



**HAL**  
open science

# Distributionally robust polynomial chance-constraints under mixture ambiguity sets

Jean B Lasserre, Tillmann Weisser

► **To cite this version:**

Jean B Lasserre, Tillmann Weisser. Distributionally robust polynomial chance-constraints under mixture ambiguity sets. *Mathematical Programming, Series A*, 2021, 185, pp.409–453. 10.1007/s10107-019-01434-8. hal-01755147v2

**HAL Id: hal-01755147**

**<https://laas.hal.science/hal-01755147v2>**

Submitted on 21 Nov 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# DISTRIBUTIONALLY ROBUST POLYNOMIAL CHANCE-CONSTRAINTS UNDER MIXTURE AMBIGUITY SETS

JEAN B. LASSERRE AND TILLMANN WEISSER

ABSTRACT. Given  $\mathbf{X} \subset \mathbb{R}^n$ ,  $\varepsilon \in (0, 1)$ , a parametrized family of probability distributions  $(\mu_{\mathbf{a}})_{\mathbf{a} \in \mathbf{A}}$  on  $\Omega \subset \mathbb{R}^p$ , we consider the feasible set  $\mathbf{X}_\varepsilon^* \subset \mathbf{X}$  associated with the *distributionally robust* chance-constraint

$$\mathbf{X}_\varepsilon^* = \{\mathbf{x} \in \mathbf{X} : \text{Prob}_\mu[f(\mathbf{x}, \omega) > 0] > 1 - \varepsilon, \forall \mu \in \mathcal{M}_{\mathbf{a}}\},$$

where  $\mathcal{M}_{\mathbf{a}}$  is the set of all possible mixtures of distributions  $\mu_{\mathbf{a}}$ ,  $\mathbf{a} \in \mathbf{A}$ . For instance and typically, the family  $\mathcal{M}_{\mathbf{a}}$  is the set of all mixtures of Gaussian distributions on  $\mathbb{R}$  with mean and standard deviation  $\mathbf{a} = (a, \sigma)$  in some compact set  $\mathbf{A} \subset \mathbb{R}^2$ . We provide a sequence of inner approximations  $\mathbf{X}_\varepsilon^d = \{\mathbf{x} \in \mathbf{X} : w_d(\mathbf{x}) < \varepsilon\}$ ,  $d \in \mathbb{N}$ , where  $w_d$  is a polynomial of degree  $d$  whose vector of coefficients is an optimal solution of a semidefinite program. The size of the latter increases with the degree  $d$ . We also obtain the strong and highly desirable asymptotic guarantee that  $\lambda(\mathbf{X}_\varepsilon^* \setminus \mathbf{X}_\varepsilon^d) \rightarrow 0$  as  $d$  increases, where  $\lambda$  is the Lebesgue measure on  $\mathbf{X}$ . Same results are also obtained for the more intricate case of distributionally robust “joint” chance-constraints.

## 1. INTRODUCTION

**Motivation.** In many optimization and control problems uncertainty is often modeled by a noise  $\omega \in \Omega \subset \mathbb{R}^p$  (following some probability distribution  $\mu$ ), which interacts with the decision variable of interest<sup>1</sup>  $\mathbf{x} \in \mathbf{X} \subset \mathbb{R}^n$  via some feasibility constraint of the form  $f(\mathbf{x}, \omega) > 0$  for some function  $f : \mathbf{X} \rightarrow \mathbb{R}$ . In the *robust* approach one imposes the constraint  $\mathbf{x} \in \mathbf{X}^R := \{\mathbf{x} \in \mathbf{X} : f(\mathbf{x}, \omega) > 0, \forall \omega \in \Omega\}$  on the decision variable  $\mathbf{x}$ . However, sometimes the resulting set  $\mathbf{X}^R$  of robust decisions can be quite small or even empty.

On the other hand, if one knows the probability distribution  $\mu$  of the noise  $\omega \in \Omega$ , then a more appealing *probabilistic* approach is to tolerate a violation of the feasibility constraint  $f(\mathbf{x}, \omega) > 0$ , provided that this violation occurs with small probability  $\varepsilon > 0$ , fixed *à priori*. That is, one imposes the less conservative chance-constraint  $\text{Prob}_\mu(f(\mathbf{x}, \omega) > 0) > 1 - \varepsilon$ , which results in the larger “feasible set” of decisions

$$(1.1) \quad \mathbf{X}_\varepsilon^\mu := \{\mathbf{x} \in \mathbf{X} : \text{Prob}_\mu(f(\mathbf{x}, \omega) > 0) > 1 - \varepsilon\}.$$

In both the robust and probabilistic cases, handling  $\mathbf{X}^R$  or  $\mathbf{X}_\varepsilon^\mu$  can be quite challenging and one is interested in respective approximations that are easier to handle. There is a rich literature on chance-constrained programming since Charnes and

---

\*The work of the two authors was supported by the European Research Council (ERC) via an ERC-Advanced Grant for the # 666981 project TAMING.

<sup>1</sup>As quoted from R. Henrion, *the biggest challenge from the algorithmic and theoretical points of view arise in chance constraints where the random and decision variables cannot be decoupled*. <https://www.stoprog.org/what-stochastic-programming>

Cooper [8], Miller [36], and the interested reader is referred to Henrion [18, 19], Dabbene [5], Li et al. [34], Prékopa [39] and Shapiro [41] for a general overview of chance constraints in optimization and control. Of particular interest are uncertainty models and methods that allow to define tractable approximations of  $\mathbf{X}_\varepsilon^\mu$ . Therefore an important issue is to analyze under which conditions on  $f$ ,  $\mu$  and the threshold  $\varepsilon$ , the resulting chance constraint (1.1) defines a convex set  $\mathbf{X}_\varepsilon^\mu$ ; see e.g. Henrion and Strugarek [19], van Ackooij [1], Nemirovski and Shapiro [38], Wang et al. [42], and the recent work of van Ackooij and Malick [2]. For instance, in [2] the authors consider joint chance-constraint ( $f(\mathbf{x}, \omega) \in \mathbb{R}^k$ ) and show that  $\mathbf{X}_\varepsilon^\mu$  is convex for sufficiently small  $\varepsilon$  if  $\omega$  is an elliptical random vector and  $f$  is convex in  $\mathbf{x}$ . A different approach to model chance constraints was considered by the first author in [27]. It uses the general Moment-Sums-of-Squares (Moment-SOS) methodology described in [30]. Related earlier work by Jasour et al. [23] have also used this Moment-SOS approach to solve some control problems with probabilistic constraints.

However, one well-sounded critic to these probabilistic approaches is that it relies on the knowledge of the exact distribution  $\mu$  of the noise  $\omega$ , which in many cases may not be a realistic assumption. Therefore modeling the uncertainty via a single *known* probability distribution  $\mu$  is questionable, and may render the chance-constraint  $\text{Prob}_\mu(f(\mathbf{x}, \omega) > 0) > 1 - \varepsilon$  in (1.4) counter-productive. It would rather make sense to assume a *partial knowledge* on the *unknown* distribution  $\mu$  of the noise  $\omega \in \Omega$ .

To overcome this drawback, *distributionally robust* chance-constrained problems consider probabilistic constraints that must hold for a whole family of distributions and typically the family is characterized by the support and first and second-order moments; see for instance Delage and Ye [11], Edogan and Iyengar [14], and Zymler et al. [51]. For instance Calafiore and El Ghaoui [6] have shown that when  $f$  is bilinear then a tractable characterization via second-order cone constraints is possible. Recently Yang and Xu [49] have considered non linear optimization problems where the constraint functions are concave in the decision variables and quasi-convex in the uncertain parameters. They show that such problems are tractable if the uncertainty is characterized by its mean and variance only; in the same spirit see also Chao Duan et al. [7], Chen et al. [9], Hanasusanto et al. [16, 17], Tong et al. [45], Wang et al. [42], Xie and Ahmed [46, 47] and Zhang et al. [50] for other tractable formulations of distributionally robust chance-constrained for optimal power flow problems.

**The uncertainty framework.** We also consider a framework where only partial knowledge of the uncertainty is available. But instead of assuming knowledge of some moments like in e.g. [49], we assume that typically the distribution  $\mu$  can be *any mixture* of probability distributions  $\mu_{\mathbf{a}} \in \mathcal{P}(\Omega)$  for some family  $\{\mu_{\mathbf{a}}\}_{\mathbf{a} \in \mathbf{A}} \subset \mathcal{P}(\Omega)$  that depend on a parameter vector  $\mathbf{a} \in \mathbf{A} \subset \mathbb{R}^t$ . That is:

$$\mu(B) = \int_{\mathbf{A}} \mu_{\mathbf{a}}(B) d\varphi(\mathbf{a}), \quad \forall B \in \mathcal{B}(\Omega),$$

where  $\varphi \in \mathcal{P}(\mathbf{A})$  can be any probability distribution on  $\mathbf{A}$ . If  $d\mu_{\mathbf{a}}(\omega) = u(\omega, \mathbf{a})$  for some density  $u$ , then by Fubini-Tonelli's Theorem [13, Theorem p. 85], the above measure  $\mu$  is well-defined. For instance, for Value-at-Risk optimization (where  $f$  is bilinear) El Ghaoui et al. [15] suggest mixtures of Gaussian measures, see [15,

§1.3], and more generally, families of probability measures with convex uncertainty sets on first and second-order moments<sup>2</sup>.

Notice that in this framework *no* mean, variance or higher order moments have to be estimated. Hence it can be viewed as an alternative and/or a complement to those considered in e.g. [11, 6, 14, 49] when a good estimation of such moments is not possible. Indeed in many cases, providing a box (where the mean vector can lie) and a possible range  $\underline{\delta} \mathbf{I} \preceq \Sigma \preceq \bar{\delta} \mathbf{I}$  for the covariance matrix  $\Sigma$ , can be more realistic than providing a single mean vector and a single covariance matrix. Below are some examples of possible mixtures. In particular it should be noted that mixtures of Gaussian measures (Ex. 1.1) are used in statistics precisely because of their high modeling power. Indeed they can approximate quite well a large family of distributions of interest in applications; see e.g. Dizioa et al. [12], Marron and Wand [35], Wang et al. [44], and the recent survey of Xu [48]. Similarly, the family of measures on  $\Omega$  with SOS-densities (Ex. 1.7) discussed in de Klerk et al. [10] also have a great modeling power with nice properties.

**Example 1.1** (Mixtures of Gaussian's).  $\Omega = \mathbb{R}$ ,  $\mathbf{a} = (a, \sigma) \in \mathbf{A} := [a, \bar{a}] \times [\underline{\sigma}, \bar{\sigma}] \subset \mathbb{R}^2$ , with  $\underline{\sigma} > 0$ , and

$$d\mu_{\mathbf{a}}(\omega) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\omega - a)^2}{2\sigma^2}\right) d\omega,$$

that is  $\mu$  is a mixture of Gaussian probability measures with mean-deviation couple  $(a, \sigma) \in \mathbf{A}$ .

**Example 1.2** (Mixtures of Exponential's).  $\Omega = \mathbb{R}_+$ ,  $\mathbf{a} = a \in \mathbf{A} := [a, \bar{a}] \subset \mathbb{R}$  with  $a > 0$ , and

$$d\mu_{\mathbf{a}}(\omega) = \frac{1}{a} \exp(-\omega/a) d\omega,$$

that is,  $\mu$  is a mixture of exponential probability measures with parameter  $1/a$ ,  $a \in \mathbf{A}$ .

**Example 1.3** (Mixtures of elliptical's). In [2] the authors have considered chance-constraints for the class of elliptical random vectors. In our framework and in the univariate case,  $\Omega = \mathbb{R}$ ,  $\mathbf{a} = (a, \sigma) \in \mathbf{A} := [a, \bar{a}] \times [\underline{\sigma}, \bar{\sigma}] \subset \mathbb{R}^2$ , with  $\underline{\sigma} > 0$ . Let  $\theta : \mathbb{R}^+ \rightarrow \mathbb{R}$  be such that  $\int_0^\infty t^k d\theta(t) < \infty$  for all  $k$ , and let  $p = \int_{\mathbb{R}} \theta(t^2) dt$ . Then:

$$d\mu_{\mathbf{a}}(\omega) = \frac{1}{p\sigma} \theta\left(\frac{(\omega - a)^2}{\sigma^2}\right) d\omega, \quad \mathbf{a} \in \mathbf{A}.$$

**Example 1.4** (Mixtures of Poisson's).  $\Omega = \mathbb{N}$  and  $\mathbf{a} = a \in \mathbf{A} := [a, \bar{a}] \subset \mathbb{R}$  with  $a > 0$ , and

$$\mu_{\mathbf{a}}(k) = \exp(-a) \frac{a^k}{k!}, \quad k = 0, 1, \dots; \quad \mathbf{a} \in \mathbf{A},$$

that is,  $\mu$  is a mixture of Poisson probability measures with parameter  $a \in \mathbf{A}$ .

**Example 1.5** (Mixtures of Binomial's). Let  $\Omega = \{0, 1, \dots, N\}$  and  $\mathbf{a} = a \in \mathbf{A} := [a, \bar{a}] \subset [0, 1]$ , and

$$\mu_{\mathbf{a}}(k) = \binom{N}{k} a^k (1-a)^{N-k}, \quad k = 0, 1, \dots, N.$$

---

<sup>2</sup>This more general uncertainty framework can also be analyzed with our approach (hence with arbitrary polynomial  $f$ ).

**Example 1.6.** With  $\Omega = \mathbb{R}$  one is given a finite family of probability measures  $(\mu_i)_{i=1,\dots,p} \subset \mathcal{P}(\Omega)$ . Then  $\mathbf{a} = a \in \mathbf{A} := \{1, \dots, p\} \subset \mathbb{R}$ , and

$$d\mu(\omega) = \sum_{a=1}^p \lambda_a d\mu_a(\omega); \quad \sum_{a \in \mathbf{A}} \lambda_a = 1; \quad \lambda_a \geq 0,$$

that is,  $\mu$  is a finite convex combination of the probability measures  $(\mu_{\mathbf{a}})$ .

**Example 1.7** (Mixtures of SOS densities). Recently, de Klerk et al. [10] have introduced measures with SOS densities to model some distributionally robust optimization problems because of their high modeling power. This family also fits our framework. Let  $\phi$  be a reference measure on  $\Omega$  with known moments. Then  $\mathbf{A} = \Sigma[\mathbf{x}]_d$  and  $d\mu_{\mathbf{a}}(\omega) = p(\omega) d\phi(\omega)$  for some  $p \in \Sigma[\omega]_d$  such that  $\int_{\Omega} p d\phi = 1$ , where  $\Sigma[\omega]_d$  is the space of SOS polynomials of degree at most  $2d$ . A measure  $\mu_{\mathbf{a}} \in \mathcal{M}_{\mathbf{a}}$  is parametrized by the Gram matrix  $\mathbf{a} = \mathbf{X} \succeq 0$  of its density  $p \in \Sigma[\omega]_d$ . For illustration purpose consider the univariate case  $\Omega \subset \mathbb{R}$  and  $p \in \Sigma[\omega]_2$ . Then

$$p(\omega) = \begin{bmatrix} 1 \\ \omega \\ \omega^2 \end{bmatrix}^T \cdot \begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{12} & X_{22} & X_{23} \\ X_{13} & X_{23} & X_{33} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ \omega \\ \omega^2 \end{bmatrix}.$$

and for every integer  $k$ , the moment  $\int_{\Omega} \omega^k d\mu_{\mathbf{a}}(\omega) = \int_{\Omega} \omega^k p(\omega) d\phi(\omega)$  is a linear function of the coefficients of  $p$  and hence of the parameter  $\mathbf{a} = \mathbf{X}$ . If  $\mathbf{M}_{\phi}$  denotes the moment matrix of the reference measure  $\phi$  on  $\Omega$ , then  $\mathbf{A} = \{\mathbf{X} = \mathbf{X}^T \in \mathbb{R}^{3 \times 3} : \mathbf{X} \succeq 0; \langle \mathbf{M}_{\phi}, \mathbf{X} \rangle = 1\}$  is a compact and convex basic semi-algebraic set.<sup>3</sup> Additional bound constraints on moments are easily included as linear constraints on the coefficients of  $p$ .

All families  $\mathcal{M}_{\mathbf{a}}$  described above share an important property that is crucial for our purpose: Namely, all moments of  $\mu_{\mathbf{a}} \in \mathcal{M}_{\mathbf{a}}$  are *polynomials* in the parameter  $\mathbf{a} \in \mathbf{A}$ . That is, for each  $\beta$ ,  $\int_{\Omega} \omega^{\beta} d\mu_{\mathbf{a}}(\omega) = p_{\beta}(\mathbf{a})$  for some polynomial  $p_{\beta} \in \mathbb{R}[\mathbf{a}]$ .

**Remark 1.8.** Another possible and related ambiguity set is to consider the family of measures  $\mu$  on  $\Omega$  whose first and second-order moments  $\mathbf{a} = (\mathbf{m}, \boldsymbol{\sigma})$  belong to some prescribed set  $\mathbf{A}$ . The resulting ambiguity set has been already used in several contributions like e.g. [9, 15, 16, 17, 51]; however, as discussed in [10], the resulting ambiguity set might be too overly conservative (even in the case where  $\mathbf{A}$  is the singleton  $\{(\mathbf{m}, \boldsymbol{\sigma})\}$ ). The methodology developed in this paper also applies with some ad-hoc modification; see §5.1.

In this uncertainty framework one now has to consider the set:

$$(1.2) \quad \mathcal{M}_{\mathbf{a}} := \left\{ \int_{\mathbf{A}} d\mu_{\mathbf{a}}(\omega) \varphi(d\mathbf{a}) : \varphi \in \mathcal{P}(\mathbf{A}) \right\},$$

where  $\mathcal{P}(\mathbf{A})$  is the set of probability measures on  $\mathbf{A}$ , and in a *distributionally robust chance-constraint approach*, with  $\varepsilon > 0$  fixed, one considers the set:

$$(1.3) \quad \mathbf{X}_{\varepsilon}^* := \{ \mathbf{x} \in \mathbf{X} : \text{Prob}_{\mu}(f(\mathbf{x}, \omega) > 0) > 1 - \varepsilon, \quad \forall \mu \in \mathcal{M}_{\mathbf{a}} \},$$

as new feasible set of decision variables. In general  $\mathbf{X}_{\varepsilon}^*$  is non-convex and can even be disconnected. Therefore obtaining accurate approximations of  $\mathbf{X}_{\varepsilon}^*$  is a difficult

<sup>3</sup>Write the characteristic polynomial of  $\mathbf{X}$  in the form  $t \mapsto t^3 - p_2(\mathbf{X})t^2 + p_1(\mathbf{X})t - p_0(\mathbf{X})$  with  $p_2, p_1, p_0 \in \mathbb{R}[\mathbf{X}]$ . Then  $\mathbf{A} = \{\mathbf{X} : p_i(\mathbf{X}) \geq 0, i = 0, 1, 2; \langle \mathbf{M}_{\phi}, \mathbf{X} \rangle = 1\}$ .

challenge. Our ultimate goal is to replace optimization problems in the general form:

$$(1.4) \quad \min_{\mathbf{x}} \{ u(\mathbf{x}) : \mathbf{x} \in C \cap \mathbf{X}_\varepsilon^* \},$$

(where  $u$  is a polynomial) with

$$(1.5) \quad \min_{\mathbf{x}} \{ u(\mathbf{x}) : \mathbf{x} \in C \cap W \},$$

where the uncertain parameter  $\omega$  has disappeared from the description of a suitable inner approximation  $W$  of  $\mathbf{X}_\varepsilon^*$ , with  $W := \{ \mathbf{x} \in \mathbf{X} : w(\mathbf{x}) \leq 0 \}$  for some polynomial  $w$ . So if  $C$  is a basic semi-algebraic set<sup>4</sup> then (1.5) is a standard polynomial optimization problem. Of course the resulting optimization problem (1.5) may still be hard to solve because the set  $C \cap W$  is not convex in general. But this may be the price to pay for avoiding a too conservative formulation of the problem. However, since in formulation (1.5) one has got rid of the disturbance parameter  $\omega$ , one may then apply the arsenal of Non Linear Programming algorithms to get a local minimizer of (1.5). If  $n$  is not too large or if some sparsity is present in problem (1.5) one may even run a hierarchy of semidefinite relaxations to approximate its global optimal value; for more details on the latter, the interested reader is referred to [30].

**Contribution.** We consider approximating  $\mathbf{X}_\varepsilon^*$  as a challenging mathematical problem and explore whether it is possible to solve it under minimal assumptions on  $f$  and/or  $\mathcal{M}_\mathbf{a}$ . Thus the focus is more on the existence and definition of a prototype “algorithm” (or approximation scheme) rather than on its “scalability”. Of course the latter issue of scalability is important for practical applications and we hope that the present contribution will provide insights on how to develop more “tractable” (but so more conservative) versions.

Our contribution is *not* in the line of research concerned with “tractable approximations” of (1.3) (e.g. under some restrictions on  $f$  and/or the family  $\mathcal{M}_\mathbf{a}$ ) and should be viewed as complementary to previously cited contributions whose focus is on scalability.

Given  $\varepsilon > 0$  fixed, a family  $\mathcal{M}_\mathbf{a}$  (e.g. as in Examples 1.1, 1.2, 1.3, 1.4, 1.5, 1.6) and an arbitrary polynomial  $f$ , our main contribution is to provide rigorous and accurate inner approximations  $(\mathbf{X}_\varepsilon^d)_{d \in \mathbb{N}} \subset \mathbf{X}_\varepsilon^*$  of the set  $\mathbf{X}_\varepsilon^*$ , that converge to  $\mathbf{X}_\varepsilon^*$  in a precise sense as  $d$  increases. More precisely:

(i) We provide a nested sequence of *inner* approximations of the set  $\mathbf{X}_\varepsilon^*$  in (1.3), in the form:

$$(1.6) \quad \mathbf{X}_\varepsilon^d := \{ \mathbf{x} \in \mathbf{X} : w_d(\mathbf{x}) < \varepsilon \}, \quad d \in \mathbb{N},$$

where  $w_d$  is a polynomial of degree at most  $d$ , and  $\mathbf{X}_\varepsilon^d \subset \mathbf{X}_\varepsilon^{d+1} \subset \mathbf{X}_\varepsilon^*$  for every  $d$ .

(ii) We obtain the strong and highly desirable asymptotic guarantee:

$$(1.7) \quad \lim_{d \rightarrow \infty} \lambda(\mathbf{X}_\varepsilon^* \setminus \mathbf{X}_\varepsilon^d) = 0,$$

where  $\lambda$  is the Lebesgue measure on  $\mathbf{X}$ . To the best of our knowledge it is the first result of this kind at this level of generality. Importantly, the “volume” convergence

---

<sup>4</sup>A basic semi-algebraic set of  $\mathbb{R}^n$  is of the form  $\{ \mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geq 0, j \in J \}$  for finitely many polynomials  $(g_j)_{j \in J}$ .

(1.7) is obtained with *no* assumption of convexity on the set  $\mathbf{X}_\varepsilon^*$  (and indeed in general  $\mathbf{X}_\varepsilon^*$  is not convex).

(iii) Last but not least, the same approach is valid with same conclusions for the more intricate case of *joint* chance-constraints, that is, probabilistic constraints of the form

$$\text{Prob}_\mu(f_j(\mathbf{x}, \omega) > 0, \quad j = 1, \dots, s_f) > 1 - \varepsilon, \quad \forall \mu \in \mathcal{M}_\mathbf{a},$$

(for some given polynomials  $(f_j) \subset \mathbb{R}[\mathbf{x}, \omega]$ ). Remarkably, such constraints which are notoriously difficult to handle in general, are relatively easy to incorporate in our formulation.

We emphasize that our approach is a non trivial extension of the numerical scheme proposed in [27] for approximating  $\mathbf{X}_\varepsilon^*$  when  $\mathcal{M}_\mathbf{a}$  is a singleton (i.e., for standard chance-constraints).

**Methodology.** The approach that we propose for determining the set  $\mathbf{X}_\varepsilon^d$  defined in (1.6) is very similar in spirit to that in [21] and [23], and a non trivial extension of the more recent work [27] where only a single distribution  $\mu$  is considered. It is an additional illustration of the versatility of the Generalized Moment Problem (GMP) model and the moment-SOS approach outside the field of optimization. Indeed we also define an infinite-dimensional LP problem  $\mathbf{P}$  in an appropriate space of measures and a sequence of semidefinite relaxations  $(\mathbf{P}_d)_{d \in \mathbb{N}}$  of  $\mathbf{P}$ , whose associated monotone sequence of optimal values  $(\rho_d)_{d \in \mathbb{N}}$  converges to the optimal value  $\rho$  of  $\mathbf{P}$ . An optimal solution of this LP is a measure  $\phi^*$  on  $\mathbf{X} \times \Omega$ . In its disintegration  $\hat{\phi}^*(d\omega|\mathbf{x}) d\mathbf{x}$ , the conditional probability  $\hat{\phi}^*(d\omega|\mathbf{x})$  is a measure  $\mu_{\mathbf{a}(\mathbf{x})}$ , for some  $\mathbf{a}(\mathbf{x}) \in \mathbf{A}$ , which identifies the worst-case distribution at  $\mathbf{x} \in \mathbf{X}$ .

At an optimal solution of the dual of the semidefinite relaxation  $(\mathbf{P}_d)$ , we obtain a polynomial  $w_d$  of degree  $2d$  whose sub-level set  $\{\mathbf{x} \in \mathbf{X} : w_d(\mathbf{x}) < 0\}$  is precisely the desired approximation  $\mathbf{X}_\varepsilon^d$  of  $\mathbf{X}_\varepsilon^*$  in (1.3); in fact the sets  $(\mathbf{X}_\varepsilon^d)_{d \in \mathbb{N}}$  provide a sequence of *inner* approximations of  $\mathbf{X}_\varepsilon^*$ .

As in [27], the support  $\Omega$  of  $\mu \in \mathcal{M}_\mathbf{a}$  and the set  $\{(\mathbf{x}, \omega) : f(\mathbf{x}, \omega) > 0\}$  are *not* required to be compact, which includes the important case where  $\mu$  can be a mixture of normal or exponential distributions.

As already mentioned, our methodology is *not* a straightforward extension of the work in [27] where  $\mathcal{M}_\mathbf{a}$  is a singleton and the moments of this distribution are assumed to be known. Indeed in the present framework and in contrast to the singleton case treated in [27], we do *not* know a sequence of moments because we do not know the exact distribution  $\mu$  of the noise  $\omega$ . For instance, some measurability issues (e.g. existence of a measurable selector) not present in [27], arise. Also and in contrast to [27], we cannot define outer approximations by passing to the complement of  $\mathbf{X}_\varepsilon^*$ .

Importantly, we also describe how to accelerate the convergence of our approximation scheme. It consists of adding additional constraints in our relaxation scheme, satisfied at every feasible solution of the infinite-dimensional LP. These additional constraints come from a specific application of Stokes' theorem in the spirit of its earlier application in [27] but more intricate and not as a direct extension. Indeed, in the framework of our infinite-dimensional LP, it is required to define a measure with support on  $\mathbf{X} \times \Omega \times \mathbf{A}$  (instead of  $\mathbf{X} \times \Omega$  in [27]) which when passing to relaxations of the LP, results in semidefinite programs of larger size

(hence more difficult to solve). However, this price to pay can be profitable because the resulting convergence is expected to be significantly faster (as experienced in other contexts).

**Pros and cons.** On a positive side, our approach solves a difficult and challenging mathematical problem as it provides a nested hierarchy of inner approximations  $(\mathbf{X}_\varepsilon^d)_{d \in \mathbb{N}}$  of  $\mathbf{X}_\varepsilon^*$  which converges to  $\mathbf{X}_\varepsilon^*$  as  $d$  increases, a highly desirable feature. Also for every (and especially small)  $d$ , whenever not empty the set  $\mathbf{X}_\varepsilon^d$  is a valid (perhaps very conservative if  $d$  is small) inner approximation of  $\mathbf{X}_\varepsilon^*$  which can be exploited in applications if needed.

On a negative side, this methodology is computationally expensive, especially to obtain a very good inner approximation  $\mathbf{X}_\varepsilon^d$  of  $\mathbf{X}_\varepsilon^*$ . Therefore and so far, for accurate approximations this approach is limited to relatively small size problems. But again, recall that this approximation problem is a difficult challenge and at least, our approach with strong asymptotic guarantees provides insights and indications on possible routes to follow if one wishes to scale the method to address larger size problems. For instance, an interesting issue not discussed here is to investigate whether sparsity patterns already exploited in polynomial optimization (e.g. as in Waki et al. [43]) can be exploited in this context.

## 2. NOTATION, DEFINITIONS AND PRELIMINARIES

**2.1. Notation and definitions.** Let  $\mathbb{R}[\mathbf{x}]$  be the ring of polynomials in the variables  $\mathbf{x} = (x_1, \dots, x_n)$  and  $\mathbb{R}[\mathbf{x}]_d$  be the vector space of polynomials of degree at most  $d$  whose dimension is  $s(d) := \binom{n+d}{n}$ . For every  $d \in \mathbb{N}$ , let  $\mathbb{N}_d^n := \{\alpha \in \mathbb{N}^n : |\alpha| (= \sum_i \alpha_i) \leq d\}$ , and let  $\mathbf{v}_d(\mathbf{x}) = (\mathbf{x}^\alpha)$ ,  $\alpha \in \mathbb{N}_d^n$ , be the vector of monomials, i.e., the canonical basis  $(\mathbf{x}^\alpha)_{|\alpha| \leq d}$  of  $\mathbb{R}[\mathbf{x}]_d$ . A polynomial  $p \in \mathbb{R}[\mathbf{x}]_d$  can be written

$$\mathbf{x} \mapsto p(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha \mathbf{x}^\alpha = \langle \mathbf{p}, \mathbf{v}_d(\mathbf{x}) \rangle,$$

for some vector of coefficients  $\mathbf{p} = (p_\alpha) \in \mathbb{R}^{s(d)}$ . For a real symmetric matrix  $\mathbf{A}$  the notation  $\mathbf{A} \succeq 0$  (resp.  $\mathbf{A} \succ 0$ ) stands for  $\mathbf{A}$  is positive semidefinite (psd) (resp. positive definite (pd)). Denote by  $\Sigma[\mathbf{x}]_d$  the convex cone of polynomials that are sums-of-squares (SOS) of degree at most  $2d$ , i.e.,

$$p \in \Sigma[\mathbf{x}]_d \iff p(\mathbf{x}) = \sum_k p_k(\mathbf{x})^2, \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

for finitely many polynomials  $p_k \in \mathbb{R}[\mathbf{x}]_d$ . The convex cone  $\Sigma[\mathbf{x}]_d$  is semidefinite representable. Indeed:  $p \in \Sigma[\mathbf{x}]_d$  if and only if there exists a real symmetric metric  $\mathbf{X} \succeq 0$  of size  $s(d)$  such that  $p_\alpha = \langle \mathbf{X}, \mathbf{B}_\alpha \rangle$  for all  $\alpha \in \mathbb{N}^n$ , and where the symmetric matrices  $(\mathbf{B}_\alpha)_{\alpha \in \mathbb{N}_{2d}^n}$  come from writing:

$$\mathbf{v}_d(\mathbf{x}) \mathbf{v}_d(\mathbf{x})^T = \sum_{\alpha \in \mathbb{N}_{2d}^n} \mathbf{x}^\alpha \mathbf{B}_\alpha.$$

For more details the interested reader is referred to [30].

Given a closed set  $\mathcal{X} \subset \mathbb{R}^\ell$ , denote by  $\mathcal{B}(\mathcal{X})$  the Borel  $\sigma$ -field of  $\mathcal{X}$ ,  $\mathcal{P}(\mathcal{X})$  the space of probability measures on  $\mathcal{X}$  and by  $\mathcal{B}(\mathcal{X})$  the space of bounded measurable functions on  $\mathcal{X}$ . We also denote by  $\mathcal{M}(\mathcal{X})$  the space of finite signed Borel measures on  $\mathcal{X}$  and by  $\mathcal{M}_+(\mathcal{X})$  its subset of finite (positive) measures on  $\mathcal{X}$ .



**Moment matrix.** Given a sequence  $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ , let  $L_{\mathbf{y}} : \mathbb{R}[\mathbf{x}] \rightarrow \mathbb{R}$  be the linear (Riesz) functional

$$f (= \sum_{\alpha} f_{\alpha} \mathbf{x}^{\alpha}) \mapsto L_{\mathbf{y}}(f) := \sum_{\alpha} f_{\alpha} y_{\alpha}.$$

Given  $\mathbf{y}$  and  $d \in \mathbb{N}$ , the *moment matrix* associated with  $\mathbf{y}$ , is the real symmetric  $s(d) \times s(d)$  matrix  $\mathbf{M}_d(\mathbf{y})$  with rows and columns indexed in  $\mathbb{N}_d^n$  and with entries

$$\mathbf{M}_d(\mathbf{y})(\alpha, \beta) := L_{\mathbf{y}}(\mathbf{x}^{\alpha+\beta}) = y_{\alpha+\beta}, \quad \alpha, \beta \in \mathbb{N}_d^n.$$

Equivalently  $\mathbf{M}_d(\mathbf{y}) = L_{\mathbf{y}}(\mathbf{v}_d(\mathbf{x})\mathbf{v}_d(\mathbf{x})^T)$  where  $L_{\mathbf{y}}$  is applied entrywise.

**Example 2.1.** For illustration, consider the case  $n = 2$ ,  $d = 2$ . Then:

$$\mathbf{M}_2(\mathbf{y}) = \begin{pmatrix} y_{00} & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{pmatrix}.$$

A sequence  $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n}$  has a *representing measure*  $\mu$  if  $y_\alpha = \int \mathbf{x}^\alpha d\mu$  for all  $\alpha \in \mathbb{N}^n$ ; if  $\mu$  is unique then  $\mu$  is said to be *moment determinate*.

A necessary condition for the existence of such a  $\mu$  is that  $\mathbf{M}_d(\mathbf{y}) \succeq 0$  for all  $d$ . Except for in the univariate case  $n = 1$ , this is only a necessary condition. However the following sufficient condition in [30, Theorem 3.13] is very useful:

**Lemma 2.2.** ([30]) If  $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n}$  satisfies  $\mathbf{M}_d(\mathbf{y}) \succeq 0$  for all  $d = 0, 1, \dots$  and

$$(2.1) \quad \sum_{k=1}^{\infty} L_{\mathbf{y}}(x_i^{2k})^{-1/2k} = +\infty, \quad i = 1, \dots, n,$$

then  $\mathbf{y}$  has a representing measure, and in addition  $\mu$  is moment determinate.

Condition (2.1) due to Nussbaum is the multivariate generalization of its earlier univariate version due to Carleman; see e.g. [30].

**Localizing matrix.** Given a sequence  $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ , and a polynomial  $g \in \mathbb{R}[\mathbf{x}]$ , the *localizing moment matrix* associated with  $\mathbf{y}$  and  $g$ , is the real symmetric  $s(d) \times s(d)$  matrix  $\mathbf{M}_d(g\mathbf{y})$  with rows and columns indexed in  $\mathbb{N}_d^n$  and with entries

$$\begin{aligned} \mathbf{M}_d(g\mathbf{y})(\alpha, \beta) &:= L_{\mathbf{y}}(g(\mathbf{x}) \mathbf{x}^{\alpha+\beta}) \\ &= \sum_{\gamma} g_{\gamma} y_{\alpha+\beta+\gamma}, \quad \alpha, \beta \in \mathbb{N}_d^n. \end{aligned}$$

Equivalently  $\mathbf{M}_d(g\mathbf{y}) = L_{\mathbf{y}}(g(\mathbf{x})\mathbf{v}_d(\mathbf{x})\mathbf{v}_d(\mathbf{x})^T)$  where  $L_{\mathbf{y}}$  is applied entrywise.

**Example 2.3.** For illustration, consider the case  $n = 2$ ,  $d = 1$ . Then the localizing matrix associated with  $\mathbf{y}$  and  $g = x_1 - x_2$ , is:

$$\mathbf{M}_1(g\mathbf{y}) = \begin{pmatrix} y_{10} - y_{01} & y_{20} - y_{11} & y_{11} - y_{02} \\ y_{20} - y_{11} & y_{20} - y_{21} & y_{21} - y_{12} \\ y_{11} - y_{02} & y_{21} - y_{12} & y_{12} - y_{02} \end{pmatrix}.$$

**Disintegration.** Given a probability measure  $\mu$  on a cartesian product  $\mathcal{X} \times \mathcal{Y}$  of topological spaces, we may decompose  $\mu$  into its marginal  $\mu_{\mathbf{x}}$  on  $\mathcal{X}$  and a stochastic kernel (or conditional probability measure)  $\hat{\mu}(d\mathbf{y}|\cdot)$  on  $\mathcal{Y}$  given  $\mathcal{X}$ , that is:

- For every  $\mathbf{x} \in \mathcal{X}$ ,  $\hat{\mu}(d\mathbf{y}|\mathbf{x}) \in \mathcal{P}(\mathcal{Y})$ , and
- For every  $B \in \mathcal{B}(\mathcal{Y})$ , the function  $\mathbf{x} \mapsto \hat{\mu}(B|\mathbf{x})$  is measurable.

Then

$$\mu(A \times B) = \int_A \hat{\mu}(B|\mathbf{x}) \mu_{\mathbf{x}}(d\mathbf{x}), \quad \forall A \in \mathcal{B}(\mathcal{X}), B \in \mathcal{B}(\mathcal{Y}).$$

**2.2. The family  $\mathcal{M}_{\mathbf{a}}$ .** Let  $\Omega \subset \mathbb{R}^p$  be the “noise” (or disturbance) space. Let  $\mathbf{A} \subset \mathbb{R}^t$  be a compact set and for every  $\mathbf{a} \in \mathbf{A}$ , let  $\mu_{\mathbf{a}} \in \mathcal{P}(\Omega)$ .

**Assumption 2.4.** *The set  $\{\mu_{\mathbf{a}} : \mathbf{a} \in \mathbf{A}\} \subset \mathcal{P}(\Omega)$  satisfies the following:*

- (i) *For every  $B \in \mathcal{B}(\Omega)$ , the function  $\mathbf{a} \mapsto \mu_{\mathbf{a}}(B)$  is measurable.*
- (ii) *For every  $\beta \in \mathbb{N}$ :*

$$(2.2) \quad \int_{\Omega} \omega^{\beta} d\mu_{\mathbf{a}}(\omega) = p_{\beta}(\mathbf{a}), \quad \forall \mathbf{a} \in \mathbf{A},$$

*for some polynomial  $p_{\beta} \in \mathbb{R}[\mathbf{a}]$ .*

- (iii) *For every  $\mathbf{a} \in \mathbf{A}$  and every polynomial  $g \in \mathbb{R}[\omega]$ ,  $\mu_{\mathbf{a}}(\{\omega : g(\omega) = 0\}) = 0$ .*

- (iv) *For every bounded measurable (resp. bounded continuous) function  $q$  on  $\mathbf{X} \times \Omega$ , the function*

$$(\mathbf{x}, \mathbf{a}) \mapsto Q(\mathbf{x}, \mathbf{a}) := \int_{\Omega} q(\mathbf{x}, \omega) d\mu_{\mathbf{a}}(\omega),$$

*is bounded measurable (resp. bounded continuous) on  $\mathbf{X} \times \mathbf{A}$ .*

For instance if  $d\mu_{\mathbf{a}}(\omega) = \theta(\mathbf{a}, \omega) d\omega$  for some measurable density  $\theta(\mathbf{a}, \cdot)$  on  $\Omega$ , then Assumption 2.4(i) follows from Fubini-Tonelli’s Theorem [13], moreover Assumption 2.4(iii) is also satisfied. Assumption 2.4(ii) is satisfied in all Examples 1.1-1.7, as well as in their multivariate extensions. Assumption 2.4(iv) is satisfied in Example 1.1, 1.2, 1.3, 1.6, 1.7, and their natural multivariate extensions. For instance, in Example 1.1 where  $\mathcal{M}_{\mathbf{a}}$  is the set of all possible mixtures of univariate Gaussian probability distributions with mean and standard deviation  $(a, \sigma) \in \mathbf{A}$ , the function

$$Q(\mathbf{x}, \mathbf{a}) := \int_{\Omega} q(\mathbf{x}, \omega) d\mu_{\mathbf{a}} = \frac{1}{\sqrt{2\pi}} \int_{\Omega} q(\mathbf{x}, (\sigma\omega + a)) \exp(-\omega^2/2) d\omega,$$

is bounded measurable (resp. continuous) in  $\mathbf{a} \in \mathbf{A}$  whenever  $q$  is bounded measurable (resp. continuous) on  $\mathbf{X} \times \Omega$ .

If the disturbance space  $\Omega$  is non-compact, we need in addition:

**Assumption 2.5.** *(If  $\Omega \subset \mathbb{R}^p$  is unbounded):*

*There exists  $c, \gamma > 0$  such that for every  $i = 1, \dots, p$ :*

$$(2.3) \quad \sup_{\mathbf{a} \in \mathbf{A}} \int_{\Omega} \exp(c|\omega_i|) d\mu_{\mathbf{a}}(\omega) < \gamma.$$

Note that this assumption is satisfied in Example 1.1, 1.2, 1.3, 1.6, 1.7, and their natural multivariate extensions.

**Definition 2.6.** *The set  $\mathcal{M}_{\mathbf{a}} \subset \mathcal{P}(\Omega)$  is the space of all possible mixtures of probability measures  $\mu_{\mathbf{a}}$ ,  $\mathbf{a} \in \mathbf{A}$ . That is,  $\mu \in \mathcal{M}_{\mathbf{a}}$  if and only if there exists  $\varphi \in \mathcal{P}(\mathbf{A})$  such that*

$$(2.4) \quad \mu(B) = \int_{\mathbf{A}} \mu_{\mathbf{a}}(B) d\varphi(\mathbf{a}), \quad \forall B \in \mathcal{B}(\Omega).$$

(By Assumption 2.4(i),  $\mu$  is well-defined.) In particular  $\mu_{\mathbf{a}} \in \mathcal{M}_{\mathbf{a}}$  for all  $\mathbf{a} \in \mathbf{A}$ .

Let  $\mathbf{X} \subset \mathbb{R}^n$  and  $\mathbf{A} \subset \mathbb{R}^t$  be compact basic semi-algebraic sets and let  $\lambda$  be the Lebesgue measure on  $\mathbf{X}$ , scaled to be a probability measure on  $\mathbf{X}$ . We assume that  $\mathbf{X}$  is simple enough so that all moments of  $\lambda$  are easily calculated or available in closed-form. Typically  $\mathbf{X}$  is a box, a simplex, an ellipsoid, etc. The set  $\Omega \subset \mathbb{R}^p$  is also a basic semi-algebraic set not necessarily compact (for instance it can be  $\mathbb{R}^p$  or the positive orthant  $\mathbb{R}_+^p$ ).

**Definition 2.7.** Given a measure  $\psi$  on  $\mathbf{X} \times \mathbf{A}$  define  $\psi'$  on  $\mathbf{X} \times \mathbf{A} \times \Omega$  by:

$$\psi'(B, C, D) = \int_{B \times C} \mu_{\mathbf{a}}(D) d\psi(\mathbf{x}, \mathbf{a}), \quad B \in \mathcal{B}(\mathbf{X}), C \in \mathcal{B}(\mathbf{A}), D \in \mathcal{B}(\Omega),$$

which is well defined by Assumption 2.4(i). Its marginal  $\psi'_{\mathbf{x}, \mathbf{a}}$  on  $\mathbf{X} \times \mathbf{A}$  is  $\psi$  and  $\hat{\psi}'(\cdot | \mathbf{x}, \mathbf{a}) = \mu_{\mathbf{a}}$ , for all  $(\mathbf{x}, \mathbf{a}) \in \mathbf{X} \times \mathbf{A}$ .

Recall that  $\mathcal{B}(\mathbf{X} \times \Omega)$  is the space of bounded measurable functions on  $\mathbf{X} \times \Omega$ . Define the linear mapping  $\mathbf{T} : \mathcal{B}(\mathbf{X} \times \Omega) \rightarrow \mathcal{B}(\mathbf{X} \times \mathbf{A})$  by:

$$(2.5) \quad g \mapsto \mathbf{T}g(\mathbf{x}, \mathbf{a}) := \int_{\Omega} g(\mathbf{x}, \omega) d\mu_{\mathbf{a}}(\omega), \quad \mathbf{a} \in \mathbf{A},$$

which is well-defined by Assumption 2.4(iv). Therefore one may define the *adjoint* linear mapping  $\mathbf{T}^* : \mathcal{M}(\mathbf{X} \times \mathbf{A}) \rightarrow \mathcal{M}(\mathbf{X} \times \Omega)$  by:

$$\begin{aligned} \langle g, \mathbf{T}^* \psi \rangle &= \left( \int_{\mathbf{X} \times \Omega} g(\mathbf{x}, \omega) d\mathbf{T}^* \psi(\mathbf{x}, \omega) \right) = \langle \mathbf{T}g, \psi \rangle \\ &= \int_{\mathbf{X} \times \mathbf{A}} \left( \int_{\Omega} g(\mathbf{x}, \omega) d\mu_{\mathbf{a}}(\omega) \right) d\psi(\mathbf{x}, \mathbf{a}), \end{aligned}$$

for all  $g \in \mathcal{B}(\mathbf{X} \times \Omega)$  and all  $\psi \in \mathcal{M}(\mathbf{X} \times \mathbf{A})$ .

**Lemma 2.8.** Let  $\mathbf{T}$  be as in (2.5). Then for every  $\psi \in \mathcal{M}_+(\mathbf{X} \times \mathbf{A})$ ,  $\mathbf{T}^* \psi = \psi'_{\mathbf{x}, \omega}$  where  $\psi'_{\mathbf{x}, \omega}$  is the marginal on  $\mathbf{X} \times \Omega$  of the measure  $\psi' \in \mathcal{M}_+(\mathbf{X} \times \mathbf{A} \times \Omega)$  in Definition 2.7.

*Proof.* Let  $g \in \mathcal{B}(\mathbf{X} \times \Omega)$  be fixed. Then:

$$\begin{aligned} \langle g, \mathbf{T}^* \psi \rangle &= \langle \mathbf{T}g, \psi \rangle = \int_{\mathbf{X} \times \mathbf{A}} \left( \int_{\Omega} g(\mathbf{x}, \omega) \mu_{\mathbf{a}}(d\omega) \right) d\psi(\mathbf{x}, \mathbf{a}) \\ &= \int_{\mathbf{X} \times \mathbf{A}} \left( \int_{\Omega} g(\mathbf{x}, \omega) \hat{\psi}'(d\omega | \mathbf{x}, \mathbf{a}) \right) d\psi(\mathbf{x}, \mathbf{a}) \\ &= \int_{\mathbf{X} \times \mathbf{A} \times \Omega} g(\mathbf{x}, \omega) d\psi'(\mathbf{x}, \mathbf{a}, \omega) \\ &= \int_{\mathbf{X} \times \Omega} g(\mathbf{x}, \omega) d\psi'_{\mathbf{x}, \omega}(\mathbf{x}, \omega) = \langle g, \psi'_{\mathbf{x}, \omega} \rangle. \end{aligned}$$

As this holds for every  $g \in \mathcal{B}(\mathbf{X} \times \Omega)$ , it follows that  $\mathbf{T}^* \psi = \psi'_{\mathbf{x}, \omega}$ .  $\square$

**Lemma 2.9.** Let Assumption 2.4 hold. Then the mapping  $\mathbf{T}$  in (2.5) extends to polynomials. Moreover,  $\mathbf{T}(\mathbb{R}[\mathbf{x}, \omega]) \subset \mathbb{R}[\mathbf{x}, \mathbf{a}]$  and

$$(2.6) \quad \langle h, \mathbf{T}^* \psi \rangle = \langle \mathbf{T}h, \psi \rangle, \quad \forall h \in \mathbb{R}[\mathbf{x}, \omega], \forall \psi \in \mathcal{M}_+(\mathbf{X} \times \mathbf{A}).$$

*Proof.* Let  $h \in \mathbb{R}[\mathbf{x}, \omega]$  be fixed, arbitrary and write  $h(\mathbf{x}, \omega) = \sum_{\alpha, \beta} h_{\alpha, \beta} \mathbf{x}^\alpha \omega^\beta$ . Then by Assumption 2.4(ii):

$$\mathbf{T}h(\mathbf{x}, \mathbf{a}) := \sum_{\alpha, \beta} h_{\alpha, \beta} \mathbf{x}^\alpha \int_{\Omega} \omega^\beta d\mu_{\mathbf{a}} = \sum_{\alpha, \beta} h_{\alpha, \beta} \mathbf{x}^\alpha p_\beta(\mathbf{a}) \in \mathbb{R}[\mathbf{x}, \mathbf{a}].$$

Hence  $\langle \mathbf{T}h, \psi \rangle = \int_{\mathbf{X} \times \mathbf{A}} \sum_{\alpha, \beta} h_{\alpha, \beta} \mathbf{x}^\alpha p_\beta(\mathbf{a}) d\psi(\mathbf{x}, \mathbf{a})$ . Next, recalling the definition of  $\psi'$  and  $\psi'_{\mathbf{x}, \omega}$ :

$$\begin{aligned} \int_{\mathbf{X} \times \mathbf{A}} \sum_{\alpha, \beta} h_{\alpha, \beta} \mathbf{x}^\alpha p_\beta(\mathbf{a}) d\psi(\mathbf{x}, \mathbf{a}) &= \int_{\mathbf{X} \times \mathbf{A}} \left( \int_{\Omega} h(\mathbf{x}, \omega) \mu_{\mathbf{a}}(d\omega) \right) d\psi(\mathbf{x}, \mathbf{a}) \\ &= \int_{\mathbf{X} \times \Omega \times \mathbf{A}} h(\mathbf{x}, \omega) d\psi'(\mathbf{x}, \omega, \mathbf{a}) \\ &= \int_{\mathbf{X} \times \Omega} h(\mathbf{x}, \omega) d\psi'_{\mathbf{x}, \omega}(\mathbf{x}, \omega) \\ &= \int_{\mathbf{X} \times \Omega} h(\mathbf{x}, \omega) d\mathbf{T}^* \psi = \langle h, \mathbf{T}^* \psi \rangle, \end{aligned}$$

which yields (2.6).  $\square$

### 3. AN IDEAL INFINITE-DIMENSIONAL LP PROBLEM

Let  $\mathbf{X} \subset \mathbb{R}^n$ ,  $\Omega \subset \mathbb{R}^p$  and  $\mathbf{A} \subset \mathbb{R}^t$  be basic semi-algebraic sets. The sets  $\mathbf{X}$  and  $\mathbf{A}$  are assumed to be compact. The ambiguity set  $\mathcal{M}_{\mathbf{a}}$  is defined in (1.2). Let  $f \in \mathbb{R}[\mathbf{x}, \omega]$  be a given polynomial and with  $\varepsilon > 0$  fixed, consider the set  $\mathbf{X}_\varepsilon^*$  defined in (1.2). Let

$$(3.1) \quad \mathbf{K} := \{(\mathbf{x}, \omega) \in \mathbf{X} \times \Omega : f(\mathbf{x}, \omega) \leq 0\}$$

$$(3.2) \quad \mathbf{K}_{\mathbf{x}} := \{\omega \in \Omega : (\mathbf{x}, \omega) \in \mathbf{K}\}, \quad \mathbf{x} \in \mathbf{X}.$$

**3.1. Basic idea and link with [27].** Suppose for the moment that  $\mathcal{M}_{\mathbf{a}}$  is the singleton  $\{\mu\}$ . To approximate the set  $\mathbf{X}_\varepsilon^*$  in (1.3) from inside, the basic idea in [27] is to consider the infinite-dimensional LP:

$$(3.3) \quad \rho = \sup_{\phi \in \mathcal{M}_+(\mathbf{K})} \{\phi(\mathbf{K}) : \phi \leq \lambda \otimes \mu\}.$$

A dual of (3.3) is the infinite-dimensional LP:

$$(3.4) \quad \rho^* = \inf_{w \in \mathbb{R}[\mathbf{x}, \omega]} \left\{ \int_{\mathbf{X} \times \Omega} w(\mathbf{x}, \omega) d(\lambda \otimes \mu) : w \geq 1 \text{ on } \mathbf{K}; w \geq 0 \text{ on } \mathbf{X} \times \Omega \right\}.$$

It is proved in [27] that (3.3) has a unique optimal solution  $d\phi^* = \mathbf{1}_{\mathbf{K}}(\mathbf{x}, \omega) d(\mu \otimes \lambda)$  and  $\rho = \rho^* = \int_{\mathbf{X}} \mu(\mathbf{K}_{\mathbf{x}}) \lambda(d\mathbf{x})$ . In addition, for every feasible solution  $w \in \mathbb{R}[\mathbf{x}, \omega]$ , let  $h \in \mathbb{R}[\mathbf{x}]$  be the polynomial  $\mathbf{x} \mapsto \int_{\Omega} w(\mathbf{x}, \omega) d\mu(\omega)$ . Since  $w \geq 0$  on  $\mathbf{X} \times \Omega$  and  $w \geq 1$  on  $\mathbf{K}$  then  $h(\mathbf{x}) \geq \mu(\mathbf{K}_{\mathbf{x}})$  for all  $\mathbf{x} \in \mathbf{X}$  and therefore  $\{\mathbf{x} : h(\mathbf{x}) < \varepsilon\} \subset \{\mathbf{x} : \mu(\mathbf{K}_{\mathbf{x}}) < \varepsilon\} = \mathbf{X}_\varepsilon^*$ . Further, [27] defines a hierarchy of semidefinite relaxations of (3.3) such that optimal solutions of their associated semidefinite duals provide polynomials  $h_d \in \mathbb{R}[\mathbf{x}]_{2d}$  of increasing degree  $d$  with the property that  $\lambda(\mathbf{X}_\varepsilon^* \setminus \{\mathbf{x} : h_d(\mathbf{x}) < \varepsilon\}) \rightarrow 0$  as  $d \rightarrow \infty$ . This was possible because *one knows exactly and in closed form* all moments of the product measure  $\lambda \otimes \mu$  on  $\mathbf{X} \times \Omega$ , a crucial ingredient of the semidefinite relaxations (14) in [27].

When  $\mathcal{M}_{\mathbf{a}}$  is not the singleton  $\{\mu\}$ , a similar approach as the one above is quite more involved because:

- One does *not* know the exact distribution of the noise  $\omega$ ; it can be any mixture of  $(\mu_{\mathbf{a}})_{\mathbf{a} \in \mathbf{A}}$ .
- For each  $\mathbf{x} \in \mathbf{X}$  such that  $\mathbf{K}_{\mathbf{x}} \neq \emptyset$  one needs to identify the worst-case distribution  $\mu_{\mathbf{x}} \in \mathcal{M}_{\mathbf{a}}$ , i.e.,  $\mu_{\mathbf{a}(\mathbf{x})}$  where  $\mathbf{a}(\mathbf{x}) = \arg \max_{\mathbf{a} \in \mathbf{A}} \mu_{\mathbf{a}}(\mathbf{K}_{\mathbf{x}})$ .

**Brief informal sketch of the strategy.** To identify the worst-case distribution  $\mu_{\mathbf{x}}$ , we introduce an unknown distribution  $\psi$  on  $\mathbf{X} \times \mathbf{A}$  with marginal  $\lambda$  on  $\mathbf{X}$ . It can be disintegrated into  $\hat{\psi}(d\mathbf{a}|\mathbf{x})\lambda(d\mathbf{x})$  and the goal is to compute  $\psi^* = \hat{\psi}^*(d\mathbf{a}|\mathbf{x})\lambda(d\mathbf{x})$ , with  $\hat{\psi}^*(d\mathbf{a}|\mathbf{x}) = \delta_{\mathbf{a}(\mathbf{x})}$ . But recall that to provide numerical approximations, somehow we need to access the moments of the involved measures, i.e., we need access to  $\gamma_{\beta}(\mathbf{a}) = \int_{\Omega} \omega^{\beta} d\mu_{\mathbf{a}}(\omega)$  for all  $\beta \in \mathbb{N}^p$ , and all  $\mathbf{a} \in \mathbf{A}$ . Fortunately, by Assumption 2.4(ii),  $\gamma_{\beta}(\mathbf{a}) = p_{\beta}(\mathbf{a})$  for some known polynomial  $p_{\beta} \in \mathbb{R}[\mathbf{a}]$ . Then given  $\psi$ :

$$\begin{aligned} \underbrace{\int_{\mathbf{X}} \mathbf{x}^{\alpha} p_{\beta}(\mathbf{a}) d\psi(\mathbf{x}, \mathbf{a})}_{\text{moments of } \psi} &= \int_{\mathbf{X}} \mathbf{x}^{\alpha} \left( \int_{\mathbf{A}} p_{\beta}(\mathbf{a}) \hat{\psi}(d\mathbf{a}|\mathbf{x}) \right) \lambda(d\mathbf{x}) \\ &= \int_{\mathbf{X}} \mathbf{x}^{\alpha} \left( \int_{\mathbf{A}} \left( \int_{\Omega} \omega^{\beta} d\mu_{\mathbf{a}}(\omega) \right) \hat{\psi}(d\mathbf{a}|\mathbf{x}) \right) \lambda(d\mathbf{x}) \\ &= \int_{\mathbf{X}} \mathbf{x}^{\alpha} \left( \int_{\Omega} \omega^{\beta} \mu(d\omega) \right) \lambda(d\mathbf{x}) \end{aligned}$$

where  $\mu(B) = \int_{\mathbf{A}} \mu_{\mathbf{a}}(B) \hat{\psi}(d\mathbf{a}|\mathbf{x})$  for all  $B \in \mathcal{B}(\Omega)$ , i.e.,  $\mu \in \mathcal{M}_{\mathbf{a}}$ . That is, by playing with  $\psi$  (more precisely its conditional  $\hat{\psi}(d\mathbf{a}|\mathbf{x})$ ) one may explore for each  $\mathbf{x} \in \mathbf{X}$ , all possible mixtures of distributions  $\mu_{\mathbf{a}}$ ,  $\mathbf{a} \in \mathbf{A}$ . Importantly, moments of such distributions are expressed in terms of moments of  $\psi$ . This is exactly what we need for computing approximations in the spirit of [27]. The linear mapping  $\mathbf{T}$  in (2.5) is the tool to link measures  $\phi$  on  $\mathbf{X} \times \Omega$  with measures  $\psi$  on  $\mathbf{X} \times \mathbf{A}$ .

As one may see, what precedes is a non trivial extension of the approach in [27]; in addition, the above informal derivation requires some measurability conditions, which by Lemma 3.1 below are guaranteed to hold.

**Lemma 3.1.** *For each  $\mathbf{x} \in \mathbf{X}$  there exist measurable mappings  $\mathbf{x} \mapsto \mathbf{a}(\mathbf{x}) \in \mathbf{A}$  and  $\mathbf{x} \mapsto \kappa(\mathbf{x}) \in \mathbb{R}$  such that:*

$$(3.5) \quad \kappa(\mathbf{x}) = \max\{\mu(\mathbf{K}_{\mathbf{x}}) : \mu \in \mathcal{M}_{\mathbf{a}}\} = \max\{\mu_{\mathbf{a}}(\mathbf{K}_{\mathbf{x}}) : \mathbf{a} \in \mathbf{A}\} = \mu_{\mathbf{a}(\mathbf{x})}(\mathbf{K}_{\mathbf{x}}).$$

The proof is postponed to §7.1. Observe that for all  $\mathbf{x} \in \mathbf{X}$ ,  $\kappa(\mathbf{x}) = \mu_{\mathbf{a}(\mathbf{x})}(\mathbf{K}_{\mathbf{x}}) = 0$  whenever  $\mathbf{K}_{\mathbf{x}} = \emptyset$ . Next, recall that for every  $\mathbf{x} \in \mathbf{X}$ :

$$\kappa(\mathbf{x}) = \max_{\mu} \{\mu(\mathbf{K}_{\mathbf{x}}) : \mu \in \mathcal{M}_{\mathbf{a}}\} = \max_{\mu} \{\text{Prob}_{\mu}(f(\mathbf{x}, \omega) \leq 0) : \mu \in \mathcal{M}_{\mathbf{a}}\}.$$

and so with  $\mathbf{K}_{\mathbf{x}}^c := \Omega \setminus \mathbf{K}_{\mathbf{x}}$ ,

$$\mu_{\mathbf{a}(\mathbf{x})}(\mathbf{K}_{\mathbf{x}}^c) = \min_{\mu} \{\mu(\mathbf{K}_{\mathbf{x}}^c) : \mu \in \mathcal{M}_{\mathbf{a}}\} = \min_{\mu} \{\text{Prob}_{\mu}(f(\mathbf{x}, \omega) > 0) : \mu \in \mathcal{M}_{\mathbf{a}}\}.$$

Consequently, the set  $\mathbf{X}_{\varepsilon}^*$  in (1.3) also reads:

$$\mathbf{X}_{\varepsilon}^* = \{\mathbf{x} \in \mathbf{X} : \mu_{\mathbf{a}(\mathbf{x})}(\mathbf{K}_{\mathbf{x}}^c) > 1 - \varepsilon\} = \{\mathbf{x} \in \mathbf{X} : \kappa(\mathbf{x}) < \varepsilon\}.$$

We next consider a certain infinite dimension LP with an important property: Namely, any feasible solution of its dual (also an infinite dimensional LP) provides the coefficients of some polynomial  $w \in \mathbb{R}[\mathbf{x}]$  such that  $\{\mathbf{x} : w(\mathbf{x}) < \varepsilon\} \subset \mathbf{X}_{\varepsilon}^*$ ; see Theorem 3.3 below. It is the analogue in the present context of the LP (12) in [27].

**3.2. An ideal infinite-dimensional LP.** Let  $\lambda$  be the Lebesgue measure on  $\mathbf{X}$ , normalized to a probability measure, and consider the infinite-dimensional linear program (LP):

$$(3.6) \quad \begin{aligned} \rho = \sup_{\phi, \psi} \quad & \{ \phi(\mathbf{K}) : \phi \leq \mathbf{T}^* \psi; \quad \psi_{\mathbf{x}} = \lambda, \\ & \phi \in \mathcal{M}_+(\mathbf{K}), \psi \in \mathcal{P}(\mathbf{X} \times \mathbf{A}) \}, \end{aligned}$$

where  $\mathbf{T}^*$  is defined in Lemma 2.8.

**Theorem 3.2.** *The infinite dimensional LP (3.6) has optimal value  $\rho = \int_{\mathbf{X}} \kappa(\mathbf{x}) d\mathbf{x}$ . Moreover, the feasible pair  $(\phi^*, \psi^*)$  with*

$$d\phi^*(\mathbf{x}, \omega) := \mathbf{1}_{\mathbf{K}}(\mathbf{x}, \omega) \mu_{\mathbf{a}(\mathbf{x})}(d\omega) d\lambda(\mathbf{x}), \quad d\psi^*(\mathbf{x}, \mathbf{a}) = \delta_{\mathbf{a}(\mathbf{x})}(d\mathbf{a}) \lambda(d\mathbf{x}),$$

is an optimal solution of (3.6), with  $\mathbf{x} \mapsto (\mathbf{a}(\mathbf{x}), \kappa(\mathbf{x}))$  as in Lemma 3.1.

*Proof.* Let  $(\phi, \psi)$  be an arbitrary feasible solution. Then

$$\begin{aligned} \phi(\mathbf{K}) = \langle \mathbf{1}, \phi \rangle & \leq \langle \mathbf{1}_{\mathbf{K}}, \mathbf{T}^* \psi \rangle = \langle \mathbf{T} \mathbf{1}_{\mathbf{K}}, \psi \rangle = \int_{\mathbf{X} \times \mathbf{A}} \mu_{\mathbf{a}(\mathbf{x})}(\mathbf{K}_{\mathbf{x}}) d\psi(\mathbf{x}, \mathbf{a}) \\ & = \int_{\mathbf{X}} \left( \int_{\mathbf{A}} \mu_{\mathbf{a}(\mathbf{x})}(\mathbf{K}_{\mathbf{x}}) \hat{\psi}(d\mathbf{a}|\mathbf{x}) \right) \lambda(d\mathbf{x}) \\ & \leq \int_{\mathbf{X}} \mu_{\mathbf{a}(\mathbf{x})}(\mathbf{K}_{\mathbf{x}}) \lambda(d\mathbf{x}) \quad [\text{by Lemma 3.1}] \\ & = \int_{\mathbf{X}} \kappa(\mathbf{x}) \lambda(d\mathbf{x}). \end{aligned}$$

Next let  $d\phi^*(\mathbf{x}, \omega) = \mathbf{1}_{\mathbf{K}}(\mathbf{x}, \omega) \mu_{\mathbf{a}(\mathbf{x})}(d\omega) \lambda(d\mathbf{x})$  and  $d\psi^*(\mathbf{x}, \mathbf{a}) = \delta_{\mathbf{a}(\mathbf{x})}(d\mathbf{a}) \lambda(d\mathbf{x})$  with  $\mathbf{x} \mapsto \mathbf{a}(\mathbf{x})$  as in Lemma 3.1. Then  $\phi^* \in \mathcal{M}_+(\mathbf{K})$  and  $\phi^* \leq \mathbf{T}^* \psi^*$ . Moreover:

$$\phi^*(\mathbf{K}) = \int_{\mathbf{X}} \left( \int_{\mathbf{K}_{\mathbf{x}}} d\mu_{\mathbf{a}(\mathbf{x})}(\omega) \right) \lambda(d\mathbf{x}) = \int_{\mathbf{X}} \kappa(\mathbf{x}) \lambda(d\mathbf{x})$$

as  $\mu_{\mathbf{a}(\mathbf{x})}(\mathbf{K}_{\mathbf{x}}) = 0$  whenever  $\mathbf{K}_{\mathbf{x}} = \emptyset$ .  $\square$

Hence, Theorem 3.2 states that in an optimal solution  $(\phi^*, \psi^*)$  of the LP (3.6), at every  $\mathbf{x} \in \mathbf{X}$ , the conditional probability  $\hat{\phi}^*(\cdot|\mathbf{x}) := \mu_{\mathbf{a}(\mathbf{x})} \in \mathcal{M}_{\mathbf{a}}$  identifies the *worst case* noise distribution  $\mu_{\mathbf{a}(\mathbf{x})}$  in  $\mathcal{M}_{\mathbf{a}}$ , that is, the one which maximizes  $\text{Prob}_{\mu}(f(\mathbf{x}, \omega) \leq 0)$  over  $\mathcal{M}_{\mathbf{a}}$ , hence which minimizes  $\text{Prob}_{\mu}(f(\mathbf{x}, \omega) > 0)$  over  $\mathcal{M}_{\mathbf{a}}$ .

**A dual of (3.6).** Recall that by Lemma 2.9, the mapping  $\mathbf{T}$  extends to polynomials, and so consider the infinite dimensional LP:

$$(3.7) \quad \begin{aligned} \rho^* = \inf_{h, w} \quad & \left\{ \int_{\mathbf{X}} w d\lambda : \quad h(\mathbf{x}, \omega) \geq 1 \quad \text{on } \mathbf{K}, \right. \\ & w(\mathbf{x}) - \mathbf{T}h(\mathbf{x}, \mathbf{a}) \geq 0 \quad \text{on } \mathbf{X} \times \mathbf{A}, \\ & h \geq 0 \quad \text{on } \mathbf{X} \times \Omega, \\ & w \in \mathbb{R}[\mathbf{x}]; h \in \mathbb{R}[\mathbf{x}, \omega] \}. \end{aligned}$$

**Theorem 3.3.** *The infinite dimensional LP (3.7) is a dual of (3.6) that is,  $\rho^* \geq \rho$ . In addition, for every feasible solution  $(w, h)$  of (3.7):*

$$(3.8) \quad w(\mathbf{x}) \geq \kappa(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbf{X},$$

with  $\mathbf{x} \mapsto \kappa(\mathbf{x})$  as in Lemma 3.1, and so for every  $\varepsilon > 0$ :

$$(3.9) \quad \mathbf{X}_w := \{\mathbf{x} : w(\mathbf{x}) < \varepsilon\} \subset \{\mathbf{x} : \mu_{\mathbf{a}(\mathbf{x})}(\mathbf{K}_{\mathbf{x}}^c) > 1 - \varepsilon\} = \mathbf{X}_{\varepsilon}^*.$$

Moreover, suppose that there is no duality gap, i.e.,  $\rho^* = \rho$ , and let  $(w_n, h_n)$  be a minimizing sequence of (3.7). Then with  $\|\cdot\|_1$  the norm of  $L_1(\mathbf{X}, \lambda)$ :

$$(3.10) \quad \lim_{n \rightarrow \infty} \|w_n - \kappa\|_1 = 0, \quad \text{and} \quad \lim_{n \rightarrow \infty} \lambda(\mathbf{X}_\varepsilon^* \setminus \mathbf{X}_{w_n}) = 0.$$

The proof is postponed to §7.2. Theorem 3.3 is still abstract and the challenge is to define a practical method to compute  $\rho$  and to *construct* effectively, inner approximations of  $\mathbf{X}_\varepsilon^*$  which converge to  $\mathbf{X}_\varepsilon^*$ , as the approximations  $\mathbf{X}_{w_n}$  in (3.10).

#### 4. A HIERARCHY OF SEMIDEFINITE RELAXATIONS

In this section we provide a numerical scheme to approximate from above the optimal value  $\rho$  of the infinite-dimensional LP (3.6) and its dual (3.7). In addition, from an optimal solution of the approximation of the dual (3.7), one is able to construct *effectively* inner approximations of  $\mathbf{X}_\varepsilon^*$  which converge to  $\mathbf{X}_\varepsilon^*$ , as the approximations  $\mathbf{X}_{w_n}$  in (3.10).

As a preliminary, we first show that the measure  $\mathbf{T}\psi^*$  in the constraint  $\phi \leq \mathbf{T}^*\psi$ , and identified as  $\psi'_{\mathbf{x}, \omega}$  in Lemma 2.8, can be “handled” through its moments.

**Lemma 4.1.** *Let  $\nu \in \mathcal{P}(\mathbf{X} \times \Omega)$  and  $\psi \in \mathcal{P}(\mathbf{X} \times \mathbf{A})$  with  $\psi_{\mathbf{x}} = \lambda$ , and let*

$$(4.1) \quad \int_{\mathbf{X} \times \Omega} \mathbf{x}^\alpha \omega^\beta d\nu(\mathbf{x}, \omega) = \int_{\mathbf{X} \times \mathbf{A}} \mathbf{x}^\alpha p_\beta(\mathbf{a}) d\psi(\mathbf{x}, \mathbf{a}), \quad \forall \alpha \in \mathbb{N}^n, \beta \in \mathbb{N}^p.$$

Then  $\nu = \mathbf{T}^*\psi$ .

The proof is postponed to §7.3.

**4.1. A hierarchy of semidefinite relaxations of (3.6).** The compact basic semi-algebraic sets  $\mathbf{X}$  and  $\mathbf{A}$  and the basic semi-algebraic set  $\Omega$  are defined by:

$$(4.2) \quad \mathbf{X} := \{\mathbf{x} : g_j(\mathbf{x}) \geq 0, j = 1, \dots, m\}$$

$$(4.3) \quad \mathbf{A} := \{\mathbf{a} \in \mathbb{R}^t : q_\ell(\mathbf{a}) \geq 0, \ell = 1, \dots, L\}$$

$$(4.4) \quad \Omega := \{\omega \in \mathbb{R}^p : s_\ell(\omega) \geq 0, \ell = 1, \dots, \bar{s}\},$$

for some polynomials  $(g_j) \subset \mathbb{R}[\mathbf{x}]$ ,  $(q_\ell) \subset \mathbb{R}[\mathbf{a}]$  and  $(s_\ell) \subset \mathbb{R}[\omega]$ . In particular if  $\bar{s} = 0$  then  $\Omega = \mathbb{R}^p$ .

Let  $d_j := \lceil \deg(g_j)/2 \rceil$ ,  $d'_\ell := \lceil \deg(q_\ell)/2 \rceil$ ,  $j = 1, \dots, m$ ,  $\ell = 1, \dots, L$ . Also let  $d'_\ell = \lceil \deg(s_\ell)/2 \rceil$ ,  $\ell = 1, \dots, \bar{s}$ . For notational convenience we also define  $g_{m+1}(\mathbf{x}, \omega) = -f(\mathbf{x}, \omega)$  with  $d_{m+1} := \lceil \deg(f)/2 \rceil$ . As  $\mathbf{X}$  and  $\mathbf{A}$  are compact,  $\mathbf{X} \subset \{\mathbf{x} : \|\mathbf{x}\|^2 \leq M\}$  and  $\mathbf{A} \subset \{\mathbf{a} : \|\mathbf{a}\|^2 \leq M\}$  for some  $M$  sufficiently large. Therefore with no loss of generality we may and will assume that

$$(4.5) \quad g_1(\mathbf{x}) = M - \|\mathbf{x}\|^2, \quad q_1(\mathbf{a}) = M - \|\mathbf{a}\|^2,$$

for some  $M$  sufficiently large. Similarly if  $\Omega$  in (4.4) is compact then we may and will also assume that  $s_1(\omega) = M - \|\omega\|^2$ . This will be very useful as it ensures compactness of the feasible sets of the semidefinite relaxations defined below.

Next, recall that by Assumption 2.4(ii), for every  $\beta \in \mathbb{N}^p$ ,  $\int_{\Omega} \omega^\beta d\mu_{\mathbf{a}}(\omega) = p_\beta(\mathbf{a})$ , for all  $\mathbf{a} \in \mathbf{A}$ , for some polynomial  $p_\beta \in \mathbb{R}[\mathbf{a}]$ .

Consider the following hierarchy of semidefinite programs indexed by  $d \geq d_{\min} \in \mathbb{N}$ , where  $2d_{\min}$  is the largest degree appearing in the polynomials that describe  $\mathbf{K}$ ,  $\Omega$ , and  $\mathbf{A}$ :

$$(4.6) \quad \begin{aligned} \rho_d = & \sup_{\mathbf{y}, \mathbf{u}, \mathbf{v}} y_{00} \\ \text{s.t. } & L_{\mathbf{y}+\mathbf{u}}(\mathbf{x}^\alpha \omega^\beta) - L_{\mathbf{v}}(\mathbf{x}^\alpha p_\beta(\mathbf{a})) = 0, \quad |\alpha + \deg(p_\beta)| \leq 2d, \\ & L_{\mathbf{v}}(\mathbf{x}^\alpha) = \lambda_\alpha, \quad \alpha \in \mathbb{N}_{2d}^n, \\ & \mathbf{M}_d(\mathbf{y}), \mathbf{M}_d(\mathbf{u}), \mathbf{M}_d(\mathbf{v}) \succeq 0, \\ & \mathbf{M}_{d-d_{m+1}}(g_{m+1} \mathbf{y}) \succeq 0, \\ & \mathbf{M}_{d-d_j}(g_j \mathbf{y}), \mathbf{M}_{d-d_j}(g_j \mathbf{u}), \mathbf{M}_d(g_j \mathbf{v}) \succeq 0, \quad j = 1, \dots, m, \\ & \mathbf{M}_{d-d_\ell^1}(s_\ell \mathbf{y}), \mathbf{M}_{d-d_\ell^1}(s_\ell \mathbf{u}) \succeq 0, \quad \ell = 1, \dots, \bar{s}, \\ & \mathbf{M}_{d-d_\ell'}(q_\ell \mathbf{v}) \succeq 0, \quad \ell = 1, \dots, L, \end{aligned}$$

where  $\mathbf{y} = (y_{\alpha\beta})$ ,  $\mathbf{u} = (u_{\alpha\beta})$ ,  $(\alpha, \beta) \in \mathbb{N}^n \times \mathbb{N}^p$  and  $\mathbf{v} = (v_{\alpha\eta})$ ,  $(\alpha, \eta) \in \mathbb{N}^n \times \mathbb{N}^t$ .

**Proposition 4.2.** *The semidefinite program (4.6) is a relaxation of (3.6), i.e.,  $\rho_d \geq \rho$  for all  $d \geq d_{\min}$ . In addition  $\rho_{d+1} \leq \rho_d$  for all  $d \geq d_{\min}$ .*

*Proof.* That  $\rho_{d+1} \leq \rho_d$  is straightforward as more constraints are added as  $d$  increases. Next, let  $\phi, \psi$  be an arbitrary feasible solution of (3.6) and let  $\mathbf{y} = (y_{\alpha\beta})$ ,  $\mathbf{u} = (u_{\alpha\beta})$ ,  $(\alpha, \beta) \in \mathbb{N}^n \times \mathbb{N}^p$  and  $\mathbf{v} = (v_{\alpha\eta})$ ,  $(\alpha, \eta) \in \mathbb{N}^n \times \mathbb{N}^t$  be the moments sequences of the measure  $\phi$ ,  $\mathbf{T}^* \psi - \phi$  and  $\psi$ , respectively. In the following we first show that  $\mathbf{y}$ ,  $\mathbf{u}$ , and  $\mathbf{v}$  are feasible for (4.6). Necessarily  $\mathbf{M}_d(\mathbf{y}) \succeq 0$  because for every vector  $\mathbf{p} \in \mathbb{R}^{s(d)}$  (with  $s(d) := \binom{n+d}{n}$ ),

$$\langle \mathbf{p}, \mathbf{M}_d(\mathbf{y}) \mathbf{p} \rangle = \int_{\mathbf{K}} p(\mathbf{x}, \omega)^2 d\phi(\mathbf{x}, \omega) \geq 0,$$

where  $p \in \mathbb{R}[\mathbf{x}]_d$  has  $\mathbf{p}$  as vector of coefficients. With similar arguments, one can show that  $\mathbf{M}_d(\mathbf{u}) \succeq 0$  and  $\mathbf{M}_d(\mathbf{v}) \succeq 0$ . Next, as  $\phi$  is supported on  $\mathbf{K}$ , for every  $\mathbf{p} \in \mathbb{R}^{s(d-d_j)}$ :

$$\langle \mathbf{p}, \mathbf{M}_{d-d_j}(g_j \mathbf{y}) \mathbf{p} \rangle = \int_{\mathbf{K}} p(\mathbf{x}, \omega)^2 g_j(\mathbf{x}) d\phi(\mathbf{x}, \omega) \geq 0,$$

because  $g_j(\mathbf{x}) \geq 0$  whenever  $(\mathbf{x}, \omega) \in \mathbf{K}$ . Hence  $\mathbf{M}_d(g_j \mathbf{y}) \succeq 0$  (and similarly  $\mathbf{M}_{d-d_j}(g_j \mathbf{u}), \mathbf{M}_{d-d_j}(g_j \mathbf{v}) \succeq 0$ ). Similarly for every  $\mathbf{p} \in \mathbb{R}^{s(d-d_\ell^1)}$ :

$$\langle \mathbf{p}, \mathbf{M}_{d-d_\ell^1}(s_\ell \mathbf{y}) \mathbf{p} \rangle = \int_{\mathbf{K}} p(\mathbf{x}, \omega)^2 s_\ell(\omega)(\mathbf{x}) d\phi(\mathbf{x}, \omega) \geq 0,$$

because  $s_\ell(\omega) \geq 0$  whenever  $(\mathbf{x}, \omega) \in \mathbf{K}$ . Hence  $\mathbf{M}_{d-d_\ell^1}(s_\ell \mathbf{y}) \succeq 0$  (and similarly  $\mathbf{M}_{d-d_\ell^1}(s_\ell \mathbf{u}) \succeq 0$ ). Finally, as  $\psi$  is supported on  $\mathbf{X} \times \mathbf{A}$  then for every  $\mathbf{p} \in \mathbb{R}^{s(d-d_\ell')}$ :

$$\langle \mathbf{p}, \mathbf{M}_{d-d_\ell'}(q_\ell \mathbf{v}) \mathbf{p} \rangle = \int_{\mathbf{X} \times \mathbf{A}} p(\mathbf{x}, \omega)^2 q_\ell(\mathbf{a}) d\psi(\mathbf{x}, \mathbf{a}) \geq 0,$$

because  $q_\ell(\mathbf{a}) \geq 0$  whenever  $(\mathbf{x}, \mathbf{a}) \in \mathbf{X} \times \mathbf{A}$ . Hence  $\mathbf{M}_{d-d_\ell'}(q_\ell \mathbf{v}) \succeq 0$ .

Next, as  $\psi_{\mathbf{x}} = \lambda$ , then for every  $\alpha \in \mathbb{N}_{2d}^n$ :

$$L_{\mathbf{v}}(\mathbf{x}^\alpha \omega^0) = v_{\alpha 0} = \int_{\mathbf{X} \times \mathbf{A}} \mathbf{x}^\alpha d\psi(\mathbf{x}, \mathbf{a}) = \int_{\mathbf{X}} \mathbf{x}^\alpha d\lambda(\mathbf{x}) = \lambda_\alpha.$$



Finally, as  $\mathbf{T}^*\psi \geq \phi$  (equivalently  $\mathbf{T}^*\psi = \phi + \nu$  with  $\nu$  being the slack measure  $\mathbf{T}^*\psi - \phi$ ), then for every  $(\alpha, \beta) \in \mathbb{N}^n \times \mathbb{N}^p$  with  $|\alpha| + \deg(p_\beta) \leq 2d$ :

$$\begin{aligned} L_{\mathbf{v}}(\mathbf{x}^\alpha p_\beta(\mathbf{a})) &= \int_{\mathbf{X} \times \mathbf{A}} \mathbf{x}^\alpha p_\beta(\mathbf{a}) d\psi(\mathbf{x}, a) = \int_{\mathbf{X} \times \mathbf{A}} \mathbf{T}(\mathbf{x}^\alpha \omega^\beta) d\psi(\mathbf{x}, \mathbf{a}) \\ &= \int_{\mathbf{X} \times \Omega} \mathbf{x}^\alpha \omega^\beta d(\mathbf{T}^*\psi)(\mathbf{x}, \omega) = \int_{\mathbf{X} \times \Omega} \mathbf{x}^\alpha \omega^\beta d(\phi + \nu) \\ &= L_{\mathbf{y}+\mathbf{u}}(\mathbf{x}^\alpha \omega^\beta), \end{aligned}$$

and therefore  $(\mathbf{y}, \mathbf{u}, \mathbf{v})$  is a feasible solution of (4.6). Finally, we show that  $\rho_d \geq \rho$ . To that end, note that  $\mathbf{y}_{00} = \int_{\mathbf{K}} d\phi = \phi(\mathbf{K})$ , i.e.,  $\rho_d \geq \phi(\mathbf{K})$ , and in particular  $\rho_d \geq \phi^*(\mathbf{K}) = \rho$  for the optimal solution  $\phi^*$  of (3.6).  $\square$

By Proposition 4.2 (4.6) defines a *hierarchy* of semidefinite programs whose associated sequence of optimal values  $(\rho_d)_{d \geq d_{\min}}$  is nonnegative and monotone non increasing.

**Size of (4.6).** The semidefinite program involves moment matrices of size  $\tau_1 := \binom{n+p+d}{d}$  and  $\tau_2 = \binom{n+t+d}{d}$ . So even though (4.6) can be solved in time polynomial in its input size, the computational burden is rapidly prohibitive, and in view of the present status of semidefinite solvers, this approach is limited to relaxations for which  $\max[\tau_1, \tau_2] \leq 10^3$ , i.e., for modest size problems. However, recall that even the first relaxations provide inner approximations of  $\mathbf{X}_\varepsilon^*$ .

**4.2. The dual of the semidefinite relaxations (4.6).** The dual of (4.6) is an SDP which has the following high-level interpretation in terms of SOS positivity certificates of size parametrized by  $d$ :

$$(4.7) \quad \begin{aligned} \rho_d^* &= \inf_{h, w, \sigma_j^i} \int_{\mathbf{X}} w(\mathbf{x}) d\lambda(\mathbf{x}) : \\ \text{s.t. } & h(\mathbf{x}, \omega) - 1 = \sum_{j=0}^{m+1} \sigma_j^1 g_j + \sum_{\ell=1}^{\bar{s}} \sigma_\ell^1 s_\ell, \quad \forall(\mathbf{x}, \omega); \\ & h(\mathbf{x}, \omega) = \sum_{j=0}^m \sigma_j^2 g_j + \sum_{\ell=1}^{\bar{s}} \sigma_\ell^2 s_\ell, \quad \forall(\mathbf{x}, \omega); \\ & w(\mathbf{x}) - \sum_{\alpha, \beta} h_{\alpha\beta} \mathbf{x}^\alpha p_\beta(\mathbf{a}) = \sum_{j=0}^m \sigma_j^3 g_j + \sum_{\ell=1}^L \sigma_\ell^3 q_\ell, \quad \forall(\mathbf{x}, \mathbf{a}); \\ & \deg(h), \deg(w) \leq 2d; \sigma_j^1 \in \Sigma[\mathbf{x}, \omega]_{d-d_j}, j = 0, \dots, m+1, \\ & \sigma_j^2 \in \Sigma[\mathbf{x}, \omega]_{d-d_j}, \sigma_j^3 \in \Sigma[\mathbf{x}, \mathbf{a}]_{d-d_j}; j = 0, \dots, m, \\ & \sigma_\ell^1, \sigma_\ell^2 \in \Sigma[\omega]_{d-d'_\ell}; \sigma_\ell^3 \in \Sigma[\mathbf{x}, \mathbf{a}]_{d-d'_\ell}, \ell = 1, \dots, L, \end{aligned}$$

where  $h(\mathbf{x}, \omega) = \sum_{|\alpha+\beta| \leq 2d} h_{\alpha\beta} \mathbf{x}^\alpha \omega^\beta$ , and  $w(\mathbf{x}) = \sum_{|\alpha| \leq 2d} w_\alpha \mathbf{x}^\alpha$ .

In compact form, (4.7) is the high level interpretation of the dual SDP of (4.6) in terms of SOS positivity certificates of size parametrized by  $d$ . Indeed:

- The dual variable  $h_{\alpha\beta}$  associated with the equality constraint  $L_{\mathbf{y}+\mathbf{u}}(\mathbf{x}^\alpha \omega^\beta) = L_{\mathbf{v}}(\mathbf{x}^\alpha p_\beta(\mathbf{a}))$  is the coefficient of  $\mathbf{x}^\alpha \omega^\beta$  for the polynomial  $h \in \mathbb{R}[\mathbf{x}, \omega]_{2d}$  in (4.7).
- Similarly, the dual variable  $w_\alpha$  associated with the equality constraint  $L_{\mathbf{v}}(\mathbf{x}^\alpha) = \lambda_\alpha$  is the coefficient of  $\mathbf{x}^\alpha$  for the polynomial  $w \in \mathbb{R}[\mathbf{x}]_{2d}$  in (4.7).

- $\sigma_j^1 \in \Sigma[\mathbf{x}, \omega]_{d-d_j}$  (resp.  $\sigma_j^2 \in \Sigma[\mathbf{x}, \omega]_{d-d_j}$ ) is the SOS polynomial associated with the matrix dual variable  $\mathbf{X}_j^1 \succeq 0$  (resp.  $\mathbf{X}_j^2 \succeq 0$ ) associated with the semidefinite constraint  $\mathbf{M}_{d-d_j}(g_j \mathbf{y}) \succeq 0$  (resp.  $\mathbf{M}_{d-d_j}(g_j \mathbf{u}) \succeq 0$ ) of (4.6).
- $\sigma_\ell^1 \in \Sigma[\mathbf{x}, \omega]_{d-d'_\ell}$  (resp.  $\sigma_\ell^2 \in \Sigma[\mathbf{x}, \omega]_{d-d'_\ell}$ ) is the SOS polynomial associated with the matrix dual variable  $\mathbf{Z}_j^1 \succeq 0$  (resp.  $\mathbf{Z}_j^2 \succeq 0$ ) associated with the semidefinite constraint  $\mathbf{M}_{d-d'_\ell}(s_\ell \mathbf{y}) \succeq 0$  (resp.  $\mathbf{M}_{d-d'_\ell}(s_\ell \mathbf{u}) \succeq 0$ ) of (4.6).
- $\sigma_j^3 \in \Sigma[\mathbf{x}, \mathbf{a}]_{d-d_j\ell}$  (resp.  $\sigma_j^3 \in \Sigma[\mathbf{x}, \mathbf{a}]_{d-d'_\ell}$ ) is the SOS polynomial associated with the matrix dual variable  $\mathbf{W}_j^1 \succeq 0$  (resp.  $\mathbf{W}_j^2 \succeq 0$ ) associated with the semidefinite constraint  $\mathbf{M}_{d-d_j}(g_j \mathbf{v}) \succeq 0$  (resp.  $\mathbf{M}_{d-d'_\ell}(s_\ell \mathbf{v}) \succeq 0$ ) of (4.6).

The SDP (4.7) is a *reinforcement* of the infinite-dimensional dual (3.7) in which the positivity constraints have been replaced with SOS *positivity certificates* à la Putinar (see [30]). For instance, the positivity constraint “ $h \geq 1$  on  $\mathbf{K}$ ” in (3.7) becomes in (4.7) the stronger:

$$h(\mathbf{x}, \omega) - 1 = \sum_{j=0}^{m+1} \sigma_j^1 g_j + \sum_{\ell=1}^{\bar{s}} \sigma_\ell^1 s_\ell, \quad \forall (\mathbf{x}, \omega),$$

for some SOS polynomials  $(\sigma_j^1)$  and  $(\sigma_\ell^1)$ .

**Theorem 4.3.** *Let Assumption 2.4 (and Assumption 2.5 as well if  $\Omega$  is unbounded) hold and assume that  $\mathbf{K}, \mathbf{A}, \mathbf{X} \times \Omega$  and  $\mathbf{X} \times \Omega \setminus \mathbf{K}$  all have nonempty interior. Then:*

- Slater’s condition holds for (4.6) and so strong duality holds. That is, for every  $d \geq d_{\min}$ ,  $\rho_d^* = \rho_d$  and (4.7) has an optimal solution  $(h_d, w_d)$ .*
- Next, define  $\mathbf{X}_\varepsilon^d := \{\mathbf{x} \in \mathbf{X} : w_d(\mathbf{x}) < \varepsilon\}$ . Then  $\mathbf{X}_\varepsilon^d \subset \mathbf{X}_\varepsilon^*$  for every  $d \geq d_{\min}$ . In addition, if  $\lim_{d \rightarrow \infty} \rho_d = \rho$  then with  $\mathbf{x} \mapsto \kappa(\mathbf{x})$  as in Lemma 3.1:*

$$(4.8) \quad \lim_{d \rightarrow \infty} \|w_d(\mathbf{x}) - \kappa(\mathbf{x})\|_{L_1(\mathbf{X}, \lambda)} = 0 \quad \text{and} \quad \lim_{d \rightarrow \infty} \lambda(\mathbf{X}_\varepsilon^* \setminus \mathbf{X}_\varepsilon^d) = 0.$$

A proof is postponed to §7.4. Note that so far, optimal solutions of (4.7) provide us with a hierarchy of inner approximations  $\mathbf{X}_\varepsilon^d \subset \mathbf{X}_\varepsilon^*$ ,  $d \geq d_{\min}$ . In addition, if the approximation scheme (4.6) is such that  $\lim_{d \rightarrow \infty} \rho_d = \rho$ , then Theorem 4.3 states that the inner approximations  $(\mathbf{X}_\varepsilon^d)$  have the additional strong asymptotic property (4.8) which in turn implies the highly desirable convergence result (4.8).

So to obtain (4.8) we need to ensure that  $\lim_{d \rightarrow \infty} \rho_d = \rho$  as  $d \rightarrow \infty$ .

**Theorem 4.4.** *Let Assumption 2.4 hold, and if  $\Omega$  is unbounded let Assumption 2.5 hold as well. Consider the hierarchy of semidefinite programs (4.6) with associated monotone sequence of optimal values  $(\rho_d)_{d \geq d_{\min}}$ . Then for each  $d \geq d_{\min}$  there is an optimal solution  $(\mathbf{y}^d, \mathbf{u}^d, \mathbf{v}^d)$  and*

$$\rho_d = y_{00}^d \downarrow \phi^*(\mathbf{K}) = \rho, \quad \text{as } d \rightarrow \infty,$$

where  $\phi^*$  is part of an optimal solution  $(\phi^*, \psi^*)$  of (3.6).

A proof is postponed to §7.5.

**4.3. Accelerating convergence via Stokes.** In previous works of a similar flavor but for volume computation in [21] and [27, 33], it was observed that the convergence  $\rho_d \rightarrow \rho$  was rather slow. In our framework, by inspection of the dual (3.7), a potential slow convergence may arise as one tries to approximate from above a discontinuous function (the indicator function  $1_{\mathbf{K}}$  of a compact set  $\mathbf{K}$ ) by polynomials, and therefore one is faced with an annoying Gibb’s phenomenon. The trick

proposed in [21] resulted in a significant acceleration of the convergence but at the price of losing its monotonicity (a highly desirable feature). This motivated the other strategy proposed in [27, 33], based on Stokes' theorem, which also resulted in a significantly faster convergence, but this time without losing its monotonicity.

In this section we provide a means to accelerate the convergence  $\rho_d \rightarrow \rho$  in Theorem 4.4, also based on Stokes' theorem applied to the optimal solution  $\phi^*$  of (3.6). However, its implementation is much more complicated than in [27, 33] as it requires introducing an additional measure in the LP (3.6).

It works when the probability measures  $(\mu_{\mathbf{a}})_{\mathbf{a} \in \mathbf{A}}$ , satisfy some additional property: For all  $\mathbf{a} \in \mathbf{A}$ , either,

- $d\mu_{\mathbf{a}}(\omega) = s(\omega, \mathbf{a})$  for some polynomial  $s \in \mathbb{R}[\omega, \mathbf{a}]$ , or
- $d\mu_{\mathbf{a}}(\omega) = q_0(\mathbf{a}) s_1(\omega, \mathbf{a}) \exp(s_2(\omega, \mathbf{a})) d\omega$ , where  $q_0$  is a rational function and  $\omega \mapsto s_i(\omega, \mathbf{a})$  belongs to  $\mathbb{R}[\omega]$ ,  $i = 1, 2$ . In addition, for every fixed  $\mathbf{a}$ , and  $i = 1, \dots, p$ ,  $\partial s_2(\omega, \mathbf{a}) / \partial \omega_i$  is a rational function of  $\omega$ .

We detail the procedure for the case where  $\Omega = \mathbb{R}$ ,  $\mathbf{a} = (a, \sigma) \in \mathbf{A} := [\underline{a}, \bar{a}] \times [\underline{\sigma}, \bar{\sigma}] \subset \mathbb{R}^2$ , and

$$d\mu_{\mathbf{a}}(\omega) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(\omega - a)^2}{2\sigma^2}\right) d\omega, \quad \mathbf{a} = (a, \sigma) \in \mathbf{A},$$

i.e.,  $\mathcal{M}_{\mathbf{a}}$  is the family of all possible mixtures of Gaussian probability measures on  $\mathbb{R}$  with mean  $a \in [\underline{a}, \bar{a}]$  and standard deviation  $\sigma \in [\underline{\sigma}, \bar{\sigma}]$ . Recall that  $\mathbf{K}_{\mathbf{x}} = \{\omega \in \mathbb{R} : f(\mathbf{x}, \omega) \leq 0\}$ . For every  $\mathbf{a} \in \mathbf{A}$ , an extended version of Stokes' theorem in [33, Lemma 3.1, p. 151], yields:

$$\int_{\mathbf{K}_{\mathbf{x}}} \frac{\partial (\omega^\beta f(\mathbf{x}, \omega) \exp(\frac{-(\omega - a)^2}{2\sigma^2}))}{\partial \omega} d\omega = 0, \quad \beta = 0, 1, \dots$$

for every  $\mathbf{x} \in \mathbf{X}$ . That is:

$$(4.9) \quad \int_{\mathbf{K}_{\mathbf{x}}} q_\beta(\mathbf{x}, \omega, \mathbf{a}) d\mu_{\mathbf{a}}(\omega) = 0, \quad \forall \mathbf{x} \in \mathbf{X}, \forall \beta = 0, 1, \dots,$$

where  $q_\beta \in \mathbb{R}[\mathbf{x}, \omega, \mathbf{a}]$  reads:

$$q_\beta(\mathbf{x}, \omega, \mathbf{a}) := \frac{\sigma^2 \partial (\omega^\beta f(\mathbf{x}, \omega))}{\partial \omega} - \omega^\beta f(\mathbf{x}, \omega) (\omega - a).$$

This in turn implies that for every  $\mathbf{a} \in \mathbf{A}$ ,

$$(4.10) \quad \int_{\mathbf{X}} \mathbf{x}^\alpha \left( \mathbf{a}^\gamma \int_{\mathbf{K}_{\mathbf{x}}} q_\beta(\mathbf{x}, \omega, \mathbf{a}) d\mu_{\mathbf{a}}(\omega) \right) d\lambda(\mathbf{x}) = 0,$$

for all  $\alpha \in \mathbb{N}^n$ ,  $\gamma \in \mathbb{N}^2$ , and  $\beta = 0, 1, \dots$ . In particular, with  $\mathbf{a}(\mathbf{x})$  as in Lemma 3.1:

$$\int_{\mathbf{X}} \mathbf{x}^\alpha \mathbf{a}(\mathbf{x})^\gamma \left( \int_{\mathbf{K}_{\mathbf{x}}} q_\beta(\mathbf{x}, \omega, \mathbf{a}(\mathbf{x})) d\mu_{\mathbf{a}(\mathbf{x})}(\omega) \right) d\lambda(\mathbf{x}) = 0,$$

or equivalently:

$$(4.11) \quad \int_{\mathbf{K}} \mathbf{x}^\alpha \mathbf{a}(\mathbf{x})^\gamma q_\beta(\mathbf{x}, \omega, \mathbf{a}(\mathbf{x})) d\phi^*(\mathbf{x}, \omega) = 0,$$

for all  $\alpha \in \mathbb{N}^n$ ,  $\gamma \in \mathbb{N}^2$ ,  $\beta = 0, 1, \dots$ .

Therefore in the infinite-dimensional LP (3.6) we may add the linear "generalized" moment constraints (4.11) because they are satisfied at an optimal solution

$\phi^*$ . However, the function  $(\mathbf{x}, \omega) \mapsto \mathbf{a}(\mathbf{x})^\gamma q_\beta(\mathbf{x}, \omega, \mathbf{a}(\mathbf{x}))$  is *not* a polynomial and these constraints cannot be implemented directly.

To overcome this problem, we introduce an additional measure  $\nu$  on  $\mathbf{K} \times \mathbf{A}$  and impose that its marginal  $\nu_{\mathbf{x}, \omega}$  on  $\mathbf{K}$  is  $\phi$  and its marginal  $\nu_{\mathbf{x}, \mathbf{a}}$  on  $\mathbf{X} \times \mathbf{A}$  is dominated by  $\psi$ . That is  $\nu_{\mathbf{x}, \omega} = \phi$  and  $\nu_{\mathbf{x}, \mathbf{a}} \leq \psi$ . On this newly introduced measure we now can impose the (Stokes) moment constraints:

$$(4.12) \quad \int_{\mathbf{X} \times \mathbf{A} \times \Omega} \mathbf{x}^\alpha \mathbf{a}^\gamma q_\beta(\mathbf{x}, \omega, \mathbf{a}) d\nu(\mathbf{x}, \omega, \mathbf{a}) = 0, \quad \forall (\alpha, \beta, \gamma) \in \mathbb{N}^n \times \mathbb{N}^p \times \mathbb{N}^2.$$

Recall that at an optimal solution  $(\phi^*, \psi^*)$  of (3.6),  $\hat{\psi}^*(d\mathbf{a}|\mathbf{x}) = \delta_{\mathbf{a}(\mathbf{x})}$  for all  $\mathbf{x} \in \mathbf{X}$  and so every feasible solution of the form  $(\phi^*, \psi^*, \nu)$  (hence with  $\nu_{\mathbf{x}, \omega} = \phi^*$  and  $\nu_{\mathbf{x}, \mathbf{a}} \leq \psi^*$ ) satisfies  $\hat{\nu}_{\mathbf{x}, \mathbf{a}}(d\mathbf{a}|\mathbf{x}) = \delta_{\mathbf{a}(\mathbf{x})}$  for all  $\mathbf{x} \in \text{supp}(\nu_{\mathbf{x}})$ , i.e., for all  $\mathbf{x} \in \text{supp}(\phi_{\mathbf{x}}^*)$ .<sup>5</sup> Disintegrating  $\nu$  as  $\hat{\nu}(d\mathbf{a}|\mathbf{x}, \omega) d\nu_{\mathbf{x}, \omega}(\mathbf{x}, \omega)$  yields

$$\begin{aligned} 0 &= \int_{\mathbf{K}} \mathbf{x}^\alpha \left( \int_{\mathbf{A}} \mathbf{a}^\gamma q_\beta(\mathbf{x}, \omega, \mathbf{a}) \hat{\nu}(d\mathbf{a}|\mathbf{x}, \omega) \right) d\nu_{\mathbf{x}, \omega}(\mathbf{x}, \omega), \\ &= \int_{\mathbf{K}} \mathbf{x}^\alpha \mathbf{a}(\mathbf{x})^\gamma q_\beta(\mathbf{x}, \omega, \mathbf{a}(\mathbf{x})) d\nu_{\mathbf{x}, \omega}(\mathbf{x}, \omega), \\ &= \int_{\mathbf{K}} \mathbf{x}^\alpha \mathbf{a}(\mathbf{x})^\gamma q_\beta(\mathbf{x}, \omega, \mathbf{a}(\mathbf{x})) d\phi^*(\mathbf{x}, \omega), \quad \forall \alpha \in \mathbb{N}^n, \gamma \in \mathbb{N}^2, \beta \in \mathbb{N}, \end{aligned}$$

which is (4.11). So in the semidefinite relaxation (4.6), we introduce the additional vector  $\mathbf{z}^1 = (z_{\alpha, \beta, \gamma}^1)$  and  $\mathbf{z}^2 = (z_{\alpha, \gamma}^2)$ ,  $\alpha \in \mathbb{N}^n$ ,  $\beta \in \mathbb{N}^p$ ,  $\gamma \in \mathbb{N}^t$ , (ideally the respective moments of the measure  $\nu$  and  $\psi^* - \nu_{\mathbf{x}, \mathbf{a}}$  where  $\nu$  is described above), and the constraints

$$\begin{aligned} L_{\mathbf{z}^1}(\mathbf{x}^\alpha \omega^\beta) &= L_{\mathbf{y}}(\mathbf{x}^\alpha \omega^\beta), \quad \forall |\alpha| + |\beta| \leq 2d, \\ L_{\mathbf{z}^1}(\mathbf{x}^\alpha \mathbf{a}^\gamma) + L_{\mathbf{z}^2}(\mathbf{x}^\alpha \mathbf{a}^\gamma) &= L_{\mathbf{v}}(\mathbf{x}^\alpha \mathbf{a}^\gamma), \quad \forall |\alpha| + |\gamma| \leq 2d, \\ L_{\mathbf{z}^1}(\mathbf{x}^\alpha \mathbf{a}^\gamma q_\beta(\mathbf{x}, \omega, \mathbf{a})) &= 0 \quad \forall |\alpha| + |\gamma| + \deg(q_\beta) \leq 2d. \end{aligned}$$

Of course, for the same index  $d \in \mathbb{N}$ , the resulting semidefinite relaxation is more computationally demanding as it now includes moments of the measure  $\nu$  on  $\mathbf{X} \times \mathbf{A} \times \Omega$  (whereas in (4.6) we only have moments of measures on  $\mathbf{X} \times \Omega$  and  $\mathbf{X} \times \mathbf{A}$ ). However, being more constrained its optimal value can be significantly smaller and the resulting convergence  $\rho_d \rightarrow \rho$  as  $d$  increases, can be expected to be much faster.

Actually, in this framework of mixtures of Gaussian measures, one may also replace (3.6) with:

$$(4.13) \quad \begin{aligned} \rho &= \sup_{\phi, \psi \geq 0} \{ \phi(\mathbf{K} \times \mathbf{A}) : \phi_{\mathbf{x}, \omega} \leq \mathbf{T}^* \psi; \quad \phi_{\mathbf{x}, \mathbf{a}} \leq \psi; \quad \psi_{\mathbf{x}} = \lambda, \\ &\int_{\mathbf{K} \times \mathbf{A}} \mathbf{x}^\alpha \mathbf{a}^\gamma q_\beta(\mathbf{x}, \omega, \mathbf{a}) d\phi(\mathbf{x}, \mathbf{a}, \omega) = 0, \\ &\forall \alpha \in \mathbb{N}^n, \gamma \in \mathbb{N}^2, \beta \in \mathbb{N}^p; \\ &\phi \in \mathcal{M}_+(\mathbf{K} \times \mathbf{A}), \psi \in \mathcal{P}(\mathbf{X} \times \mathbf{A}) \}, \end{aligned}$$

<sup>5</sup>This is because by compactness of  $\mathbf{X}$  and  $\mathbf{A}$ , for all  $\alpha, \gamma$ ,  $\int \mathbf{x}^\alpha \mathbf{a}^\gamma d\nu_{\mathbf{x}, \mathbf{a}}(\mathbf{x}, \mathbf{a}) = \int \mathbf{x}^\alpha \mathbf{a}(\mathbf{x})^\gamma \eta(\mathbf{x}, \mathbf{a}(\mathbf{x})) \lambda(d\mathbf{x})$  for some nonnegative measurable function  $\eta \leq 1$ , and so  $\nu_{\mathbf{x}}(d\mathbf{x}) = \eta(\mathbf{x}, \mathbf{a}(\mathbf{x})) \lambda(d\mathbf{x})$ . This in turn implies that for every  $\gamma$  and almost all  $\mathbf{x} \in \mathbf{X}$ ,  $\int \mathbf{a}^\gamma \hat{\nu}_{\mathbf{x}, \mathbf{a}}(d\mathbf{a}|\mathbf{x}) = \mathbf{a}(\mathbf{x})^\gamma = \int \mathbf{a}^\gamma \delta_{\mathbf{a}(\mathbf{x})}(d\mathbf{a})$ . Hence  $\hat{\nu}_{\mathbf{x}, \mathbf{a}}(d\mathbf{a}|\mathbf{x}) = \delta_{\mathbf{a}(\mathbf{x})}(d\mathbf{a})$  for all  $\mathbf{x} \in \text{supp}(\nu_{\mathbf{x}})$ .

where now  $\phi$  is a measure on  $\mathbf{K} \times \mathbf{A}$  (instead of  $\mathbf{K}$  before). The dual of (4.13) reads:

$$(4.14) \quad \rho^* = \inf_{w,h,\theta,s} \left\{ \int_{\mathbf{X}} w(\mathbf{x}) d\lambda(\mathbf{x}) : \right. \\ \left. \begin{aligned} & h(\mathbf{x}, \omega) + \theta(\mathbf{x}, \mathbf{a}) \\ & + \sum_{\alpha,\beta,\gamma} s_{\alpha\beta\gamma} \mathbf{x}^\alpha \mathbf{a}^\gamma q_\beta(\mathbf{x}, \mathbf{a}, \omega) \geq 1, \quad \text{on } \mathbf{K} \times \mathbf{A} \\ & w(\mathbf{x}) - Th(\mathbf{x}, \mathbf{a}) - \theta(\mathbf{x}, \mathbf{a}) \geq 0, \quad \text{on } \mathbf{X} \times \mathbf{A} \\ & h \geq 0 \text{ on } \mathbf{X} \times \Omega; \quad \theta \geq 0 \text{ on } \mathbf{X} \times \mathbf{A}, \\ & w \in \mathbb{R}[\mathbf{x}], \quad h \in \mathbb{R}[\mathbf{x}, \omega], \quad \theta \in \mathbb{R}[\mathbf{x}, \mathbf{a}], \quad s \in \mathbb{R}[\mathbf{x}, \mathbf{a}, \omega] \end{aligned} \right\}.$$

In particular, let  $(w, h, \theta, s)$  be an arbitrary feasible solution of (4.14). Then for every  $\mathbf{x} \in \mathbf{X}$ , and with  $\mathbf{a}(\mathbf{x})$  as in Lemma 3.1,

$$\begin{aligned} w(\mathbf{x}) &\geq Th(\mathbf{x}, \mathbf{a}(\mathbf{x})) + \theta(\mathbf{x}, \mathbf{a}(\mathbf{x})) = \int_{\Omega} h(\mathbf{x}, \omega) d\mu_{\mathbf{a}(\mathbf{x})}(\omega) + \theta(\mathbf{x}, \mathbf{a}(\mathbf{x})) \\ &= \int_{\Omega} (h(\mathbf{x}, \omega) + \theta(\mathbf{x}, \mathbf{a}(\mathbf{x}))) d\mu_{\mathbf{a}(\mathbf{x})}(\omega) \\ &\geq \int_{\mathbf{K}_x} (h(\mathbf{x}, \omega) + \theta(\mathbf{x}, \mathbf{a}(\mathbf{x}))) d\mu_{\mathbf{a}(\mathbf{x})}(\omega) \\ &= \int_{\mathbf{K}_x} \underbrace{[h(\mathbf{x}, \omega) + \theta(\mathbf{x}, \mathbf{a}(\mathbf{x})) + \sum_{\alpha,\beta,\gamma} s_{\alpha\beta\gamma} \mathbf{x}^\alpha \mathbf{a}(\mathbf{x})^\gamma q_\beta(\mathbf{x}, \mathbf{a}(\mathbf{x}), \omega)]}_{\geq 1 \text{ on } \mathbf{K}_x} d\mu_{\mathbf{a}(\mathbf{x})}(\omega) \\ &\quad - \sum_{\alpha,\beta,\gamma} s_{\alpha\beta\gamma} \mathbf{x}^\alpha \mathbf{a}(\mathbf{x})^\gamma q_\beta(\mathbf{x}, \mathbf{a}(\mathbf{x}), \omega) d\mu_{\mathbf{a}(\mathbf{x})}(\omega) \\ &\geq \int_{\mathbf{K}_x} d\mu_{\mathbf{a}(\mathbf{x})} - \sum_{\alpha,\beta,\gamma} s_{\alpha\beta\gamma} \mathbf{x}^\alpha \mathbf{a}(\mathbf{x})^\gamma \underbrace{\int_{\mathbf{K}_x} q_\beta(\mathbf{x}, \mathbf{a}(\mathbf{x}), \omega) d\mu_{\mathbf{a}(\mathbf{x})}(\omega)}_{= 0 \text{ by (4.9)}} \\ &\geq \kappa(\mathbf{x}). \end{aligned}$$

One may show that in (4.13) and (4.14),

$$\rho = \rho^* = \int_{\mathbf{X}} \kappa(\mathbf{x}) d\lambda(\mathbf{x}),$$

and (4.8) in Theorem 4.3 also holds for an arbitrary minimizing sequence of (4.14).

The hierarchy of associated SDP-relaxations (i.e. the analogues for (4.13) of (4.6) for (3.6)) and its dual hierarchy are obtained in an obvious manner by truncation of the infinite sequences. At step  $d$  of the latter hierarchy we also obtain a polynomial  $w_d$  with same properties as in Theorem 4.3

**4.4. Numerical experiments.** In the illustrative numerical experiments described below we have restricted to mixtures of univariate Gaussian variables  $\mu_{\mathbf{a}}$  (hence with  $\Omega = \mathbb{R}$  and  $\mathbf{a} = (\text{mean}, \text{deviation}) \in \mathbf{A} \subset \mathbb{R}^2$ ). To implement the semidefinite relaxations (4.13) we have used the GloptiPoly software [20] dedicated to solving the Generalized Moment Problem. The resulting SDPs are solved using version 8.1 of Mosek [37].

We discuss three examples chosen to a) illustrate the effect (and efficiency) of Stokes constraints, b) compare the approximations with the real feasible set  $\mathbf{X}_\varepsilon^*$  in

(1.3) (approximated with intensive simulations), and c) show the behavior of the approximations for different violation probabilities.

**Approximations with and without Stokes.** In order to illustrate the difference in quality of the approximation of  $\mathbf{X}_\varepsilon^*$  when using or not using Stokes constraints, consider the example where  $\mathbf{X} = [-1, 1]$ ,  $f(\mathbf{x}, \omega) = \omega - \mathbf{x}$ ,  $\mathbf{A} = [-0.1, 0.1] \times [0.8, 1]$ , i.e., we consider univariate Gaussian measures with mean approximately 0 and deviation slightly less than 1. For every fixed  $\mathbf{x}$ , due to the simple expression of  $f$  we can express  $\text{Prob}_{\mu_{\mathbf{a}}}(\{\omega \in \Omega : g(\mathbf{x}, \omega) < 0\})$  as an analytic expression in  $\mathbf{a}$ . It is hence relatively easy to obtain a good estimation of  $\rho(\mathbf{x}) := \min_{\mathbf{a} \in \mathbf{A}} \text{Prob}_{\mu_{\mathbf{a}}}(\{\omega \in \Omega : g(\mathbf{x}, \omega) < 0\}) = 1 - \kappa(\mathbf{x})$  by sampling over  $\mathbf{a}$  and taking the minimum. In Figure 1 is displayed  $\mathbf{x} \mapsto \rho(\mathbf{x})$  in black and two different approximations  $w_d$  computed for relaxation orders  $d = 8$  in blue and  $d = 12$  in red. The dashed lines are the polynomials corresponding to problem formulations including Stokes constraints.

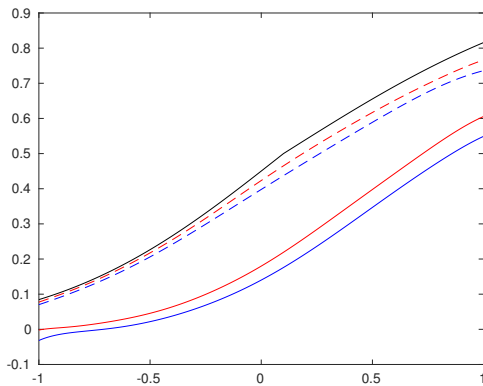


FIGURE 1. Approximation of  $\rho(\mathbf{x})$  (black) by polynomials  $1 - w_8(\mathbf{x})$  (blue) and  $1 - w_{12}(\mathbf{x})$  (red), dashed/solid lines correspond to with/without Stokes constraints

As a first remark observe that, in accordance with the theoretic results, all approximations are underestimators of  $\rho(\mathbf{x})$ . However, the approximations  $(1 - w_d)$  computed with Stokes constraints are much closer to  $\rho$  than the ones computed without. The former approximations are particularly close to  $\rho$  for significant values of violation probability, i.e., for small probabilities on the vertical axis. For higher probabilities they degrade (but are still quite good). This can be due to the non-differentiability of  $\rho$  at  $\mathbf{x} = 0.1$ . In order to display  $\mathbf{X}_\varepsilon^*$  and its  $\mathbf{X}_\varepsilon^d := \{\mathbf{x} \in \mathbf{X} : w_d(\mathbf{x}) \leq \varepsilon\}$ , e.g., for a violation probability of 30% ( $\varepsilon = 0.3$ ) one looks at the sets  $\{\mathbf{x} \in \mathbf{X} : \rho(\mathbf{x}) \geq 0.7\}$  and  $\{\mathbf{x} \in \mathbf{X} : 1 - w_d(\mathbf{x}) \geq 0.7\}$  with  $w_d$  an optimal solution of the dual for the analogue of the step- $d$  relaxation of (4.13). This yields approximately that the interval  $[0.62, 1]$  is the true feasible set. With Stokes, the approximations  $w_8$  and  $w_{12}$  yield the respective intervals  $[0.85, 1]$  and  $[0.75, 1]$  while the approximations without Stokes provide an empty interval.

**Inner approximations from various relaxations.** As seen in the previous example, Stokes constraints are essential for the performance of our approach. In this

section we therefore only report results using these additional constraints. In the second illustrative example,  $\mathbf{X} = [-1, 1]^2$ ,  $f(\mathbf{x}, \omega) = 2\omega \mathbf{x}_2^2 - 2\omega \mathbf{x}_1^2 - 1$  and mean and deviation as in the example before in an environment of 0 and 0.9 respectively. In Figure 2 we plot the feasible set  $\mathbf{X}_\varepsilon^*$  and its approximations ( $\mathbf{X}_\varepsilon^d$ ) for a violation level of 10% ( $\varepsilon = 0.1$ ).

The feasible set is approximated as follows. We discretize  $\mathbf{X}$  into 200 and  $\mathbf{A}$  into 100 steps in each direction respectively. For each point  $\mathbf{x}$  and each combination of parameters  $\mathbf{a}$  we draw 1000 realizations of  $\omega$  from the normal distribution described by  $\mathbf{a}$ . The point  $\mathbf{x}$  is considered to be feasible whenever for each  $\mathbf{a}$ ,  $f(\mathbf{x}, \omega)$  is positive for at least 900 out of the 1000 realizations of  $\omega$ . This simulation takes about 8600 seconds (without the authors claiming to be experts for Monte Carlo simulations) whereas the approximations for  $d = 8, 10, 12$  take 5, 43, and 482 seconds respectively.

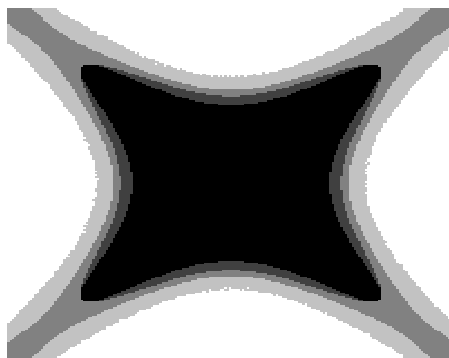


FIGURE 2. Monte Carlo simulation (light grey) of  $\mathbf{X}_\varepsilon^*$  and inner approximations  $\mathbf{X}_\varepsilon^d$  for  $d = 8, 10, 12$ , in decreasing intensity

Inspection of Figure 2 reveals that the feasible set  $\mathbf{X}_\varepsilon^*$  is non-convex. Already the lowest approximation  $\mathbf{X}_\varepsilon^8$  (black) is able to capture this behavior. The next approximation  $\mathbf{X}_\varepsilon^{10}$  (dark grey) is already a bit larger and  $\mathbf{X}_\varepsilon^{12}$  (medium grey) captures a significant part of  $\mathbf{X}_\varepsilon^*$  ( $\approx 74\%$ ). Its computation time is 18 times faster than the one required for the Monte Carlo simulation of  $\mathbf{X}_\varepsilon^*$ . In addition, and in contrast to the approximation via Monte Carlo,  $\mathbf{X}_\varepsilon^d$  is guaranteed to be inside the true feasible set.

**Inner approximations with different violation levels.** In the third example,  $\mathbf{X} = [-1, 1]^3$ ,  $\Omega = \mathbb{R}$ ,  $f(\mathbf{x}, \omega) = -2\omega x_1^2 + 2\omega x_2^2 - 2\omega x_3^2 - 1$ . We compute the inner approximations  $\mathbf{X}_\varepsilon^d$  for  $d = 8, 10, 12$ . To compute the Monte Carlo approximation of  $\mathbf{X}_\varepsilon^*$  in a reasonable time, we fix the mean of the distribution to 0 and the standard deviation  $\sigma$  is taken in the interval  $\mathbf{A} = [0.4, 0.6]$ . For Monte Carlo we discretize  $\mathbf{X}$  and  $\mathbf{A}$  in 100 steps in each direction and draw again 1000 realizations of  $\omega$  for each point and each  $\sigma$ . This simulation takes about 2277 seconds. In the first example we have already seen that the polynomial approximations  $w_d$  are quite good for large violation probabilities. In Table 1 we compare the “volume” of our approximations against the Monte Carlo simulation, i.e., the ratio of the number

$(r, \text{time}) \setminus \varepsilon$	50%	25%	12.5%	6.25%	3.125%
8 (30s)	96.94%	83.07%	69.70%	22.72%	0%
10 (107s)	99.91%	86.70%	73.21%	73.79%	2.48%
12 (633s)	100.0%	90.13%	79.94%	61.31%	27.98%

TABLE 1. Polynomial approximations vs Monte Carlo simulation.

of points admissible for our approximations over the number of points admissible in Monte Carlo. As the polynomial approximations are inner approximations, we expect the rati to be less than one (assuming that Monte Carlo is accurate).

Again the polynomial approximations ( $w_d(x_1, x_2, x_3)$ ) are computed significantly faster than the Monte Carlo approximation  $\rho(x_1, x_2, x_3)$ . As in the first example, for large  $\varepsilon$  the approximations are pretty exact. However, for all relaxation orders  $d$  the quality of approximation decreases with  $\varepsilon$ , and eventually  $\mathbf{X}_{0.03125}^8 = \emptyset$ . However we should not forget that good approximations with small  $\varepsilon$  are difficult to achieve in any case. Therefore it is quite interesting that we can retrieve almost 30% of  $\mathbf{X}_{0.03125}^*$  with  $\mathbf{X}_{0.03125}^{12}$  and using moments up to order 12 only.

## 5. EXTENSIONS

With some *ad-hoc* adjustments, the framework presented in this paper can be extended to consider problems with:

- only first- and second-order moments knowledge (no information about the distributions contributing to the mixture), and
- distributionally robust *joint* chance-constraints.

**5.1. Modeling with only first and second order moments.** As mentioned in Remark 1.8, another possible and related ambiguity set is to consider the family  $\mathcal{M}_{\mathbf{a}}$  of measures on  $\Omega$  whose only first and second-order moments  $\mathbf{a} = (\mathbf{m}, \Sigma)$  belong to some prescribed set  $\mathbf{A}$ . The approach described in this paper also works with the following modifications.

Suppose that  $\omega$  follows some unknown distribution on  $\Omega \subset \mathbb{R}^p$  whose first and second order moments  $\mathbf{a} = (\mathbf{m}, \Sigma) \in \mathbf{A}$ , e.g.  $\mathbf{A} = [\underline{\mathbf{m}}, \overline{\mathbf{m}}] \times \Theta$  where  $\Theta := \{\Sigma : \underline{\delta}I \preceq \Sigma \preceq \overline{\delta}I\}$ , for some  $\underline{\delta} > 0$ . Then  $\mathbf{A}$  is basic semi-algebraic set in  $\mathbb{R}^{p(p+3)/2}$  defined by  $4p$  polynomial inequalities. For instance, with

$$\mathbf{m} = (m_1, m_2); \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix},$$

$$\mathbf{A} = \left\{ \begin{array}{l} \underline{m}_i \leq m_i \leq \overline{m}_i, \quad i = 1, 2; \quad 2\underline{\delta} \leq \sigma_{11} + \sigma_{22} \leq 2\overline{\delta}; \\ (\overline{\delta} - \sigma_{11})(\overline{\delta} - \sigma_{22}) - \sigma_{12}^2 \geq 0; \quad (\sigma_{11} - \underline{\delta})(\sigma_{22} - \underline{\delta}) - \sigma_{12}^2 \geq 0. \end{array} \right\}.$$



The infinite-dimensional LP (3.6) now becomes:

$$\begin{aligned} \rho = \sup_{\phi, \nu, \mu, \psi} \{ & \phi(\mathbf{K}) : \phi + \nu = \mu; \quad \mu_{\mathbf{x}} = \psi_{\mathbf{x}} = \lambda; \\ & \int \mathbf{x}^\alpha \omega_i d\phi + \int \mathbf{x}^\alpha \omega_i d\nu = \int \mathbf{x}^\alpha m_i d\psi, \quad \alpha \in \mathbb{N}^n, i = 1, \dots, p \\ & \int \mathbf{x}^\alpha \omega_i \omega_j d\phi + \int \mathbf{x}^\alpha \omega_i \omega_j d\nu = \int \mathbf{x}^\alpha \sigma_{ij} d\psi, \quad \alpha \in \mathbb{N}^n, 1 \leq i < j \leq p \\ & \phi \in \mathbf{M}_+(\mathbf{K}), \psi, \mu \in \mathbf{M}_+(\mathbf{X} \times \boldsymbol{\Omega}), \psi \in \mathbf{M}_+(\mathbf{X} \times \mathbf{A}) \}. \end{aligned}$$

Then semidefinite relaxations analogues of (4.6) are defined in the obvious way and their associated monotone sequence of optimal values  $(\rho_d)_{d \in \mathbb{N}}$  converges to  $\rho$  as  $d$  increases. As for (4.7), from an optimal solution of their dual one provides inner approximation of  $(\mathbf{X}_\varepsilon^d)_{d \in \mathbb{N}}$  of  $\mathbf{X}_\varepsilon^*$  and analogues of Theorem 4.3 and 4.4 also hold.

**5.2. Joint chance-constraints.** The case of *joint* chance-constraints, i.e., when several probabilistic constraints

$$(5.1) \quad \text{Prob}_\mu(f_j(\mathbf{x}, \omega) > 0, j = 1, \dots, s_f) > 1 - \varepsilon,$$

are considered jointly, is in general significantly more complicated than its relaxation which considers them individually, i.e.,

$$\text{Prob}_\mu(f_j(\mathbf{x}, \omega) > 0) > 1 - \varepsilon, \quad j = 1, \dots, s_f.$$

For instance, tractable formulations valid for individual chance-constraints may not be valid any more for joint chance-constraints.

We next show that joint chance-constraints (5.1) can be modelled in our framework, relatively easily. Instead of the set  $\mathbf{K}$  in (3.1) we now consider the sets:

$$(5.2) \quad \mathbf{K}^j := \{(\mathbf{x}, \omega) \in \mathbf{X} \times \boldsymbol{\Omega} : f_j(\mathbf{x}, \omega) \leq 0\}, \quad j = 1, \dots, s_f.$$

$$(5.3) \quad \mathbf{K}_{\mathbf{x}}^j := \{\omega \in \boldsymbol{\Omega} : (\mathbf{x}, \omega) \in \mathbf{K}^j\}, \quad j = 1, \dots, s_f; \quad \forall \mathbf{x} \in \mathbf{X}$$

$$(5.4) \quad \mathbf{K} := \{(\mathbf{x}, \omega) \in \mathbf{X} \times \boldsymbol{\Omega} : (\mathbf{x}, \omega) \in \bigcup_{j=1}^{s_f} \mathbf{K}^j\},$$

$$(5.5) \quad \mathbf{K}_{\mathbf{x}} := \{\omega \in \boldsymbol{\Omega} : (\mathbf{x}, \omega) \in \bigcup_{j=1}^{s_f} \mathbf{K}_{\mathbf{x}}^j\}, \quad \forall \mathbf{x} \in \mathbf{X}$$

All results of §3, i.e., Theorem 3.2 and Theorem 3.3, remain valid with now  $\mathbf{K}$  and  $\mathbf{K}_{\mathbf{x}}$ ,  $\mathbf{x} \in \mathbf{X}$ , as in (5.4) and (5.5) respectively. Indeed Lemma 3.1 remains valid with  $\mathbf{K}_{\mathbf{x}}$  as (5.5). (In particular we still have  $\mu_{\mathbf{a}}(\partial \mathbf{K}_{\mathbf{x}}) = 0$  for all  $\mathbf{a} \in \mathbf{A}$ , as now the boundary  $\partial \mathbf{K}_{\mathbf{x}}$  is contained in a finite union of zero sets of polynomials.)

What is not obvious is how to define the analogues of the semidefinite relaxations (3.6) because  $\mathbf{K}$  is *not* a *basic* semi-algebraic set any more. It is a finite union  $\bigcup_{j=1}^{s_f} \mathbf{K}^j$  of basic semi-algebraic sets with overlaps.

The analogue of the infinite-dimensional LP (3.6) reads:

$$(5.6) \quad \hat{\rho} = \sup_{\phi_j, \psi \geq 0} \left\{ \sum_{j=1}^{s_f} \int_{\mathbf{K}^j} d\phi_j : \sum_{j=1}^{s_f} \phi_j \leq \mathbf{T}^* \psi; \quad \psi_{\mathbf{x}} = \lambda, \right. \\ \left. \phi_j \in \mathcal{M}_+(\mathbf{K}^j), \psi \in \mathcal{P}(\mathbf{X} \times \mathbf{A}) \right\},$$

where  $\mathbf{T}^*$  is defined in Lemma 2.8. It is important to emphasize that even though the sets  $\mathbf{K}^j$  overlap, we do *not* require that the measures  $\phi_j$  are mutually singular.

The dual of (5.6) reads:

$$(5.7) \quad \hat{\rho}^* = \inf_{h,w} \left\{ \int_{\mathbf{X}} w d\lambda : \begin{array}{l} h(\mathbf{x}, \omega) \geq 1 \quad \text{on } \mathbf{K}^j, j = 1, \dots, s_f \\ w(\mathbf{x}) - \mathbf{T}h(\mathbf{x}, \mathbf{a}) \geq 0 \quad \text{on } \mathbf{X} \times \mathbf{A}; \quad h \geq 0 \text{ on } \mathbf{X} \times \mathbf{\Omega}, \\ w \in \mathbb{R}[\mathbf{x}]; h \in \mathbb{R}[\mathbf{x}, \omega] \end{array} \right\}.$$

**Theorem 5.1.** *Let  $\mathbf{K}^j$ ,  $j = 1, \dots, s_f$ , and  $\mathbf{K}$  be as in (5.2) and (5.4) respectively. Then:*

- (i) *The optimal value  $\rho$  of (3.6) and the optimal value  $\hat{\rho}$  of (5.6) are identical.*
- (ii) *Define the functions:*

$$(5.8) \quad (\mathbf{x}, \omega) \mapsto \theta_j(\mathbf{x}, \omega) := \frac{1}{|\{\ell \in \{1, \dots, s_f\} : (\mathbf{x}, \omega) \in \mathbf{K}_\ell\}|}, \quad j = 1, \dots, s_f,$$

and let  $(\phi^*, \psi^*)$  be an optimal solution of (5.6). Then  $(\phi_1^*, \dots, \phi_{s_f}^*, \psi^*)$  with

$$(5.9) \quad d\phi_j^*(\mathbf{x}, \omega) := \mathbf{1}_{\mathbf{K}^j}(\mathbf{x}, \omega) \theta_j(\mathbf{x}, \omega) d\phi^*(\mathbf{x}, \omega), \quad j = 1, \dots, s_f,$$

is an optimal solution of (5.6) and  $\phi^* = \sum_j \phi_j^*$ .

*Proof.* Let  $(\phi_1, \dots, \phi_{s_f}, \psi)$  be an arbitrary feasible solution of (5.6) and let  $\phi := \sum_{j=1}^{s_f} \phi_j$ . Then  $\phi \in \mathcal{M}(\mathbf{K})_+$  and  $\phi \leq \mathbf{T}^*\psi$ . Therefore  $(\phi, \psi)$  is feasible for (3.6). Moreover, as  $\text{supp}(\phi_j) \subset \mathbf{K}^j \subset \mathbf{K}$  for all  $j = 1, \dots, s_f$ ,

$$\sum_{j=1}^{s_f} \int_{\mathbf{K}^j} d\phi_j = \sum_{j=1}^{s_f} \int_{\mathbf{K}} d\phi_j = \int_{\mathbf{K}} d\left(\sum_{j=1}^{s_f} \phi_j\right) = \int_{\mathbf{K}} d\phi = \phi(\mathbf{K}),$$

which shows that  $\hat{\rho} \leq \rho$ . To prove the reverse inequality consider the functions  $(\theta_j)$  in (5.8), and let  $(\phi^*, \psi^*) \in \mathcal{M}_+(\mathbf{K}) \times \mathcal{M}_+(\mathbf{X} \times \mathbf{A})$  be an optimal solution of (3.6). Observe that

$$\sum_{j=1}^{s_f} \theta_j(\mathbf{x}, \omega) = 1, \quad \forall (\mathbf{x}, \omega) \in \mathbf{K}.$$

Let  $(\phi_j^*)$ ,  $j := 1, \dots, s_f$ , be as in (5.9). Then  $\phi_j^* \in \mathcal{M}_+(\mathbf{K}_j)$ ,  $j = 1, \dots, s_f$ , and  $\sum_{j=1}^{s_f} \phi_j^*(\mathbf{K}_j) = \phi^*(\mathbf{K})$ . Hence  $\rho \leq \hat{\rho}$ . Therefore  $(\phi_1^*, \dots, \phi_{s_f}^*, \psi^*)$  is an optimal solution of (5.6) and  $\phi^* = \sum_j \phi_j^*$ .  $\square$

As a consequence of Theorem 5.1, Theorem 3.3 also holds with now  $\mathbf{K}_\mathbf{x}$  as in (5.5). In the original proof just use  $\phi = \sum_j \phi_j$  and the definitions of  $\mathbf{K}$  and  $\mathbf{K}_\mathbf{x}$  in (5.4)-(5.5).

**5.3. Semidefinite relaxations.** We briefly describe the semidefinite relaxations of the LP (5.6), which are the analogues of (4.6) for the LP (3.6). For every  $j = 1, \dots, s_f$  let  $g_j^* := -f_j$  and let  $d_j^* := \lceil \deg(g_j^*)/2 \rceil$ . Let  $2d_{\min}$  to be the largest degree appearing in the polynomials that describe  $\mathbf{K}, \mathbf{\Omega}, \mathbf{A}$ , and consider the semidefinite

programs indexed by  $d \geq d_{\min}$ .

$$\begin{aligned}
(5.10) \quad \rho_d = & \sup_{\mathbf{y}^j, \mathbf{u}, \mathbf{v}} \sum_{j=1}^{s_f} y_{00}^j \\
\text{s.t.} \quad & L_{\mathbf{y}^1 + \dots + \mathbf{y}^{s_f} + \mathbf{u}}(\mathbf{x}^\alpha \omega^\beta) - L_{\mathbf{v}}(\mathbf{x}^\alpha p_\beta(\mathbf{a})) = 0, \quad |\alpha + \deg(p_\beta)| \leq 2d, \\
& L_{\mathbf{v}}(\mathbf{x}^\alpha) = \lambda_\alpha, \quad \alpha \in \mathbb{N}_{2d}^n, \\
& \mathbf{M}_d(\mathbf{y}^j), \mathbf{M}_d(\mathbf{u}), \mathbf{M}_d(\mathbf{v}) \succeq 0, \\
& \mathbf{M}_{d-d_j^*}(g_j^* \mathbf{y}^j) \succeq 0, \quad j = 1, \dots, s_f, \\
& \mathbf{M}_{d-d_\ell}(g_\ell \mathbf{y}^j), \mathbf{M}_{d-d_\ell}(g_\ell \mathbf{u}), \mathbf{M}_d(g_\ell \mathbf{v}) \succeq 0, \\
& \ell = 1, \dots, m; \quad j = 1, \dots, s_f, \\
& \mathbf{M}_{d-d_\ell^1}(s_\ell \mathbf{y}^j), \mathbf{M}_{d-d_\ell^1}(s_\ell \mathbf{u}) \succeq 0, \\
& \ell = 1, \dots, \bar{s}; \quad j = 1, \dots, s_f, \\
& \mathbf{M}_{d-d_\ell'}(q_\ell \mathbf{v}) \succeq 0, \quad \ell = 1, \dots, L,
\end{aligned}$$

where  $\mathbf{y}^j = (y_{\alpha\beta}^j)$ ,  $\mathbf{u} = (v_{\alpha\beta})$ ,  $(\alpha, \beta) \in \mathbb{N}^n \times \mathbb{N}^p$ ,  $j = 1, \dots, s_f$ , and  $\mathbf{v} = (v_{\alpha\eta})$ ,  $(\alpha, \eta) \in \mathbb{N}^n \times \mathbb{N}^t$ .

The dual of (5.10) is a reinforcement of (5.7) and its interpretation in terms of SOS positivity certificates of size parametrized by  $d$  (the analogue of (4.7)) reads:

$$\begin{aligned}
(5.11) \quad \rho_d^* = & \inf_{h, w, \sigma_j^i} \int_{\mathbf{X}} w(\mathbf{x}) d\lambda(\mathbf{x}) : \\
\text{s.t.} \quad & h(\mathbf{x}, \omega) - 1 = \sum_{\ell=0}^m \sigma_{j\ell}^1 g_\ell + \sigma_{j, m+1}^1 g_j^* + \sum_{\ell'=1}^{\bar{s}} \sigma_{j\ell'}^1 s_{\ell'}, \quad \forall(\mathbf{x}, \omega); \\
& \text{for all } j = 1, \dots, s_f; \\
& h(\mathbf{x}, \omega) = \sum_{\ell=0}^m \sigma_\ell^2 g_\ell + \sum_{\ell'=1}^{\bar{s}} \sigma_{\ell'}^2 s_{\ell'}, \quad \forall(\mathbf{x}, \omega); \\
& w(\mathbf{x}) - \sum_{\alpha, \beta} h_{\alpha\beta} \mathbf{x}^\alpha p_\beta(\mathbf{a}) = \sum_{\ell=0}^m \sigma_\ell^3 g_\ell + \sum_{\ell'=1}^L \sigma_{\ell'}^3 q_{\ell'}, \quad \forall(\mathbf{x}, \mathbf{a}); \\
& \deg(h), \deg(w) \leq 2d; \quad \sigma_\ell^1 \in \Sigma[\mathbf{x}, \omega]_{d-d_\ell}, \quad j = 0, \dots, m \\
& \sigma_{j\ell}^1, \sigma_\ell^2 \in \Sigma[\mathbf{x}, \omega]_{d-d_\ell}; \quad \sigma_{j, m+1}^1 \in \Sigma[\mathbf{x}, \omega]_{d-d_j^*}, \quad \ell = 0, \dots, m; \quad j = 1, \dots, s_f \\
& \sigma_\ell^3 \in \Sigma[\mathbf{x}, \mathbf{a}]_{d-d_\ell}; \quad \ell = 0, \dots, m; \quad \sigma_{\ell'}^3 \in \Sigma[\mathbf{x}, \mathbf{a}]_{d-d_{\ell'}'}; \quad \ell' = 1, \dots, L \\
& \sigma_{j\ell'}^1, \sigma_{\ell'}^2 \in \Sigma[\mathbf{x}, \omega]_{d-d_{\ell'}'}, \quad j = 1, \dots, s_f; \quad \ell' = 1, \dots, L
\end{aligned}$$

where  $h(\mathbf{x}, \omega) = \sum_{|\alpha+\beta| \leq 2d} h_{\alpha\beta} \mathbf{x}^\alpha \omega^\beta$ , and  $w(\mathbf{x}) = \sum_{|\alpha| \leq 2d} w_\alpha \mathbf{x}^\alpha$ .

For every  $d \geq d_{\min}$ ,  $\rho_d \geq \hat{\rho}$  ( $= \rho$ ) and Theorem 4.4 is valid for the hierarchy of semidefinite relaxations (5.10). Next, under the same assumptions of non-empty interior for  $\mathbf{X}, \mathbf{A}, \Omega, \mathbf{K}, \mathbf{K}^j$  and  $(\mathbf{X} \times \Omega) \setminus \mathbf{K}$ , Theorem 4.3 is also valid for the dual hierarchy (5.11).

## 6. CONCLUSION

Computing or even approximating the feasible set associated with a distributionally-robust chance-constraint is a challenging problem. We have described a systematic

numerical scheme which provides a monotone sequence (a hierarchy) of inner approximations, all in the form  $\{\mathbf{x} \in \mathbf{X} : w_d(\mathbf{x}) < \varepsilon\}$  for some polynomial of increasing degree  $d$ , with strong asymptotic guarantees as  $d$  increases. To the best of our knowledge it is the first result of this type at this level of generality. Of course this comes with a price as the polynomial which defines each approximation is obtained by solving a semidefinite program whose size increases with its degree. Therefore and so far, this approach is limited to problems of small dimension (except perhaps if some sparsity can be exploited). So in its present form this contribution should be considered as complementary (rather than a competitor) to other algorithmic approaches where scalability is of primary importance. However it may also provide useful insights and a benchmark (for small dimension problems) for the latter approaches.

## 7. APPENDIX

**Lemma 7.1.** *Under Assumption 2.5, every  $\mu \in \mathcal{M}_{\mathbf{a}}$  is moment determinate.*

*Proof.* As  $\mu \in \mathcal{M}_{\mathbf{a}}$ , there exists  $\varphi \in \mathcal{P}(\mathbf{A})$  such that

$$\mu(B) = \int_{\Omega} \mu_{\mathbf{a}}(B) d\varphi(\mathbf{a}), \quad B \in \mathcal{B}(\Omega).$$

By Assumption 2.5, there exists  $c, \gamma$  such that for every  $i = 1, \dots, p$ ,

$$\int_{\Omega} \exp(c|\omega_i|) d\mu_{\mathbf{a}}(\omega) < \gamma, \quad \forall \mathbf{a} \in \mathbf{A},$$

and therefore

$$\sup_{i=1, \dots, p} \int_{\Omega} \exp(c|\omega_i|) d\mu(\omega) < \gamma.$$

As  $(c\omega_i)^{2j}/(2j)! < \exp(|\omega_i|)$ , one obtains  $\int_{\Omega} \omega_i^{2j} d\mu(\omega) < c^{-2j}\gamma(2j)!$  for all  $j = 1, 2, \dots$  and all  $i = 1, \dots, p$ . But this implies

$$\sum_{i=1}^{\infty} \left( \int_{\Omega} \omega_i^{2j} d\mu(\omega) \right)^{-1/2j} = +\infty, \quad i = 1, \dots, p,$$

that is,  $\mu$  satisfies Carleman's condition (2.2), and so is moment determinate.  $\square$

### 7.1. Proof of Lemma 3.1.

*Proof.* With  $\mathbf{x} \in \mathbf{X}$  fixed, let  $\theta := \sup_{\mathbf{a} \in \mathbf{A}} \mu_{\mathbf{a}}(\mathbf{K}_{\mathbf{x}})$ . For every  $\mu \in \mathcal{M}_{\mathbf{a}}$ , there exists  $\varphi \in \mathcal{P}(\mathbf{A})$  such that  $\mu(\mathbf{K}_{\mathbf{x}}) = \int_{\mathbf{A}} \mu_{\mathbf{a}}(\mathbf{K}_{\mathbf{x}}) d\varphi(\mathbf{a})$  and so  $\mu(\mathbf{K}_{\mathbf{x}}) \leq \theta$  for all  $\mu \in \mathcal{M}_{\mathbf{a}}$ . Conversely  $\sup_{\mu \in \mathcal{M}_{\mathbf{a}}} \mu(\mathbf{K}_{\mathbf{x}}) \geq \theta$  because  $\mu_{\mathbf{a}} \in \mathcal{M}_{\mathbf{a}}$  for all  $\mathbf{a} \in \mathbf{A}$ . Next let  $v : \mathbf{X} \times \mathbf{A} \rightarrow [0, 1]$  be given by  $v(\mathbf{x}, \mathbf{a}) := \mu_{\mathbf{a}}(\mathbf{K}_{\mathbf{x}})$ . By construction  $0 \leq v \leq 1$  on  $\mathbf{X} \times \mathbf{A}$  and if  $v(\mathbf{x}, \cdot)$  is upper-semicontinuous on  $\mathbf{A}$  for every  $\mathbf{x} \in \mathbf{X}$ , then by [29, Proposition 4.4, p. 2018] there exists a *measurable selector*  $\mathbf{x} \mapsto \mathbf{a}(\mathbf{x}) \in \mathbf{A}$ ,  $\mathbf{x} \in \mathbf{X}$ , such that  $v(\mathbf{x}, \mathbf{a}(\mathbf{x})) = \max\{v(\mathbf{x}, \mathbf{a}) : \mathbf{a} \in \mathbf{A}\}$ , that is, the desired result (3.5) holds.

So it remains to prove that  $v(\mathbf{x}, \cdot)$  is upper-semicontinuous on  $\mathbf{A}$  for every  $\mathbf{x} \in \mathbf{X}$ . In fact we even prove that  $v(\mathbf{x}, \cdot)$  is continuous on  $\mathbf{A}$  for every  $\mathbf{x} \in \mathbf{X}$ . So let  $(a_n)_{n \in \mathbb{N}} \subset \mathbf{A}$  with  $a_n \rightarrow \mathbf{a} \in \mathbf{A}$  as  $n \rightarrow \infty$ . Let  $q$  be an arbitrary bounded continuous function on  $\Omega$ . Then by Assumption 2.4(iv),

$$\lim_{n \rightarrow \infty} \int_{\Omega} q d\mu_{a_n}(\omega) = \lim_{n \rightarrow \infty} Q(a_n) = Q(\mathbf{a}) = \int_{\Omega} q d\mu_{\mathbf{a}}(\omega),$$

which proves that  $\mu_{\mathbf{a}_n} \Rightarrow \mu_{\mathbf{a}}$  as  $n \rightarrow \infty$  (where  $\Rightarrow$  denotes the weak convergence of probability measures ; see Billingsley [4]). In addition, in view of the definition of  $\mathbf{K}_{\mathbf{x}}$  in (3.2), its boundary  $\partial\mathbf{K}_{\mathbf{x}}$  is contained in the zero set of some polynomials and therefore, by Assumption 2.4(iii),  $\mu_{\mathbf{a}_n}(\partial\mathbf{K}_{\mathbf{x}}) = \mu_{\mathbf{x}}(\partial\mathbf{K}_{\mathbf{x}}) = 0$  for all  $n$  (i.e.  $\partial\mathbf{K}_{\mathbf{x}}$  is a  $\mu_{\mathbf{a}_n}$ -continuity set). Hence by the Portmanteau theorem [4, Theorem 2.1, p. 11] it follows that

$$\lim_{n \rightarrow \infty} \mu_{\mathbf{a}_n}(\mathbf{K}_{\mathbf{x}}) = \lim_{n \rightarrow \infty} v(\mathbf{x}, \mathbf{a}_n) = \mu_{\mathbf{a}}(\mathbf{K}_{\mathbf{x}}) = v(\mathbf{x}, \mathbf{a}),$$

i.e.,  $v(\mathbf{x}, \cdot)$  is continuous on  $\mathbf{A}$  for every  $\mathbf{x} \in \mathbf{X}$ . In addition  $\kappa(\mathbf{x}) = v(\mathbf{x}, \mathbf{a}(\mathbf{x}))$  is also measurable.  $\square$

## 7.2. Proof of Theorem 3.3.

*Proof.* *Weak duality* holds because for every feasible solution  $(w, h)$  of (3.7) and  $(\phi, \psi)$  of (3.6), one has:

$$\begin{aligned} \int_{\mathbf{X}} w d\lambda &= \int_{\mathbf{X} \times \mathbf{A}} w(\mathbf{x}) d\psi(\mathbf{x}, \mathbf{a}) \geq \int_{\mathbf{X} \times \mathbf{A}} \mathbf{T}h(\mathbf{x}, \mathbf{a}) d\psi(\mathbf{x}, \mathbf{a}) \\ &= \int_{\mathbf{X} \times \Omega} h(\mathbf{x}, \omega) \mathbf{T}^* \psi(d(\mathbf{x}, \omega)) \\ &\geq \int_{\mathbf{K}} h(\mathbf{x}, \omega) d\phi(\mathbf{x}, \omega) \\ &\geq \int_{\mathbf{K}} d\phi = \phi(\mathbf{K}). \end{aligned}$$

Moreover let  $\psi^*$  be an optimal solution of (3.6) as in Theorem 3.2, so that  $\psi^* = \delta_{\mathbf{a}(\mathbf{x})} \lambda(d\mathbf{x})$ . Then for every  $\mathbf{x} \in \mathbf{X}$ :

$$\begin{aligned} w(\mathbf{x}) &\geq \int_{\mathbf{A}} \mathbf{T}h(\mathbf{x}, \mathbf{a}) \psi^*(d\mathbf{a}|\mathbf{x}) = \int_{\Omega} h(\mathbf{x}, \omega) d\mu_{\mathbf{a}(\mathbf{x})}(\omega) \\ &\geq \int_{\mathbf{K}_{\mathbf{x}}} h(\mathbf{x}, \omega) d\mu_{\mathbf{a}(\mathbf{x})}(\omega) \\ &\geq \int_{\mathbf{K}_{\mathbf{x}}} d\mu_{\mathbf{a}(\mathbf{x})}(\omega) = \kappa(\mathbf{x}), \end{aligned}$$

i.e.,  $w(\mathbf{x}) \geq \kappa(\mathbf{x})$  for all  $\mathbf{x} \in \mathbf{X}$ . In particular

$$\{\mathbf{x} \in \mathbf{X} : w(\mathbf{x}) < \varepsilon\} \subset \{\mathbf{x} \in \mathbf{X} : \kappa(\mathbf{x}) < \varepsilon\} = \mathbf{X}_{\varepsilon}^*.$$

Next, if there is no duality gap, i.e., if  $\rho = \rho^*$ , then for a minimizing sequence  $(w_n, h_n)$  of (3.7),

$$\lim_{n \rightarrow \infty} \int_{\mathbf{X}} (w_n(\mathbf{x}) - \kappa(\mathbf{x})) d\lambda(\mathbf{x}) = \rho^* - \rho = 0,$$

that is,  $w_n$  converges to  $\kappa(\mathbf{x})$  in  $L_1(\mathbf{X}, \lambda)$ . Finally, by Ash [3, Theorem 2.5.1], convergence in  $L_1(\mathbf{X}, \lambda)$  implies convergence in  $\lambda$ -measure, and so, for every fixed  $0 < \ell \in \mathbb{N}$ ,

$$(7.1) \quad \lim_{n \rightarrow \infty} \lambda \{\mathbf{x} \in \mathbf{X} : |w_n(\mathbf{x}) - \kappa(\mathbf{x})| > 1/\ell\} = 0.$$

Next, observe that

$$\mathbf{X}_\varepsilon^* = \{\mathbf{x} \in \mathbf{X} : \kappa(\mathbf{x}) < \varepsilon\} = \bigcup_{\ell=1}^{\infty} \underbrace{\{\mathbf{x} \in \mathbf{X} : \kappa(\mathbf{x}) < \varepsilon - 1/\ell\}}_{:=R_\ell}$$

and so  $\lambda(\mathbf{X}_\varepsilon^*) = \lim_{\ell \rightarrow \infty} \lambda(R_\ell)$ . Next

$$\lambda(R_\ell) = \lambda(R_\ell \cap \{\mathbf{x} : w_n(\mathbf{x}) < \varepsilon\}) + \lambda(R_\ell \cap \{\mathbf{x} : w_n(\mathbf{x}) \geq \varepsilon\}).$$

By convergence in measure (7.1),  $\lim_{n \rightarrow \infty} \lambda(R_\ell \cap \{\mathbf{x} : w_n(\mathbf{x}) \geq \varepsilon\}) = 0$ . Hence

$$\lambda(R_\ell) = \lim_{n \rightarrow \infty} \lambda(R_\ell \cap \{\mathbf{x} : w_n(\mathbf{x}) < \varepsilon\}) \leq \lim_{n \rightarrow \infty} \lambda(\{\mathbf{x} : w_n(\mathbf{x}) < \varepsilon\}) \leq \lambda(\mathbf{X}_\varepsilon^*),$$

and as  $\lambda(R_\ell) \rightarrow \lambda(\mathbf{X}_\varepsilon^*)$ ,  $\lim_{n \rightarrow \infty} \lambda(\{\mathbf{x} : w_n(\mathbf{x}) < \varepsilon\}) = \lambda(\mathbf{X}_\varepsilon^*)$ .  $\square$

### 7.3. Proof of Lemma 4.1.

*Proof.* If  $\Omega$  is compact then it follows from the definition of  $\mathbf{T}$  and  $\mathbf{T}^*$ . For the general case where  $\Omega$  is not necessarily compact, disintegrate  $\nu$  and  $\psi$  as

$$d\nu(\mathbf{x}, \Omega) = \hat{\nu}(d\omega|\mathbf{x}) \nu_{\mathbf{x}}(d\mathbf{x}), \quad d\psi(\mathbf{x}, \mathbf{a}) = \hat{\psi}(d\mathbf{a}|\mathbf{x}) \psi_{\mathbf{x}}(d\mathbf{x}).$$

By (4.1) with  $\beta = 0$ ,  $\int_{\mathbf{X}} \mathbf{x}^\alpha \nu_{\mathbf{x}}(d\mathbf{x}) = \int_{\mathbf{X}} \mathbf{x}^\alpha \psi_{\mathbf{x}}(d\mathbf{x})$  for all  $\alpha \in \mathbb{N}^n$ , and as  $\mathbf{X}$  is compact it follows that  $\psi_{\mathbf{x}} = \nu_{\mathbf{x}}$ . Next, fix  $\beta \in \mathbb{N}^p$ . Then for every  $\alpha \in \mathbb{N}^n$

$$\int_{\mathbf{X}} \mathbf{x}^\alpha \left( \int_{\Omega} \omega^\beta \hat{\nu}(d\omega|\mathbf{x}) \right) \nu_{\mathbf{x}}(d\mathbf{x}) = \int_{\mathbf{X}} \mathbf{x}^\alpha \left( \int_{\mathbf{A}} p_\beta(\mathbf{a}) \hat{\psi}(d\mathbf{a}|\mathbf{x}) \right) \nu_{\mathbf{x}}(d\mathbf{x}),$$

and again as  $\mathbf{X}$  is compact this implies

$$\int_{\Omega} \omega^\beta \hat{\nu}(d\omega|\mathbf{x}) = \int_{\mathbf{A}} p_\beta(\mathbf{a}) \hat{\psi}(d\mathbf{a}|\mathbf{x}), \quad \forall \mathbf{x} \in \mathbf{X} \setminus B_\beta,$$

where  $B_\beta \in \mathcal{B}(\mathbf{X})$  is such that  $\nu_{\mathbf{x}}(B_\beta) = 0$ . As  $\beta \in \mathbb{N}^p$  was arbitrary,

$$\int_{\Omega} \omega^\beta \hat{\nu}(d\omega|\mathbf{x}) = \int_{\mathbf{A}} p_\beta(\mathbf{a}) \hat{\psi}(d\mathbf{a}|\mathbf{x}), \quad \forall \beta \in \mathbb{N}^p, \forall \mathbf{x} \in \mathbf{X} \setminus B^*,$$

where  $\nu_{\mathbf{x}}(B^*) = \nu_{\mathbf{x}}(\bigcup_{\beta} B_\beta) = 0$ . Next, define the measure  $\varphi_{\mathbf{x}}$  on  $\Omega$  by

$$\varphi_{\mathbf{x}}(B) = \int_{\mathbf{A}} \mu_{\mathbf{a}}(B) \hat{\psi}(d\mathbf{a}|\mathbf{x}), \quad \forall \mathbf{x} \in \mathbf{X} \setminus B^*,$$

which is well defined by Assumption 2.4(i). Moreover by construction,  $\varphi_{\mathbf{x}} \in \mathcal{M}_{\mathbf{a}}$  for all  $\mathbf{x} \in \mathbf{X} \setminus B^*$ , and

$$\begin{aligned} \int_{\Omega} \omega^\beta \hat{\nu}(d\omega|\mathbf{x}) &= \int_{\mathbf{A}} p_\beta(\mathbf{a}) \hat{\psi}(d\mathbf{a}|\mathbf{x}), \quad \forall \beta \\ &= \int_{\mathbf{A}} \left( \int_{\Omega} \omega^\beta \mu_{\mathbf{a}}(d\omega) \right) \hat{\psi}(d\mathbf{a}|\mathbf{x}), \quad \forall \beta \\ &= \int_{\Omega} \omega^\beta \varphi_{\mathbf{x}}(d\omega), \quad \forall \beta. \end{aligned}$$

By Lemma 7.1,  $\varphi_{\mathbf{x}}$  is moment determinate and therefore  $\hat{\nu}(d\omega|\mathbf{x}) = \varphi_{\mathbf{x}}$  for all  $\mathbf{x} \in \mathbf{X} \setminus B^*$ . Next, let  $g \in \mathcal{B}(\mathbf{X} \times \Omega)$  be fixed arbitrary. Then:

$$\begin{aligned} \langle g, \nu \rangle &= \langle g, \hat{\nu}(d\omega|\mathbf{x}) \nu_{\mathbf{x}}(d\mathbf{x}) \rangle &= \langle g, \varphi_{\mathbf{x}}(d\omega) \nu_{\mathbf{x}}(d\mathbf{x}) \rangle \\ &= \langle g, \varphi_{\mathbf{x}}(d\omega) \psi_{\mathbf{x}}(d\mathbf{x}) \rangle \\ &= \langle \mathbf{T}g, \hat{\psi}(d\mathbf{a}|\mathbf{x}) \psi_{\mathbf{x}}(d\mathbf{x}) \rangle \\ &= \langle \mathbf{T}g, \psi \rangle = \langle g, \mathbf{T}^* \psi \rangle, \end{aligned}$$

and as it holds for all  $g \in \mathcal{B}(\mathbf{X} \times \Omega)$ ,  $\nu = \mathbf{T}^* \psi$ .  $\square$

#### 7.4. Proof of Theorem 4.3.

*Proof.* (i) We first prove that Slater's condition holds for (4.6). Observe that for all feasible solutions,  $y_{00} + u_{00} = L_{\mathbf{v}}(1) = 1$  and therefore  $0 \leq \rho_d \leq 1$ . Let  $\lambda_{\mathbf{A}}$  be the Lebesgue measure on  $\mathbf{A}$ , normalized to a probability measure. Let  $\mathbf{y}$  be the the moments of the measure  $d\phi(\mathbf{x}, \omega) = 1_{\mathbf{S}}(\mathbf{x}) d\lambda(\mathbf{x}) \otimes \hat{\mu}(d\omega)$ , where

$$\hat{\mu}(B) = \int_{\mathbf{A}} \mu_{\mathbf{a}}(B) d\lambda_{\mathbf{A}}(\mathbf{a}), \quad B \in \mathcal{B}(\Omega).$$

Similarly, let  $\nu := \lambda \otimes \hat{\mu}$  (and so  $\phi \leq \nu$ ) and let  $\psi = \lambda_{\mathbf{A}} \otimes \lambda$ . Let  $\mathbf{y}$  (resp.  $\mathbf{u}$ ) be the vector of moments of  $\phi$  (resp.  $\nu - \phi$ ) up to order  $2d$ , and let  $\mathbf{v}$  be the vector of moments of  $\psi$  up to order  $2d$ . Then  $\mathbf{M}_d(\mathbf{y}), \mathbf{M}_d(\mathbf{u}), \mathbf{M}_d(\mathbf{v}) \succ 0$ . Similarly  $\mathbf{M}_{d-d_j}(\mathbf{y}) \succ 0$ ,  $j = 0, \dots, m$ ,  $\mathbf{M}_{d-d_j}(g_j \mathbf{u}) \succ 0$ ,  $j = 1, \dots, m$ , and  $\mathbf{M}_{d-d'_\ell}(q_\ell \mathbf{v}) \succ 0$ ,  $\ell = 1, \dots, L$ , because  $\mathbf{X} \times \Omega, \mathbf{K}, \mathbf{X} \times \Omega \setminus \mathbf{K}, \mathbf{A}$  all have nonempty interior. Moreover, as

$$\begin{aligned} \int \mathbf{x}^\alpha \omega^\beta d\nu(\mathbf{x}, \omega) &= \int_{\mathbf{X}} \mathbf{x}^\alpha d\lambda(\mathbf{x}) \int_{\mathbf{A}} \left( \int_{\Omega} \omega^\beta d\mu_{\mathbf{a}}(\omega) \right) d\lambda_{\mathbf{A}}(\mathbf{a}) \\ &= \int_{\mathbf{X}} \mathbf{x}^\alpha d\lambda(\mathbf{x}) \int_{\mathbf{A}} p_\beta(\mathbf{a}) d\lambda_{\mathbf{A}}(\mathbf{a}) = \int_{\mathbf{X} \times \mathbf{A}} \mathbf{x}^\alpha p_\beta(\mathbf{a}) d\psi(\mathbf{x}, \mathbf{a}), \end{aligned}$$

we deduce  $L_{\mathbf{y}+\mathbf{u}}(\mathbf{x}^\alpha \omega^\beta) = L_{\mathbf{v}}(\mathbf{x}^\alpha p_\beta(\mathbf{a}))$ , and therefore  $(\mathbf{y}, \mathbf{u}, \mathbf{v})$  is an admissible solution of (4.6) which is strictly feasible, i.e., Slater's condition holds for (4.6) and therefore strong duality  $\rho_d = \rho_d^*$  holds. In particular, as  $\rho_d < \infty$ , (4.7) has an optimal solution  $(h_d, w_d, \sigma_j^i)$ .

(ii) Next feasibility in (4.7) implies

$$\begin{aligned} h_d(\mathbf{x}, \omega) &\geq 1, \quad \forall (\mathbf{x}, \omega) \in \mathbf{K}, \\ h_d(\mathbf{x}, \omega) &\geq 0, \quad \forall (\mathbf{x}, \omega) \in \mathbf{X} \times \Omega, \\ w_d(\mathbf{x}) - \sum_{\alpha, \beta} h_{d, \alpha, \beta} \mathbf{x}^\alpha p_\beta(\mathbf{a}) &\geq 0, \quad \forall (\mathbf{x}, \mathbf{a}) \in \mathbf{X} \times \mathbf{A}. \end{aligned}$$

Let  $(\phi^*, \psi^*)$  be optimal solution of (3.6), as in Theorem 3.2, and let  $d\nu^*(\mathbf{x}, \omega) : \mu_{\mathbf{a}(\mathbf{x})}(d\omega) \lambda(d\mathbf{x})$ . Let  $\mathbf{x} \in \mathbf{X}$  be fixed. Integrating the first w.r.t.  $\nu_{\mathbf{x}}^*(d\omega|\mathbf{x}) (= \mu_{\mathbf{a}(\mathbf{x})}(d\omega))$ , the third one w.r.t.  $\psi^*(d\mathbf{a}|\mathbf{x}) (= \delta_{\mathbf{a}(\mathbf{x})})$ , and using the second inequality yields:

$$\begin{aligned} w_d(\mathbf{x}) + \sum_{\alpha, \beta} h_{d, \alpha, \beta} \mathbf{x}^\alpha \left( \int_{\Omega} \omega^\beta d\mu_{\mathbf{a}(\mathbf{x})}(\omega) - \underbrace{\int_{\mathbf{A}} p_\beta(\mathbf{a}) d\psi^*(d\mathbf{a}|\mathbf{x})}_{= \int_{\Omega} \omega^\beta d\mu_{\mathbf{a}(\mathbf{x})}(\omega)} \right) \\ \geq \begin{cases} \int_{\mathbf{K}_{\mathbf{x}}} \nu_{\mathbf{x}}^*(d\omega|\mathbf{x}) = \kappa(\mathbf{x}), & \text{if } \mathbf{K}_{\mathbf{x}} \neq \emptyset, \\ 0 (= \kappa(\mathbf{x})) & \text{otherwise.} \end{cases} \end{aligned}$$

In other words

$$(7.2) \quad w_d(\mathbf{x}) \geq \kappa(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbf{X}.$$

Therefore

$$\{\mathbf{x} \in \mathbf{X} : w_d(\mathbf{x}) < \varepsilon\} \subseteq \{\mathbf{x} \in \mathbf{X} : \kappa(\mathbf{x}) < \varepsilon\} = \mathbf{X}_\varepsilon^*.$$

Next if  $\lim_{d \rightarrow \infty} \rho_d = \rho$  then

$$\lim_{d \rightarrow \infty} \int_{\mathbf{X}} w_d d\lambda = \lim_{d \rightarrow \infty} \rho_d^* = \lim_{d \rightarrow \infty} \rho_d = \int_{\mathbf{X}} \kappa(\mathbf{x}) \lambda(d\mathbf{x}),$$

which yields

$$\int_{\mathbf{X}} (w_d(\mathbf{x}) - \kappa(\mathbf{x})) d\lambda(\mathbf{x}) \rightarrow 0 \quad \text{as } d \rightarrow \infty,$$

which combined with (7.2), yields  $w_d \rightarrow \kappa$  in  $L_1(\mathbf{X}, \lambda)$ .  $\square$

#### 7.5. Proof of Theorem 4.4.

*Proof.* We prove Theorem 4.4 for the case where  $\Omega$  is unbounded as the arguments also work for the bounded case (but without Assumption 2.4 and 2.5).

Let  $(\mathbf{y}^k, \mathbf{u}^k, \mathbf{v}^k)$  be a maximizing sequence of (4.6). For every  $i = 1, \dots, p$  and  $j \in \mathbb{N}$ , let  $\beta(i, j) \in \mathbb{N}^p$  be such that  $\beta(i, j)_k = 2j\delta_{k=i}$ . Observe that for every feasible solution  $(\mathbf{y}, \mathbf{u}, \mathbf{v})$  of (4.6), and from a consequence of Assumption 2.5 (see the proof of Lemma 7.1):

$$(7.3) \quad \begin{aligned} L_{\mathbf{y}}(\omega_i^{2j}) + L_{\mathbf{u}}(\omega_i^{2j}) &= L_{\mathbf{v}}(p_{\beta(i,j)}(\mathbf{a})) \\ &= \int_{\mathbf{A}} p_{\beta(i,j)}(\mathbf{a}) d\lambda_{\mathbf{A}}(\mathbf{a}) \leq c^{-2j} \gamma(2j)! \end{aligned}$$

for all  $j = 1, \dots, d$  and all  $i = 1, \dots, p$ . Therefore  $L_{\mathbf{y}}(\omega_i^{2j}) \leq c^{-2j} \gamma(2j)!$  and  $L_{\mathbf{u}}(\omega_i^{2j}) \leq c^{-2j} \gamma(2j)!$  for all  $j = 1, \dots, d$  and all  $i = 1, \dots, n$ . As  $\mathbf{A}$  is compact and  $q_1(\mathbf{a}) = M - \|\mathbf{a}\|^2$ , it follows that  $L_{\mathbf{y}}(\omega_i^{2j}) \leq M^j$ ,  $j = 1, \dots, d$ ,  $i = 1, \dots, L$ . By the same argument using now  $g_1(\mathbf{x}) = M - \|\mathbf{x}\|^2$ ,  $L_{\mathbf{y}}(x_i^{2j}) \leq M^j$ ,  $L_{\mathbf{u}}(x_i^{2j}) \leq M^j$ , and  $L_{\mathbf{v}}(x_i^{2j}) \leq M^j$ ,  $j = 1, \dots, d$ ,  $i = 1, \dots, L$ .

Next, as  $y_{00} \leq 1$ ,  $\mathbf{u}_{00} \leq 1$  and  $v_{00} = 1$ , and  $\mathbf{M}_d(\mathbf{y}), \mathbf{M}_d(\mathbf{u}), \mathbf{M}_d(\mathbf{v}) \succeq 0$ , by invoking [32, ], we obtain

$$|y_{\alpha\beta}| \leq \max[1, \max_i [L_{\mathbf{y}}(x_i^{2d}), L_{\mathbf{y}}(\omega_i^{2d})]] = \tau_1,$$

$$|u_{\alpha\beta}| \leq \max[1, \max_i [L_{\mathbf{y}}(x_i^{2d}), L_{\mathbf{u}}(\omega_i^{2d})]] = \tau_2; \quad |v_{\alpha\beta}| \leq \max[1, M^d] = \tau_3,$$

for all  $(\alpha, \beta)$ . This implies that the feasible set of (4.6) is compact and so (4.6) has an optimal solution  $(\mathbf{y}^d, \mathbf{u}^d, \mathbf{v}^d)$  for every  $d \geq d_{\min}$ .

For every  $d$  let  $\tau_d := \max[\tau_1, \tau_2, \tau_3]$ . Next, by completing with zeros, consider the finite vectors  $\mathbf{y}^d$ ,  $\mathbf{u}^d$  and  $\mathbf{v}^d$  has infinite sequences. As  $|y_{\alpha\beta}| \leq \tau_d$ ,  $|u_{\alpha\beta}| \leq \tau_d$  and  $|v_{\alpha\eta}| \leq \tau_d$ , whenever  $|\alpha + \beta| \leq 2d$  and  $|\alpha + \eta| \leq 2d$ , by a standard argument <sup>6</sup>

<sup>6</sup>Let  $(u^n)_{n \in \mathbb{N}}$  be a sequence of infinite sequences such that  $\sup_n |u_i^n| < \tau_i$  for all  $i = 1, \dots$ , and let  $\hat{u}_i^n := u_i^n / \tau_i$ , for all  $i, n$ . Then  $(\hat{u}^n) \subset \mathbf{B}_1$  where  $\mathbf{B}_1$  is the unit ball of  $\ell_\infty$ . By weak-star sequential compactness of  $\mathbf{B}_1$ , there is a subsequence  $(n_k)$  and  $\hat{u} \in \mathbf{B}_1$  such that  $\hat{u}^{n_k} \rightarrow \hat{u}$  for the  $\sigma(\ell_\infty, \ell_1)$  weak-star topology of  $\ell_\infty$ . In particular,  $\hat{u}_i^{n_k} \rightarrow \hat{u}_i$ , as  $k \rightarrow \infty$ , for all  $i = 1, \dots$ , which implies  $u_i^{n_k} \rightarrow \tau_i \hat{u}_i$ , as  $k \rightarrow \infty$ , for all  $i = 1, \dots$



there exists a subsequence  $(d_k)_{k \in \mathbb{N}}$  and infinite sequences  $\mathbf{y}^* = (y_{\alpha\beta}^*)$ ,  $\mathbf{u}^* = (u_{\alpha\beta}^*)$ , and  $\mathbf{v}^* = (v_{\alpha\eta})$ ,  $\alpha \in \mathbb{N}^n$ ,  $\beta \in \mathbb{N}^p$ , and  $\eta \in \mathbb{N}^L$ , such that

$$(7.4) \quad \lim_{k \rightarrow \infty} y_{\alpha\beta}^{d_k} = y_{\alpha\beta}^*, \quad \lim_{k \rightarrow \infty} u_{\alpha\beta}^{d_k} = u_{\alpha\beta}^*, \quad \alpha \in \mathbb{N}^n, \beta \in \mathbb{N}^p,$$

$$(7.5) \quad \lim_{k \rightarrow \infty} v_{\alpha\eta}^{d_k} = v_{\alpha\eta}^*, \quad \alpha \in \mathbb{N}^n, \eta \in \mathbb{N}^L.$$

Next, fix  $r \in \mathbb{N}$  arbitrary. By (7.4) and (7.5),  $\mathbf{M}_r(\mathbf{y}^*) \succeq 0$ ,  $\mathbf{M}_r(\mathbf{u}^*) \succeq 0$  and  $\mathbf{M}_r(\mathbf{v}^*) \succeq 0$ . In addition:

$$L_{\mathbf{y}^*}(\omega_i^{2j})^{-1/2j} \geq c\gamma^{-1/2j} (2j!)^{-1/2j}, \quad \forall i, j,$$

and similarly

$$L_{\mathbf{u}^*}(\omega_i^{2j})^{-1/2j} \geq c\gamma^{-1/2j} (2j!)^{-1/2j}, \quad \forall i, j.$$

Also

$$L_{\mathbf{y}^*}(x_i^{2j})^{-1/2j}, L_{\mathbf{u}^*}(x_i^{2j})^{-1/2j}, L_{\mathbf{v}^*}(x_i^{2j})^{-1/2j} \geq M^{-1/2} \quad \forall i, j.$$

Therefore as  $\sum_{j=1}^{\infty} (2j!)^{-1/2j} = +\infty$ , one obtains

$$(7.6) \quad \sum_{j=1}^{\infty} L_{\mathbf{y}^*}(\omega_i^{2j})^{-1/2j} = +\infty, \quad \text{and} \quad \sum_{j=1}^{\infty} L_{\mathbf{y}^*}(x_i^{2j})^{-1/2j} = +\infty, \quad \forall i, j,$$

and similarly for  $\mathbf{u}^*$ ,  $\mathbf{v}^*$ . In summary, the three sequences  $\mathbf{y}^*$ ,  $\mathbf{u}^*$  and  $\mathbf{v}^*$  satisfy the multivariate Carleman's condition (2.1). As  $\mathbf{M}_d(\mathbf{y}^*)$ ,  $\mathbf{M}_d(\mathbf{u}^*)$ ,  $\mathbf{M}_d(\mathbf{v}^*) \succeq 0$ , they have a representing measure  $\phi^*$ ,  $\varphi^*$  and  $\psi^*$  respectively on  $\mathbb{R}^n \times \mathbb{R}^p$ ,  $\mathbb{R}^n \times \mathbb{R}^p$  and  $\mathbb{R}^n \times \mathbb{R}^t$ .

Next, as (4.5) holds, the quadratic module of  $\mathbb{R}[\mathbf{x}, \mathbf{a}]$  generated by the polynomials  $(g_j, q_\ell)$  is Archimedean. Therefore, as  $\mathbf{M}_d(g_j \mathbf{v}^*) \succeq 0$  and  $\mathbf{M}_d(q_\ell \mathbf{v}^*) \succeq 0$  for all  $d$  and all  $j = 1, \dots, m$ ,  $\ell = 1, \dots, L$ , by Putinar's Theorem [40], the measure  $\psi^*$  is supported on  $\mathbf{X} \times \mathbf{A}$ . Also, as (4.5) holds, the quadratic module of  $\mathbb{R}[\mathbf{x}]$  generated by the polynomials  $(g_j)$  is Archimedean. Hence the marginal  $\phi_{\mathbf{x}}^*$  of  $\phi$  is supported on  $\mathbf{X}$ . If  $\Omega$  is compact (and as then  $s_1(\omega) = M - \|\omega\|^2$ ) a similar argument shows that  $\phi^*$  is supported on  $\mathbf{K}$ .

If  $\Omega$  is not compact, then by (7.3) and (7.4) we have  $L_{\mathbf{y}^*}(\omega_i^{2j}) \leq c^{-2j} \gamma (2j!)$  for all  $i, j$ . As  $\mathbf{M}_d(s_\ell \mathbf{y}^*) \succeq 0$  for all  $\ell$ , and  $\mathbf{M}_d(-f \mathbf{y}^*) \succeq 0$  for all  $d$ , then by [28, Theorem 2.2, p. 2494]<sup>7</sup>,  $f(\mathbf{x}, \omega) \leq 0$  and  $s_\ell(\omega) \geq 0$ ,  $\ell = 1, \dots, \bar{s}$ , on the support of  $\phi^*$ . That is,  $\text{support}(\phi^*) \subset \mathbf{K}$ .

Hence  $(\phi^*, \psi^*)$  is a feasible solution of (3.6). But as  $\phi^*(\mathbf{K}) = \lim_{d \rightarrow \infty} \rho_d \geq \rho$ , we conclude that  $(\phi^*, \psi^*)$  is an optimal solution with value  $\rho$ .  $\square$

## 7.6. Verifying Assumption 2.5 and Assumption 2.5.

7.6.1. **A is a finite set.** In this case  $\Omega = \mathbb{R}^p$  and  $\mathbf{A} = \{1, \dots, \kappa\}$ . Then Assumption 2.4 holds and Assumption 2.5 holds whenever it holds for each individual  $\mu_i$ ,  $i = 1, \dots, \kappa$ . The set  $\mathcal{M}_{\mathbf{a}}$  can be identified with the simplex  $\Delta = \{\lambda \in \mathbb{R}^\kappa : \sum_i \lambda = 1, \lambda \geq 0\}$ . In Theorem 3.2, the conditional probability  $\hat{\phi}^*(d\mathbf{a}|\mathbf{x})$ ,  $\mathbf{x} \in \mathbf{S}$ , of the optimal solution  $\phi^*$ , identifies the worst-case distribution  $\mu_{\mathbf{a}(\mathbf{x})} \in \mathbf{A}$  for every  $\mathbf{x} \in \mathbf{S}$ .

<sup>7</sup>In the proof of Theorem 2.2 in [28],  $c = 1$ , but the proof can be extended easily to arbitrary  $c > 0$

7.6.2. *Mixture of Multivariate Gaussian distributions.* In the general case  $\Omega = \mathbb{R}^p$ ,  $\mathbf{a} = (\theta, \Sigma)$  with  $\underline{\theta} \leq \theta \leq \bar{\theta}$ , where  $\underline{\theta}, \bar{\theta} \in \mathbb{R}^p$ , and  $\Sigma = \Sigma^T = (\sigma_{ij}) \in \mathbb{R}^{p \times p}$ , with  $\underline{\delta} \mathbf{I} \preceq \Sigma \preceq \bar{\delta} \mathbf{I}$  and  $\underline{\delta} > 0$ . That is,

$$d\mu_{\mathbf{a}}(\omega) = \frac{1}{\sqrt{(2\pi)^p \det(\Sigma)}} \exp\left(-\frac{1}{2}(\omega - \theta)\Sigma^{-1}(\omega - \theta)\right) d\omega,$$

The measurability condition in Assumption 2.4(i) follows from Fubini-Tonelli's theorem. Assumption 2.4(ii) is also satisfied. For instance, the fourth-order central moments read (see e.g. [https://en.wikipedia.org/wiki/Multivariate\\_normal\\_distribution#Higher\\_moments](https://en.wikipedia.org/wiki/Multivariate_normal_distribution#Higher_moments)):

$$\int (\omega_i - \theta_i)^4 d\mu_{\mathbf{a}}(\omega) = 3\sigma_{ii}^2; \quad \int (\omega_i - \theta_i)^3 (\omega_j - \theta_j) d\mu_{\mathbf{a}}(\omega) = 3\sigma_{ii}\sigma_{ij}$$

$$\int (\omega_i - \theta_i)^2 (\omega_j - \theta_j)^2 d\mu_{\mathbf{a}}(\omega) = \sigma_{ii}\sigma_{jj} + 2\sigma_{ij}^2,$$

$$\int (\omega_i - \theta_i)^2 (\omega_j - \theta_j) (\omega_k - \theta_k) d\mu_{\mathbf{a}}(\omega) = \sigma_{ii}\sigma_{jk} + 2\sigma_{ij}\sigma_{ik},$$

$$\int (\omega_i - \theta_i) (\omega_j - \theta_j) (\omega_k - \theta_k) (\omega_\ell - \theta_\ell) d\mu_{\mathbf{a}}(\omega) = \sigma_{ij}\sigma_{k\ell} + \sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk},$$

and higher-order central moments are homogeneous polynomials in the entries of  $\Sigma$ . This immediately implies that non-central moments are polynomials in  $(\sigma_{ij})$  and  $\theta$ . Assumption 2.4(iii) is also straightforward as  $\mu_{\mathbf{a}}$  has a density w.r.t.  $d\omega$ , everywhere positive. Concerning Assumption 2.4(iv), let  $h$  be bounded continuous on  $\mathbf{X} \times \Omega$ . With the change of variable  $\mathbf{u} = \Sigma^{-1/2}(\omega - \theta)$  one has

$$H(\mathbf{x}, \mathbf{a}) := \int_{\Omega} h(\mathbf{x}, \omega) d\mu_{\mathbf{a}}(\omega) = \frac{1}{\sqrt{(2\pi)^p}} \int_{\Omega} h(\mathbf{x}, \Sigma^{1/2} \mathbf{u} + \theta) \exp\left(-\frac{1}{2}\|\mathbf{u}\|^2\right) d\mathbf{u},$$

and since  $h$  is bounded and continuous, it follows that  $H$  is continuous in  $(\mathbf{x}, \mathbf{a}) \in \mathbf{X} \times \mathbf{A}$ . Finally, Assumption 2.5 also holds.

7.6.3. *Mixture of exponential distributions.* In this case  $\Omega = \mathbb{R}_+^p$ ,  $\mathbf{a} = (a_1, \dots, a_p)$  with  $\mathbf{a} \in \mathbf{A} := [\underline{\mathbf{a}}, \bar{\mathbf{a}}]$ ,  $\underline{\mathbf{a}} > 0$ , and

$$d\mu_{\mathbf{a}}(\omega) = \left( \prod_{i=1}^p \frac{1}{a_i} \right) \exp\left(-\sum_{i=1}^p \frac{\omega_i}{a_i}\right) d\omega = \otimes_{i=1}^p d\mu_{a_i}(\omega_i)$$

with  $d\mu_{a_i}(\omega_i) = \frac{1}{a_i} \exp(-\omega_i/a_i) d\omega_i$ ,  $i = 1, \dots, p$ .

Again, the measurability condition in Assumption 2.4(i) follows from Fubini-Tonelli's theorem. Then for Assumption 2.4(ii),

$$\int_{\Omega} \omega^\beta d\mu_{\mathbf{a}}(\omega) = \prod_{i=1}^p \left( \frac{1}{a_i} \int_{\mathbb{R}_+} \omega_i^{\beta_i} \exp(\omega_i/a_i) d\omega_i \right) = \mathbf{a}^\beta \prod_{i=1}^p \beta_i! \in \mathbb{R}[\mathbf{a}],$$

and Assumption 2.4(iii) also holds. Like for the Gaussian, and after the change of variable  $u_i = \omega_i/a_i$ ,  $i = 1, \dots, p$ , one shows easily that Assumption 2.4(iv) holds. Finally for Assumption 2.5,

$$\int_{\Omega} \exp(c|\omega|) d\mu_{\mathbf{a}}(\omega) = \left( \prod_{i=1}^p \frac{1}{a_i} \right) \exp\left(-\sum_{i=1}^p \omega_i \left(\frac{1}{a_i} - c\right)\right) d\omega < \infty,$$

whenever  $c < 1/a_i$  for all  $i = 1, \dots, p$ .

7.6.4. *Mixture of elliptical's.* Assumption 2.4(ii) holds for Example 1.3. For instance with  $\theta : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that  $\int_{\mathbb{R}} t^k \theta(t^2) dt < \infty$  for all  $k$ , and  $s = \int_{\mathbb{R}} \theta(t^2) dt$ , let

$$d\mu_{\mathbf{a}}(\omega) = \frac{1}{s\sigma} \theta((\omega - a)^2 / \sigma^2) d\omega, \quad \mathbf{a} = (a, \sigma) \in \mathbf{A}.$$

Then

$$\int \omega^j d\mu_{\mathbf{a}}(\omega) = \frac{1}{s} \int_{\mathbb{R}} (\sigma t + a)^j \theta(t^2) dt = p_j(a, \sigma), \quad j = 0, 1, \dots,$$

7.6.5. *Mixture of Poisson's.* Assumption 2.4(ii) holds for Example 1.4. For instance

$$p_1(\mathbf{a}) = \int_{\Omega} \omega d\mu_{\mathbf{a}}(\omega) = a; \quad p_2(\mathbf{a}) = \int_{\Omega} \omega^2 d\mu_{\mathbf{a}}(\omega) = a^2 + a.$$

7.6.6. *Mixture of Binomial's.* Assumption 2.4(ii) holds for Example 1.5. For instance

$$p_1(\mathbf{a}) = \int_{\Omega} \omega d\mu_{\mathbf{a}}(\omega) = N a; \quad p_2(\mathbf{a}) = \int_{\Omega} \omega^2 d\mu_{\mathbf{a}}(\omega) = N a (1 - a).$$

## REFERENCES

- [1] W. van Ackooij. Convexity Statements for Linear Probability Constraints with Gaussian Technology Matrices and Copulae Correlated Rows
- [2] W. van Ackooij, J. Malick. Eventual convexity of probability constraints with elliptical distributions, *Math. Program.*, pp. 1–27, 2018.
- [3] R. Ash. *Real Analysis and Probability*, Academic Press Inc., Boston, USA, 1972
- [4] P. Billingsley. *Convergence of Probability Measures*, John Wiley & Sons, New York, 1968
- [5] G. Calafiore, F. Dabbene, Probabilistic and Randomized Methods for Design under Uncertainty, G. Calafiore and F. Dabbene (Eds.), Springer, 2006.
- [6] G.C. Calafiore, L. El Ghaoui. On Distributionally Robust Chance-Constrained Linear Programs, *J. Optim. Theory Appl.* 130, pp. 1–22, 2006.
- [7] Chao Duan, Wanliang Fang, Lin Jiang, Li Yao, Jun Liu. Distributionally Robust Chance-Constrained Voltage-Concerned DC-OPF with Wasserstein Metric.
- [8] A. Charnes, W.W. Cooper. Chance constrained programming. *Manag. Sci.* 6, pp. 73–79, 1959
- [9] W. Chen, M. Sim, J. Sun, C.P. Teo. From CVaR to uncertainty set: Implications in joint chance-constrained optimization. *Operations research*, 58(2), 470–485, 2010
- [10] E. de Klerk, D. Kuhn, K. Postek. Distributionally robust optimization with polynomial densities: theory, models and algorithms, [arXiv:1805.03588](https://arxiv.org/abs/1805.03588)
- [11] E. Delage, Y. Ye. Distributionally robust optimization under moment uncertainty with applications to data-driven problems. *Operations Research* 58, pp. 595–612, 2010
- [12] M. Di Zioa, U. Guarneraa, R. Roccib. A mixture of mixture models for a classification problem: The unity measure error, *Computational Statistics and Data Analysis* 51, pp. 2573–2585, 2007
- [13] J.L. Doob. *Measure Theory*, Springer-Verlag, 1994, New York.
- [14] E. Erdogan, G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Math. Program. Sér. B* 107, pp. 37–61, 2006
- [15] L. El Ghaoui, M. Oks, F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach, *Oper. Res.* 51, pp. 543–556, 2003
- [16] G.A. Hanasusanto, V. Roitch, D. Kuhn, W. Wiesemann. Ambiguous joint chance constraints under mean and dispersion information. *Oper. Res.* 65, pp. 751–767, 2017
- [17] G.A. Hanasusanto, V. Roitch, D. Kuhn, W. Wiesemann. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Math. Program.* 151, pp. 35–62, 2015.
- [18] R. Henrion. Structural Properties of Linear Probabilistic Constraints, *Optimization* 56, pp. 425–440, 2007.
- [19] R. Henrion, C. Strugarek. Convexity of chance constraints with independent random variables. *Computational Optimization and Applications* 41, pp. 263–276, 2008.

- [20] D. Henrion, J.B. Lasserre, J. Lofberg. Gloptipoly 3: moments, optimization and semidefinite programming, *Optim. Methods and Softwares* 24, pp. 761–779, 2009.
- [21] D. Henrion, J. B. Lasserre, C. Savorgnan, Approximate volume and integration for basic semi-algebraic sets, *SIAM Review* 51, pp. 722–743, 2009.
- [22] D. Henrion, M. Korda, Convex computation of the region of attraction of polynomial control systems. *IEEE Trans. Aut. Control* 59, pp. 297–312, 2014.
- [23] A. M. Jasour, N.S. Aybat, C. M. Lagoa, Semidefinite programming for chance constrained optimization over semi-algebraic sets, *SIAM J. Optim.* 25, No 3, pp. 1411–1440, 2015
- [24] A. Jasour, C. Lagoa, Convex constrained semialgebraic volume optimization: Application in systems and control, 2017, [arXiv:1701.08910](https://arxiv.org/abs/1701.08910)
- [25] R. Jiang, Y. Guan. Data-driven chance constrained stochastic program. *Math. Program.* 158, 291–327, 2016
- [26] M. Korda, D. Henrion, C. N. Jones, Convex computation of the maximum controlled invariant set for polynomial control systems, *SIAM J. Control Optim.* 52(5):2944–2969, 2014.
- [27] J.B. Lasserre, Representation of chance-constraints with strong asymptotic properties, *IEEE Control Systems Letters* 1, pp. 50–55, 2017.
- [28] J.B. Lasserre, The K-moment problem for continuous linear functionals, *Trans. Amer. Math. Soc.* 365, pp. 2489–2504, 2013.
- [29] J.B. Lasserre, A “Joint+Marginal” approach to parametric polynomial optimization, *SIAM J. Optim.* 20, pp. 1995–2022, 2010.
- [30] J. B. Lasserre, Moments, positive polynomials and their applications. Imperial College Press, London, 2010.
- [31] J.B. Lasserre, Lebesgue decomposition in action via semidefinite relaxations, *Adv. Comput. Math.* 42, pp. 1129–1148, 2016.
- [32] J.B. Lasserre, T. Netzer. SOS approximations of nonnegative polynomials via simple high degree perturbations”. *Math. Zeitschrift* 256, pp. 99–112, 2006.
- [33] J.B. Lasserre, Computing gaussian and exponential measures of semi-algebraic sets, *Adv. Appl. Math.* 91, pp. 137–163, 2017.
- [34] P. Li, M. Wendt, G. Wozny, A Probabilistically Constrained Model Predictive Controller, *Automatica* 38, pp. 1171–1176, 2002.
- [35] J.S. Marron, M.P. Wand. Exact mean integrated squared error, *The Ann. Statist.* 20, pp. 712–736, 1992
- [36] B. Miller, H. Wagner. Chance-constrained programming with joint constraints. *Oper. Res.* 13, pp. 930–945, 1965 34.
- [37] Mosek Aps, Mosek Matlab Toolbox, 2017.
- [38] A. Nemirovski, A. Shapiro. Convex approximations of chance constrained programs. *SIAM J. Optim.* 17, pp. 969–996, 2006
- [39] A. Prékopa, Probabilistic Programming, in *Stochastic Programming*, A. Ruszczyński and A. Shapiro (Eds.), *Handbooks in Operations Research and Management Science* Volume 10, pp. 267–351, 2003.
- [40] M. Putinar, Positive polynomials on compact semi-algebraic sets, *Indiana Univ. Math. J.* 42, pp. 969–984, 1993.
- [41] A. Shapiro, D. Dentcheva, A. Ruszczyński, *Lecture on Stochastic Programming: Modeling and Theory*, 2nd Ed., SIAM, Philadelphia, 2014.
- [42] S. Wang, J. Li, C. Peng. Distributionally robust chance-constrained program surgery planning with downstream resource, *Proceedings 2017 International Conference on Service Systems and Service Management*, Dalian, China, June 2017.
- [43] H. Waki, S. Kim, M. Kojima, M. Muramatsu, Sums of squares and semidefinite programming relaxations for polynomial optimization problems with structured sparsity, *SIAM J. Optim.* 17, pp. 218–242, 2006.
- [44] S.I. Wang, A.T. Chaganti, P. Liang. Estimating Mixture Models via Mixtures of Polynomials *Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS’15)*, Montreal, December 2015, pp. 487–495
- [45] X. Tong, H. Sun, X. Luo, Q. Zheng, Distributionally robust chance constrained optimization for economic dispatch in renewable energy integrated systems, *J. Global Optim.* 70, pp. 131–158, 2018.
- [46] W. Xie, S. Ahmed. Distributionally robust chance constrained optimal power flow with renewables: A conic reformulation, *IEEE Trans. Power Systems*, 2017.

- [47] W. Xie, S. Ahmed. On deterministic reformulations of distributionally robust joint chance constrained optimization problems. *SIAM J. Optim.* 28, pp. 1151-1182, 2018.
- [48] D. Xu. The Applications of Mixtures of Normal Distributions in Empirical Finance: A Selected Survey, Working Papers 0904, University of Waterloo, Department of Economics, revised September 2009.
- [49] Wenzhuo Yang, Huan Xu. Distributionally Robust Chance Constraints for Non-Linear Uncertainties, *Math. Program.* 155, pp. 231–265, 2016.
- [50] Zhang, Y., Shen, S., J. Mathieu. Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads, *IEEE Trans. Power Systems* 32, pp. 1378–1388, 2016.
- [51] S. Zymler, D. Kuhn, B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Math. Program.* 137, pp. 167-198, 2013

LAAS-CNRS AND INSTITUTE OF MATHEMATICS, UNIVERSITY OF TOULOUSE, LAAS, 7 AVENUE DU COLONEL ROCHE, 31077 TOULOUSE CÉDEX 4, FRANCE, TEL: +33561336415

*E-mail address:* [lasserre@laas.fr](mailto:lasserre@laas.fr)

LAAS-CNRS, UNIVERSITY OF TOULOUSE, LAAS, 7 AVENUE DU COLONEL ROCHE, 31077 TOULOUSE CÉDEX 4, FRANCE, TEL: +33561336441

*E-mail address:* [tweisser@laas.fr](mailto:tweisser@laas.fr)