begin{bmatrix} 0.8103 \\ 1.0000 \\ 0.4343 \end{bmatrix}$
(iii)	-0.9749	$\begin{bmatrix} 0.0171 \\ 0.2829 \\ 0.6999 \end{bmatrix}$	$\begin{bmatrix} 0.0937 \\ 0.7096 \\ 0.1967 \end{bmatrix}$	$\begin{bmatrix} 0.0000 \\ 0.4666 \\ 0.1370 \end{bmatrix}, \begin{bmatrix} 0.5616 \\ 0.0000 \\ 0.1122 \end{bmatrix}, \begin{bmatrix} 0.8450 \\ 1.0000 \\ 0.4474 \end{bmatrix}$

$$u = (1.0000, 0.5803, 0.4606, 0.4808, 0.3938, 0.2696, 0.2579, 0.2838, 0.2109, 0.3293, 0.2913, 0.1870, 0.1821, 0.1662, 0.1091, 0.1793, 0.2027, 0.1235, 0.1361, 0.2560).$$

(ii). For  $N = 100$ , we get that

$$l = (1.0000, 0.4935, 0.3799, 0.4135, 0.3150, 0.1828, 0.2065, 0.1975, 0.1745, 0.2459, 0.2261, 0.1061, 0.1268, 0.0924, 0.0837, 0.1280, 0.1195, 0.0926, 0.1102, 0.1709),$$

$$u = (1.0000, 0.5882, 0.4545, 0.5182, 0.4156, 0.2529, 0.2838, 0.2833, 0.2294, 0.3545, 0.3178, 0.1768, 0.1941, 0.1565, 0.1242, 0.1844, 0.2035, 0.1451, 0.1570, 0.2716).$$

(iii). For  $N = 200$ , we get that

$$l = (1.0000, 0.4803, 0.4177, 0.4157, 0.3170, 0.1957, 0.2253, 0.2508, 0.1784, 0.2580, 0.2310, 0.1274, 0.1459, 0.1170, 0.0875, 0.1348, 0.1719, 0.0998, 0.1048, 0.1886),$$

$$u = (1.0000, 0.5647, 0.4698, 0.5137, 0.3939, 0.2712, 0.2738, 0.2883, 0.2250, 0.3387, 0.3097, 0.1904, 0.1950, 0.1662, 0.1300, 0.1889, 0.2062, 0.1396, 0.1510, 0.2470).$$

Applying Algorithm 4.4, we get the optimal value  $F^*$  and the optimizer  $(x_0^*, x^*)$  in the initial loop  $k = 2$  for each case. The computational results are given in Table 1. Since  $x_0^* = F^*$  and  $y^*$  admits a measure  $\mu = \theta_1\delta_{u_1} + \theta_2\delta_{u_2} + \theta_3\delta_{u_3}$ , we only list  $F^*$ ,  $\theta = (\theta_1, \theta_2, \theta_3)$  and  $u_1, u_2, u_3$  for convenience. As the sample size increases, the optimal value of  $F^*$  improves. This indicates that the ambiguity set can be estimated by sampling averages and the accuracy increases as the sampling size increases.



## 6 Conclusions and Discussions

This paper studies distributionally robust optimization when the ambiguity set is given by moment constraints. The DROM has a deterministic objective, some constraints on the decision variable and a worst case expectation constraint. The distributionally robust min-max optimization is a special case of DROM. The objective and constraints are assumed to be polynomial functions in the decision variable. Under the SOS-convexity assumption, we show that the DROM is equivalent to a linear conic optimization problem with moment constraints, as well as the psd polynomial conic condition. The Moment-SOS relaxation method (i.e., Algorithm 4.4) is proposed to solve the linear conic optimization. The method can deal with moments of any order. Moreover, it not only returns the optimal value and optimizers for the original DROM, but also gives the measure that achieves the worst case expectation constraint. Under some general assumptions (e.g., the archimedeaness), we proved the asymptotic and finite convergence of the proposed method (see Theorems 4.6, 4.7 and 4.8). Numerical examples, as well as some applications, are given to show how it solves DROM problems.

The distributionally robust optimization is attracting broad interests in various applications. There is much future work to do. In this paper, we assumed the random function  $h(x, \xi)$  is linear in the decision variable  $x$ . How can we solve the DROM if  $h(x, \xi)$  is not linear in  $x$ ? To prove the DROM (3.1) is equivalent to the linear conic optimization (3.13), we assumed the objective and constraints are SOS-convex. When they are not SOS-convex, how can we get equivalent linear conic optimization for (3.1)? They are important future work.

**Author Contributions** The authors have done analysis and computational results.

**Data Availability** The paper does not analyse or generate any datasets, because the work proceeds within a theoretical and mathematical approach.

## Declarations

**Conflict of interest** The authors have no relevant financial or non-financial interests to disclose.

**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

1. Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: Robust Optim., vol. 28. Princeton University Press, New Jersey (2009)
2. Ben-Tal, A., Den Hertog, D., et al.: Robust solutions of optimization problems affected by uncertain probabilities. *Manag. Sci.* **59**(2), 341–357 (2013)
3. Ben-Tal, A., Nemirovski, A.: Lectures on modern convex optimization. SIAM, Philadelphia, PA (2012)
4. Bertsimas, D., Brown, D., Caramanis, C.: Theory and applications of robust optimization. *SIAM Rev.* **53**(3), 464–501 (2011)

5. Bertsimas, D., Sim, M., Zhang, M.: Adaptive distributionally robust optimization. *Manag. Sci.* **65**(2), 604–618 (2019)
6. Birge, J., Louveaux, F.: Introduction to stochastic programming. Springer, Newyork (2011)
7. Chen, Y., Sun, H., Xu, H.: Decomposition and discrete approximation methods for solving two-stage distributionally robust optimization problems. *Comput. Optim. Appl.* **78**(1), 205–238 (2021)
8. Chen, Z., Sim, M., Xu, H.: Distributionally robust optimization with infinitely constrained ambiguity sets. *Oper. Res.* **67**(5), 1328–1344 (2019)
9. Curto, R., Fialkow, L.: Recursiveness, positivity, and truncated moment problems. *Houston J. Math.* **17**, 603–635 (1991)
10. Curto, R., Fialkow, L.: Truncated  $K$ -moment problems in several variables. *J. Oper. Theory* **54**, 189–226 (2005)
11. Delage, E., Ye, Y.: Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* **58**(3), 595–612 (2010)
12. Duchi, J., Namkoong, H.: Learning models with uniform performance via distributionally robust optimization, Preprint, (2018). [arXiv:1810.08750](https://arxiv.org/abs/1810.08750)
13. Esfahani, P., Kuhn, D.: Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Math. Program.* **171**(1–2), 115–166 (2018)
14. Ghadimi, S., Lan, G.: Stochastic first- and zeroth-order methods for nonconvex stochastic programming. *SIAM J. Optim.* **23**(4), 2341–2368 (2013)
15. Guo, B., Nie, J., Yang, Z.: Learning diagonal gaussian mixture models and incomplete tensor decompositions. *Vietnam J. Math.* **50**(2), 421–446 (2022)
16. Gürbüzbalaban, M., Ruszczyński, A., Zhu, L.: A stochastic subgradient method for distributionally robust non-convex learning, Preprint, (2020). [arXiv:2006.04873](https://arxiv.org/abs/2006.04873)
17. Hanasusanto, G., Roitch, V., et al.: A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Math. Program.* **151**(1), 35–62 (2015)
18. Helton, J.W., Nie, J.: Semidefinite representation of convex sets. *Math. Program.* **122**(1), 21–64 (2010)
19. Henrion, D., Lasserre, J.: Detecting global optimality and extracting solutions in GloptiPoly, Positive polynomials in control, 293–310. Springer, Berlin, Heidelberg (2005)
20. Henrion, D., Lasserre, J., Loeferberg, J.: GloptiPoly3: moments, optimization and semidefinite programming. *Optim. Methods Softw.* **24**(4–5), 761–779 (2009)
21. Henrion, D., Korda, M., Lasserre, J.: The moment-SOS hierarchy. World Scientific, Singapore (2020)
22. de Klerk, E., Kuhn, D., Postek, K.: Distributionally robust optimization with polynomial densities: theory, models and algorithms. *Math. Program.* **181**, 265–296 (2020)
23. Krein, M., Louvish, D.: The Markov moment problem and extremal problems. American Mathematical Society, Washington, D.C (1977)
24. Lan, G.: First-order and stochastic optimization methods for machine learning. Springer, Newyork (2020)
25. Lasserre, J.: Global optimization with polynomials and the problem of moments. *SIAM J. Optim.* **11**, 796–817 (2001)
26. Lasserre, J.: Convexity in semi-algebraic geometry and polynomial optimization. *SIAM J. Optim.* **19**, 1995–2014 (2009)
27. Lasserre, J.: Moment, Positive Polynomials and Their Applications. Imperial College Press, London (2009)
28. Lasserre, J., Weisser, T.: Distributionally robust polynomial chance-constraints under mixture ambiguity sets. *Math. Program.* **185**(1), 409–453 (2021)
29. Laurent, M.: Sums of squares, moment matrices and optimization over polynomials, Emerging Applications of Algebraic Geometry of IMA Volumes in Mathematics and its Applications 149, pp. 157–270, Springer, (2009)
30. Lofberg, J.: YALMIP: A toolbox for modeling and optimization in MATLAB, IEEE international conference on robotics and automation (IEEE Cat. No. 04CH37508), IEEE, (2004)
31. Mevissen, M., Ragnoli, E., Yu, J.: Data-driven distributionally robust polynomial optimization. *Adv. Neural Inf. Process. Syst.* **26**, 37–45 (2013)
32. Mohri, M., Sivek, G., Suresh A.: Agnostic federated learning, Preprint, (2019). [arXiv:1902.00146](https://arxiv.org/abs/1902.00146)
33. Netzer, T.: On semidefinite representations of non-closed sets. *Linear Algebra Appl.* **432**, 3072–3078 (2010)
34. Nie, J.: Certifying convergence of Lasserre’s hierarchy via flat truncation. *Math. Program.* **142**(1–2), 485–510 (2013)
35. Nie, J.: The  $\mathcal{A}$ -truncated  $K$ -moment problem. *Found. Comput. Math.* **14**(6), 1243–1276 (2014)
36. Nie, J.: Optimality conditions and finite convergence of Lasserre’s hierarchy. *Math. Program.* **146**(1–2), 97–121 (2014)

37. Nie, J.: The hierarchy of local minimums in polynomial optimization. *Math. Program.* **151**(2), 555–583 (2015)
38. Nie, J.: Linear optimization with cones of moments and nonnegative polynomials. *Math. Program.* **153**(1), 247–274 (2015)
39. Nie, J., Yang, L., Zhong, S.: Stochastic polynomial optimization. *Optim. Methods Softw.* **35**(2), 329–347 (2020)
40. Pflug, G., Wozabal, D.: Ambiguity in portfolio selection. *Quantitative Finance* **7**(4), 435–442 (2007)
41. Postek, K., Ben-Tal, A., et al.: Robust optimization with ambiguous stochastic constraints under mean and dispersion information. *Oper. Res.* **66**(3), 814–833 (2018)
42. Putinar, M.: Positive polynomials on compact semi-algebraic sets. *Indiana Univ. Math. J.* **42**(3), 969–984 (1993)
43. Rachev, S., Römisch, W.: Quantitative stability in stochastic programming: the method of probability metrics. *Math. Oper. Res.* **27**(4), 792–818 (2002)
44. Rahimian, H., Mehrotra, S.: Distributionally robust optimization: A review, Preprint, (2019). [arXiv:1908.05659](https://arxiv.org/abs/1908.05659)
45. Ruszczyński, A., Shapiro, A., (Eds.), *Stochastic programming models*, Handbook in Operations Research and Management Science, Elsevier Science, Amsterdam, (2003)
46. Shapiro, A., Ahmed, S.: On a class of minimax stochastic programs. *SIAM J. Optim.* **14**(4), 1237–1249 (2004)
47. Shapiro, A., Dentcheva, D., Ruszczyński, A.: *Lectures on stochastic programming: modeling and theory*. SIAM, Philadelphia (2014)
48. Sturm, J.: Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Methods Softw.* **11**(1–4), 625–653 (1999)
49. Sun, H., Shapiro, A., Chen, X.: Distributionally robust stochastic variational inequalities. *Math. Program.* (2022). <https://doi.org/10.1007/s10107-022-01889-2>
50. Wiesemann, W., Kuhn, D., Sim, M.: Distributionally robust convex optimization. *Oper. Res.* **62**(6), 1358–1376 (2014)
51. Xu, H., Liu, Y., Sun, H.: Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods. *Math. Program.* **169**(2), 489–529 (2018)
52. Yang, Y., Wu, W.: A distributionally robust optimization model for real-time power dispatch in distribution networks. *IEEE Trans. Smart Grid* **10**(4), 3743–3752 (2018)
53. Zhang, J., Xu, H., Zhang, L.: Quantitative stability analysis for distributionally robust optimization with moment constraints. *SIAM Optim.* **26**(3), 1855–1862 (2016)
54. Zhang, Z., Ahmed, S., Lan, G.: Efficient algorithms for distributionally robust stochastic optimization with discrete scenario support, Preprint, (2019). [arXiv:1909.11216](https://arxiv.org/abs/1909.11216)
55. Zhu, S., Fukushima, Masao: Worst-case conditional value-at-risk with application to robust portfolio management. *Oper. Res.* **57**(5), 1155–1168 (2009)
56. Zymler, S., Kuhn, D., Rustem, B.: Distributionally robust joint chance constraints with second-order moment information. *Math. Program.* **137**(1–2), 167–198 (2013)