

# A "JOINT+MARGINAL" APPROACH TO PARAMETRIC POLYNOMIAL OPTIMIZATION

JEAN B. LASSERRE

ABSTRACT. Given a compact parameter set  $\mathbf{Y} \subset \mathbb{R}^p$ , we consider polynomial optimization problems ( $\mathbf{P}_{\mathbf{y}}$ ) on  $\mathbb{R}^n$  whose description depends on the parameter  $\mathbf{y} \in \mathbf{Y}$ . We assume that one can compute all moments of some probability measure  $\varphi$  on  $\mathbf{Y}$ , absolutely continuous with respect to the Lebesgue measure (e.g.  $\mathbf{Y}$  is a box or a simplex and  $\varphi$  is uniformly distributed). We then provide a hierarchy of semidefinite relaxations whose associated sequence of optimal solutions converges to the moment vector of a probability measure that encodes all information about all global optimal solutions  $\mathbf{x}^*(\mathbf{y})$  of  $\mathbf{P}_{\mathbf{y}}$ , as  $\mathbf{y} \in \mathbf{Y}$ . In particular, one may approximate as closely as desired any polynomial functional of the optimal solutions, like e.g. their  $\varphi$ -mean. In addition, using this knowledge on moments, the measurable function  $\mathbf{y} \mapsto x_k^*(\mathbf{y})$  of the  $k$ -th coordinate of optimal solutions, can be estimated, e.g. by maximum entropy methods. Also, for a boolean variable  $x_k$ , one may approximate as closely as desired its persistency  $\varphi(\{\mathbf{y} : x_k^*(\mathbf{y}) = 1\})$ , i.e. the probability that in an optimal solution  $\mathbf{x}^*(\mathbf{y})$ , the coordinate  $x_k^*(\mathbf{y})$  takes the value 1. At last but not least, from an optimal solution of the dual semidefinite relaxations, one provides a sequence of polynomial (resp. piecewise polynomial) lower approximations with  $L_1(\varphi)$  (resp. almost uniform) convergence to the optimal value function.

## 1. INTRODUCTION

Roughly speaking, given a set parameters  $\mathbf{Y}$  and an optimization problem whose description depends on  $\mathbf{y} \in \mathbf{Y}$  (call it  $\mathbf{P}_{\mathbf{y}}$ ), *parametric optimization* is concerned with the behavior and properties of the optimal value as well as primal (and possibly dual) optimal solutions of  $\mathbf{P}_{\mathbf{y}}$ , when  $\mathbf{y}$  varies in  $\mathbf{Y}$ . This is a quite challenging problem and in general one may obtain information locally around some nominal value  $\mathbf{y}_0$  of the parameter. There is a vast and rich literature on the topic and for a detailed treatment, the interested reader is referred to e.g. Bonnans and Shapiro [4] and the many references therein. Sometimes, in the context of optimization with data uncertainty, some probability distribution  $\varphi$  on the parameter set  $\mathbf{Y}$  is available and in this context one is also interested in e.g. the distribution of the optimal value, optimal solutions, all viewed as random variables. In particular, for discrete optimization problems where cost coefficients are random variables with joint distribution  $\varphi$ , some bounds on the expected optimal value have been obtained. More recently Natarajan et al. [17] extended the earlier work in [3] to even

---

1991 *Mathematics Subject Classification.* 65 D15, 65 K05, 46 N10, 90 C22.

*Key words and phrases.* Parametric and polynomial optimization; semidefinite relaxations.

provide a convex optimization problem for computing the so-called *persistence* values<sup>1</sup> of (discrete) variables, for a particular distribution  $\varphi^*$  in a certain set  $\Theta$  of distributions. However, this convex formulation requires knowledge of the convex hull of a discrete set and approximations are needed. The approach is nicely illustrated on a discrete choice problem and a stochastic knapsack problem. For more details on persistence in discrete optimization, the interested reader is referred to [17] and the references therein.

In the context of polynomial equations whose coefficients are themselves polynomials of some parameter  $\mathbf{y} \in \mathbf{Y}$ , some specific "parametric" methods exist. For instance, one may compute symbolically once and for all, what is called a *comprehensive* Gröbner basis, i.e., a fixed basis that is a Gröbner basis for all  $\mathbf{y} \in \mathbf{Y}$ ; see Weispfenning [25] and more recently Rostalski [19] for more details. Then when needed, one may compute the solutions for a specific value of the parameter  $\mathbf{y}$ , e.g. by the eigenvalue method of Möller and Stetter [16, 22]. However, one still needs to apply the latter method for each value of the parameter  $\mathbf{y}$ . A similar two-step approach is also proposed for homotopy (instead of Gröbner bases) methods in [19].

The purpose of this paper is to show that in one restricts to the case of *polynomial* parametric optimization then all information about the optimal value and optimal solutions can be obtained, or at least, approximated as closely as desired.

**Contribution.** We here restrict our attention to parametric polynomial optimization, that is, when  $\mathbf{P}_{\mathbf{y}}$  is described by *polynomial* equality and inequality constraints on both the parameter vector  $\mathbf{y}$  and the optimization variables  $\mathbf{x}$ . Moreover, the set  $\mathbf{Y}$  is restricted to be a compact basic semi-algebraic set of  $\mathbb{R}^p$ , and preferably a set sufficiently simple so that one may obtain the moments of some probability measure on  $\mathbf{Y}$ , absolutely continuous with respect to the Lebesgue measure. For instance if  $\mathbf{Y}$  is a simple set (like a simplex, a box) one may choose  $\varphi$  to be the probability measure uniformly distributed on  $\mathbf{Y}$ ; typical  $\mathbf{Y}$  candidates are polyhedra. Or sometimes, in the context of optimization with data uncertainty,  $\varphi$  is already specified. We also suppose that  $\mathbf{P}_{\mathbf{y}}$  has a unique optimal solution for almost all values of the parameter  $\mathbf{y} \in \mathbf{Y}$ . In this specific context we are going to show that one may get insightful information on the set of all global optimal solutions of  $\mathbf{P}_{\mathbf{y}}$ , via what we call a "*Joint+marginal*" approach. Our contribution is as follows:

(a) Call  $J(\mathbf{y})$  (resp.  $\mathbf{X}_{\mathbf{y}}^* \in \mathbb{R}^n$ ) the optimal value (resp. the set of optimal solutions) of  $\mathbf{P}_{\mathbf{y}}$  for the value  $\mathbf{y} \in \mathbf{Y}$  of the parameter. We first define an infinite-dimensional optimization problem  $\mathbf{P}$  whose optimal value is exactly  $\rho = \int_{\mathbf{Y}} J(\mathbf{y}) d\varphi(\mathbf{y})$ . Any optimal solution of  $\mathbf{P}_{\mathbf{y}}$  is a probability measure  $\mu^*$  on  $\mathbb{R}^n \times \mathbb{R}^p$  with marginal  $\varphi$  on  $\mathbb{R}^p$ . It turns out that  $\mu^*$  encodes all information on the optimal solutions  $\mathbf{X}_{\mathbf{y}}^*$ ,  $\mathbf{y} \in \mathbf{Y}$ . Whence the name "*Joint+marginal*" as  $\mu^*$  is a *joint* distribution of  $\mathbf{x}$  and  $\mathbf{y}$ , and  $\varphi$  is the *marginal* of  $\mu^*$  on  $\mathbb{R}^p$ .

(b) Next, we provide a hierarchy of semidefinite relaxations of  $\mathbf{P}$  with associated sequence of optimal values  $(\rho_i)_i$ , in the spirit of the hierarchy defined in [13]. An optimal solution of the  $i$ -th semidefinite relaxation is a sequence  $\mathbf{z}^i = (z_{\alpha\beta}^i)$  indexed in the monomial basis  $(\mathbf{x}^\alpha \mathbf{y}^\beta)$  of the subspace  $\mathbb{R}[\mathbf{x}, \mathbf{y}]_i$  of polynomials of degree

<sup>1</sup>Given a 0 – 1 optimization problem  $\max\{\mathbf{c}'\mathbf{x} : \mathbf{x} \in \mathcal{X} \cap \{0, 1\}^n\}$  and a distribution  $\varphi$  on  $\mathbf{c}$ , the persistence value of the variable  $x_i$  is  $\text{Prob}_\varphi(x_i^* = 1)$  at an optimal solution  $\mathbf{x}^*(\mathbf{c}) = (x_i^*)$ .

at most  $2i$ . If for almost all  $\mathbf{y} \in \mathbf{Y}$ ,  $\mathbf{P}_{\mathbf{y}}$  has a unique *global* optimal solution  $\mathbf{x}^*(\mathbf{y}) \in \mathbb{R}^n$ , then as  $i \rightarrow \infty$ ,  $\mathbf{z}^i$  converges pointwise to the sequence of moments of  $\mu^*$  defined in (a). In particular, one obtains the distribution of the optimal solution  $\mathbf{x}^*(\mathbf{y})$ , and therefore, one may approximate as closely as desired any polynomial functional of the solution  $\mathbf{x}^*(\mathbf{y})$ , like e.g. the  $\varphi$ -mean or variance of  $\mathbf{x}^*(\mathbf{y})$ .

In addition, if the optimization variable  $x_k$  is boolean then one may approximate as closely as desired its *persistence*  $\varphi(\{\mathbf{y} : x_k^*(\mathbf{y}) = 1\})$  (i.e., the probability that  $x_k^*(\mathbf{y}) = 1$  in an optimal solution  $\mathbf{x}^*(\mathbf{y})$ ), as well as a necessary and sufficient condition for this persistence to be 1.

(c) Finally, let  $e(k) \in \mathbb{N}^n$  be the vector  $(\delta_{j=k})_j$ . Then as  $i \rightarrow \infty$ , and for every  $\beta \in \mathbb{N}^p$ , the sequence  $(z_{e(k)\beta}^i)$  converges to  $z_{k\beta}^* := \int_{\mathbf{Y}} \mathbf{y}^\beta g_k(\mathbf{y}) d\varphi(\mathbf{y})$  for the measurable function  $\mathbf{y} \mapsto g_k(\mathbf{y}) := x_k^*(\mathbf{y})$ . In other words, the sequence  $(z_{k\beta}^*)_{\beta \in \mathbb{N}^p}$  is the moment sequence of the measure  $d\psi(y) := x_k^*(\mathbf{y}) d\varphi(\mathbf{y})$  on  $\mathbf{Y}$ . And so, the  $k$ -th coordinate function  $\mathbf{y} \mapsto x_k^*(\mathbf{y})$  of optimal solutions of  $\mathbf{P}_{\mathbf{y}}$ ,  $\mathbf{y} \in \mathbf{Y}$ , can be estimated, e.g. by maximum entropy methods. Of course, the latter estimation is not pointwise but it still provides useful information on optimal solutions, e.g. the shape of the function  $\mathbf{y} \mapsto x_k^*(\mathbf{y})$ , especially if the function  $x_k^*(\cdot)$  is continuous, as illustrated on some simple examples. For instance, for parametric polynomial equations, one may use this estimation of  $\mathbf{x}^*(\mathbf{y})$  as an initial point for Newton's method for any given value of the parameter  $\mathbf{y}$ .

Finally, the computational complexity of the above methodology is roughly the same as the moment approach described in [13] for an optimization problem with  $n+p$  variables since we consider the joint distribution of the  $n$  variables  $\mathbf{x}$  and the  $p$  parameters  $\mathbf{y}$ . Hence, the approach is particularly interesting when the number of parameters is small, say 1 or 2. In addition, in the latter case the max-entropy estimation has been shown to be very efficient in several examples in the literature; see e.g. [5, 23, 24]. However, in view of the present status of SDP solvers, if no sparsity or symmetry is taken into account as proposed in e.g. [14], the approach is limited to small to medium size polynomial optimization problems.

But this computational price may not seem that high in view of the ambitious goal of the approach. After all, keep in mind that by applying the moment approach to a single  $(n+p)$ -variables problem, one obtains information on global optimal solutions of an  $n$ -variables problem that depends on  $p$  parameters, that is, one approximates  $n$  functions of  $p$  variables!

## 2. A RELATED LINEAR PROGRAM

Let  $\mathbb{R}[\mathbf{x}, \mathbf{y}]$  denote the ring of polynomials in the variables  $\mathbf{x} = (x_1, \dots, x_n)$ , and the variables  $\mathbf{y} = (y_1, \dots, y_p)$ , whereas  $\mathbb{R}[\mathbf{x}, \mathbf{y}]_k$  denotes its subspace of polynomials of degree at most  $k$ . Let  $\Sigma[\mathbf{x}, \mathbf{y}] \subset \mathbb{R}[\mathbf{x}, \mathbf{y}]$  denote the subset of polynomials that are sums of squares (in short s.o.s.). For a real symmetric matrix  $\mathbf{A}$  the notation  $\mathbf{A} \succeq 0$  stands for  $\mathbf{A}$  is positive semidefinite.

**The parametric optimization problem.** Let  $\mathbf{Y} \subset \mathbb{R}^p$  be a compact set, called the *parameter set*, and let  $f, h_j : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}$ ,  $j = 1, \dots, m$ , be continuous. For each  $\mathbf{y} \in \mathbf{Y}$ , fixed, consider the following optimization problem:

$$(2.1) \quad J(\mathbf{y}) := \inf_{\mathbf{x}} \{ f_{\mathbf{y}}(\mathbf{x}) : h_{\mathbf{y}j}(\mathbf{x}) \geq 0, j = 1, \dots, m \}$$

where the functions  $f_{\mathbf{y}}, h_{\mathbf{y}j} : \mathbb{R}^n \rightarrow \mathbb{R}$  are defined via:

$$\left. \begin{array}{l} \mathbf{x} \mapsto f_{\mathbf{y}}(\mathbf{x}) \quad := \quad f(\mathbf{x}, \mathbf{y}) \\ \mathbf{x} \mapsto h_{\mathbf{y}j}(\mathbf{x}) \quad := \quad h_j(\mathbf{x}, \mathbf{y}), \quad j = 1, \dots, m \end{array} \right\} \quad \forall \mathbf{x} \in \mathbb{R}^n, \forall \mathbf{y} \in \mathbb{R}^p.$$

Next, let  $\mathbf{K} \subset \mathbb{R}^n \times \mathbb{R}^p$  be the set:

$$(2.2) \quad \mathbf{K} := \{(\mathbf{x}, \mathbf{y}) : \mathbf{y} \in \mathbf{Y}; \quad h_j(\mathbf{x}, \mathbf{y}) \geq 0, \quad j = 1, \dots, m\},$$

and for each  $\mathbf{y} \in \mathbf{Y}$ , let

$$(2.3) \quad \mathbf{K}_{\mathbf{y}} := \{\mathbf{x} \in \mathbb{R}^n : h_{\mathbf{y}j}(\mathbf{x}) \geq 0, \quad j = 1, \dots, m\}.$$

The interpretation is as follows:  $\mathbf{Y}$  is a set of parameters and for each instance  $\mathbf{y} \in \mathbf{Y}$  of the parameter, one wishes to compute an optimal *decision* vector  $\mathbf{x}^*(\mathbf{y})$  that solves problem (2.1). Let  $\varphi$  be a Borel probability measure on  $\mathbf{Y}$ , with a positive density with respect to the Lebesgue measure on  $\mathbb{R}^p$ . For instance choose for  $\varphi$  the probability measure

$$\varphi(B) := \left( \int_{\mathbf{Y}} d\mathbf{y} \right)^{-1} \int_{\mathbf{Y} \cap B} d\mathbf{y}, \quad \forall B \in \mathcal{B}(\mathbb{R}^p),$$

uniformly distributed on  $\mathbf{Y}$ . Sometimes, e.g. in the context of optimization with data uncertainty,  $\varphi$  is already specified.

We will use  $\varphi$  (or more precisely, its moments) to get information on the distribution of optimal solutions  $\mathbf{x}^*(\mathbf{y})$  of  $\mathbf{P}_{\mathbf{y}}$ , viewed as random vectors.

In the rest of the paper we assume that for every  $\mathbf{y} \in \mathbf{Y}$ , the set  $\mathbf{K}_{\mathbf{y}}$  in (2.3) is nonempty.

**2.1. A related infinite-dimensional linear program.** Let  $\mathbf{M}(\mathbf{K})$  be the set of finite Borel measures on  $\mathbf{K}$ , and consider the following infinite-dimensional linear program  $\mathbf{P}$ :

$$(2.4) \quad \rho := \inf_{\mu \in \mathbf{M}(\mathbf{K})} \left\{ \int_{\mathbf{K}} f d\mu : \pi\mu = \varphi \right\}$$

where  $\pi\mu$  denotes the marginal of  $\mu$  on  $\mathbb{R}^p$ , that is,  $\pi\mu$  is a probability measure on  $\mathbb{R}^p$  defined by

$$\pi\mu(B) := \mu(\mathbb{R}^n \times B), \quad \forall B \in \mathcal{B}(\mathbb{R}^p).$$

Notice that  $\mu(\mathbf{K}) = 1$  for any feasible solution  $\mu$  of  $\mathbf{P}$ . Indeed, as  $\varphi$  is a probability measure and  $\pi\mu = \varphi$  one has  $1 = \varphi(\mathbf{Y}) = \mu(\mathbb{R}^n \times \mathbb{R}^p) = \mu(\mathbf{K})$ .

Recall that for two Borel spaces  $X, Y$ , the graph  $\text{Gr}\psi \subset X \times Y$  of a set-valued mapping  $\psi : X \rightarrow Y$  is the set

$$\text{Gr}\psi := \{(\mathbf{x}, \mathbf{y}) : \mathbf{x} \in X; \mathbf{y} \in \psi(\mathbf{x})\}.$$

If  $\psi$  is measurable then any measurable function  $h : X \rightarrow Y$  with  $h(\mathbf{x}) \in \psi(\mathbf{x})$  for every  $\mathbf{x} \in X$ , is called a (measurable) *selector*.

**Lemma 2.1.** *Let both  $\mathbf{Y} \subset \mathbb{R}^p$  and  $\mathbf{K}$  in (2.2) be compact. Then the set-valued mapping  $\mathbf{y} \mapsto \mathbf{K}_{\mathbf{y}}$  is Borel-measurable. In addition:*

- (a) *The mapping  $\mathbf{y} \mapsto J(\mathbf{y})$  is measurable.*
- (b) *There exists a measurable selector  $g : \mathbf{Y} \rightarrow \mathbf{K}_{\mathbf{y}}$  such that  $J(\mathbf{y}) = f(g(\mathbf{y}), \mathbf{y})$  for every  $\mathbf{y} \in \mathbf{Y}$ .*

*Proof.* As  $\mathbf{K}$  and  $\mathbf{Y}$  are both compact, the set valued mapping  $\mathbf{y} \mapsto \mathbf{K}_{\mathbf{y}} \subset \mathbb{R}^n$  is compact-valued. Moreover, the graph of  $\mathbf{K}_{\mathbf{y}}$  is by definition the set  $\mathbf{K}$ , which is a Borel subset of  $\mathbb{R}^n \times \mathbb{R}^p$ . Hence, by [11, Proposition D.4],  $\mathbf{K}_{\mathbf{y}}$  is a measurable function from  $\mathbf{Y}$  to the space of nonempty compact subsets of  $\mathbb{R}^n$ , topologized by the Hausdorff metric. Next, since  $\mathbf{x} \mapsto f_{\mathbf{y}}(\mathbf{x})$  is continuous for every  $\mathbf{y} \in \mathbf{Y}$ , (a) and (b) follows from e.g. [11, Proposition D.5].  $\square$

**Theorem 2.2.** *Let both  $\mathbf{Y} \subset \mathbb{R}^p$  and  $\mathbf{K}$  in (2.2) be compact and assume that for every  $\mathbf{y} \in \mathbf{Y}$ , the set  $\mathbf{K}_{\mathbf{y}} \subset \mathbb{R}^n$  in (2.3) is nonempty. Let  $\mathbf{P}$  be the optimization problem (2.4) and let  $\mathbf{X}_{\mathbf{y}}^* := \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}, \mathbf{y}) = J(\mathbf{y})\}$ ,  $\mathbf{y} \in \mathbf{Y}$ . Then:*

(a)  $\rho = \int_{\mathbf{Y}} J(\mathbf{y}) d\varphi(\mathbf{y})$  and  $\mathbf{P}$  has an optimal solution.

(b) For every optimal solution  $\mu^*$  of  $\mathbf{P}$ , and for almost all  $\mathbf{y} \in \mathbf{Y}$ , there is a probability measure  $\psi^*(d\mathbf{x} | \mathbf{y})$  on  $\mathbf{X}_{\mathbf{y}}^*$  such that:

$$(2.5) \quad \mu^*(C \times B) = \int_{B \cap \mathbf{Y}} \psi^*(C \cap \mathbf{X}_{\mathbf{y}}^* | \mathbf{y}) d\varphi(\mathbf{y}), \quad \forall B \in \mathcal{B}(\mathbb{R}^p), C \in \mathcal{B}(\mathbb{R}^n).$$

(c) Assume that for almost  $\mathbf{y} \in \mathbf{Y}$ , the set of minimizers of  $\mathbf{X}_{\mathbf{y}}^*$  is the singleton  $\{\mathbf{x}^*(\mathbf{y})\}$  for some  $\mathbf{x}^*(\mathbf{y}) \in \mathbf{K}_{\mathbf{y}}$ . Then there is a measurable mapping  $g : \mathbf{Y} \rightarrow \mathbf{K}_{\mathbf{y}}$  such that

$$(2.6) \quad g(\mathbf{y}) = \mathbf{x}^*(\mathbf{y}) \text{ for every } \mathbf{y} \in \mathbf{Y}; \quad \rho = \int_{\mathbf{Y}} f(g(\mathbf{y}), \mathbf{y}) d\varphi(\mathbf{y}),$$

and for every  $\alpha \in \mathbb{N}^n$ , and  $\beta \in \mathbb{N}^p$ :

$$(2.7) \quad \int_{\mathbf{K}} \mathbf{x}^\alpha \mathbf{y}^\beta d\mu^*(\mathbf{x}, \mathbf{y}) = \int_{\mathbf{Y}} \mathbf{y}^\beta g(\mathbf{y})^\alpha d\varphi(\mathbf{y}).$$

*Proof.* (a) As  $\mathbf{K}$  is compact then so is  $\mathbf{K}_{\mathbf{y}}$  for every  $\mathbf{y} \in \mathbf{Y}$ . Next, as  $\mathbf{K}_{\mathbf{y}} \neq \emptyset$  for every  $\mathbf{y} \in \mathbf{Y}$  and  $f$  is continuous, the set  $\mathbf{X}_{\mathbf{y}}^* := \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}, \mathbf{y}) = J(\mathbf{y})\}$  is nonempty for every  $\mathbf{y} \in \mathbf{Y}$ . Let  $\mu$  be any feasible solution of  $\mathbf{P}$  and so by definition, its marginal on  $\mathbb{R}^p$  is just  $\varphi$ . Since  $\mathbf{X}_{\mathbf{y}}^* \neq \emptyset$ ,  $\forall \mathbf{y} \in \mathbf{Y}$ , one has  $f_{\mathbf{y}}(\mathbf{x}) \geq J(\mathbf{y})$  for all  $\mathbf{x} \in \mathbf{K}_{\mathbf{y}}$  and all  $\mathbf{y} \in \mathbf{Y}$ . So,  $f(\mathbf{x}, \mathbf{y}) \geq J(\mathbf{y})$  for all  $(\mathbf{x}, \mathbf{y}) \in \mathbf{K}$  and therefore

$$\int_{\mathbf{K}} f d\mu \geq \int_{\mathbf{K}} J(\mathbf{y}) d\mu = \int_{\mathbf{Y}} J(\mathbf{y}) d\varphi,$$

which proves that  $\rho \geq \int_{\mathbf{Y}} J(\mathbf{y}) d\varphi$ .

On the other hand, recall that  $\mathbf{K}_{\mathbf{y}} \neq \emptyset$ ,  $\forall \mathbf{y} \in \mathbf{Y}$ . Consider the set-valued mapping  $\mathbf{y} \mapsto \mathbf{X}_{\mathbf{y}}^* \subset \mathbf{K}_{\mathbf{y}}$ . As  $f$  is continuous and  $\mathbf{K}$  is compact, then  $\mathbf{X}_{\mathbf{y}}^*$  is compact-valued. In addition, as  $f_{\mathbf{y}}$  is continuous, by [11, D6] (or [20]) there exists a measurable selector  $g : \mathbf{Y} \rightarrow \mathbf{X}_{\mathbf{y}}^*$  (and so  $f(g(\mathbf{y}), \mathbf{y}) = J(\mathbf{y})$ ). Therefore, for every  $\mathbf{y} \in \mathbf{Y}$ , let  $\psi_{\mathbf{y}}^*$  be the Dirac probability measure with support on the singleton  $g(\mathbf{y}) \in \mathbf{X}_{\mathbf{y}}^*$ , and let  $\mu$  be the probability measure on  $\mathbf{K}$  defined by:

$$\mu(C, B) := \int_B 1_C(g(\mathbf{y})) \varphi(d\mathbf{y}), \quad \forall B \in \mathcal{B}(\mathbb{R}^p), C \in \mathcal{B}(\mathbb{R}^n).$$

(The measure  $\mu$  is well-defined because  $g$  is measurable.) Then  $\mu$  is feasible for  $\mathbf{P}$  and

$$\begin{aligned} \rho \leq \int_{\mathbf{K}} f d\mu &= \int_{\mathbf{Y}} \left[ \int_{\mathbf{K}_{\mathbf{y}}} f(\mathbf{x}, \mathbf{y}) d\delta_{g(\mathbf{y})} \right] d\varphi(\mathbf{y}) \\ &= \int_{\mathbf{Y}} f(g(\mathbf{y}), \mathbf{y}) d\varphi(\mathbf{y}) = \int_{\mathbf{Y}} J(\mathbf{y}) d\varphi(\mathbf{y}), \end{aligned}$$

which shows that  $\mu$  is an optimal solution of  $\mathbf{P}$  and  $\rho = \int_{\mathbf{Y}} J(\mathbf{y}) d\varphi(\mathbf{y})$ .

(b) Let  $\mu^*$  be an arbitrary optimal solution of  $\mathbf{P}$ , hence supported on  $\mathbf{K}_{\mathbf{y}} \times \mathbf{Y}$ . Therefore, as  $\mathbf{K}$  is contained in the cartesian product  $\mathbb{R}^p \times \mathbb{R}^n$ , the probability measure  $\mu^*$  can be disintegrated as

$$\mu^*(C, B) := \int_{B \cap \mathbf{Y}} \psi^*(C \cap \mathbf{K}_{\mathbf{y}} | \mathbf{y}) d\varphi(\mathbf{y}), \quad \forall B \in \mathcal{B}(\mathbb{R}^p), C \in \mathcal{B}(\mathbb{R}^n),$$

where for all  $\mathbf{y} \in \mathbf{Y}$ ,  $\psi^*(\cdot | \mathbf{y})$  is a probability measure on  $\mathbf{K}_{\mathbf{y}}$ . (The object  $\psi^*(\cdot | \cdot)$  is called a stochastic kernel; see e.g. [8, p. 88–89] or [11, D8].) Hence from (a),

$$\begin{aligned} \rho = \int_{\mathbf{Y}} J(\mathbf{y}) d\varphi(\mathbf{y}) &= \int_{\mathbf{K}} f(\mathbf{x}, \mathbf{y}) d\mu^*(\mathbf{x}, \mathbf{y}) \\ &= \int_{\mathbf{Y}} \left( \int_{\mathbf{K}_{\mathbf{y}}} f(\mathbf{x}, \mathbf{y}) \psi^*(d\mathbf{x} | \mathbf{y}) \right) d\varphi(\mathbf{y}). \end{aligned}$$

Therefore, using  $f(\mathbf{x}, \mathbf{y}) \geq J(\mathbf{y})$  on  $\mathbf{K}$ ,

$$0 = \int_{\mathbf{Y}} \left( \int_{\mathbf{K}_{\mathbf{y}}} \underbrace{J(\mathbf{y}) - f(\mathbf{x}, \mathbf{y})}_{\leq 0} \psi^*(d\mathbf{x} | \mathbf{y}) \right) d\varphi(\mathbf{y}),$$

which implies  $\psi^*(\mathbf{X}^*(\mathbf{y}) | \mathbf{y}) = 1$  for almost all  $\mathbf{y} \in \mathbf{Y}$ .

(c) Let  $g : \mathbf{Y} \rightarrow \mathbf{K}_{\mathbf{y}}$  be the measurable mapping of Lemma 2.1(b). As  $J(\mathbf{y}) = f(g(\mathbf{y}), \mathbf{y})$  and  $(g(\mathbf{y}), \mathbf{y}) \in \mathbf{K}$  then necessarily  $g(\mathbf{y}) \in \mathbf{X}_{\mathbf{y}}^*$  for every  $\mathbf{y} \in \mathbf{Y}$ . Next, let  $\mu^*$  be an optimal solution of  $\mathbf{P}$ , and let  $\alpha \in \mathbb{N}^n$ ,  $\beta \in \mathbb{N}^p$ . Then

$$\begin{aligned} \int_{\mathbf{K}} \mathbf{x}^{\alpha} \mathbf{y}^{\beta} d\mu^*(\mathbf{x}, \mathbf{y}) &= \int_{\mathbf{Y}} \mathbf{y}^{\beta} \left( \int_{\mathbf{X}_{\mathbf{y}}^*} \mathbf{x}^{\alpha} \psi^*(d\mathbf{x} | \mathbf{y}) \right) d\varphi(\mathbf{y}) \\ &= \int_{\mathbf{Y}} \mathbf{y}^{\beta} g(\mathbf{y})^{\alpha} d\varphi(\mathbf{y}), \end{aligned}$$

the desired result.  $\square$

An optimal solution  $\mu^*$  of  $\mathbf{P}$  encodes *all* information on the optimal solutions  $\mathbf{x}^*(\mathbf{y})$  of  $\mathbf{P}_{\mathbf{y}}$ . For instance, let  $\mathbf{B}$  be a given Borel set of  $\mathbb{R}^n$ . Then from Theorem 2.2,

$$\text{Prob}(\mathbf{x}^*(\mathbf{y}) \in \mathbf{B}) = \mu^*(\mathbf{B} \times \mathbb{R}^p) = \int_{\mathbf{Y}} \psi^*(\mathbf{B} | \mathbf{y}) d\varphi(\mathbf{y}),$$

with  $\psi^*$  as in Theorem 2.2(b).

Consequently, if one knows an optimal solution  $\mu^*$  of  $\mathbf{P}$  then one may evaluate functionals on the solutions of  $\mathbf{P}_{\mathbf{y}}$ ,  $\mathbf{y} \in \mathbf{Y}$ . That is, assuming that for almost all

$\mathbf{y} \in \mathbf{Y}$ , problem  $\mathbf{P}_{\mathbf{y}}$  has a unique optimal solution  $\mathbf{x}^*(\mathbf{y})$ , and given a measurable mapping  $h : \mathbb{R}^n \rightarrow \mathbb{R}^q$ , one may evaluate the functional

$$\int_{\mathbf{Y}} h(\mathbf{x}^*(\mathbf{y})) d\varphi(\mathbf{y}).$$

For instance, with  $\mathbf{x} \mapsto h(\mathbf{x}) := \mathbf{x}$  one obtains the *mean* vector  $E_{\varphi}(\mathbf{x}^*(\mathbf{y})) := \int_{\mathbf{Y}} \mathbf{x}^*(\mathbf{y}) d\varphi(\mathbf{y})$  of optimal solutions  $\mathbf{x}^*(\mathbf{y})$ ,  $\mathbf{y} \in \mathbf{Y}$ .

**Corollary 2.3.** *Let both  $\mathbf{Y} \subset \mathbb{R}^p$  and  $\mathbf{K}$  in (2.2) be compact. Assume that for every  $\mathbf{y} \in \mathbf{Y}$ , the set  $\mathbf{K}_{\mathbf{y}} \subset \mathbb{R}^n$  in (2.3) is nonempty, and for almost all  $\mathbf{y} \in \mathbf{Y}$ , the set  $\mathbf{X}_{\mathbf{y}}^* := \{\mathbf{x} \in \mathbf{K}_{\mathbf{y}} : J(\mathbf{y}) = f(\mathbf{x}, \mathbf{y})\}$  is the singleton  $\{\mathbf{x}^*(\mathbf{y})\}$ . Then for every measurable mapping  $h : \mathbb{R}^n \rightarrow \mathbb{R}^q$ ,*

$$(2.8) \quad \int_{\mathbf{Y}} h(\mathbf{x}^*(\mathbf{y})) d\varphi(\mathbf{y}) = \int_{\mathbf{K}} h(\mathbf{x}) d\mu^*(\mathbf{x}, \mathbf{y}).$$

where  $\mu^*$  is an optimal solution of  $\mathbf{P}$ .

*Proof.* By Theorem 2.2(c)

$$\int_{\mathbf{K}} h(\mathbf{x}) d\mu^*(\mathbf{x}, \mathbf{y}) = \int_{\mathbf{Y}} \left[ \int_{\mathbf{X}_{\mathbf{y}}^*} h(\mathbf{x}) \psi^*(d\mathbf{x} | \mathbf{y}) \right] d\varphi(\mathbf{y}) = \int_{\mathbf{Y}} h(\mathbf{x}^*(\mathbf{y})) d\varphi(\mathbf{y}).$$

□

**2.2. Duality.** Consider the following infinite-dimensional linear program  $\mathbf{P}^*$ :

$$(2.9) \quad \rho^* := \sup_{p \in \mathbb{R}[\mathbf{y}]} \int_{\mathbf{Y}} p d\varphi$$

$$f(\mathbf{x}, \mathbf{y}) - p(\mathbf{y}) \geq 0 \quad \forall (\mathbf{x}, \mathbf{y}) \in \mathbf{K}.$$

Then  $\mathbf{P}^*$  is a *dual* of  $\mathbf{P}$ .

**Lemma 2.4.** *Let both  $\mathbf{Y} \subset \mathbb{R}^p$  and  $\mathbf{K}$  in (2.2) be compact and let  $\mathbf{P}$  and  $\mathbf{P}^*$  be as in (2.4) and (2.9) respectively. Then there is no duality gap, i.e.,  $\rho = \rho^*$ .*

*Proof.* For a topological space  $\mathcal{X}$  denote by  $C(\mathcal{X})$  the space of bounded continuous functions on  $\mathcal{X}$ . Let  $\mathcal{M}(\mathbf{K})$  be the vector space of finite signed Borel measures on  $\mathbf{K}$  (and so  $\mathbf{M}(\mathbf{K})$  is its positive cone). Let  $\pi : \mathcal{M}(\mathbf{K}) \rightarrow \mathcal{M}(\mathbf{Y})$  be defined by  $(\pi\mu)(B) = \mu((\mathbb{R}^n \times B) \cap \mathbf{K})$  for all  $B \in \mathcal{B}(\mathbf{Y})$ , with adjoint mapping  $\pi^* : C(\mathbf{Y}) \rightarrow C(\mathbf{K})$  defined as

$$(\mathbf{x}, \mathbf{y}) \mapsto (\pi^*h)(\mathbf{x}, \mathbf{y}) := h(\mathbf{y}), \quad \forall h \in C(\mathbf{Y}).$$

Put (2.4) in the framework of infinite-dimensional linear programs on vector spaces, as described in e.g. [1]. That is:

$$\rho = \inf_{\mu \in \mathcal{M}(\mathbf{K})} \{ \langle f, \mu \rangle : \pi\mu = \varphi, \mu \geq 0 \},$$

with dual:

$$\tilde{\rho} = \sup_{h \in C(\mathbf{Y})} \{ \langle h, \varphi \rangle : f - \pi^*h \geq 0 \text{ on } \mathbf{K} \}.$$

One first proves that  $\rho = \tilde{\rho}$  and then  $\tilde{\rho} = \rho^*$ .

By [1, Theor. 3.10], to get  $\rho = \tilde{\rho}$ , it suffices to prove that the set  $D := \{(\pi\mu, \langle f, \mu \rangle) : \mu \in \mathbf{M}(\mathbf{K})\}$  is closed for the respective weak  $\star$  topologies  $\sigma(\mathcal{M}(\mathbf{Y}) \times$

$\mathbb{R}, C(\mathbf{Y}) \times \mathbb{R}$ ) and  $\sigma(\mathcal{M}(\mathbf{K}), C(\mathbf{K}))$ . Therefore consider a converging sequence  $\pi\mu_n \rightarrow a$  with  $\mu_n \in \mathbf{M}(\mathbf{K})$ . The sequence  $(\mu_n)$  is uniformly bounded because

$$\mu_n(\mathbf{K}) = (\pi\mu_n)(\mathbf{Y}) = \langle 1, \pi\mu_n \rangle \rightarrow \langle 1, a \rangle = a(\mathbf{Y}).$$

But by the Banach-Alaoglu Theorem (see e.g. [2]), the bounded closed sets of  $\mathbf{M}(\mathbf{K})$  are compact in the weak  $\star$  topology. And so  $\mu_{n_k} \rightarrow \mu$  for some  $\mu \in \mathbf{M}(\mathbf{K})$  and some subsequence  $(n_k)$ . Next, observe that for  $h \in C(\mathbf{Y})$  arbitrary,

$$\langle h, \pi\mu_{n_k} \rangle = \langle \pi^*h, \mu_{n_k} \rangle \rightarrow \langle \pi^*h, \mu \rangle = \langle h, \pi\mu \rangle,$$

where we have used that  $\pi^*h \in C(\mathbf{K})$ . Hence combining the above with  $\pi\mu_{n_k} \rightarrow a$ , we obtain  $\pi\mu = a$ . Similarly,  $\langle f, \mu_{n_k} \rangle \rightarrow \langle f, \mu \rangle$  because  $f \in C(\mathbf{K})$ . Hence  $D$  is closed and the desired result  $\rho = \tilde{\rho}$  follows.

We next prove that  $\tilde{\rho} = \rho^*$ . Given  $\epsilon > 0$  fixed arbitrary, there is a function  $h_\epsilon \in C(\mathbf{Y})$  such that  $f - h_\epsilon \geq 0$  on  $\mathbf{K}$  and  $\int h_\epsilon d\varphi \geq \tilde{\rho} - \epsilon$ . By compactness of  $\mathbf{Y}$  and the Stone-Weierstrass theorem, there is  $p_\epsilon \in \mathbb{R}[\mathbf{y}]$  such that  $\sup_{\mathbf{y} \in \mathbf{Y}} |h_\epsilon(\mathbf{y}) - p_\epsilon(\mathbf{y})| \leq \epsilon$ . Hence the polynomial  $\tilde{p}_\epsilon := p_\epsilon - \epsilon$  is feasible with value  $\int_{\mathbf{Y}} \tilde{p}_\epsilon d\varphi \geq \tilde{\rho} - 3\epsilon$ , and as  $\epsilon$  was arbitrary, the result  $\tilde{\rho} = \rho^*$  follows.  $\square$

As next shown, optimal or nearly optimal solutions of  $\mathbf{P}^*$  provide us with polynomial lower approximations of the optimal value function  $\mathbf{y} \mapsto J(\mathbf{y})$  that converges to  $J(\cdot)$  in the  $L_1(\varphi)$  norm. Moreover, one may also obtain a piecewise polynomial approximation that converges to  $J(\cdot)$  almost uniformly. (Recall that a sequence of measurable functions  $(g_n)$  on a measure space  $(\mathbf{Y}, \mathcal{B}(\mathbf{Y}), \varphi)$  converges to  $g$  *almost uniformly* if and only if for every  $\epsilon > 0$ , there is a set  $A \in \mathcal{B}(\mathbf{Y})$  such that  $\varphi(A) < \epsilon$  and  $g_n \rightarrow g$  uniformly on  $\mathbf{Y} \setminus A$ .)

**Corollary 2.5.** *Let both  $\mathbf{Y} \subset \mathbb{R}^p$  and  $\mathbf{K}$  in (2.2) be compact and assume that for every  $\mathbf{y} \in \mathbf{Y}$ , the set  $\mathbf{K}_{\mathbf{y}}$  is nonempty. Let  $\mathbf{P}^*$  be as in (2.9). If  $(p_i)_{i \in \mathbb{N}} \subset \mathbb{R}[\mathbf{y}]$  is a maximizing sequence of (2.9) then*

$$(2.10) \quad \int_{\mathbf{Y}} |J(\mathbf{y}) - p_i(\mathbf{y})| d\varphi \rightarrow 0 \quad \text{as } i \rightarrow \infty.$$

Moreover, define the functions  $(\tilde{p}_i)$  as follows:

$$\tilde{p}_0 := p_0, \quad \mathbf{y} \mapsto \tilde{p}_i(\mathbf{y}) := \max[\tilde{p}_{i-1}(\mathbf{y}), p_i(\mathbf{y})], \quad i = 1, 2, \dots$$

Then  $\tilde{p}_i \rightarrow J(\cdot)$  almost uniformly.

*Proof.* By Lemma 2.4, we already know that  $\rho^* = \rho$  and so

$$\int_{\mathbf{Y}} p_i(\mathbf{y}) d\varphi(\mathbf{y}) \uparrow \rho^* = \rho = \int_{\mathbf{Y}} J(\mathbf{y}) d\varphi.$$

Next by feasibility of  $p_i$  in (2.9)

$$f(\mathbf{x}, \mathbf{y}) \geq p_i(\mathbf{y}) \quad \forall (\mathbf{x}, \mathbf{y}) \in \mathbf{K} \Rightarrow \inf_{\mathbf{x} \in \mathbf{K}_{\mathbf{y}}} f(\mathbf{x}, \mathbf{y}) = J(\mathbf{y}) \geq p_i(\mathbf{y}) \quad \forall \mathbf{y} \in \mathbf{Y}.$$

Hence (2.10) follows from  $p_i(\mathbf{y}) \leq J(\mathbf{y})$  on  $\mathbf{Y}$ .

Next, with  $\mathbf{y} \in \mathbf{Y}$  fixed, the sequence  $(\tilde{p}_i(\mathbf{y}))_i$  is obviously monotone non decreasing and bounded above by  $J(\mathbf{y})$ , hence with a limit  $p^*(\mathbf{y}) \leq J(\mathbf{y})$ . Therefore  $\tilde{p}_i$  has the pointwise limit  $\mathbf{y} \mapsto p^*(\mathbf{y}) \leq J(\mathbf{y})$ . Also, by the Montone convergence



theorem,  $\int_{\mathbf{Y}} \tilde{p}_i(\mathbf{y}) d\varphi(\mathbf{y}) \rightarrow \int_{\mathbf{Y}} p^*(\mathbf{y}) d\varphi(\mathbf{y})$ . This latter fact combined with (2.10) and  $p_i(\mathbf{y}) \leq \tilde{p}_i(\mathbf{y}) \leq J(\mathbf{y})$  yields

$$0 = \int_{\mathbf{Y}} (J(\mathbf{y}) - p^*(\mathbf{y})) d\varphi(\mathbf{y}),$$

which in turn implies that  $p^*(\mathbf{y}) = J(\mathbf{y})$  for almost all  $\mathbf{y} \in \mathbf{Y}$ . Therefore  $\tilde{p}_i(\mathbf{y}) \rightarrow J(\mathbf{y})$  for almost all  $\mathbf{y} \in \mathbf{Y}$ . And so, by Egoroff's Theorem [2, Theor. 2.5.5],  $\tilde{p}_i \rightarrow J(\cdot)$  almost uniformly.  $\square$

### 3. A HIERARCHY OF SEMIDEFINITE RELAXATIONS

In general, solving the infinite-dimensional problem  $\mathbf{P}$  and getting an optimal solution  $\mu^*$  is impossible. One possibility is to use numerical discretization schemes on a box containing  $\mathbf{K}$ ; see for instance [12]. But in the present context of parametric optimization, if one selects finitely many *grid* points  $(\mathbf{x}, \mathbf{y}) \in \mathbf{K}$ , one is implicitly considering solving (or rather approximating)  $\mathbf{P}_{\mathbf{y}}$  for finitely many points  $\mathbf{y}$  in a grid of  $\mathbf{Y}$ , which we want to avoid. To avoid this numerical discretization scheme we will use specific features of  $\mathbf{P}$  when its data  $f$  (resp.  $\mathbf{K}$ ) is a polynomial (resp. a compact basic semi-algebraic set).

Therefore in this section we are now considering a *polynomial* parametric optimization problem, a special case of (2.1) as we assume the following:

- $f \in \mathbb{R}[\mathbf{x}, \mathbf{y}]$  and  $h_j \in \mathbb{R}[\mathbf{x}, \mathbf{y}]$ , for every  $j = 1, \dots, m$ .
- $\mathbf{K}$  is compact and  $\mathbf{Y} \subset \mathbb{R}^p$  is a compact basic semi-algebraic set.

Hence the set  $\mathbf{K} \subset \mathbb{R}^n \times \mathbb{R}^p$  in (2.2) is a compact basic semi-algebraic set. We also assume that there is a probability measure  $\varphi$  on  $\mathbf{Y}$ , absolutely continuous with respect to the Lebesgue measure, whose moments  $\gamma = (\gamma_\beta)$ ,  $\beta \in \mathbb{N}^p$ , are available. As already mentioned, if  $\mathbf{Y}$  is a simple set (like e.g. a simplex or a box) then one may choose  $\varphi$  to be the probability measure uniformly distributed on  $\mathbf{Y}$ , for which all moments can be computed easily. Sometimes, in the context of optimization with data uncertainty, the probability measure  $\varphi$  is already specified and in this case we assume that its moments  $\gamma = (\gamma_\beta)$ ,  $\beta \in \mathbb{N}^p$ , are available.

**3.1. Notation and preliminaries.** Let  $\mathbb{N}_i^n := \{\alpha \in \mathbb{N}^n : |\alpha| \leq i\}$  with  $|\alpha| = \sum_i \alpha_i$ . With a sequence  $\mathbf{z} = (z_{\alpha\beta})$ ,  $\alpha \in \mathbb{N}^n, \beta \in \mathbb{N}^p$ , indexed in the canonical basis  $(\mathbf{x}^\alpha \mathbf{y}^\beta)$  of  $\mathbb{R}[\mathbf{x}, \mathbf{y}]$ , let  $L_{\mathbf{z}} : \mathbb{R}[\mathbf{x}, \mathbf{y}] \rightarrow \mathbb{R}$  be the linear mapping:

$$f (= \sum_{\alpha\beta} f_{\alpha\beta}(\mathbf{x}, \mathbf{y})) \mapsto L_{\mathbf{z}}(f) := \sum_{\alpha\beta} f_{\alpha\beta} z_{\alpha\beta}, \quad f \in \mathbb{R}[\mathbf{x}, \mathbf{y}].$$

*Moment matrix.* The moment matrix  $\mathbf{M}_i(\mathbf{z})$  associated with a sequence  $\mathbf{z} = (z_{\alpha\beta})$ , has its rows and columns indexed in the canonical basis  $(\mathbf{x}^\alpha \mathbf{y}^\beta)$ , and with entries.

$$\mathbf{M}_i(\mathbf{z})(\alpha, \beta), (\delta, \gamma) = L_{\mathbf{z}}(\mathbf{x}^\alpha \mathbf{y}^\beta \mathbf{x}^\delta \mathbf{y}^\gamma) = z_{(\alpha+\delta)(\beta+\gamma)},$$

for every  $\alpha, \delta \in \mathbb{N}_i^n$  and every  $\beta, \gamma \in \mathbb{N}_i^p$ .

*Localizing matrix.* Let  $q$  be the polynomial  $(\mathbf{x}, \mathbf{y}) \mapsto q(\mathbf{x}, \mathbf{y}) := \sum_{u,v} q_{uv} \mathbf{x}^u \mathbf{y}^v$ . The localizing matrix  $\mathbf{M}_i(q, \mathbf{z})$  associated with  $q \in \mathbb{R}[\mathbf{x}, \mathbf{y}]$  and a sequence  $\mathbf{z} = (z_{\alpha\beta})$ , has its rows and columns indexed in the canonical basis  $(\mathbf{x}^\alpha \mathbf{y}^\beta)$ , and with entries.

$$\begin{aligned} \mathbf{M}_i(q, \mathbf{z})(\alpha, \beta), (\delta, \gamma) &= L_{\mathbf{z}}(q(\mathbf{x}, \mathbf{y}) \mathbf{x}^\alpha \mathbf{y}^\beta \mathbf{x}^\delta \mathbf{y}^\gamma) \\ &= \sum_{u \in \mathbb{N}^n, v \in \mathbb{N}^p} q_{uv} z_{(\alpha+\delta+u)(\beta+\gamma+v)}, \end{aligned}$$

for every  $\alpha, \delta \in \mathbb{N}_i^n$  and every  $\beta, \gamma \in \mathbb{N}_i^p$ .

A sequence  $\mathbf{z} = (z_{\alpha\beta}) \subset \mathbb{R}$  has a *representing* finite Borel measure supported on  $\mathbf{K}$  if there exists a finite Borel measure  $\mu$  such that

$$z_{\alpha\beta} = \int_{\mathbf{K}} \mathbf{x}^\alpha \mathbf{y}^\beta d\mu, \quad \forall \alpha \in \mathbb{N}^n, \beta \in \mathbb{N}^p.$$

The next important result states a necessary and sufficient condition when  $\mathbf{K}$  is compact and its defining polynomials  $(h_k) \subset \mathbb{R}[\mathbf{x}, \mathbf{y}]$  satisfy some condition.

**Assumption 3.1.** *Let  $(h_j)_{j=1}^t \subset \mathbb{R}[\mathbf{x}, \mathbf{y}]$  be a given family of polynomials. There is some  $N$  such that the quadratic polynomial  $(\mathbf{x}, \mathbf{y}) \mapsto N - \|(\mathbf{x}, \mathbf{y})\|^2$  can be written*

$$N - \|(\mathbf{x}, \mathbf{y})\|^2 = \sigma_0 + \sum_{k=1}^t \sigma_k h_k,$$

for some s.o.s. polynomials  $(\sigma_j)_{j=1}^t \subset \Sigma[\mathbf{x}, \mathbf{y}]$ .

**Theorem 3.2.** *Let  $\mathbf{K} := \{(\mathbf{x}, \mathbf{y}) : h_k(\mathbf{x}, \mathbf{y}) \geq 0, j = 1, \dots, t\}$  and let  $(h_k)_{k=1}^t$  satisfy Assumption 3.1. A sequence  $\mathbf{z} = (z_{\alpha\beta})$  has a representing measure on  $\mathbf{K}$  if and only if:*

$$\mathbf{M}_i(\mathbf{z}) \succeq 0; \quad \mathbf{M}_i(h_k \mathbf{z}) \succeq 0, \quad k = 0, \dots, t.$$

Theorem 3.2 is a direct consequence of Putinar's Positivstellensatz [18] and [21]. Of course, when Assumption 3.1 holds then  $\mathbf{K}$  is compact. On the other hand, if  $\mathbf{K}$  is compact and one knows a bound  $N$  for  $\|(\mathbf{x}, \mathbf{y})\|$  on  $\mathbf{K}$  then it suffices to add the redundant quadratic constraint  $h_{t+1}(\mathbf{x}, \mathbf{y}) := N^2 - \|(\mathbf{x}, \mathbf{y})\|^2 \geq 0$  to the definition of  $\mathbf{K}$ , and Assumption 3.1 holds.

**3.2. Semidefinite relaxations.** To compute (or at least, approximate) the optimal value  $\rho$  of problem  $\mathbf{P}$  in (2.4), we now provide a hierarchy of semidefinite relaxations in the spirit of those defined in [13].

Let  $\mathbf{K} \subset \mathbb{R}^n \times \mathbb{R}^p$  be as in (2.2), and let  $\mathbf{Y} \subset \mathbb{R}^p$  be the compact semi-algebraic set defined by:

$$(3.1) \quad \mathbf{Y} := \{\mathbf{y} \in \mathbb{R}^p : h_k(\mathbf{y}) \geq 0, \quad k = m+1, \dots, t\}$$

for some polynomials  $(h_k)_{k=m+1}^t \in \mathbb{R}[\mathbf{y}]$ ; let  $v_k := \lceil (\deg h_k)/2 \rceil$  for every  $k = 1, \dots, t$ . Next, let  $\gamma = (\gamma_\beta)$  with

$$\gamma_\beta = \int_{\mathbf{Y}} \mathbf{y}^\beta d\varphi(\mathbf{y}), \quad \forall \beta \in \mathbb{N}^p,$$

be the moments of a probability measure  $\varphi$  on  $\mathbf{Y}$ , absolutely continuous with respect to the Lebesgue measure, and let  $i_0 := \max[\lceil (\deg f)/2 \rceil, \max_k v_k]$ . For  $i \geq i_0$ , consider the following semidefinite relaxations:

$$(3.2) \quad \begin{aligned} \rho_i = \inf_{\mathbf{z}} \quad & L_{\mathbf{z}}(f) \\ \text{s.t.} \quad & \mathbf{M}_i(\mathbf{z}) \succeq 0 \\ & \mathbf{M}_{i-v_j}(h_j \mathbf{z}) \succeq 0, \quad j = 1, \dots, t \\ & L_{\mathbf{z}}(\mathbf{y}^\beta) = \gamma_\beta, \quad \forall \beta \in \mathbb{N}_i^p. \end{aligned}$$

**Theorem 3.3.** *Let  $\mathbf{K}, \mathbf{Y}$  be as (2.2) and (3.1) respectively, and let  $(h_k)_{k=1}^t$  satisfy Assumption 3.1. Assume that for every  $\mathbf{y} \in \mathbf{Y}$  the set  $\mathbf{K}_{\mathbf{y}}$  is nonempty, and for almost all  $\mathbf{y} \in \mathbf{Y}$ ,  $J(\mathbf{y})$  is attained at a unique optimal solution  $\mathbf{x}^*(\mathbf{y})$ . Consider the semidefinite relaxations (3.2). Then:*

(a)  $\rho_i \uparrow \rho$  as  $i \rightarrow \infty$ .

(b) *Let  $\mathbf{z}^i$  be a nearly optimal solution of (3.2), e.g. such that  $L_{\mathbf{z}^i}(f) \leq \rho_i + 1/i$ , and let  $g : \mathbf{Y} \rightarrow \mathbf{K}_{\mathbf{y}}$  be the measurable mapping in Theorem 2.2(c). Then*

$$(3.3) \quad \lim_{i \rightarrow \infty} z_{\alpha\beta}^i = \int_{\mathbf{Y}} \mathbf{y}^\beta g(\mathbf{y})^\alpha d\varphi(\mathbf{y}), \quad \forall \alpha \in \mathbb{N}^n, \beta \in \mathbb{N}^p.$$

In particular, for every  $k = 1, \dots, n$ ,

$$(3.4) \quad \lim_{i \rightarrow \infty} z_{e(k)\beta}^i = \int_{\mathbf{Y}} \mathbf{y}^\beta g_k(\mathbf{y}) d\varphi(\mathbf{y}), \quad \forall \beta \in \mathbb{N}^p,$$

where  $e(k) = (\delta_{j=k})_j \in \mathbb{N}^n$ .

The proof is postponed to Section 4.

**Remark 3.4.** Observe that if  $\rho_i = +\infty$  for some index  $i$  in the hierarchy (and hence for all  $i' \geq i$ ), then the set  $\mathbf{K}_{\mathbf{y}}$  is empty for all  $\mathbf{y}$  in some Borel set of  $\mathbf{Y}$  with  $\varphi(\mathbf{Y}) > 0$ . Conversely, one may prove that if  $\mathbf{K}_{\mathbf{y}}$  is empty for all  $\mathbf{y}$  in some Borel set of  $\mathbf{Y}$  with  $\varphi(\mathbf{Y}) > 0$ , then necessarily  $\rho_i = +\infty$  for all  $i$  sufficiently large. In other words, the hierarchy of semidefinite relaxations (3.2) may also provide a certificate of emptiness of  $\mathbf{K}_{\mathbf{y}}$  for some Borel set of  $\mathbf{Y}$  with positive Lebesgue measure.

**3.3. The dual semidefinite relaxations.** The dual of the semidefinite relaxation (3.2) reads:

$$(3.5) \quad \begin{aligned} \rho_i^* &= \sup_{p, (\sigma_j)} \int_{\mathbf{Y}} p d\varphi \\ \text{s.t.} \quad & f - p = \sigma_0 + \sum_{j=1}^t \sigma_j h_j \\ & p \in \mathbb{R}[\mathbf{y}]; \sigma_j \subset \Sigma[\mathbf{x}, \mathbf{y}], \quad j = 1, \dots, t \\ & \deg p \leq 2i, \deg \sigma_j h_j \leq 2i, \quad j = 1, \dots, t \end{aligned}$$

Observe that (3.5) is a strengthening of (2.9) as one restricts to polynomials  $p \in \mathbb{R}[\mathbf{y}]$  of degree at most  $2i$  and the nonnegativity of  $f-p$  in (2.9) is replaced with a stronger requirement in (3.5). Therefore  $\rho_i^* \leq \rho^*$  for every  $i$ .

**Theorem 3.5.** *Let  $\mathbf{K}, \mathbf{Y}$  be as (2.2) and (3.1) respectively, and let  $(h_k)_{k=1}^t$  satisfy Assumption 3.1. Assume that for every  $\mathbf{y} \in \mathbf{Y}$  the set  $\mathbf{K}_{\mathbf{y}}$  is nonempty, and consider the semidefinite relaxations (3.5). Then:*

(a)  $\rho_i^* \uparrow \rho$  as  $i \rightarrow \infty$ .

(b) *Let  $(p_i, (\sigma_j^i))$  be a nearly optimal solution of (3.5), e.g. such that  $\int_{\mathbf{Y}} p_i d\varphi \geq \rho_i^* - 1/i$ . Then  $p_i \leq J(\cdot)$  and*

$$(3.6) \quad \lim_{i \rightarrow \infty} \int_{\mathbf{Y}} (J(\mathbf{y}) - p_i(\mathbf{y})) d\varphi(\mathbf{y}) = 0$$

Moreover if one defines

$$\tilde{p}_0 := p_0, \quad \mathbf{y} \mapsto \tilde{p}_i(\mathbf{y}) := \max[\tilde{p}_{i-1}(\mathbf{y}), p_i(\mathbf{y})], \quad i = 1, 2, \dots,$$

then  $\tilde{p}_i \rightarrow J(\cdot)$  almost uniformly on  $\mathbf{Y}$ .

*Proof.* Recall that by Lemma 2.4,  $\rho = \rho^*$ . Moreover let  $(p_k) \subset \mathbb{R}[\mathbf{y}]$  be a maximizing sequence of (2.9) as in Corollary 2.5 with value  $s_k := \int p_k d\varphi$ , and let  $p'_k := p_k - 1/k$  for every  $k$  so that  $f - p'_k > 1/k$  on  $\mathbf{K}$ . By Theorem 3.2, there exist s.o.s. polynomials  $(\sigma_j^k) \subset \Sigma[\mathbf{x}, \mathbf{y}]$  such that  $f - p'_k = \sigma_0^k + \sum_j \sigma_j^k h_j$ . Letting  $d_k$  be the maximum degree of  $\sigma_0$  and  $\sigma_j h_j$ ,  $j = 1, \dots, t$ , it follows that  $(s_k - 1/k, (\sigma_j^k))$  is a feasible solution of (3.5) with  $i := d_k$ . Hence  $\rho^* \geq \rho_{d_k}^* \geq s_k - 1/k$  and the result (a) follows because  $s_k \rightarrow \rho^*$ , and the sequence  $\rho_i^*$  is monotone. Then (b) follows from Corollary 2.5.  $\square$

Hence Theorem 3.5 provides a lower polynomial approximation  $p_i \in \mathbb{R}[\mathbf{y}]$  of the optimal value function  $J(\cdot)$ . Its degree is bounded by  $2i$ , the order of the moments  $(\gamma_\beta)$  of  $\varphi$  taken into account in the semidefinite relaxation (3.5). Moreover one may even define a piecewise polynomial lower approximation  $\tilde{p}_i$  that converges almost uniformly to  $J(\cdot)$  on  $\mathbf{Y}$ .

**Functionals of the optimal solutions.** Theorem 3.3 provides a mean of approximating any polynomial functional on the optimal solutions of  $\mathbf{P}_{\mathbf{y}}$ ,  $\mathbf{y} \in \mathbf{Y}$ . Indeed,

**Corollary 3.6.** *Let  $\mathbf{K}, \mathbf{Y}$  be as (2.2) and (3.1) respectively, and let  $(h_k)_{k=1}^t$  satisfy Assumption 3.1. Assume that for every  $\mathbf{y} \in \mathbf{Y}$  the set  $\mathbf{K}_{\mathbf{y}}$  is nonempty, and for almost all  $\mathbf{y} \in \mathbf{Y}$ ,  $J(\mathbf{y})$  is attained at a unique optimal solution  $\mathbf{x}^*(\mathbf{y}) \in \mathbf{X}_{\mathbf{y}}^*$ . Let  $h \in \mathbb{R}[\mathbf{x}]$ ,*

$$\mathbf{x} \mapsto h(\mathbf{x}) := \sum_{\alpha \in \mathbb{N}^n} h_\alpha \mathbf{x}^\alpha,$$

and let  $\mathbf{z}^i$  be a nearly optimal solution of the semidefinite relaxations (3.2).

Then, for  $i$  sufficiently large,

$$\int_{\mathbf{Y}} h(\mathbf{x}^*(\mathbf{y})) d\varphi(\mathbf{y}) \approx \sum_{\alpha \in \mathbb{N}^n} h_\alpha z_{\alpha 0}^i.$$

*Proof.* The proof is an immediate consequence of Theorem 3.3 and Corollary 2.3.  $\square$

**3.4. Persistence for Boolean variables.** One interesting and potentially useful application is in Boolean optimization. Indeed suppose that for some subset  $I \subseteq \{1, \dots, n\}$ , the variables  $(x_i)$ ,  $i \in I$ , are boolean, that is, the definition of  $\mathbf{K}$  in (2.2) includes the constraints  $x_i^2 - x_i = 0$ , for every  $i \in I$ .

Then for instance, one might be interested to determine whether in an optimal solution  $\mathbf{x}^*(\mathbf{y})$  of  $\mathbf{P}_{\mathbf{y}}$ , and for some index  $i \in I$ , one has  $x_i^*(\mathbf{y}) = 1$  (or  $x_i^*(\mathbf{y}) = 0$ ) for almost all values of the parameter  $\mathbf{y} \in \mathbf{Y}$ . In [3, 17] the probability that  $x_k^*(\mathbf{y})$  is 1 is called the *persistence* of the boolean variable  $x_k^*(\mathbf{y})$ .

**Corollary 3.7.** *Let  $\mathbf{K}, \mathbf{Y}$  be as in (2.2) and (3.1) respectively. Let  $(h_k)_{k=1}^t$  satisfy (3.1). Assume that for every  $\mathbf{y} \in \mathbf{Y}$  the set  $\mathbf{K}_{\mathbf{y}}$  is nonempty. Let  $\mathbf{z}^i$  be a nearly optimal solution of the semidefinite relaxations (3.2). Then for  $k \in I$  fixed.*

(a)  $x_k^*(\mathbf{y}) = 1$  for almost all  $\mathbf{y} \in \mathbf{Y}$ , only if  $\lim_{i \rightarrow \infty} z_{e^{(k)}0}^i = 1$ .

(b)  $x_k^*(\mathbf{y}) = 0$  for almost all  $\mathbf{y} \in \mathbf{Y}$ , only if  $\lim_{i \rightarrow \infty} z_{e^{(k)}0}^i = 0$ .

Assume that for almost all  $\mathbf{y} \in \mathbf{Y}$ ,  $J(\mathbf{y})$  is attained at a unique optimal solution  $\mathbf{x}^*(\mathbf{y}) \in \mathbf{X}_{\mathbf{y}}^*$ . Then  $\text{Prob}(x_k^*(\mathbf{y}) = 1) = \lim_{i \rightarrow \infty} z_{e^{(k)}0}^i$ , and so:

- (c)  $x_k^*(\mathbf{y}) = 1$  for almost all  $\mathbf{y} \in \mathbf{Y}$ , if and only if  $\lim_{i \rightarrow \infty} z_{e(k)0}^i = 1$ .  
 (d)  $x_k^*(\mathbf{y}) = 0$  for almost all  $\mathbf{y} \in \mathbf{Y}$ , if and only if  $\lim_{i \rightarrow \infty} z_{e(k)0}^i = 0$ .

*Proof.* (a) The *only if part*. Let  $\alpha := e(k) \in \mathbb{N}^n$ . From the proof of Theorem 3.3, there is a subsequence  $(i_l)_l \subset (i)_i$  such that

$$\lim_{l \rightarrow \infty} z_{e(k)0}^{i_l} = \int_{\mathbf{K}} x_k d\mu^*,$$

where  $\mu^*$  is an optimal solution of  $\mathbf{P}$ . Hence, by Theorem 2.2(b),  $\mu^*$  can be disintegrated into  $\psi^*(d\mathbf{x}|\mathbf{y})d\varphi(\mathbf{y})$  where  $\psi^*(\cdot|\mathbf{y})$  is a probability measure on  $\mathbf{X}_{\mathbf{y}}^*$  for every  $\mathbf{y} \in \mathbf{Y}$ . Therefore,

$$\begin{aligned} \lim_{l \rightarrow \infty} z_{e(k)0}^{i_l} &= \int_{\mathbf{Y}} \left( \int_{\mathbf{X}_{\mathbf{y}}^*} x_k \psi^*(d\mathbf{x}|\mathbf{y}) \right) d\varphi(\mathbf{y}), \\ &= \int_{\mathbf{Y}} \psi^*(\mathbf{X}_{\mathbf{y}}^*|\mathbf{y}) d\varphi(\mathbf{y}) \quad [\text{because } x_k^* = 1 \text{ in } \mathbf{X}_{\mathbf{y}}^*] \\ &= \int_{\mathbf{Y}} d\varphi(\mathbf{y}) = 1, \end{aligned}$$

and as the subsequence  $(i_l)_l$  was arbitrary, the whole sequence  $(z_{e(k)0}^i)$  converges to 1, the desired result. The proof of (b) being exactly the same is omitted.

Next, if for every  $\mathbf{y} \in \mathbf{Y}$ ,  $J(\mathbf{y})$  is attained at a singleton, by Theorem 3.3(b),

$$\begin{aligned} \lim_{i \rightarrow \infty} z_{e(k)0}^i &= \int_{\mathbf{Y}} x_k^*(\mathbf{y}) d\varphi(\mathbf{y}) = \varphi(\{\mathbf{y} : x_k^*(\mathbf{y}) = 1\}) \\ &= \text{Prob}(x_k^*(\mathbf{y}) = 1), \end{aligned}$$

from which (c) and (d) follow.  $\square$

**3.5. Estimating the density  $g(\mathbf{y})$ .** By Corollary 3.6, one may approximate any polynomial functional of the optimal solutions, like for instance the mean, variance, etc .. (with respect to the probability measure  $\varphi$ ). However, one may also wish to approximate (in some sense) the "curve"  $\mathbf{y} \mapsto g_k(\mathbf{y})$ , that is, the surface described by the  $k$ -th coordinate  $\mathbf{x}_k^*(\mathbf{y})$  of the optimal solution  $\mathbf{x}^*(\mathbf{y})$  when  $\mathbf{y}$  varies in  $\mathbf{Y}$ .

So let  $g : \mathbf{Y} \rightarrow \mathbb{R}^n$  be the measurable mapping in Theorem 3.3 and suppose that one knows some lower bound vector  $\mathbf{a} = (a_k) \in \mathbb{R}^n$ , where:

$$a_k \leq \inf \{ x_k : (\mathbf{x}, \mathbf{y}) \in \mathbf{K} \}, \quad k = 1, \dots, n.$$

Then for every  $k = 1, \dots, n$ , the measurable function  $\hat{g}_k : \mathbf{Y} \rightarrow \mathbb{R}^n$  defined by

$$(3.7) \quad \mathbf{y} \mapsto \hat{g}_k(\mathbf{y}) := g_k(\mathbf{y}) - a_k, \quad \mathbf{y} \in \mathbf{Y},$$

is nonnegative and integrable with respect to  $\varphi$ .

Hence for every  $k = 1, \dots, n$ , one may consider  $d\lambda := \hat{g}_k d\varphi$  as a Borel measure on  $\mathbf{Y}$  with unknown density  $\hat{g}_k$  with respect to  $\varphi$ , but with known moments  $\mathbf{u} = (u_\beta)$ . Indeed, using (3.4),

$$\begin{aligned} u_\beta &:= \int_{\mathbf{Y}} \mathbf{y}^\beta d\lambda(\mathbf{y}) = -a_k \int_{\mathbf{Y}} \mathbf{y}^\beta d\varphi(\mathbf{y}) + \int_{\mathbf{Y}} \mathbf{y}^\beta g_k(\mathbf{y}) d\varphi(\mathbf{y}) \\ (3.8) \quad &= -a_k \gamma_\beta + z_{e(k)\beta}, \quad \forall \beta \in \mathbb{N}^p, \end{aligned}$$

where for every  $k = 1, \dots, n$ ,

$$z_{e(k)\beta} = \lim_{i \rightarrow \infty} z_{e(k)\beta}^i, \quad \forall \beta \in \mathbb{N}^p,$$

with  $\mathbf{z}^i$  being an optimal (or nearly optimal) solution of the semidefinite relaxation (3.2).

Hence we are now faced with a density estimation problem, that is: Given the sequence of moments  $\gamma_\beta = \int_{\mathbf{Y}} \mathbf{y}^\beta g_k(\mathbf{y}) d\varphi$ ,  $\beta \in \mathbb{N}^p$ , of the unknown nonnegative measurable function  $g_k$  on  $\mathbf{Y}$ , "estimate"  $g_k$ . One possibility is the so-called *maximum entropy* approach, briefly described in the next section.

**Maximum-entropy estimation.** We briefly describe the *maximum entropy* estimation technique in the univariate case. The multivariate case generalizes easily. Let  $g \in L_1([0, 1])^2$  be a nonnegative function only known via the first  $2d + 1$  moments  $\mathbf{u} = (u_j)_{j=0}^{2d}$  of its associated measure  $d\varphi = g d\mathbf{x}$  on  $[0, 1]$ . (In the context of previous section, the function  $g$  to estimate is  $\mathbf{y} \mapsto g_k(\mathbf{y})$  in (3.7) from the sequence  $\mathbf{u}$  in (3.8) of its (multivariate) moments.)

From that partial knowledge one wishes (a) to provide an estimate  $h_d$  of  $g$  such that the first  $2d + 1$  moments of the measure  $h_d d\mathbf{x}$  match those of  $g d\mathbf{x}$ , and (b) analyze the asymptotic behavior of  $h_d$  when  $d \rightarrow \infty$ . This problem has important applications in various areas of physics, engineering, and signal processing in particular.

An elegant methodology is to search for  $h_d$  in a (finitely) parametrized family  $\{h_d(\lambda, x)\}$  of functions, and optimize over the unknown parameters  $\lambda$  via a suitable criterion. For instance, one may wish to select an estimate  $h_d$  that maximizes some appropriate *entropy*. Several choices of entropy functional are possible as long as one obtains a convex optimization problem in the finitely many coefficients  $\lambda_i$ 's. For more details the interested reader is referred to e.g. Borwein and Lewis [6, 7] and the many references therein.

We here choose the Boltzmann-Shannon entropy  $\mathcal{H} : L_1([0, 1]) \rightarrow \mathbb{R} \cup \{-\infty\}$ :

$$(3.9) \quad h \mapsto \mathcal{H}[h] := - \int_0^1 h(x) \ln h(x) dx,$$

a strictly concave functional. Therefore, the problem reduces to:

$$(3.10) \quad \sup_h \left\{ \mathcal{H}[h] : \int_0^1 x^j h(x) dx = u_j, \quad j = 0, \dots, 2d \right\}.$$

The structure of this infinite-dimensional convex optimization problem permits to search for an optimal solution  $h_d^*$  of the form:

$$(3.11) \quad x \mapsto h_d^*(x) = \exp \sum_{j=0}^{2d} \lambda_j^* x^j,$$

and so  $\lambda^*$  is an optimal solution of the finite-dimensional unconstrained convex problem

$$\theta(\mathbf{u}) := \sup_{\lambda} \langle \mathbf{u}, \lambda \rangle - \int_0^1 \exp \left( \sum_{j=0}^{2d} \lambda_j x^j \right) dx.$$

Notice that the above function  $\theta$  is just the Legendre-Fenchel transform of the convex function  $\lambda \mapsto \int_0^1 \exp \sum_{j=0}^{2d} \lambda_j x^j dx$ .

---

<sup>2</sup> $L_1([0, 1])$  denote the Banach space of integrable functions on the interval  $[0, 1]$  of the real line, equipped with the norm  $\|g\|_1 = \int_0^1 |b(\mathbf{x})| d\mathbf{x}$ .

An optimal solution can be calculated by applying first-order methods, in which case the gradient  $\nabla v_d$  of the function

$$\lambda \mapsto v_d(\lambda) := \langle \mathbf{u}, \lambda \rangle - \int_0^1 \exp \left( \sum_{j=0}^{2d} \lambda_j x^j \right) dx,$$

is provided by:

$$\frac{\partial v_d(\lambda)}{\partial \lambda_k} = u_k - \int_0^1 x^k \exp \left( \sum_{j=0}^{2d} \lambda_j x^j \right) dx, \quad k = 0, \dots, 2d + 1.$$

If one applies second-order methods, e.g. Newton's method, then computing the Hessian  $\nabla^2 v_d$  at current iterate  $\lambda$ , reduces to computing

$$\frac{\partial^2 v_d(\lambda)}{\partial \lambda_k \partial \lambda_j} = - \int_0^1 x^{k+j} \exp \left( \sum_{j=0}^{2d} \lambda_j x^j \right) dx, \quad k, j = 0, \dots, 2d + 1.$$

In such simple cases like a box  $[a, b]$  (or  $[a, b]^n$  in the multivariate case) such quantities can be approximated quite accurately via cubature formula as described in e.g. [9]. In particular, several cubature formula behave very well for exponentials of polynomials as shown in e.g. Bender et al. [5]. An alternative with no cubature formula is also proposed in [15].

One has the following convergence result which follows directly from [6, Theor. 1.7 and p. 259].

**Proposition 3.8.** *Let  $0 \leq g \in L_1([0, 1])$  and for every  $d \in \mathbb{N}$ , let  $h_d^*$  in (3.11) be an optimal solution of (3.10). Then, as  $d \rightarrow \infty$ ,*

$$\int_0^1 \psi(\mathbf{y}) (h_d^*(\mathbf{y}) - g(\mathbf{y})) dx \rightarrow 0,$$

for every bounded measurable function  $\psi : [0, 1] \rightarrow \mathbb{R}$  which is continuous almost everywhere

Hence, the max-entropy estimate we obtain is not a pointwise estimate of  $g$ , and so, at some points of  $[0, 1]$  the max-entropy density  $h_d^*$  and the density  $g$  to estimate may differ significantly. However, for sufficiently large  $d$ , both curves of  $h_d^*$  and  $g$  are close to each other. In our context, recall that  $g$  is for instance  $y \mapsto x_k^*(y)$ , and so in general, for fixed  $y$ ,  $h_d^*(y)$  is close to  $x_k^*(y)$  and might be chosen for the  $k$ -coordinate of an initial point  $\mathbf{x}$ , input of a local minimization algorithm to find the global minimizer  $\mathbf{x}^*(y)$ .

**3.6. Illustrative examples.** In this section we provide some simple illustrative examples. To show the potential of the approach we have voluntarily chosen very simple examples for which one knows the solutions exactly so as to compare the results we obtain with the exact optimal value and optimal solutions. The semidefinite relaxations (3.2) were implemented by using the software package Gloptipoly [10]. The max-entropy estimate  $h_d^*$  of  $g_k$  was computed by using Newton's method, where at each iterate  $(\lambda^{(k)}, h_d(\lambda^{(k)}))$ :

$$\lambda^{(k+1)} = \lambda^{(k)} - (\nabla^2 v_d(\lambda^{(k)}))^{-1} \nabla v_d(\lambda^{(k)}).$$

**Example 1.** For illustration purpose, consider the toy example where  $\mathbf{Y} := [0, 1]$ ,

$$\mathbf{K} := \{(x, y) : 1 - x^2 + y^2 \geq 0; x, y \in \mathbf{Y}\} \subset \mathbb{R}^2, \quad (x, y) \mapsto f(x, y) := -x^2y.$$

Hence for each value of the parameter  $y \in \mathbf{Y}$ , the unique optimal solution is  $x^*(y) := \sqrt{1 - y^2}$ . And so in Theorem 3.3(b),  $y \mapsto g(y) = \sqrt{1 - y^2}$ .

Let  $\varphi$  be the probability measure uniformly distributed on  $[0, 1]$ . Therefore,

$$\rho = \int_0^1 J(y) d\varphi(y) = - \int_0^1 y(1 - y^2) dy = -1/4.$$

Solving (3.2) with  $i := 3$ , that is, with moments up to order 6, one obtains the optimal value  $-0.250146$ . Solving (3.2) with  $i := 4$ , one obtains the optimal value  $-0.25001786$  and the moment sequence

$$\mathbf{z} = (1, 0.7812, 0.5, 0.6604, 0.3334, 0.3333, 0.5813, 0.25, 0.1964, 0.25, 0.5244, 0.2, 0.1333, \\ 0.1334, 0.2, 0.4810, 0.1667, 0.0981, 0.0833, 0.0983, 0.1667)$$

Observe that

$$z_{1k} - \int_0^1 y^k \sqrt{1 - y^2} dy \approx O(10^{-6}), \quad k = 0, \dots, 4, \\ z_{1k} - \int_0^1 y^k \sqrt{1 - y^2} dy \approx O(10^{-5}), \quad k = 5, 6, 7.$$

Using a max-entropy approach to approximate the density  $y \mapsto g(y)$  on  $[0, 1]$ , with the first 5 moments  $z_{1k}$ ,  $k = 0, \dots, 4$ , we find that the optimal function  $h_4^*$  in (3.11) is obtained with

$$\lambda^* = (-0.1564, 2.5316, -12.2194, 20.3835, -12.1867).$$

Both curves of  $g$  and  $h_4^*$  are displayed in Figure 1. Observe that with only 5 moments, the max-entropy solution  $h_4^*$  approximates  $g$  relatively well, even if it differs significantly at some points. Indeed, the shape of  $h_4^*$  resembles very much that of  $g$ .

Finally, from an optimal solution of (3.5) one obtains for  $p \in \mathbb{R}[y]$ , the degree-8 univariate polynomial

$$y \mapsto p(y) = -0.0004 - 0.9909y - 0.0876y^2 + 1.4364y^3 - 1.2481y^4 \\ + 2.1261y^5 - 2.1309y^6 + 1.1593y^7 - 0.2641y^8$$

and Figure 2 displays the curve  $y \mapsto J(y) - p(y)$  on  $[0, 1]$ . One observes that  $J \geq p$  and the maximum difference is about  $3.10^{-4}$  close to 0 and much less for  $y \geq 0.1$ , a good precision with only 8 moments.

**Example 2.** Again with  $\mathbf{Y} := [0, 1]$ , let

$$\mathbf{K} := \{(\mathbf{x}, y) : 1 - x_1^2 - x_2^2 \geq 0\} \subset \mathbb{R}^2, \quad (\mathbf{x}, y) \mapsto f(\mathbf{x}, y) := yx_1 + (1 - y)x_2.$$

For each value of the parameter  $y \in \mathbf{Y}$ , the unique optimal solution  $\mathbf{x}^* \in \mathbf{K}$  satisfies

$$(x_1^*(y))^2 + (x_2^*(y))^2 = 1; \quad (x_1^*(y))^2 = \frac{y^2}{y^2 + (1 - y)^2}, \quad (x_2^*(y))^2 = \frac{(1 - y)^2}{y^2 + (1 - y)^2},$$

with optimal value

$$J(y) = -\frac{y^2}{\sqrt{y^2 + (1 - y)^2}} - \frac{(1 - y)^2}{\sqrt{y^2 + (1 - y)^2}} = -\sqrt{y^2 + (1 - y)^2}.$$



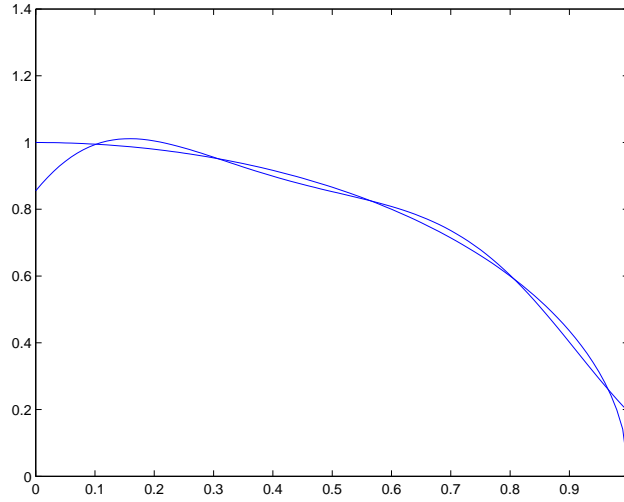


FIGURE 1. Example 1:  $g(y) = \sqrt{1 - y^2}$  versus  $h_4^*(y)$

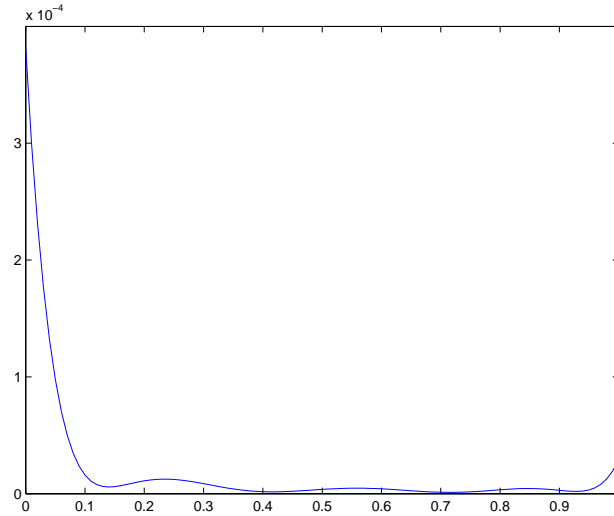


FIGURE 2. Example 1:  $J(y) - p(y)$  on  $[0, 1]$

So in Theorem 3.3(b),

$$y \mapsto g_1(y) = \frac{-y}{\sqrt{y^2 + (1-y)^2}}, \quad y \mapsto g_2(y) = \frac{y-1}{\sqrt{y^2 + (1-y)^2}},$$

and with  $\varphi$  being the probability measure uniformly distributed on  $[0, 1]$ ,

$$\rho = \int_0^1 J(y) d\varphi(y) = - \int_0^1 \sqrt{y^2 + (1-y)^2} dy \approx -0.81162$$

Solving (3.2) with  $i := 3$ , that is, with moments up to order 6, one obtains  $\rho_3 \approx -0.8117$  with  $\rho_3 - \rho \approx O(10^{-5})$ . Solving (3.2) with  $i := 4$ , one obtains  $\rho_4 \approx -0.81162$  with  $\rho_4 - \rho \approx O(10^{-6})$ , and the moment sequence  $(z_{k10})$ ,  $k = 0, 1, 2, 3, 4$ :

$$z_{k10} = (-0.6232, -0.4058, -0.2971, -0.2328, -0.1907),$$

and

$$z_{k10} - \int_0^1 y^k g_1(y) dy \approx O(10^{-5}), \quad k = 0, \dots, 4.$$

Using a max-entropy approach to approximate the density  $y \mapsto -g_1(y)$  on  $[0, 1]$ , with the first 5 moments  $z_{1k}$ ,  $k = 0, \dots, 4$ , we find that the optimal function  $h_4^*$  in (3.11) is obtained with

$$\lambda^* = (-3.61284, 15.66153266 - 29.43090127.326347 - 9.9884452).$$

and we find that

$$z_{k10} + \int_0^1 y^k h_4^*(y) dy \approx O(10^{-11}), \quad k = 0, \dots, 4.$$

In Figure 3 are displayed the two functions  $-g_1$  and  $h_4^*$ , and one observes a very good concordance.

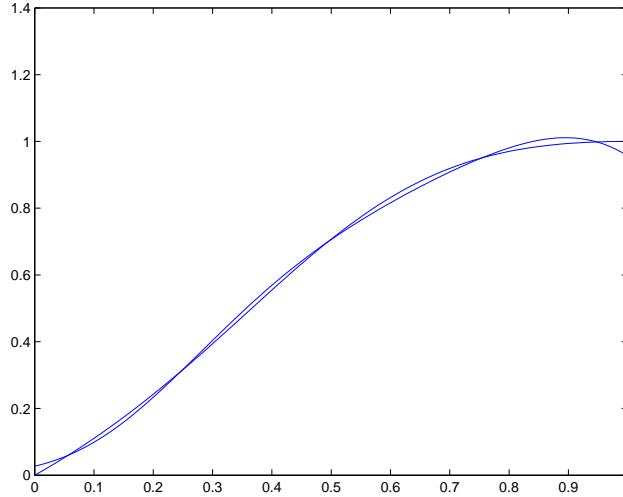


FIGURE 3. Example 2:  $h_4^*(y)$  versus  $-g_1(y) = y/\sqrt{y^2 + (1-y)^2}$

Finally, from an optimal solution of (3.5) one obtains for  $p \in \mathbb{R}[y]$ , the degree-8 univariate polynomial

$$\begin{aligned} x \mapsto p(y) := & -1.0000 + 0.9983y - 0.4537y^2 - 0.9941y^3 + 2.2488y^4 - 7.6739y^5 \\ & + 11.8448y^6 - 7.9606y^7 + 1.9903y^8 \end{aligned}$$

and Figure 4 displays the curve  $y \mapsto J(y) - p(y)$  on  $[0, 1]$ . One observes that  $J \geq p$  and the maximum difference is about  $10^{-4}$ , a good precision with only 8 moments.

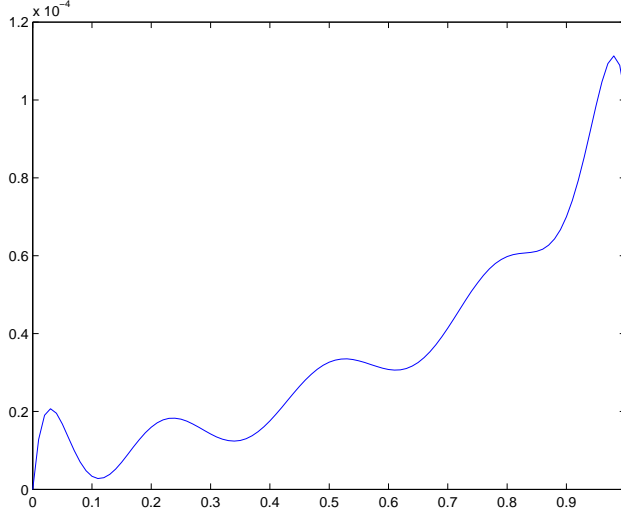


FIGURE 4. Example 2:  $J(y) - p(y)$  on  $[0, 1]$

**Example 3.** In this example one has  $\mathbf{Y} = [0, 1]$ ,  $(\mathbf{x}, y) \mapsto f(\mathbf{x}, y) := yx_1 + (1-y)x_2$ , and

$$\mathbf{K} := \{(\mathbf{x}, y) : yx_1^2 + x_2^2 - y \leq 0; x_1^2 + yx_2^2 - y \leq 0\}.$$

That is, for each  $y \in \mathbf{Y}$  the set  $\mathbf{K}_y$  is the intersection of two ellipsoids. It is easy to check that  $1 + x_i^*(y) \geq 0$  for all  $y \in \mathbf{Y}$ ,  $i := 1, 2$ . With  $i = 4$  the max-entropy estimate  $y \mapsto h_4^*(y)$  for  $1 + x_1^*(y)$  is obtained with

$$\lambda^* = (-0.2894, 1.7192, -19.8381, 36.8285, -18.4828),$$

whereas the max-entropy estimate  $y \mapsto h_4^*(y)$  for  $1 + x_2^*(y)$  is obtained with

$$\lambda^* = (-0.1018, -3.0928, 4.4068, 1.7096, -7.5782).$$

Figure 5 displays the curves of  $x_1^*(y)$  and  $x_2^*(y)$ , as well as the constraint  $h_1(\mathbf{x}^*(y), y)$ . Observe that  $h_1(\mathbf{x}^*(y), y) \approx 0$  on  $[0, 1]$  which means that for almost all  $y \in [0, 1]$ , at an optimal solution  $\mathbf{x}^*(y)$ , the constraint  $h_1 \leq 0$  is saturated. Figure 6 displays the curves of  $h_1(\mathbf{x}^*(y), y)$  and  $h_2(\mathbf{x}^*(y), y)$ .

**Example 4.** This time  $\mathbf{Y} = [0, 1]$ ,  $(\mathbf{x}, y) \mapsto f(\mathbf{x}, y) := (1 - 2y)(x_1 + x_2)$ , and

$$\mathbf{K} := \{(\mathbf{x}, y) : yx_1^2 + x_2^2 - y = 0; x_1^2 + yx_2^2 - y = 0\}.$$

That is, for each  $y \in \mathbf{Y}$  the set  $\mathbf{K}_y$  is the intersection of two ellipses, and

$$\mathbf{x} = \left( \pm \sqrt{\frac{y}{1+y}}, \pm \sqrt{\frac{y}{1+y}} \right); \quad J(y) = -2|1 - 2y| \sqrt{\frac{y}{1+y}}.$$

With  $i = 4$  the max-entropy estimate  $y \mapsto h_4^*(y)$  for  $1 + x_1^*(y)$  is obtained with

$$\lambda^* = (0.3071151, -12.51867, 43.215907, -46.985733, 16.395944).$$

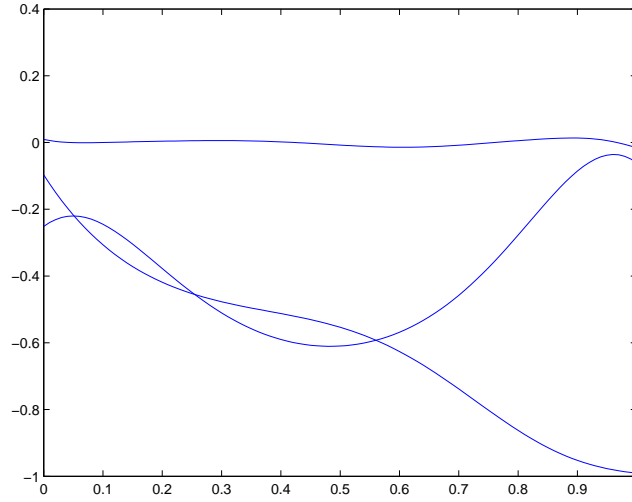


FIGURE 5. Example 3:  $x_1^*(y)$ ,  $x_2^*(y)$  and  $h_1(\mathbf{x}^*(y), y)$  on  $[0, 1]$

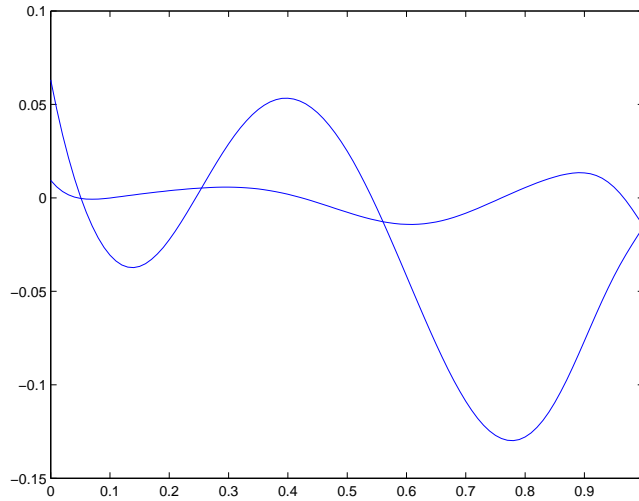
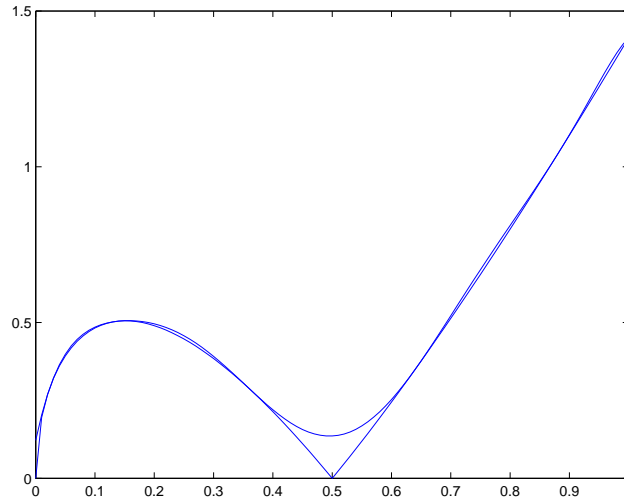
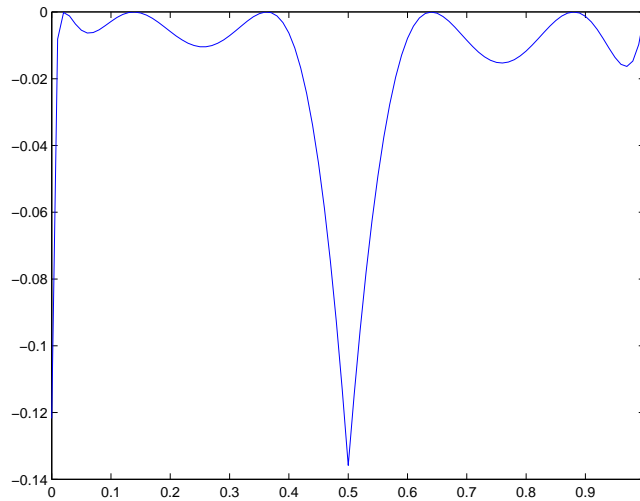


FIGURE 6. Example 3:  $h_1(\mathbf{x}^*(y), y)$  and  $h_2(\mathbf{x}^*(y), y)$  on  $[0, 1]$

In Figure 7 are displayed the curves  $y \mapsto -p(y)$  and  $y \mapsto -J(y)$ , whereas in Figure 8 is displayed the curve  $y \mapsto p(y) - J(y)$ . One may see that  $p$  is a good lower approximation of  $J$  even with only 8 moments.

FIGURE 7. Example 4:  $-p(y)$  and  $-J(y)$  on  $[0, 1]$ FIGURE 8. Example 4: the curve  $p(y) - J(y)$  on  $[0, 1]$ 

On the other hand, in Figure 9 is displayed  $h_4^*(y)$  versus  $x_1^*(y)$  where the latter is  $-\sqrt{y/(1+y)}$  on  $[0, 1/2]$  and  $\sqrt{y/(1+y)}$  on  $[1/2, 1]$ . Here we see that the discontinuity of  $x_1^*(y)$  is difficult to approximate "pointwise" with few moments, and

despite a very good precision on the five first moments. Indeed:

$$\int_0^1 y^k (h_4^*(y) - 1) dx - \int_0^1 y^k x_1^*(y) dx = O(10^{-14}), \quad k = 0, \dots, 4.$$

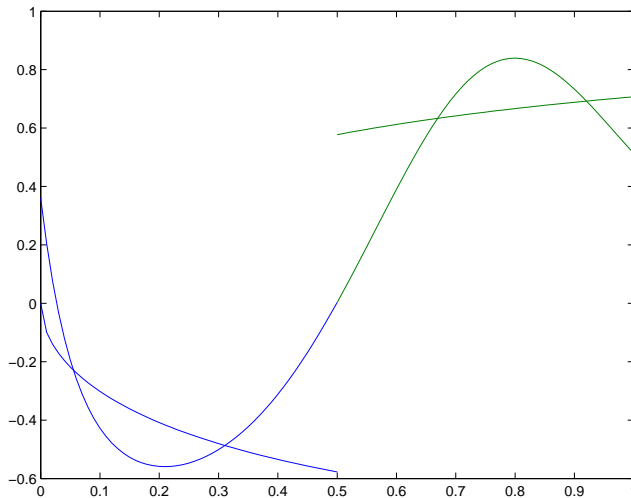


FIGURE 9. Example 4:  $h_4^*(y) - 1$  and  $x_1^*(y)$  on  $[0, 1]$

We end up this section with the case where the density  $g_k$  to estimate is a step function which would be the case in an optimization problem  $\mathbf{P}_y$  with boolean variables (e.g. the variable  $x_k$  takes values in  $\{0, 1\}$ ).

**Example 5.** Assume that with a single parameter  $y \in [0, 1]$ , the density  $g_k$  to estimate is the step function.

$$y \mapsto g_k(y) := \begin{cases} 1 & \text{if } y \in [0, 1/3] \cup [2/3, 1] \\ 0 & \text{otherwise.} \end{cases}$$

The max-entropy estimate  $h_4^*$  in (3.11) with 5 moments is obtained with

$$\lambda^* = [-0.6547367219.170724 - 115.39354192.4493171655 - 96.226948865],$$

and we have

$$\int_0^1 y^k h_4^*(y) dy - \int_0^1 y^k dg_k(y) \approx O(10^{-8}), \quad k = 0, \dots, 4.$$

In particular, the persistency  $\int_0^1 g_k(y) dy = 2/3$  of the variable  $x_k^*(y)$ , is very well approximated (up to  $10^{-8}$  precision) by  $\int h_4^*(y) dy$ , with only 5 moments.

Of course, in this case and with only 5 moments, the density  $h_4^*$  is not a good pointwise approximation of the step function  $g_k$ ; however its "shape" reveals the two steps of value 1 separated by a step of value 0. A better pointwise approximation would require more moments.

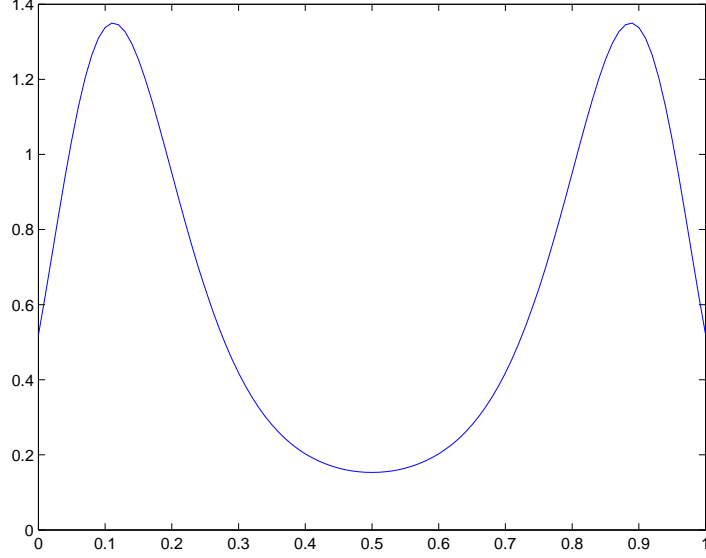


FIGURE 10. Example 5:  $g_k(y) = 1_{[0,1/3] \cup [2/3,1]}$  versus  $h_4^*(y)$

#### 4. APPENDIX

##### Proof of Theorem 3.3.

We already know that  $\rho_i \leq \rho$  for all  $i \geq i_0$ . We also need to prove that  $\rho_i > -\infty$  for sufficiently large  $i$ . Let  $Q \subset \mathbb{R}[\mathbf{x}, \mathbf{y}]$  be the *quadratic module* generated by the polynomials  $\{h_j\} \subset \mathbb{R}[\mathbf{x}, \mathbf{y}]$  that define  $\mathbf{K}$ , i.e.,

$$Q := \{ \sigma \in \mathbb{R}[\mathbf{x}, \mathbf{y}] : \sigma = \sigma_0 + \sum_{j=1}^t \sigma_j h_j \text{ with } \{\sigma_j\}_{j=0}^t \subset \Sigma[\mathbf{x}, \mathbf{y}] \}.$$

In addition, let  $Q(l) \subset Q$  be the set of elements  $\sigma \in Q$  which have a representation  $\sigma_0 + \sum_{j=0}^t \sigma_j h_j$  for some s.o.s. family  $\{\sigma_j\} \subset \Sigma^2$  with  $\deg \sigma_0 \leq 2l$  and  $\deg \sigma_j h_j \leq 2l$  for all  $j = 1, \dots, t$ .

Let  $i \in \mathbb{N}$  be fixed. As  $\mathbf{K}$  is compact, there exists  $N$  such that  $N \pm \mathbf{x}^\alpha \mathbf{y}^\beta > 0$  on  $\mathbf{K}$ , for all  $\alpha \in \mathbb{N}^n$  and  $\beta \in \mathbb{N}^p$ , with  $|\alpha + \beta| \leq 2i$ . Therefore, under Assumption 3.1(ii), the polynomial  $N \pm \mathbf{x}^\alpha \mathbf{y}^\beta$  belongs to  $Q$ ; see Putinar [18]. But there is even some  $l(i)$  such that  $N \pm \mathbf{x}^\alpha \mathbf{y}^\beta \in Q(l(i))$  for every  $|\alpha + \beta| \leq 2i$ . Of course we also have  $N \pm \mathbf{x}^\alpha \mathbf{y}^\beta \in Q(l)$  for every  $|\alpha + \beta| \leq 2i$ , whenever  $l \geq l(i)$ . Therefore, let us take  $l(i) \geq i_0$ . For every feasible solution  $\mathbf{z}$  of  $\mathbf{Q}_{l(i)}$  one has

$$|z_{\alpha\beta}| = |L_{\mathbf{z}}(\mathbf{x}^\alpha \mathbf{y}^\beta)| \leq N, \quad \forall |\alpha + \beta| \leq 2i.$$

This follows from  $z_0 = 1$ ,  $\mathbf{M}_{l(i)}(\mathbf{z}) \succeq 0$  and  $\mathbf{M}_{l(i)-v_j}(h_j \mathbf{z}) \succeq 0$ , which implies

$$N z_0 \pm z_{\alpha\beta} = L_{\mathbf{z}}(N \pm \mathbf{x}^\alpha \mathbf{y}^\beta) = L_{\mathbf{z}}(\sigma_0) + \sum_{j=1}^t L_{\mathbf{z}}(\sigma_j h_j) \geq 0$$

for some  $\{\sigma_j\} \subset \Sigma[\mathbf{x}, \mathbf{y}]$  with  $\deg \sigma_j h_j \leq 2l(i)$ . In particular,  $L_{\mathbf{z}}(f) \geq -N \sum_{\alpha, \beta} |f_{\alpha\beta}|$ , which proves that  $\rho_{l(i)} > -\infty$ , and so  $\rho_i > -\infty$  for all sufficiently large  $i$ .

From what precedes, and with  $k \in \mathbb{N}$  arbitrary, let  $l(k) \geq k$  and  $N_k$  be such that

$$(4.1) \quad N_k \pm \mathbf{x}^\alpha \mathbf{y}^\beta \in Q(l(k)) \quad \forall \alpha \in \mathbb{N}^n, \beta \in \mathbb{N}^p \text{ with } |\alpha + \beta| \leq 2k.$$

Let  $i \geq l(i_0)$ , and let  $\mathbf{z}^i$  be a nearly optimal solution of (3.2) with value

$$(4.2) \quad \rho_i \leq L_{\mathbf{z}^i}(f) \leq \rho_i + \frac{1}{i} \quad \left( \leq \rho + \frac{1}{i} \right).$$

Fix  $k \in \mathbb{N}$ . Notice that from (4.1), for every  $i \geq l(k)$ , one has

$$|L_{\mathbf{z}^i}(\mathbf{x}^\alpha \mathbf{y}^\beta)| \leq N_k z_0 = N_k, \quad \forall \alpha \in \mathbb{N}^n, \beta \in \mathbb{N}^p \text{ with } |\alpha + \beta| \leq 2k.$$

Therefore, for all  $i \geq l(i_0)$ ,

$$(4.3) \quad |z_{\alpha\beta}^i| = |L_{\mathbf{z}^i}(\mathbf{x}^\alpha \mathbf{y}^\beta)| \leq N'_k, \quad \forall \alpha \in \mathbb{N}^n, \beta \in \mathbb{N}^p \text{ with } |\alpha + \beta| \leq 2k,$$

where  $N'_k = \max[N_k, V_k]$ , with

$$V_k := \max_{\alpha, \beta, i} \{ |z_{\alpha\beta}^i| : |\alpha + \beta| \leq 2k; \quad l(i_0) \leq i \leq l(k) \}.$$

Complete each vector  $\mathbf{z}^i$  with zeros to make it an infinite bounded sequence in  $l_\infty$ , indexed in the canonical basis  $(\mathbf{x}^\alpha \mathbf{y}^\beta)$  of  $\mathbb{R}[\mathbf{x}, \mathbf{y}]$ . In view of (4.3),

$$(4.4) \quad |z_{\alpha\beta}^i| \leq N'_k \quad \forall \alpha \in \mathbb{N}^n, \beta \in \mathbb{N}^p \text{ with } 2k-1 \leq |\alpha + \beta| \leq 2k,$$

and for all  $k = 1, 2, \dots$

Hence, let  $\widehat{\mathbf{z}}^i \in l_\infty$  be the new sequence defined by

$$\widehat{z}_{\alpha\beta}^i := \frac{z_{\alpha\beta}^i}{N'_k}, \quad \forall \alpha \in \mathbb{N}^n, \beta \in \mathbb{N}^p \text{ with } 2k-1 \leq |\alpha + \beta| \leq 2k, \quad \forall k = 1, 2, \dots,$$

and in  $l_\infty$ , consider the sequence  $\{\widehat{\mathbf{z}}^i\}_i$ , as  $i \rightarrow \infty$ .

Obviously, the sequence  $\{\widehat{\mathbf{z}}^i\}_i$  is in the unit ball  $B_1$  of  $l_\infty$ , and so, by the Banach-Alaoglu theorem (see e.g. Ash [2]), there exists  $\widehat{\mathbf{z}} \in B_1$ , and a subsequence  $\{i_l\}$ , such that  $\widehat{\mathbf{z}}^{i_l} \rightarrow \widehat{\mathbf{z}}$  as  $l \rightarrow \infty$ , for the weak  $\star$  topology  $\sigma(l_\infty, l_1)$  of  $l_\infty$ . In particular, pointwise convergence holds, that is,

$$\lim_{l \rightarrow \infty} \widehat{z}_{\alpha\beta}^{i_l} \rightarrow \widehat{z}_{\alpha\beta} \quad \forall \alpha \in \mathbb{N}^n, \beta \in \mathbb{N}^p.$$

Next, define

$$z_{\alpha\beta} := \widehat{z}_{\alpha\beta} \times N'_k \quad \forall \alpha \in \mathbb{N}^n, \beta \in \mathbb{N}^p \text{ with } 2k-1 \leq |\alpha + \beta| \leq 2k, \quad \forall k = 1, 2, \dots$$

The pointwise convergence  $\widehat{\mathbf{z}}^{i_l} \rightarrow \widehat{\mathbf{y}}$  implies the pointwise convergence  $\mathbf{z}^{i_l} \rightarrow \mathbf{z}$ , i.e.,

$$(4.5) \quad \lim_{l \rightarrow \infty} z_{\alpha\beta}^{i_l} \rightarrow z_{\alpha\beta} \quad \forall \alpha \in \mathbb{N}^n, \beta \in \mathbb{N}^p.$$

Next, let  $s \in \mathbb{N}$  be fixed. From the pointwise convergence (4.5) we deduce that

$$\lim_{l \rightarrow \infty} \mathbf{M}_s(z^{i_l}) = \mathbf{M}_s(\mathbf{z}) \succeq 0.$$

Similarly

$$\lim_{l \rightarrow \infty} \mathbf{M}_s(h_j \mathbf{z}^{i_l}) = \mathbf{M}_s(h_j \mathbf{z}) \succeq 0, \quad j = 1, \dots, t.$$

As  $s$  was arbitrary, we obtain

$$(4.6) \quad \mathbf{M}_s(\mathbf{y}) \succeq 0; \quad \mathbf{M}_s(h_j \mathbf{z}) \succeq 0, \quad j = 1, \dots, t; \quad s = 0, 1, 2, \dots,$$



which by Theorem 3.2 implies that  $\mathbf{z}$  is the sequence of moments of some finite measure  $\mu^*$  with support contained in  $\mathbf{K}$ . Moreover, the pointwise convergence (4.5) also implies that

$$(4.7) \quad \int_{\mathbf{Y}} \mathbf{y}^\beta d\varphi(\mathbf{y}) = \gamma_\beta = \lim_{l \rightarrow \infty} z_{0\beta}^{i_l} = z_{0\beta} = \int_{\mathbf{K}} \mathbf{y}^\beta d\mu^*, \quad \forall \beta \in \mathbb{N}^p.$$

As measures on compact sets are determinate, (4.7) implies that the marginal of  $\mu^*$  on  $\mathbb{R}^p$  is the probability measure  $\varphi$ , and so  $\mu^*$  is feasible for  $\mathbf{P}$ . Finally, combining the pointwise convergence (4.5) with (4.2) yields

$$\rho \geq \lim_{l \rightarrow \infty} \rho_{i_l} = \lim_{i \rightarrow \infty} L_{\mathbf{z}^{i_l}}(f) = L_{\mathbf{z}}(f) = \int_{\mathbf{K}} f d\mu^*,$$

which in turn yields that  $\mu^*$  is an optimal solution of  $\mathbf{P}$ . And so  $\rho_{i_l} \rightarrow \rho$  as  $l \rightarrow \infty$ . As the sequence  $(\rho_i)$  is monotone this yields the desired result (a).

(b) Next, let  $\alpha \in \mathbb{N}^n$  and  $\beta \in \mathbb{N}^p$  be fixed, arbitrary. From (4.5), we have:

$$\lim_{l \rightarrow \infty} z_{\alpha\beta}^{i_l} = z_{\alpha\beta} = \int_{\mathbf{K}} \mathbf{x}^\alpha \mathbf{y}^\beta d\mu^*,$$

and by Theorem 2.2(c)

$$\lim_{l \rightarrow \infty} z_{\alpha\beta}^{i_l} = \int_{\mathbf{K}} \mathbf{x}^\alpha \mathbf{y}^\beta d\mu^* = \int_{\mathbf{Y}} \mathbf{y}^\beta g(\mathbf{y})^\alpha d\varphi(\mathbf{y}),$$

and as the converging subsequence was arbitrary, the above convergence holds for the whole sequence  $(z_{\alpha\beta}^i)$ .  $\square$

## REFERENCES

- [1] E.J Anderson and P. Nash, *Linear Programming in Infinite-Dimensional Spaces*, John Wiley & Sons, Chichester (1987).
- [2] R. Ash, *Real Analysis and Probability*, Academic Press, San Diego (1972).
- [3] D. Bertsimas, K. Natarajan, and Chung-Piaw Teo, Persistence in discrete optimization under data uncertainty, *Math. Prog. Ser. B* **108** (2005), 251–274.
- [4] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.
- [5] C.M. Bender, L. R. Mead, and N. Papanicolaou, Maximum entropy summation of divergent perturbation series, *J. Math. Phys.* **28** (1987), 1016–1018.
- [6] J. Borwein and A.S. Lewis, On the convergence of moment problems, *Trans. Am. Math. Soc.* **325** (1991), 249–271.
- [7] J. Borwein and A.S. Lewis, Convergence of best entropy estimates, *SIAM J. Optim.* **1** (1991), 191–205.
- [8] E.G Dynkin and A.A. Yushkevich, *Controlled Markov Processes*, Springer-Verlag, New York (1979).
- [9] W. Gautschi, *Numerical Analysis: An Introduction*, Birkhäuser, Boston (1997).
- [10] D. Henrion, J. B. Lasserre and J. Lofberg, GloptiPoly 3: moments, optimization and semi-definite programming, *Optim. Methods and Softw.*, to appear.  
<http://www.laas.fr/~henrion/software/gloptipoly3/>
- [11] O. Hernández-Lerma and J. B. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer-Verlag, New York (1996).
- [12] O. Hernández-Lerma and J.B. Lasserre, Approximation schemes for infinite linear programs, *SIAM J. Optim.* **8** (1998), 973–988.
- [13] J.B. Lasserre, Global optimization with polynomials and the problem of moments, *SIAM J. Optim.* **11** (2001), 796–817.
- [14] J.B. Lasserre, Convergent SDP-relaxations in polynomial optimization with sparsity, *SIAM J. Optim.* **17** (2006), 822–843.

- [15] J.B. Lasserre, Semidefinite programming for gradient and Hessian computation in maximum entropy estimation, *Proceedings 48th IEEE CDC Conference*, New-Orleans (2007), pp. 3060–3064.
- [16] H.M. Möller and H.J. Stetter, Multivariate polynomial equations with multiple zeros solved by matrix eigenproblems, *Num. Math.* **70** (1995), 311–329.
- [17] K. Natarajan, Miao Song and Chung-Piaw Teo, Persistency and its applications in choice modelling, *Mang. Sci.* **55** (2009), 453–469.
- [18] M. Putinar, Positive polynomials on compact semi-algebraic sets, *Indiana Univ. Math. J.* **42** (1993), 969–984.
- [19] P. Rostalski, Algebraic Moments: Real root finding and related topics, PhD Thesis, Automatic Control Laboratory, ETH Zurich, Switzerland, May 2009.
- [20] M. Schäl, Conditions for optimality and for the limit of  $n$ -stage optimal policies to be optimal, *Z. Wahrs. verw. Geb.* **32** (1975), 179–196.
- [21] M. Schweighofer, Optimization of polynomials on compact semialgebraic sets, *SIAM J. Optim.* **15** (2005), 805–825.
- [22] H.J. Stetter, *Numerical Polynomial Algebra*, SIAM, Philadelphia (2004).
- [23] A. Tagliani, Entropy estimate of probability densities having assigned moments: Hausdorff case, *Appl. Math. Lett.* **15** (2002), 309–314.
- [24] A. Tagliani, Entropy estimate of probability densities having assigned moments: Stieltjes case, *Appl. Math. Comput.* **130** (2002), 201–211.
- [25] V. Weispfenning, Comprehensive Gröbner bases, *J. Symb. Comp.* **14** (1992), 1–29.

LAAS-CNRS AND INSTITUTE OF MATHEMATICS, UNIVERSITY OF TOULOUSE, LAAS, 7 AVENUE  
DU COLONEL ROCHE, 31077 TOULOUSE CÉDEX 4, FRANCE  
*E-mail address:* `lasserre@laas.fr`

