

# Semidefinite approximations of projections and polynomial images of semialgebraic sets

Victor Magron<sup>1</sup>

Didier Henrion<sup>2,3,4</sup>

Jean-Bernard Lasserre<sup>2,3</sup>

October 17, 2014

## Abstract

Given a compact semialgebraic set  $\mathbf{S} \subset \mathbb{R}^n$  and a polynomial map  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , we consider the problem of approximating the image set  $\mathbf{F} = f(\mathbf{S}) \subset \mathbb{R}^m$ . This includes in particular the projection of  $\mathbf{S}$  on  $\mathbb{R}^m$  for  $n \geq m$ . Assuming that  $\mathbf{F} \subset \mathbf{B}$ , with  $\mathbf{B} \subset \mathbb{R}^m$  being a “simple” set (e.g. a box or a ball), we provide two methods to compute certified outer approximations of  $\mathbf{F}$ . Method 1 exploits the fact that  $\mathbf{F}$  can be defined with an existential quantifier, while Method 2 computes approximations of the support of image measures. The two methods output a sequence of super-level sets defined with a single polynomial that yield explicit outer approximations of  $\mathbf{F}$ . Finding the coefficients of this polynomial boils down to computing an optimal solution of a convex semidefinite program. We provide guarantees of strong convergence to  $\mathbf{F}$  in  $L_1$  norm on  $\mathbf{B}$ , when the degree of the polynomial approximation tends to infinity. Several examples of applications are provided, together with numerical experiments.

**Keywords** Semialgebraic sets; semidefinite programming; moment relaxations; polynomial sum of squares.

## 1 Introduction

Consider a polynomial map  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $\mathbf{x} \mapsto f(\mathbf{x}) := (f_1(\mathbf{x}), \dots, f_m(\mathbf{x})) \in \mathbb{R}^m[\mathbf{x}]$  of degree  $d := \max\{\deg f_1, \dots, \deg f_m\}$  and a compact basic semialgebraic set

$$\mathbf{S} := \{\mathbf{x} \in \mathbb{R}^n : g_1^{\mathbf{S}}(\mathbf{x}) \geq 0, \dots, g_n^{\mathbf{S}}(\mathbf{x}) \geq 0\} \quad (1)$$

defined by polynomials  $g_1^{\mathbf{S}}, \dots, g_n^{\mathbf{S}} \in \mathbb{R}[\mathbf{x}]$ .

---

<sup>1</sup>Circuits and Systems Group, Department of Electrical and Electronic Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.

<sup>2</sup>CNRS; LAAS; 7 avenue du colonel Roche, F-31400 Toulouse; France.

<sup>3</sup>Université de Toulouse; LAAS, F-31400 Toulouse, France.

<sup>4</sup>Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, CZ-16626 Prague, Czech Republic

Since  $\mathbf{S}$  is compact, the image set

$$\mathbf{F} := f(\mathbf{S})$$

is included in a basic compact semialgebraic set  $\mathbf{B}$ , assumed to be “simple” (e.g. a box or a ball) and described by

$$\mathbf{B} := \{\mathbf{y} \in \mathbb{R}^m : g_1^{\mathbf{B}}(\mathbf{y}) \geq 0, \dots, g_{n_{\mathbf{B}}}^{\mathbf{B}}(\mathbf{y}) \geq 0\} \quad (2)$$

for some polynomials  $g_1^{\mathbf{B}}, \dots, g_{n_{\mathbf{B}}}^{\mathbf{B}} \in \mathbb{R}[\mathbf{y}]$ .

The purpose of this paper is to approximate  $\mathbf{F}$ , the image of  $\mathbf{S}$  under the polynomial map  $f$ , with superlevel sets of single polynomials of fixed degrees. One expects the approximation to be *tractable*, i.e. to be able to control the degree of the polynomials used to define the approximations. This appears to be quite a challenging problem since the polynomial map  $f$  and the set  $\mathbf{S}$  can be both complicated.

This problem includes two important special cases. The first problem is to approximate the projection of  $\mathbf{S}$  on  $\mathbb{R}^m$  for  $n \geq m$ . The second problem is the approximation of Pareto curves in the context of multicriteria optimization. In [MHL14], we reformulate this second problem through parametric polynomial optimization, which can be solved using a hierarchy of semidefinite approximations. The present work proposes an alternative solution via approximations of polynomial images of semialgebraic sets.

In the case of semialgebraic set projections, notice that computer algebra algorithms provide an exact description of the projection but with a very high computational cost even in small dimension. These algorithms are based on real quantifier elimination (see e.g. [Tar51, Col74, BCR98]). For state-of-the-art computer algebra algorithms for quantifier elimination, we refer the interested reader to the survey [Bas14] and the references therein. Quantifier elimination can be performed with the famous cylindrical algebraic decomposition algorithm. However, for a finite set of  $s$  polynomials in  $n$  variables, the (time) computational complexity of this algorithm is bounded by  $(sd)^{2^{O(n)}}$ , thus doubly exponential. In [GJ88] an algorithm was proposed with single exponential complexity in the case of bounded nonsingular algebraic sets. For applications that satisfy certain additional assumptions (e.g. radicality, equidimensionality, etc.) one can use the variant quantifier elimination method proposed in [HD12], which is less computationally demanding.

To avoid this high computational cost inherent to exact computations in computer algebra algorithms, we consider a less ambitious goal. On the one hand, we do not require an exact description of the projection but rather a hierarchy of outer approximations with a guarantee of convergence to the exact projection. On the other hand, the present methodology only requires moderate assumptions: 1) the set  $\mathbf{S}$  is compact and 2) either the semialgebraic set  $\mathbf{S}$  or  $\mathbf{F}$  (resp.  $\mathbf{B} \setminus \mathbf{F}$ ) has nonempty interior.

**Contribution and general methodology** We provide two methods to approximate the image of semialgebraic sets under polynomial applications.

- Method 1 consists of rewriting  $\mathbf{F}$  as a set defined with an existential quantifier. Then, one can outer approximate  $\mathbf{F}$  as closely as desired with a hierarchy of superlevel sets

of the form  $\mathbf{F}_r^1 := \{\mathbf{y} \in \mathbf{B} : q_r(\mathbf{y}) \geq 0\}$  for some polynomials  $q_r \in \mathbb{R}[\mathbf{y}]$  of increasing degrees  $2r$ .

- Method 2 consists of building a hierarchy of relaxations for the infinite dimensional moment problem whose optimal value is the volume of  $\mathbf{F}$  and whose optimum is the restriction of the Lebesgue measure on  $\mathbf{F}$ . Then, one can outer approximate  $\mathbf{F}$  as closely as desired with a hierarchy of super level sets of the form  $\mathbf{F}_r^2 := \{\mathbf{y} \in \mathbf{B} : w_r(\mathbf{y}) \geq 1\}$ , for some polynomials  $w_r \in \mathbb{R}[\mathbf{y}]$  of increasing degrees  $2r$ .

Method 1 and Method 2 share the following essential features:

1. The sets  $\mathbf{F}_r^1$  and  $\mathbf{F}_r^2$  are described with a single polynomial of degree  $2r$ .
2. Assuming non-emptiness of the interior of  $\mathbf{S}$ , resp. of  $\mathbf{F}$  and  $\mathbf{B} \setminus \mathbf{F}$ , one has  $\lim_{r \rightarrow \infty} \text{vol}(\mathbf{F}_r^1 \setminus \mathbf{F}) = 0$ , resp.  $\lim_{r \rightarrow \infty} \text{vol}(\mathbf{F}_r^2 \setminus \mathbf{F}) = 0$ , where  $\text{vol}(\cdot)$  stands for the volume or Lebesgue measure.
3. Computing the coefficient vectors of the polynomials  $(q_r)_{r \in \mathbb{N}}$ , resp.  $(w_r)_{r \in \mathbb{N}}$ , boils down to finding optimal solutions of a hierarchy of semidefinite programs. The size of these programs is parametrized by the relaxation order  $r$  and depends on the number of variables  $n$ , the number  $m$  of components of the polynomial  $f$  as well as its degree  $d$ . For the hierarchy of semidefinite programs associated with Method 1, the number of variables at step  $r$  is bounded by  $\binom{n+m+2r}{2r}$ , with  $(n^{\mathbf{S}} + n^{\mathbf{B}} + 1)$  semidefinite constraints of size at most  $\binom{n+m+r}{r}$ . Step  $r$  of the semidefinite hierarchy associated with Method 2 involves at most  $\binom{n+2rd}{2rd} + 2\binom{m+2r}{2r}$  variables,  $(n^{\mathbf{S}} + 1)$  semidefinite constraints of size at most  $\binom{n+rd}{rd}$  and  $2(n^{\mathbf{B}} + 1)$  semidefinite constraints of size at most  $\binom{m+r}{r}$ .
4. Data sparsity can be exploited to reduce the overall computational cost.

Method 1 relies on the previous study [Las13], in which the author obtains tractable approximations of sets defined with existential quantifiers. The present article provides an extension of the result of [Las13, Theorem 3.4], where one does not require anymore that some set has zero Lebesgue measure. This is mandatory to prove the volume convergence result.

In [HLS09], the authors consider the problem of approximating the volume of a general compact basic semialgebraic set. The initial problem is then reformulated as an infinite dimensional linear programming (LP) problem, whose unknown is the restriction of the Lebesgue measure on the set of interest. The main idea behind Method 2 is a similar infinite dimensional LP reformulation of the problem, whose unknown is  $\mu_1$ , the restriction of the Lebesgue measure on  $\mathbf{F}$ . One ends up in computing a finite number of moments of the measure  $\mu_0$  supported on  $\mathbf{S}$  such that the image of  $\mu_0$  under  $f$  is precisely  $\mu_1$ . Note however that there is an important novelty compared with [HLS09], in which the set under study is explicitly described as a basic compact semialgebraic set (i.e. the intersection of superlevel sets of known polynomials), whereas such a description is not known for  $\mathbf{F}$ .

**Structure of the paper** The paper is organized as follows. Section 2 recalls the basic background about polynomial sum of squares approximations, moment and localizing matrices. Section 3 presents our approximation method for existential quantifier elimination (Method 1). Section 4 is dedicated to the support of image measures (Method 2). In Section 5, we analyze the theoretical complexity of both methods and describe how the system sparsity can be exploited. Section 6 presents several examples where Method 1 and Method 2 are successfully applied.

## 2 Notation and Definitions

Let  $\mathbb{R}[\mathbf{x}]$  (resp.  $\mathbb{R}_{2r}[\mathbf{x}]$ ) be the ring of real polynomials (resp. of degree at most  $2r$ ) in the variable  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ , for  $r \in \mathbb{N}$ . With  $\mathbf{S}$  a basic semialgebraic set as in (1), we set  $r_j^{\mathbf{S}} := \lceil (\deg g_j^{\mathbf{S}})/2 \rceil, j = 1, \dots, n^{\mathbf{S}}$  and with  $\mathbf{B}$  a basic semialgebraic set as in (2), we set  $r_j^{\mathbf{B}} := \lceil (\deg g_j^{\mathbf{B}})/2 \rceil, j = 1, \dots, n^{\mathbf{B}}$ . Let  $\Sigma[\mathbf{x}]$  denote the cone of sum of squares (SOS) of polynomial, and let  $\Sigma_r[\mathbf{x}]$  denote the cone of polynomials SOS of degree at most  $2r$ , that is  $\Sigma_r[\mathbf{x}] := \Sigma[\mathbf{x}] \cap \mathbb{R}_{2r}[\mathbf{x}]$ .

For the ease of notation, we set  $g_0^{\mathbf{S}}(\mathbf{x}) := 1$  and  $g_0^{\mathbf{B}}(\mathbf{y}) := 1$ . For each  $r \in \mathbb{N}$ , let  $\mathbf{Q}_r(\mathbf{S})$  (resp.  $\mathbf{Q}_r(\mathbf{B})$ ) be the  $r$ -truncated quadratic module (a convex cone) generated by  $g_0^{\mathbf{S}}, \dots, g_{n^{\mathbf{S}}}^{\mathbf{S}}$  (resp.  $g_0^{\mathbf{B}}, \dots, g_{n^{\mathbf{B}}}^{\mathbf{B}}$ ):

$$\begin{aligned}\mathbf{Q}_r(\mathbf{S}) &:= \left\{ \sum_{j=0}^{n^{\mathbf{S}}} s_j(\mathbf{x}) g_j^{\mathbf{S}}(\mathbf{x}) : s_j \in \Sigma_{r-r_j^{\mathbf{S}}}[\mathbf{x}], j = 0, \dots, n^{\mathbf{S}} \right\}, \\ \mathbf{Q}_r(\mathbf{B}) &:= \left\{ \sum_{j=0}^{n^{\mathbf{B}}} s_j(\mathbf{y}) g_j^{\mathbf{B}}(\mathbf{y}) : s_j \in \Sigma_{r-r_j^{\mathbf{B}}}[\mathbf{y}], j = 0, \dots, n^{\mathbf{B}} \right\}.\end{aligned}$$

Now, we introduce additional notations which are required for Method 1. Let us first describe the product set

$$\mathbf{K} := \mathbf{S} \times \mathbf{B} = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{n+m} : g_1(\mathbf{x}, \mathbf{y}) \geq 0, \dots, g_{n^{\mathbf{K}}}(\mathbf{x}, \mathbf{y}) \geq 0\} \subset \mathbb{R}^{n+m}, \quad (3)$$

with  $n^{\mathbf{K}} := n^{\mathbf{S}} + n^{\mathbf{B}}$  and the polynomials  $g_j \in \mathbb{R}[\mathbf{x}, \mathbf{y}], j = 1, \dots, n^{\mathbf{K}}$  are defined by:

$$g_j(\mathbf{x}, \mathbf{y}) := \begin{cases} g_j^{\mathbf{S}}(\mathbf{x}) & \text{if } 1 \leq j \leq n^{\mathbf{S}}, \\ g_j^{\mathbf{B}}(\mathbf{y}) & \text{if } n^{\mathbf{S}} + 1 \leq j \leq n^{\mathbf{K}}. \end{cases}$$

As previously, we set  $r_j := \lceil (\deg g_j)/2 \rceil, j = 1, \dots, n^{\mathbf{K}}, g_0(\mathbf{x}, \mathbf{y}) := 1$  and  $\mathbf{Q}_r(\mathbf{K})$  stands for the  $r$ -truncated quadratic module generated by the polynomials  $g_0, \dots, g_{n^{\mathbf{K}}}$ .

To guarantee the theoretical convergence of our two methods, we need to assume the existence of the following algebraic certificates of boundedness of the sets  $\mathbf{S}$  and  $\mathbf{B}$ :

**Assumption 2.1.** There exists an integer  $j^{\mathbf{S}}$  (resp.  $j^{\mathbf{B}}$ ) such that  $g_{j^{\mathbf{S}}}^{\mathbf{S}} = g^{\mathbf{S}} := N^{\mathbf{S}} - \|\mathbf{x}\|_2^2$  (resp.  $g_{j^{\mathbf{B}}}^{\mathbf{B}} = g^{\mathbf{B}} := N^{\mathbf{B}} - \|\mathbf{y}\|_2^2$ ) for large enough positive integers  $N^{\mathbf{S}}$  and  $N^{\mathbf{B}}$ .

For every  $\alpha \in \mathbb{N}^n$  the notation  $\mathbf{x}^\alpha$  stands for the monomial  $x_1^{\alpha_1} \dots x_n^{\alpha_n}$  and for every  $r \in \mathbb{N}$ , let  $\mathbb{N}_r^n := \{\alpha \in \mathbb{N}^n : \sum_{j=1}^n \alpha_j \leq r\}$ , whose cardinality is  $\binom{n+r}{r}$ . One writes a polynomial  $p \in \mathbb{R}[\mathbf{x}, \mathbf{y}]$  as follows:

$$(\mathbf{x}, \mathbf{y}) \mapsto p(\mathbf{x}, \mathbf{y}) = \sum_{(\alpha, \beta) \in \mathbb{N}^{n+m}} p_{\alpha\beta} \mathbf{x}^\alpha \mathbf{y}^\beta,$$

and we identify  $p$  with its vector of coefficients  $\mathbf{p} = (p_{\alpha\beta})$  in the canonical basis  $(\mathbf{x}^\alpha \mathbf{y}^\beta)$ ,  $\alpha \in \mathbb{N}^n$ ,  $\beta \in \mathbb{N}^m$ .

Given a real sequence  $\mathbf{z} = (z_{\alpha\beta})$ , we define the multivariate linear functional  $\ell_{\mathbf{z}} : \mathbb{R}[\mathbf{x}, \mathbf{y}] \rightarrow \mathbb{R}$  by  $\ell_{\mathbf{z}}(p) := \sum_{\alpha\beta} p_{\alpha\beta} z_{\alpha\beta}$ , for all  $p \in \mathbb{R}[\mathbf{x}, \mathbf{y}]$ .

**Moment matrix** The *moment* matrix associated with a sequence  $\mathbf{z} = (z_{\alpha\beta})_{(\alpha, \beta) \in \mathbb{N}^{n+m}}$ , is the real symmetric matrix  $\mathbf{M}_r(\mathbf{z})$  with rows and columns indexed by  $\mathbb{N}_r^{n+m}$ , and whose entries are defined by:

$$\mathbf{M}_r(\mathbf{z})((\alpha, \beta), (\delta, \gamma)) := \ell_{\mathbf{z}}(\mathbf{x}^{\alpha+\delta} \mathbf{y}^{\beta+\gamma}), \quad \forall \alpha, \delta \in \mathbb{N}_r^n, \quad \forall \beta, \gamma \in \mathbb{N}_r^m.$$

**Localizing matrix** The *localizing* matrix associated with a sequence  $\mathbf{z} = (z_{\alpha\beta})_{(\alpha, \beta) \in \mathbb{N}^{n+m}}$  and a polynomial  $q \in \mathbb{R}[\mathbf{x}, \mathbf{y}]$  (with  $q(\mathbf{x}, \mathbf{y}) = \sum_{u,v} q_{uv} \mathbf{x}^u \mathbf{y}^v$ ) is the real symmetric matrix  $\mathbf{M}_r(q\mathbf{z})$  with rows and columns indexed by  $\mathbb{N}_r^{n+m}$ , and whose entries are defined by:

$$\mathbf{M}_r(q\mathbf{z})((\alpha, \beta), (\delta, \gamma)) := \ell_{\mathbf{z}}(q(\mathbf{x}, \mathbf{y}) \mathbf{x}^{\alpha+\delta} \mathbf{y}^{\beta+\gamma}), \quad \forall \alpha, \delta \in \mathbb{N}_r^n, \quad \forall \beta, \gamma \in \mathbb{N}_r^m.$$

We define the restriction of the Lebesgue measure on a subset  $\mathbf{A} \subset \mathbf{B}$  by  $\lambda_{\mathbf{A}}(d\mathbf{y}) := \mathbf{1}_{\mathbf{A}}(\mathbf{y}) d\mathbf{y}$ , with  $\mathbf{1}_{\mathbf{A}} : \mathbf{B} \rightarrow \{0, 1\}$  denoting the indicator function on  $\mathbf{A}$ :

$$\mathbf{1}_{\mathbf{A}}(\mathbf{y}) := \begin{cases} 1 & \text{if } \mathbf{y} \in \mathbf{A}, \\ 0 & \text{otherwise.} \end{cases}$$

The moments of the Lebesgue measure on  $\mathbf{B}$  are denoted by

$$z_{\beta}^{\mathbf{B}} := \int \mathbf{y}^{\beta} \lambda_{\mathbf{B}}(d\mathbf{y}) \in \mathbb{R}, \quad \beta \in \mathbb{N}^m \quad (4)$$

We assume that the bounding set  $\mathbf{B}$  is “simple” in the following sense:

**Assumption 2.2.** Set  $\mathbf{B}$  is such that the moments (4) are explicitly given.

## 3 Method 1: existential quantifier elimination

### 3.1 Semialgebraic sets defined with existential quantifiers

The set  $\mathbf{F} = f(\mathbf{S})$  is the image of the compact semialgebraic set  $\mathbf{S}$  under the polynomial map  $f : \mathbf{S} \rightarrow \mathbf{B}$ , thus it can be defined with an existential quantifier:

$$\mathbf{F} = \{\mathbf{y} \in \mathbf{B} : \exists \mathbf{x} \in \mathbf{S} \text{ s.t. } h_f(\mathbf{x}, \mathbf{y}) \geq 0\},$$

with

$$h_f : \mathbb{R}^{n+m} \rightarrow \mathbb{R}, \quad (\mathbf{x}, \mathbf{y}) \mapsto h_f(\mathbf{x}, \mathbf{y}) := -\|\mathbf{y} - f(\mathbf{x})\|_2^2 = -\sum_{j=1}^m (y_j - f_j(\mathbf{x}))^2.$$

Let us also define

$$h : \mathbb{R}^m \rightarrow \mathbb{R}, \quad \mathbf{y} \mapsto h(\mathbf{y}) := \sup_{\mathbf{x} \in \mathbf{S}} h_f(\mathbf{x}, \mathbf{y}).$$

**Theorem 3.1.** *There exists a sequence of polynomials  $(p_r)_{r \in \mathbb{N}} \subset \mathbb{R}[\mathbf{y}]$  such that  $p_r(\mathbf{y}) \geq h_f(\mathbf{x}, \mathbf{y})$  for all  $r \in \mathbb{N}$ ,  $\mathbf{x} \in \mathbf{S}$ ,  $\mathbf{y} \in \mathbf{B}$  and such that*

$$\lim_{r \rightarrow \infty} \int |p_r(\mathbf{y}) - h(\mathbf{y})| \lambda_{\mathbf{B}}(d\mathbf{y}) = 0. \quad (5)$$

*Proof.* The result follows readily from [Las13, Theorem 3.1 (3.4)] with the notations  $\mathbf{x} \leftarrow \mathbf{y}$ ,  $\mathbf{y} \leftarrow \mathbf{x}$ ,  $\mathbf{K} \leftarrow \mathbf{B} \times \mathbf{S}$ ,  $\mathbf{K}_{\mathbf{x}} \leftarrow \mathbf{S} \neq \emptyset$  and  $\underline{J}_f \leftarrow -h$ , which is lower semi-continuous.  $\square$

**Theorem 3.2.** *For each  $r \in \mathbb{N}$ , define  $\mathbf{F}_r := \{\mathbf{y} \in \mathbf{B} : p_r(\mathbf{y}) \geq 0\}$  where the sequence of polynomials  $(p_r)_{r \in \mathbb{N}} \subset \mathbb{R}[\mathbf{y}]$  is as in Theorem 3.1. Then  $\mathbf{F}_r \supset \mathbf{F}$  and one has*

$$\lim_{r \rightarrow \infty} \text{vol}(\mathbf{F}_r \setminus \mathbf{F}) = 0. \quad (6)$$

*Proof.* Let  $r \in \mathbb{N}$ . By assumption, one has  $p_r(\mathbf{y}) \geq h_f(\mathbf{x}, \mathbf{y})$  for all  $\mathbf{x} \in \mathbf{S}$ ,  $\mathbf{y} \in \mathbf{B}$ . Thus, one has  $p_r(\mathbf{y}) \geq h(\mathbf{y})$  for all  $\mathbf{y} \in \mathbf{B}$ , which implies that  $\mathbf{F}_r \supset \mathbf{F}$ .

It remains to prove (6). Let us define  $\mathbf{F}(k) := \{\mathbf{y} \in \mathbf{B} : h(\mathbf{y}) \geq -1/k\}$ . First, we show that

$$\lim_{k \rightarrow \infty} \text{vol} \mathbf{F}(k) = \text{vol} \mathbf{F}. \quad (7)$$

For each  $k \in \mathbb{N}$ , one has  $\mathbf{F}(k+1) \subseteq \mathbf{F}(k) \subseteq \mathbf{B}$ , thus the sequence of indicator functions  $(\mathbf{1}_{\mathbf{F}(k)})_{k \in \mathbb{N}}$  is nonincreasing and bounded. Next, let us show that for all  $\mathbf{y} \in \mathbf{B}$ ,  $\mathbf{1}_{\mathbf{F}(k)}(\mathbf{y}) \rightarrow \mathbf{1}_{\mathbf{F}}(\mathbf{y})$ , as  $k \rightarrow \infty$ :

- Let  $\mathbf{y} \in \mathbf{F}$ . By the inclusion  $\mathbf{F} \subseteq \mathbf{F}(k)$ , for each  $k \in \mathbb{N}$ ,  $\mathbf{1}_{\mathbf{F}(k)}(\mathbf{y}) = \mathbf{1}_{\mathbf{F}}(\mathbf{y}) = 1$  and the result trivially holds.
- Let  $\mathbf{y} \in \mathbf{B} \setminus \mathbf{F}$ , so there exists  $\epsilon > 0$  such that  $h(\mathbf{y}) = -\epsilon$ . Thus, there exists  $k_0 \in \mathbb{N}$  such that for all  $k \geq k_0$ ,  $\mathbf{y} \in \mathbf{B} \setminus \mathbf{F}(k)$ .

Hence,  $\mathbf{1}_{\mathbf{F}(k)}(\mathbf{y}) \rightarrow \mathbf{1}_{\mathbf{F}}(\mathbf{y})$  for each  $\mathbf{y} \in \mathbf{B}$ , as  $k \rightarrow \infty$  (monotone non-increasing). By the Monotone Convergence Theorem,  $\mathbf{1}_{\mathbf{F}(k)}(\mathbf{y}) \rightarrow \mathbf{1}_{\mathbf{F}}(\mathbf{y})$  for the  $L_1$  norm on  $\mathbf{B}$  and (7) holds.

Next, we prove that for each  $k \in \mathbb{N}$ ,

$$\lim_{r \rightarrow \infty} \text{vol} \mathbf{F}_r \leq \text{vol} \mathbf{F}(k). \quad (8)$$

By Theorem 3.1 applied to the sequence  $(p_r)_{r \in \mathbb{N}}$ , one has  $\lim_{r \rightarrow \infty} \int |p_r(\mathbf{y}) - h(\mathbf{y})| \lambda_{\mathbf{B}}(d\mathbf{y}) = 0$ . Thus, by [Ash72, Theorem 2.5.1], the sequence  $(p_r)_{r \in \mathbb{N}}$  converges to  $h$  in measure, i.e. for every  $\epsilon > 0$ ,

$$\lim_{r \rightarrow \infty} \text{vol}(\{\mathbf{y} \in \mathbf{B} : |p_r(\mathbf{y}) - h(\mathbf{y})| \geq \epsilon\}) = 0. \quad (9)$$

For every  $k \geq 1$ , observe that:

$$\begin{aligned} \text{vol } \mathbf{F}_r &= \text{vol}(\mathbf{F}_r \cap \{\mathbf{y} \in \mathbf{B} : |p_r(\mathbf{y}) - h(\mathbf{y})| \geq 1/k\}) + \\ &\quad \text{vol}(\mathbf{F}_r \cap \{\mathbf{y} \in \mathbf{B} : |p_r(\mathbf{y}) - h(\mathbf{y})| < 1/k\}). \end{aligned} \quad (10)$$

It follows from (9) that  $\lim_{r \rightarrow \infty} \text{vol}(\mathbf{F}_r \cap \{\mathbf{y} \in \mathbf{B} : |p_r(\mathbf{y}) - h(\mathbf{y})| \geq 1/k\}) = 0$ . In addition, for all  $r \in \mathbb{N}$ ,

$$\text{vol}(\mathbf{F}_r \cap \{\mathbf{y} \in \mathbf{B} : |p_r(\mathbf{y}) - h(\mathbf{y})| < 1/k\}) \leq \text{vol}(\{\mathbf{y} \in \mathbf{B} : h(\mathbf{y}) \geq -1/k\}) = \text{vol } \mathbf{F}(k). \quad (11)$$

Using both (10) and (11), and letting  $r \rightarrow \infty$ , yields (8). Thus, we have the following inequalities:

$$\text{vol } \mathbf{F} \leq \lim_{r \rightarrow \infty} \text{vol } \mathbf{F}_r \leq \text{vol } \mathbf{F}(k).$$

Using (7) and letting  $k \rightarrow \infty$  yields the desired result.  $\square$

## 3.2 Practical computation using semidefinite programming

In this section we show how the sequence of polynomials of Theorems 3.1 and 3.2 can be computed in practice. Define  $r_{\min}^{(1)} := \max\{d, r_1, \dots, r_n\}$ . For  $r \geq r_{\min}^{(1)}$ , consider the following hierarchy of semidefinite programs:

$$\begin{aligned} p_r^* &:= \inf_q \sum_{\beta \in \mathbb{N}_{2r}^m} q_\beta z_\beta^{\mathbf{B}} \\ \text{s.t.} \quad & q - h_f \in \mathbf{Q}_r(\mathbf{K}), \\ & q \in \mathbb{R}_{2r}[\mathbf{y}]. \end{aligned} \quad (12)$$

The semidefinite program dual of (12) reads:

$$\begin{aligned} d_r^* &:= \sup_{\mathbf{z}} \ell_{\mathbf{z}}(h_f) \\ \text{s.t.} \quad & \mathbf{M}_r(\mathbf{z}) \succeq 0, \\ & \mathbf{M}_{r-r_j}(g_j \mathbf{z}) \succeq 0, \quad j = 1, \dots, n^{\mathbf{K}}, \\ & \ell_{\mathbf{z}}(\mathbf{y}^\beta) = z_\beta^{\mathbf{B}}, \quad \forall \beta \in \mathbb{N}_{2r}^m. \end{aligned} \quad (13)$$

**Theorem 3.3.** *Let  $r \geq r_{\min}^{(1)}$  and suppose that Assumption 2.1 holds. Then:*

1.  $p_r^* = d_r^*$ , i.e. there is no duality gap between the semidefinite program (12) and its dual (13).

2. The semidefinite program (13) has an optimal solution. In addition, if  $\mathbf{S}$  has nonempty interior, then the semidefinite program (12) has an optimal solution  $q_r$ , and the sequence  $(q_r)_{r \in \mathbb{N}}$  converges to  $h$  in  $L_1$  norm on  $\mathbf{B}$ :

$$\lim_{r \rightarrow \infty} \int |q_r(\mathbf{y}) - h(\mathbf{y})| \lambda_{\mathbf{B}}(d\mathbf{y}) = 0. \quad (14)$$

3. Defining the set

$$\mathbf{F}_r^1 := \{\mathbf{y} \in \mathbf{B} : q_r(\mathbf{y}) \geq 0\}$$

it holds that

$$\mathbf{F}_r^1 \supset \mathbf{F}$$

and

$$\lim_{r \rightarrow \infty} \text{vol}(\mathbf{F}_r^1 \setminus \mathbf{F}) = 0.$$

*Proof.* 1. Let  $\mathcal{D}_r$  (resp.  $\mathcal{D}_r^*$ ) stand for the feasible (resp. optimal) solution set of the semidefinite program (13). First, we prove that  $\mathcal{D}_r \neq \emptyset$ . Let  $\mathbf{z} = (z_{\alpha\beta})_{(\alpha,\beta) \in \mathbb{N}_{2r}^{n+m}}$  be the sequence moments of  $\lambda_{\mathbf{K}}$ , the Lebesgue measure on  $\mathbf{K} = \mathbf{S} \times \mathbf{B}$ . Since the measure is supported on  $\mathbf{K}$ , the semidefinite constraints  $\mathbf{M}_r(\mathbf{z}) \succeq 0$ ,  $\mathbf{M}_{r-r_j}(g_j \mathbf{z}) \succeq 0$ ,  $j = 1, \dots, n^{\mathbf{K}}$  are satisfied. By construction, the marginal of  $\lambda_{\mathbf{K}}$  on  $\mathbf{B}$  is  $\lambda_{\mathbf{B}}$  and the following equality constraints are satisfied:  $\ell_{\mathbf{z}}(\mathbf{y}^\beta) = z_\beta^{\mathbf{B}}$  for all  $\beta \in \mathbb{N}_{2r}^m$ . Thus, the finite sequence  $\mathbf{z}$  lies in  $\mathcal{D}_r \neq \emptyset$ .

Note that Assumption 2.1 implies that the semidefinite constraints  $\mathbf{M}_{r-1}(g^{\mathbf{S}} \mathbf{z}) \succeq 0$  and  $\mathbf{M}_{r-1}(g^{\mathbf{B}} \mathbf{z}) \succeq 0$  both hold. Thus, the first diagonal elements of  $\mathbf{M}_{r-1}(g^{\mathbf{S}} \mathbf{z})$  and  $\mathbf{M}_{r-1}(g^{\mathbf{B}} \mathbf{z})$  are nonnegative, and since  $\ell_{\mathbf{z}}(1) = z_0^{\mathbf{B}}$ , it follows that  $\ell_{\mathbf{z}}(x_i^{2k}) \leq (N^{\mathbf{S}})^k z_0^{\mathbf{B}}$ ,  $i = 1, \dots, n$  and  $\ell_{\mathbf{z}}(y_j^{2k}) \leq (N^{\mathbf{B}})^k z_0^{\mathbf{B}}$ ,  $j = 1, \dots, m$ ,  $k = 0, \dots, r$  and we deduce from [LN07, Lemma 4.3, p. 111] that  $|z_{\alpha\beta}|$  is bounded for all  $(\alpha, \beta) \in \mathbb{N}_{2r}^{n+m}$ . Thus, the feasible set  $\mathcal{D}_r$  is compact as closed and bounded. Hence, the set  $\mathcal{D}_r^*$  is nonempty and bounded. The claim then follows from the sufficient condition of strong duality in [Trn05].

2. Assume that  $\mathbf{S}$  has nonempty interior, so  $\mathbf{K} = \mathbf{S} \times \mathbf{B}$  has also nonempty interior. Thus, the feasible solution  $\mathbf{z}$  (defined above) satisfies  $\mathbf{M}_r(\mathbf{z}) \succ 0$ ,  $\mathbf{M}_{r-r_j}(g_j \mathbf{z}) \succ 0$ ,  $j = 1, \dots, n^{\mathbf{K}}$ , which implies that Slater's condition holds for (13). Note also that the semidefinite program (12) has the trivial feasible solution  $q = 0$  since  $-h_f$  is SOS by construction. As a consequence of a now standard result of duality in semidefinite programming (see e.g. [VB94]), the semidefinite program (12) has an optimal solution  $q_r \in \mathbb{R}_{2r}[\mathbf{y}]$ .

Let us consider a sequence of polynomials  $(p_r)_{r \in \mathbb{N}} \subset \mathbb{R}[\mathbf{y}]$  as in Theorem 3.1. Now, fix  $\epsilon > 0$ . By Theorem 3.1, there exists  $r_0 \in \mathbb{N}$  such that

$$\int |p_r(\mathbf{y}) - h(\mathbf{y})| \lambda_{\mathbf{B}}(d\mathbf{y}) \leq \epsilon/2, \quad (15)$$

for all  $r \geq r_0$ . Then, observe that the polynomial  $p_r^\epsilon := p_r + \epsilon/(2 \text{vol } \mathbf{B})$  satisfies  $p_r^\epsilon(\mathbf{y}) - h_f(\mathbf{x}, \mathbf{y}) > 0$ , for all  $\mathbf{x} \in \mathbf{S}$ ,  $\mathbf{y} \in \mathbf{B}$ . Let  $r \geq \max\{\lceil (\deg p_r)/2 \rceil, r_{\min}^{(1)}\}$ . As



a consequence of Putinar's Positivstellensatz (e.g. [Las09, Section 2.5]), there exist  $s_0, \dots, s_{n^{\mathbf{K}}} \in \Sigma[\mathbf{x}, \mathbf{y}]$  such that

$$p_r^\epsilon(\mathbf{y}) - h_f(\mathbf{x}, \mathbf{y}) = \sum_{j=0}^{n^{\mathbf{K}}} s_j(\mathbf{x}, \mathbf{y}) g_j(\mathbf{x}, \mathbf{y}),$$

with  $\deg(s_j g_j) \leq 2r$  for  $j = 0, \dots, n^{\mathbf{K}}$ . And so,  $p_r^\epsilon - h_f$  lies in the  $r$ -truncated quadratic module  $\mathbf{Q}_r(\mathbf{K})$ , which implies that  $p_r^\epsilon$  is a feasible solution for Problem (12). Hence,  $q_r$  being an optimal solution of problem (12), the following holds:

$$\int q_r(\mathbf{y}) \lambda_{\mathbf{B}}(d\mathbf{y}) \leq \int p_r^\epsilon(\mathbf{y}) \lambda_{\mathbf{B}}(d\mathbf{y}) = \int [p_r(\mathbf{y}) + \epsilon/(2 \text{vol}(\mathbf{B}))] \lambda(d\mathbf{y}). \quad (16)$$

Combining (15) and (16) yields  $\int |q_r(\mathbf{y}) - h(\mathbf{y})| \lambda(d\mathbf{y}) \leq \epsilon$ , concludes the proof.

3. This is a consequence of Theorem 3.2. □

## 4 Method 2: support of image measures

Given a compact set  $\mathbf{A} \subset \mathbb{R}^n$ , let  $\mathcal{M}(\mathbf{A})$  stand for the vector space of finite signed Borel measures supported on  $\mathbf{A}$ , understood as functions from the Borel sigma algebra  $\mathcal{B}(\mathbf{A})$  to the real numbers. Let  $\mathcal{C}(\mathbf{A})$  stand for the space of continuous functions on  $\mathbf{A}$ , equipped with the sup-norm (a Banach space). Since  $\mathbf{A}$  is compact, the topological dual (i.e. the set of continuous linear functionals) of  $\mathcal{C}(\mathbf{A})$  (equipped with the sup-norm), denoted by  $\mathcal{C}(\mathbf{A})'$ , is (isometrically isomorphically identified with)  $\mathcal{M}(\mathbf{A})$  equipped with the total variation norm, denoted by  $\|\cdot\|_{\text{TV}}$ . The cone of non-negative elements of  $\mathcal{C}(\mathbf{A})$ , resp.  $\mathcal{M}(\mathbf{A})$ , is denoted by  $\mathcal{C}_+(\mathbf{A})$ , resp.  $\mathcal{M}_+(\mathbf{A})$ . Recall that  $\lambda_{\mathbf{B}}$  stand for the Lebesgue measure on  $\mathbf{B}$ . If  $\mu, \nu \in \mathcal{M}(\mathbf{A})$ , the notation  $\mu \ll \nu$  stands for  $\mu$  being absolutely continuous w.r.t.  $\nu$ , whereas the notation  $\mu \leq \nu$  means that  $\nu - \mu \in \mathcal{M}_+(\mathbf{A})$ . For background on functional analysis and measure spaces see e.g. [RF10, Section 21.5].

### 4.1 LP primal-dual conic formulation

Given a polynomial application  $f : \mathbf{S} \rightarrow \mathbf{B}$ , the pushforward or image map

$$f_{\#} : \mathcal{M}(\mathbf{S}) \rightarrow \mathcal{M}(\mathbf{B})$$

is defined such that

$$f_{\#} \mu_0(\mathbf{A}) := \mu_0(\{\mathbf{x} \in \mathbf{S} : f(\mathbf{x}) \in \mathbf{A}\})$$

for every set  $\mathbf{A} \in \mathcal{B}(\mathbf{B})$  and every measure  $\mu_0 \in \mathcal{M}(\mathbf{S})$ . The measure  $f_{\#} \mu_0 \in \mathcal{M}(\mathbf{B})$  is then called the *image measure* of  $\mu_0$  under  $f$ , see e.g. [AFP00, Section 1.5].

To approximate the image set  $\mathbf{F} = f(\mathbf{S})$ , one considers the infinite-dimensional linear programming (LP) problem:

$$\begin{aligned} p^* := \sup_{\mu_0, \mu_1, \hat{\mu}_1} \quad & \int \mu_1 \\ \text{s.t.} \quad & \mu_1 + \hat{\mu}_1 = \lambda_{\mathbf{B}}, \\ & \mu_1 = f_{\#}\mu_0, \\ & \mu_0 \in \mathcal{M}_+(\mathbf{S}), \quad \mu_1, \hat{\mu}_1 \in \mathcal{M}_+(\mathbf{B}), \end{aligned} \tag{17}$$

In the above LP, by definition of the image measure,  $\mu_1$  exists whenever  $\mu_0$  is given. The following result gives conditions for  $\mu_0$  to exist whenever  $\mu_1$  is given.

**Lemma 4.1.** *Given a measure  $\mu_1 \in \mathcal{M}_+(\mathbf{B})$ , there is a measure  $\mu_0 \in \mathcal{M}_+(\mathbf{S})$  such that  $f_{\#}\mu_0 = \mu_1$  if and only if there is no continuous function  $v \in \mathcal{C}(\mathbf{B})$  such that  $v(f(\mathbf{x})) \geq 0$  for all  $\mathbf{x} \in \mathbf{S}$  and  $\int v(\mathbf{y})d\mu_1(\mathbf{y}) < 0$ .*

*Proof.* This follows from [CK77, Theorem 6] which is an extension to locally convex topological spaces of the celebrated Farkas Lemma in finite-dimensional linear optimization. One has just to verify that the image cone  $f_{\#}(\mathcal{M}_+(\mathbf{S})) = \{f_{\#}\mu_0 : \mu_0 \in \mathcal{M}_+(\mathbf{S})\}$  is closed in the weak-star topology  $\sigma(\mathcal{M}(\mathbf{B}), \mathcal{C}(\mathbf{B}))$  of  $\mathcal{M}(\mathbf{B})$ . This, in turn, follows from continuity of  $f$  and compactness of  $\mathbf{S}$ .  $\square$

**Corollary 4.2.** *Given any measure  $\mu_1 \in \mathcal{M}_+(\mathbf{B})$  admissible for LP (17), there is a measure  $\mu_0 \in \mathcal{M}_+(\mathbf{S})$  such that  $f_{\#}\mu_0 = \mu_1$ .*

*Proof.* Since  $\mu_1 \in \mathcal{M}_+(\mathbf{B})$ ,  $\mu_1 + \hat{\mu}_1 = \lambda_{\mathbf{B}}$  and  $\text{spt } \mu_1 \subseteq \mathbf{F}$ , it holds  $0 \leq \mu_1 \ll \lambda_{\mathbf{F}}$  and by Radon-Nikodým there exists a function  $q_1 \in L_1(\lambda_{\mathbf{F}})$  such that  $d\mu_1(\mathbf{y}) = q_1(\mathbf{y})d\lambda_{\mathbf{F}}(\mathbf{y})$  and  $q_1(\mathbf{y}) \geq 0$  for all  $\mathbf{y} \in \mathbf{F}$ . The claim follows then from Lemma 4.1 since it is impossible to find a function  $v \in \mathcal{C}(\mathbf{B})$  such that  $v(\mathbf{y}) \geq 0$  for all  $\mathbf{y} \in \mathbf{F}$  while satisfying  $\int v(\mathbf{y})q_1(\mathbf{y})d\lambda_{\mathbf{F}}(\mathbf{y}) < 0$ .  $\square$

**Lemma 4.3.** *Let  $\mu_1^*$  be an optimal solution of LP (17). Then  $\mu_1^* = \lambda_{\mathbf{F}}$  and  $p^* = \text{vol } \mathbf{F}$ .*

*Proof.* From Corollary 4.2, the choice  $\mu_1^* = \lambda_{\mathbf{F}} \in \mathcal{M}_+(\mathbf{B})$  is admissible for LP (17) since there exists  $\mu_0^* \in \mathcal{M}_+(\mathbf{S})$  such that  $f_{\#}\mu_0^* = \mu_1^*$  and we can enforce  $\hat{\mu}_1^* = \lambda_{\mathbf{B}} - \mu_1^*$ . Optimality of  $\mu_1^* = \lambda_{\mathbf{F}}$  then follows exactly as in the proof of [HLS09, Theorem 3.1].  $\square$

Next, we express problem (17) as an infinite-dimensional conic problem on appropriate vector spaces. By construction, a feasible solution of problem (17) satisfies:

$$\int_{\mathbf{B}} v(\mathbf{y}) \mu_1(d\mathbf{y}) - \int_{\mathbf{S}} v(f(\mathbf{x})) \mu_0(d\mathbf{x}) = 0, \tag{18}$$

$$\int_{\mathbf{B}} w(\mathbf{y}) \mu_1(d\mathbf{y}) + \int_{\mathbf{B}} w(\mathbf{y}) \hat{\mu}_1(d\mathbf{y}) = \int_{\mathbf{B}} w(\mathbf{y}) \lambda(d\mathbf{y}), \tag{19}$$

for all continuous test functions  $v, w \in \mathcal{C}(\mathbf{B})$ .

Then, we cast problem (17) as a particular instance of a primal LP in the canonical form given in [Bar02, 7.1.1]:

$$\begin{aligned} p^* = \sup_x \quad & \langle x, c \rangle_1 \\ \text{s.t.} \quad & \mathcal{A}x = b, \\ & x \in E_1^+, \end{aligned} \tag{20}$$

with

- the vector space  $E_1 := \mathcal{M}(\mathbf{S}) \times \mathcal{M}(\mathbf{B})^2$ ;
- the vector space  $F_1 := \mathcal{C}(\mathbf{S}) \times \mathcal{C}(\mathbf{B})^2$ ;
- the duality  $\langle \cdot, \cdot \rangle_1 : E_1 \times F_1 \rightarrow \mathbb{R}$ , given by the integration of continuous functions against Borel measures, since  $E_1 = F_1'$ ;
- the decision variable  $x := (\mu_0, \mu_1, \hat{\mu}_1) \in E_1$  and the reward  $c := (0, 1, 0) \in F_1$ ;
- $E_2 := \mathcal{M}(\mathbf{B})^2$ ,  $F_2 := \mathcal{C}(\mathbf{B})^2$  and the right hand side vector  $b := (0, \lambda) \in E_2 = F_2'$ ;
- the linear operator  $\mathcal{A} : E_1 \rightarrow E_2$  given by

$$\mathcal{A}(\mu_0, \mu_1, \hat{\mu}_1) := \begin{bmatrix} -f_{\#}\mu_0 + \mu_1 \\ \mu_1 + \hat{\mu}_1 \end{bmatrix}.$$

Notice that all spaces  $E_1$ ,  $E_2$  (resp.  $F_1$ ,  $F_2$ ) are equipped with the weak topologies  $\sigma(E_1, F_1)$ ,  $\sigma(E_2, F_2)$  (resp.  $\sigma(F_1, E_1)$ ,  $\sigma(F_2, E_2)$ ). Importantly,  $\sigma(E_1, F_1)$  is the weak-star topology (since  $E_1 = F_1'$ ). Observe that  $\mathcal{A}$  is continuous with respect to the weak topology, as  $\mathcal{A}'(F_2) \subset F_1$ .

With these notations, the dual LP in the canonical form given in [Bar02, 7.1.2] reads:

$$\begin{aligned} d^* = \inf_y \quad & \langle b, y \rangle_2 \\ \text{s.t.} \quad & \mathcal{A}'y - c \in \mathcal{C}_+(\mathbf{B})^2 \end{aligned} \tag{21}$$

with

- the dual variable  $y := (v, w) \in E_2$ ;
- the (pre)-dual cone  $\mathcal{C}_+(\mathbf{B})^2$ , whose dual is  $E_1^+$ ;
- the duality pairing  $\langle \cdot, \cdot \rangle_2 : E_2 \times F_2 \rightarrow \mathbb{R}$ , with  $E_2 = F_2'$ ;
- the adjoint linear operator  $\mathcal{A}' : F_2 \rightarrow F_1$  given by

$$\mathcal{A}'(v, w) := \begin{bmatrix} -v \circ f \\ v + w \\ w \end{bmatrix}.$$

Using our original notations, the dual LP of problem (17) then reads:

$$\begin{aligned}
d^* &:= \inf_{v,w} \int w(\mathbf{y}) \lambda_{\mathbf{B}}(d\mathbf{y}) \\
\text{s.t. } & v(f(\mathbf{x})) \geq 0, \quad \forall \mathbf{x} \in \mathbf{S}, \\
& w(\mathbf{y}) \geq 1 + v(\mathbf{y}), \quad \forall \mathbf{y} \in \mathbf{B}, \\
& w(\mathbf{y}) \geq 0, \quad \forall \mathbf{y} \in \mathbf{B}, \\
& v, w \in \mathcal{C}(\mathbf{B}).
\end{aligned} \tag{22}$$

**Theorem 4.4.** *There is no duality gap between problem (17) and problem (22), i.e.  $p^* = d^*$ .*

*Proof.* This theorem follows from the “zero duality gap” result from [Bar02, Theorem 7.2], if one can prove that the cone

$$\mathcal{A}(E_1^+) := \{(\mathcal{A}x, \langle x, c \rangle_1) : x \in E_1^+\} \tag{23}$$

is closed in  $E_2 \times \mathbb{R}$ . To do so, let us consider a sequence  $(x^{(k)}) = (\mu_0^{(k)}, \mu_1^{(k)}, \hat{\mu}_1^{(k)}) \subset E_1^+$  such that  $\mathcal{A}x^{(k)} \rightarrow s = (s_1, s_2)$  and  $\langle x^{(k)}, c \rangle_1 \rightarrow t$ . Let us prove that  $(s, t) = (\mathcal{A}x^*, \langle x^*, c \rangle_1)$  for some  $x^* = (\mu_0^*, \mu_1^*, \hat{\mu}_1^*) \in E_1^+$ . As  $c = (0, 1, 0)$ , one has  $\|\mu_1^{(k)}\|_{\text{TV}} = \int_{\mathbf{B}} \mu_1^{(k)}(d\mathbf{y}) \rightarrow t (\geq 0)$ , thus  $\sup_k \|\mu_1^{(k)}\|_{\text{TV}} < \infty$ . Therefore there is a subsequence (denoted by the same indices)  $(\mu_1^{(k)})$  which converges to  $\mu_1^* \in \mathcal{M}_+(\mathbf{B})$  for the weak-star topology. In particular  $\|\mu_1^*\|_{\text{TV}} = t$ . Hence from  $\hat{\mu}_1^{(k)} + \mu_1^{(k)} \rightarrow s_2$  one deduces that  $\hat{\mu}_1^{(k)} \rightarrow s_2 - \mu_1^*$  for the weak-star topology. But then we also have  $-f_{\#}\mu_0^{(k)} \rightarrow s_1 - \mu_1^*$  in the weak-star topology of  $\mathcal{M}(\mathbf{B})$ . Therefore  $-s_1 + \mu_1^*$  is a positive measure. So let  $\mu_0^*$  be such that  $f_{\#}\mu_0^* = -s_1 + \mu_1^*$  guaranteed to exist since we have seen that  $f_{\#}(\mathcal{M}_+(S))$  is weak-star closed. Then we have  $\mathcal{A}(x^*) = (s_1, s_2)$  and  $\langle x^*, c \rangle = t$ , the desired result.  $\square$

## 4.2 Practical computation using semidefinite programming

For each  $r \geq r_{\min}^{(2)} := \max\{\lceil r_1^{\mathbf{S}}/d \rceil, \dots, \lceil r_{n^{\mathbf{S}}}^{\mathbf{S}}/d \rceil, r_1^{\mathbf{B}}, \dots, r_{n^{\mathbf{B}}}^{\mathbf{B}}\}$ , let  $\mathbf{z}_0 = (z_{0\beta})_{\beta \in \mathbb{N}_{2r}^m}$  be the finite sequence of moments up to degree  $2r$  of measure  $\mu_0$ . Similarly, let  $\mathbf{z}_1$  and  $\hat{\mathbf{z}}_1$  stand for the sequences of moments up to degree  $2r$ , respectively associated with  $\mu_1$  and  $\hat{\mu}_1$ . Problem (17) can be relaxed with the following semidefinite program:

$$\begin{aligned}
p_r^* &:= \sup_{\mathbf{z}_0, \mathbf{z}_1, \hat{\mathbf{z}}_1} z_{10} \\
\text{s.t. } & z_{1\beta} + \hat{z}_{1\beta} = z_{\beta}^{\mathbf{B}}, L_{\mathbf{z}_0}(f(\mathbf{x})^{\beta}) = z_{1\beta}, \quad \forall \beta \in \mathbb{N}_{2r}^m, \\
& \mathbf{M}_{r-d-r_j^{\mathbf{S}}}(g_j^{\mathbf{S}}\mathbf{z}_0) \succeq 0, \quad j = 0, \dots, n^{\mathbf{S}}, \\
& \mathbf{M}_{r-r_j^{\mathbf{B}}}(g_j^{\mathbf{B}}\mathbf{z}_1) \succeq 0, \mathbf{M}_{r-r_j^{\mathbf{B}}}(g_j^{\mathbf{B}}\hat{\mathbf{z}}_1) \succeq 0, \quad j = 0, \dots, n^{\mathbf{B}}.
\end{aligned} \tag{24}$$

Consider also the following semidefinite program, which is a strengthening of problem (22) and also the dual of problem (25):

$$\begin{aligned}
d_r^* &:= \inf_{v,w} \sum_{\beta \in \mathbb{N}_{2r}^m} w_\beta z_\beta^{\mathbf{B}} \\
\text{s.t. } & v \circ f \in \mathbf{Q}_{rd}(\mathbf{S}), \\
& w - 1 - v \in \mathbf{Q}_r(\mathbf{B}), \\
& w \in \mathbf{Q}_r(\mathbf{B}), \\
& v, w \in \mathbb{R}_{2r}[\mathbf{y}].
\end{aligned} \tag{25}$$

**Theorem 4.5.** *Let  $r \geq r_{\min}^{(2)}$  and suppose that both  $\mathbf{F}$  and  $\mathbf{B} \setminus \mathbf{F}$  have nonempty interior and that Assumption 2.1 holds. Then:*

1.  $p_r^* = d_r^*$ , i.e. there is no duality gap between the semidefinite program (24) and its dual (25).
2. The semidefinite program (25) has an optimal solution  $(v_r, w_r) \in \mathbb{R}_{2r}[\mathbf{y}] \times \mathbb{R}_{2r}[\mathbf{y}]$ , and the sequence  $(w_r)$  converges to  $\mathbf{1}_{\mathbf{F}}$  in  $L_1$  norm on  $\mathbf{B}$ :

$$\lim_{r \rightarrow \infty} \int |w_r(\mathbf{y}) - \mathbf{1}_{\mathbf{F}}(\mathbf{y})| \lambda_{\mathbf{B}}(d\mathbf{y}) = 0. \tag{26}$$

3. Defining the set

$$\mathbf{F}_r^2 := \{\mathbf{y} \in \mathbf{B} : w_r(\mathbf{y}) \geq 1\}$$

its holds that

$$\mathbf{F}_r^2 \supset \mathbf{F}$$

and

$$\lim_{r \rightarrow \infty} \text{vol}(\mathbf{F}_r^2 \setminus \mathbf{F}) = 0.$$

*Proof.*

1. Let  $\mu_1 = \lambda_{\mathbf{F}}$ , let  $\mu_0$  be such that  $f_{\#}\mu_0 = \mu_1$  as in Lemma ??, and let  $\hat{\mu}_1 = \lambda_{\mathbf{B}} - \mu_1$  so that  $(\mu_0, \mu_1, \hat{\mu}_1)$  is feasible for LP (17). Given  $r \geq r_{\min}^{(2)}$ , let  $\mathbf{z}_0, \mathbf{z}_1$  and  $\hat{\mathbf{z}}_1$  be the sequences of moments up to degree  $2r$  of  $\mu_0, \mu_1$  and  $\hat{\mu}_1$ , respectively. Clearly,  $(\mathbf{z}_0, \mathbf{z}_1, \hat{\mathbf{z}}_1)$  is feasible for program (24). Then, as in the proof of the first item of Theorem 3.3, the optimal solution set of the program (24) is nonempty and bounded, which by [Trn05] implies that there is no duality gap between the semidefinite program (25) and its dual (24).
2. Now, one shows that  $(\mathbf{z}_0, \mathbf{z}_1, \hat{\mathbf{z}}_1)$  is strictly feasible for program (24). Using the fact that
  - (a)  $\mathbf{F}$  (resp.  $\mathbf{B} \setminus \mathbf{F}$ ) has nonempty interior,
  - (b)  $\mathbf{z}_1$  (resp.  $\hat{\mathbf{z}}_1$ ) is the moment sequence of  $\mu_1$  (resp.  $\hat{\mu}_1$ ),

one has  $\mathbf{M}_r(g_j^{\mathbf{B}} \mathbf{z}_1) \succ 0$  (resp.  $\mathbf{M}_r(g_j^{\mathbf{B}} \hat{\mathbf{z}}_1) \succ 0$ ), for each  $j = 0, \dots, n^{\mathbf{B}}$ . Moreover,  $\mathbf{M}_r(g_j^{\mathbf{S}} \mathbf{z}_0) \succ 0$ , for all  $j = 0, \dots, n^{\mathbf{S}}$ . Otherwise, assume that there exists a nontrivial vector  $\mathbf{q}$  such that  $\mathbf{M}_r(g_j^{\mathbf{S}} \mathbf{z}_0) \mathbf{q} = 0$  for some  $j$ . As  $\mathbf{F}$  has nonempty interior, it contains an open set  $\mathbf{A} \subset \mathbb{R}^m$ . By continuity of  $f$ , the set  $f^{-1}(\mathbf{A}) := \{\mathbf{x} \in \mathbf{S} : f(\mathbf{x}) \in \mathbf{A}\}$  is an open set of  $\mathbf{S}$  and  $\mu_0(f^{-1}(\mathbf{A})) = \mu_1(\mathbf{A}) > 0$ . Then,

$$0 = \langle \mathbf{q}, \mathbf{M}_r(g_j^{\mathbf{S}} \mathbf{z}_0) \mathbf{q} \rangle = \int_{\mathbf{S}} q(\mathbf{x})^2 g_j^{\mathbf{S}}(\mathbf{x}) d\mu_0(\mathbf{x}) \geq \int_{f^{-1}(\mathbf{A})} q(\mathbf{x})^2 g_j^{\mathbf{S}}(\mathbf{x}) d\mu_0(\mathbf{x}),$$

which yields  $q(\mathbf{x})^2 g_j^{\mathbf{S}}(\mathbf{x}) = 0$  on the open set  $f^{-1}(\mathbf{A})$ , leading to a contradiction.

Therefore, as for the proof of the second item of Theorem 3.3, we conclude that the semidefinite program (25) has an optimal solution  $(v_r, w_r) \in \mathbb{R}_{2r}[\mathbf{y}] \times \mathbb{R}_{2r}[\mathbf{y}]$ .

Next, one proves that there exists a sequence of polynomials  $(w_k)_{k \in \mathbb{N}} \subset \mathbb{R}[\mathbf{y}]$  such that  $w_k(\mathbf{y}) \geq \mathbf{1}_{\mathbf{F}}(\mathbf{y})$ , for all  $\mathbf{y} \in \mathbf{B}$  and such that

$$\lim_{k \rightarrow \infty} \int |w_k(\mathbf{y}) - \mathbf{1}_{\mathbf{F}}(\mathbf{y})| \lambda_{\mathbf{B}}(d\mathbf{y}) = 0. \quad (27)$$

The set  $\mathbf{F}$  being closed, the indicator function  $\mathbf{1}_{\mathbf{F}}$  is upper semi-continuous and bounded, so there exists a non-increasing sequence of bounded continuous functions  $h_k : \mathbf{B} \rightarrow \mathbb{R}$  such that  $h_k(\mathbf{y}) \downarrow \mathbf{1}_{\mathbf{F}}(\mathbf{y})$ , for all  $\mathbf{y} \in \mathbf{B}$ , as  $k \rightarrow \infty$ . Using the Monotone Convergence Theorem,  $h_k \rightarrow \mathbf{1}_{\mathbf{F}}$  for the  $L_1$  norm. By the Stone-Weierstrass Theorem, there exists a sequence of polynomials  $(w'_k)_{k \in \mathbb{N}} \subset \mathbb{R}[\mathbf{y}]$ , such that  $\sup_{\mathbf{y} \in \mathbf{B}} |w'_k(\mathbf{y}) - h_k(\mathbf{y})| \leq 1/k$ . The polynomial  $w_k := w'_k + 1/k$  satisfies  $w_k > h_k \geq \mathbf{1}_{\mathbf{F}}$  and (27) holds.

Let us define  $\tilde{w}_k := w_k + \epsilon/(2 \text{vol } \mathbf{B})$ ,  $\tilde{v}_k := w_k - 1$ . Next, for  $r \in \mathbb{N}$  large enough, one proves that  $(\tilde{v}_k, \tilde{w}_k)$  is a feasible solution of (25). Using the fact that  $\tilde{w}_k > w_k > \mathbf{1}_{\mathbf{F}}$ , one has  $\tilde{w}_k \in \mathbf{Q}_r(\mathbf{B})$ , as a consequence of Putinar's Positivstellensatz. For each  $\mathbf{x} \in \mathbf{S}$ ,  $\tilde{v}_k(f(\mathbf{x})) = w_k(\mathbf{y}) - 1 > \mathbf{1}_{\mathbf{F}}(f(\mathbf{x})) - 1 + \epsilon/(2 \text{vol } \mathbf{B}) > 0$ , so  $\tilde{v}_k \circ f$  lies in  $\mathbf{Q}_{rd}(\mathbf{S})$ . Similarly,  $\tilde{w}_k - \tilde{v}_k - 1 \in \mathbf{Q}_r(\mathbf{B})$ . Then, one concludes using the same arguments as for (14) in the proof of the second item of Theorem 3.3.

3. Let  $\mathbf{y} \in \mathbf{F}$ . There exists  $\mathbf{x} \in \mathbf{S}$  such that  $\mathbf{y} = f(\mathbf{x})$ . Let  $(v_r, w_r) \in \mathbb{R}_{2r}[\mathbf{y}] \times \mathbb{R}_{2r}[\mathbf{y}]$  be an optimal solution of (25). By feasibility,  $w_r(\mathbf{y}) - 1 \geq v_r(\mathbf{y}) = v_r(f(\mathbf{x})) \geq 0$ . Thus,  $\mathbf{F}_r^2 \supset \mathbf{F}$ . Finally, the proof of the convergence in volume is analogous to the proof of (6) in Theorem 3.2.

□

## 5 Computational considerations

### 5.1 Complexity analysis and lifting strategy

#### 5.1.1 Method 1

First, consider the semidefinite program (13) of Method 1. For  $r \geq r_{\min}^{(1)}$ , the number of variables  $n^{(1)}$  (resp. size of semidefinite matrices  $m^{(1)}$ ) of problem (13) satisfies:

$$n^{(1)} \leq \binom{n+m+2r}{2r}.$$

Problem (13) involves  $(n^{\mathbf{B}} + n^{\mathbf{S}} + 1)$  semidefinite constraints of size  $m^{(1)}$  bounded as follows:

$$m^{(1)} \leq \binom{n+m+r}{r}.$$

#### 5.1.2 Method 2

Now, consider the semidefinite program (24). For  $r \geq r_{\min}^{(2)}$ , the number of variables  $n^{(2)}$  of Problem (24) satisfies:

$$n^{(2)} \leq \binom{n+2rd}{2rd} + 2 \binom{m+2r}{2r}.$$

Problem (24) also involves  $(n^{\mathbf{S}} + 1)$  semidefinite constraints of size at most  $\binom{n+rd}{rd}$  and  $2(n^{\mathbf{B}} + 1)$  semidefinite constraints of size at most  $\binom{m+r}{r}$ .

Due to the dependence on the degree  $d$  of the polynomial application, one observes that the number of variables (resp. constraints) can quickly become large if  $d$  is not small. An alternative formulation to limit the blowup of these relaxations is obtained by considering  $y_1, \dots, y_m$  as “lifting” variables, respectively associated with  $f_1, \dots, f_m$ , together with the following  $2m$  additional constraints:

$$g_{n_{\mathbf{S}}+j}^{\mathbf{S}}(\mathbf{x}, \mathbf{y}) := y_j - f_j(\mathbf{x}), \quad g_{n_{\mathbf{S}}+2j}^{\mathbf{S}}(\mathbf{x}, \mathbf{y}) := f_j(\mathbf{x}) - y_j, \quad j = 1, \dots, m.$$

By considering the basic compact semialgebraic set  $\mathbf{S}_{\text{lift}} \subset \mathbb{R}^{n+m}$  given by

$$\mathbf{S}_{\text{lift}} := \{(\mathbf{x}, \mathbf{y}) \in \mathbf{K} : g_{n_{\mathbf{S}}+1}^{\mathbf{S}}(\mathbf{x}, \mathbf{y}) \geq 0, \dots, g_{n_{\mathbf{S}}+2m}^{\mathbf{S}}(\mathbf{x}, \mathbf{y}) \geq 0\}, \quad (28)$$

problem (25) becomes:

$$\begin{aligned} \inf_{v,w} \quad & \sum_{\beta \in \mathbb{N}_{2r}^m} w_{\beta} z_{\beta}^{\mathbf{B}} \\ \text{s.t.} \quad & v \in \mathbf{Q}_r(\mathbf{S}_{\text{lift}}), \\ & w - 1 - v \in \mathbf{Q}_r(\mathbf{B}), \\ & w \in \mathbf{Q}_r(\mathbf{B}), \\ & v, w \in \mathbb{R}_{2r}[\mathbf{y}], \end{aligned} \quad (29)$$

which is actually equivalent to the following problem:

$$\begin{aligned}
& \inf_w \sum_{\beta \in \mathbb{N}_{2r}^m} w_\beta z_\beta^{\mathbf{B}} \\
& \text{s.t. } w - 1 \in \mathbf{Q}_r(\mathbf{S}_{\text{lift}}), \\
& \quad w \in \mathbf{Q}_r(\mathbf{B}), \\
& \quad w \in \mathbb{R}_{2r}[\mathbf{y}].
\end{aligned} \tag{30}$$

For problem (30), the minimal relaxation order is  $r_{\text{lift}}^{(2)} := \max\{\lceil \frac{d}{2} \rceil, r_1^{\mathbf{S}}, \dots, r_{n^{\mathbf{S}}}^{\mathbf{S}}, r_1^{\mathbf{B}}, \dots, r_{n^{\mathbf{B}}}^{\mathbf{B}}\}$  and the number of variables  $n_{\text{lift}}^{(2)}$  is bounded as follows:

$$n_{\text{lift}}^{(2)} \leq \binom{n + m + 2r}{2r} + \binom{m + 2r}{2r}.$$

Problem (30) involves  $(n^{\mathbf{S}} + 2m + 1)$  semidefinite constraints of size at most  $\binom{n+m+r}{r}$  and  $(n^{\mathbf{B}} + 1)$  semidefinite constraints of size at most  $\binom{m+r}{r}$ . When  $m$  is small and  $d$  is large, this seems to be a suitable choice to reduce the computational cost of the semidefinite program (24). Experimental results described further (see Table 1 in Section 6.1 and Table 2 in Section 6.4) agree with this observation.

## 5.2 Exploiting sparsity

As explained above, both Method 1 and Method 2 are computationally demanding in general. However, if the polynomials  $\mathbf{x} \mapsto f_j(\mathbf{x})$ ,  $(j = 1, \dots, m)$  have some structured sparsity, then one can still exploit sparsity in a way similar to the one described in [WKKM06, Las06] to handle problems in higher dimensions. In particular, let  $\{1, \dots, n\}$  be the union  $\bigcup_{j=1}^m I_j$  of subsets  $I_j \subseteq \{1, \dots, n\}$  and assume that for each  $j = 1, \dots, m$ , the polynomial  $f_j$  involves only variables  $\{x_i \mid i \in I_j\}$ . One also suppose that the collection  $\{I_1, \dots, I_m\}$  satisfies the so-called *running intersection property*:

**Definition 5.1.** The collection  $\{I_1, \dots, I_m\}$  of subsets  $I_j \subseteq \{1, \dots, n\}$  satisfies the running intersection property if the following holds for each  $j = 1, \dots, m - 1$ :

$$I_{j+1} \cap \bigcup_{k=1}^j I_k \subseteq I_l \quad \text{for some } l \leq j.$$

The following assumption allows to apply the sparse representation result of [Las06, Corollary 3.9] while using either Method 1 or Method 2.

**Assumption 5.2.** The index set  $\{1, \dots, n\}$  is partitioned into  $m$  disjoint sets  $I_j$ ,  $j = 1, \dots, m$  so that:

1. The collection  $\{I_1, \dots, I_m\}$  satisfies the running intersection property.
2. For each  $j = 1, \dots, n^{\mathbf{S}}$ , there exists some  $k_j$  such that the polynomial  $g_j^{\mathbf{S}}$  in (1) involves only variables  $\{x_i \mid i \in I_{k_j}\}$ .



3. In the definition (1) of  $\mathbf{S}$ , we replace the inequality constraint  $N^{\mathbf{S}} - \|\mathbf{x}\|_2^2 \geq 0$  by the  $m$  quadratic constraints:

$$N_j - \sum_{i \in I_j} x_i^2 \geq 0, \quad j = 1, \dots, m.$$

For each  $j = 1, \dots, m$ , index the variable  $y_j$  by  $n + j$  and define  $I_j^{(1)} := I_j \cup \{n + 1, \dots, n + m\}$ .

**Proposition 5.3.** *Under Assumption 5.2, the collection  $\{I_1^{(1)}, \dots, I_m^{(1)}\}$  of subsets  $I_j^{(1)} \subseteq \{1, \dots, n, n + 1, \dots, n + m\}$  satisfies the running intersection property.*

*Proof.* The collection  $\{I_1, \dots, I_m\}$  of subsets  $I_j \subseteq \{1, \dots, n\}$  satisfies the running intersection property. For each  $j = 1, \dots, m - 1$ , there exists  $l \leq j$  such that  $I_{j+1} \cap \bigcup_{k=1}^j I_k \subseteq I_l$ . Thus,  $I_{j+1}^{(1)} \cap \bigcup_{k=1}^j I_k^{(1)} = (I_{j+1} \cap \bigcup_{k=1}^j I_k) \cup \{n + 1, \dots, n + m\} \subseteq I_l \cup \{n + 1, \dots, n + m\} = I_l^{(1)}$ , the desired result.  $\square$

Then Assumption 5.2 allows one to apply the sparse representation result of [Las06, Corollary 3.9] to the semidefinite program (12) associated with Method 1. Indeed, observe that the polynomial  $(\mathbf{x}, \mathbf{y}) \mapsto h_f(\mathbf{x}, \mathbf{y})$  can be decomposed as  $h_f = \sum_{j=1}^m h_{f_j}$ , where for each  $j = 1, \dots, m$ , the polynomial  $h_{f_j}$  involves only the variables  $\{x_i \mid i \in I_j\}$  (the same variables involved in  $f_j$ ) and  $\mathbf{y}$ .

Under Assumption 5.2, this sparse representation result can also be applied for the semidefinite program (30) associated with the lifting variant of Method 2 described in Section 5.1.2. This is due to the fact that for each  $j = 1, \dots, m$ , the polynomials  $g_{n^{\mathbf{S}}+j}^{\mathbf{S}}$  and  $g_{n^{\mathbf{S}}+2j}^{\mathbf{S}}$  involve only the variables  $\{x_i \mid i \in I_j\}$  and  $\mathbf{y}$ .

## 6 Application examples

Here we present some application examples together with numerical results. In particular, this section illustrates that our methodology is a unified framework which can tackle important special cases: semialgebraic set projections (Section 6.2) and Pareto curves approximations (Section 6.3). Moreover, the framework can be extended to approximate images of semialgebraic sets under semialgebraic applications (Section 6.4).

The numerical results are given after solving either the semidefinite program (12) for Method 1, the semidefinite program (25) for Method 2 or the semidefinite program (30) for the lifting variant of Method 2, with the YALMIP toolbox [Löf04] for MATLAB. As explained in Section 5.1.2, the outer approximations obtained by Method 2 and its lifting variant are the same, but their semidefinite formulations differ.

Benchmarks are performed on an Intel Core i5 CPU (2.40 GHz) with YALMIP interfaced with the semidefinite programming solver MOSEK [Dah12].

## 6.1 Polynomial image of semialgebraic sets

**Example 1.** Consider the image of the two-dimensional unit ball  $\mathbf{S} := \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_2^2 \leq 1\}$  under the polynomial application  $f(\mathbf{x}) := (x_1 + x_1x_2, x_2 - x_1^3)/2$ . We choose  $\mathbf{B} = \mathbf{S}$  since it can be checked that  $\mathbf{F} = f(\mathbf{S}) \subset \mathbf{B}$ .

On Figure 1 resp. 2, we represent in light gray the outer approximations  $\mathbf{F}_r^1$  resp.  $\mathbf{F}_r^2$  of  $\mathbf{F}$  obtained by Method 1 resp. 2, for increasing values of the relaxation order  $r$ . On each figure, the black dots correspond to the image set of the points obtained by uniform sampling of  $\mathbf{S}$  under  $f$ . We observe that the approximations behave well around the locally convex parts of the boundary of  $\mathbf{F}$ , and that it is not straightforward to decide whether Method 1 or Method 2 provides the best approximations.

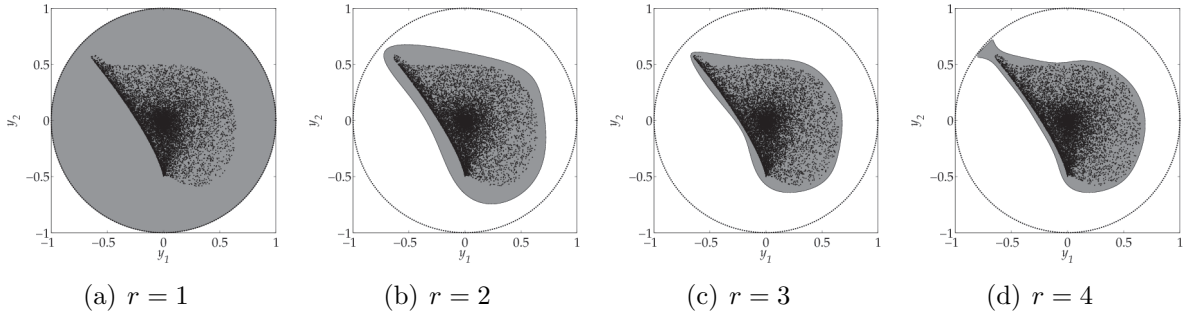


Figure 1: Outer approximations  $\mathbf{F}_r^1$  (light gray) of  $\mathbf{F}$  (black dot samples) for Example 1, for  $r = 1, 2, 3, 4$ .

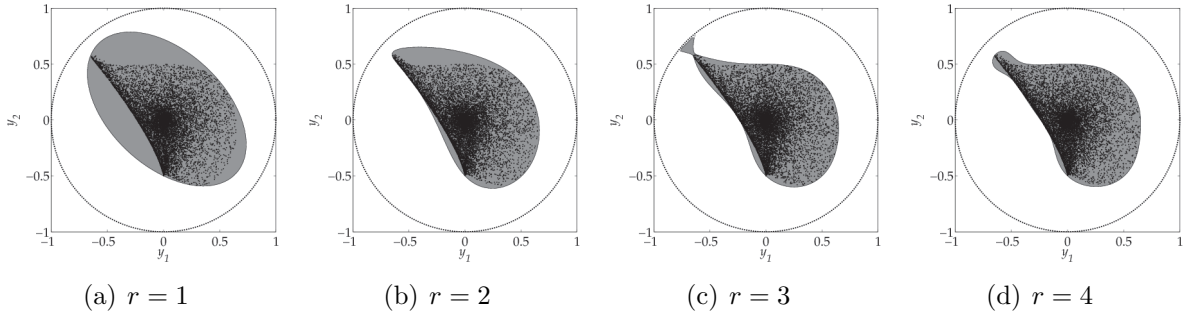


Figure 2: Outer approximations  $\mathbf{F}_r^2$  (light gray) of  $\mathbf{F}$  (black dot samples) for Example 1, for  $r = 1, 2, 3, 4$ .

We indicate in Table 1 the data related to the semidefinite programs solved by MOSEK to compute approximations of increasing degrees, while using Method 1, Method 2 and Method 2 with the lifting strategy (see Section 5.1.2 for more details). For each problem, “vars” stands for the total number of variables and “size” stands for the size of the semidefinite matrices. The computational timings of Method 2 with the lifting strategy are similar to those of Method 1, for  $r = 1, \dots, 5$ . However, for  $r = 6$  the size of the problem is significantly smaller with Method 1.

Table 1: Comparison of timing results for Example 1

relaxation order $r$		1	2	3	4	5	6
Method 1	vars	40	212	1039	4211	14028	40251
	size	30	111	350	915	1991	3822
	time (s)	0.64	0.72	0.77	1.69	8.22	40.37
Method 2	vars	286	2140	8241	22720	51166	100626
	size	129	471	1029	1803	2793	3999
	time (s)	0.65	0.74	1.54	3.4	12.89	43.74
Method 2 with lifting	vars	51	308	1499	5882	19546	56710
	size	32	157	536	1411	3128	6127
	time (s)	0.58	0.66	0.68	1.93	10.07	63.88

## 6.2 Projections of semialgebraic sets

For  $n \geq m$ , we focus on the special case of projections. Let  $f$  be the projection of  $\mathbf{S}$  with respect to the  $m$  first coordinates, i.e.  $f(\mathbf{x}) := (x_1, \dots, x_m)$ . It turns out that in this case, the semidefinite program (30) associated to the lifting variant of Method 2, has the following simpler formulation:

$$\begin{aligned}
 \inf_w \quad & \sum_{\beta \in \mathbb{N}_{2r}^m} w_\beta z_\beta^{\mathbf{B}} \\
 \text{s.t.} \quad & w - 1 \in \mathbf{Q}_r(\mathbf{S}), \\
 & w \in \mathbf{Q}_r(\mathbf{B}), \\
 & w \in \mathbb{R}_{2r}[x_1, \dots, x_m].
 \end{aligned} \tag{31}$$

**Example 2.** Consider the projection  $\mathbf{F}$  on the first two coordinates of the semialgebraic set  $\mathbf{S} := \{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x}\|_2^2 \leq 1, 1/4 - (x_1 + 1/2)^2 - x_2^2 \geq 0, 1/9 - (x_1 - 1/2)^4 - x_2^4 \geq 0\}$ , which belongs to  $\mathbf{B} := \{x \in \mathbb{R}^2 : \|\mathbf{x}\|_2^2 \leq 1\}$ .

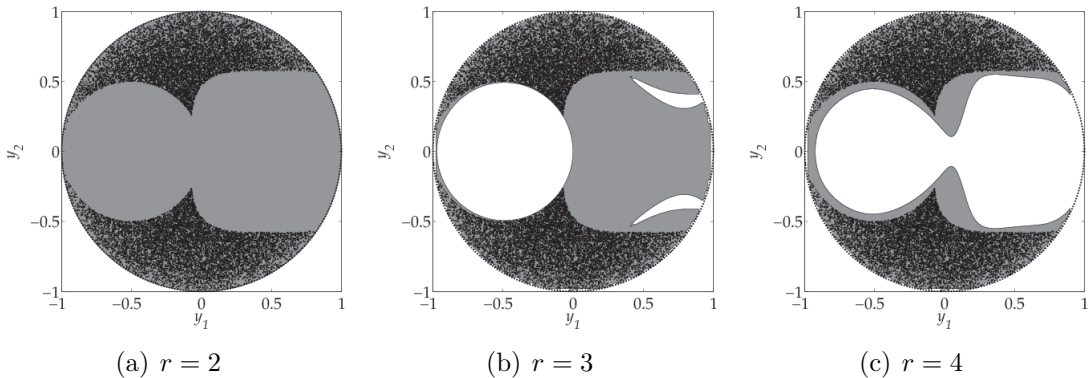


Figure 3: Outer approximations  $\mathbf{F}_r^1$  (light gray) of  $\mathbf{F}$  (black dot samples) for Example 2, for  $r = 2, 3, 4$ .

Figure 3 resp. 4 displays approximation of the projection of  $\mathbf{S}$  on the first two coordinates with Method 1 resp. 2.

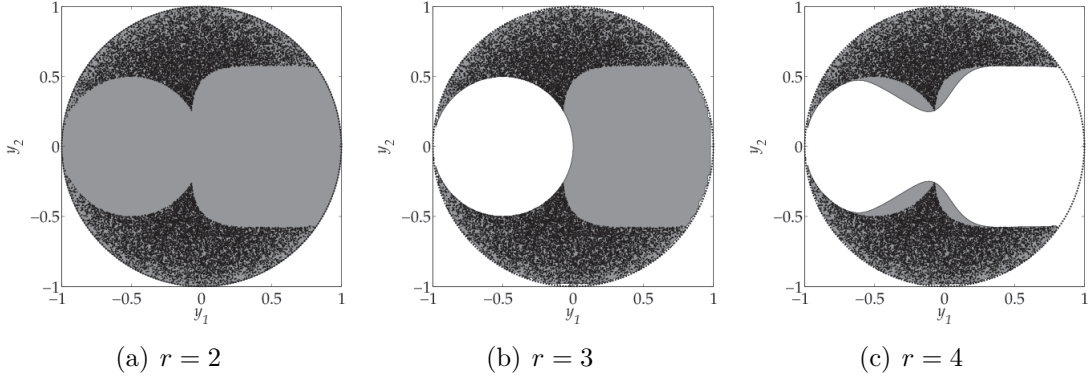


Figure 4: Outer approximations  $\mathbf{F}_r^2$  (light gray) of  $\mathbf{F}$  (black dot samples) for Example 2, for  $r = 2, 3, 4$ .

### 6.3 Approximating Pareto curves

In [MHL14], we propose a method to approximate *Pareto curves* associated with bicriteria polynomial optimization problems  $\min_{\mathbf{x} \in \mathbf{S}} \{(f_1(\mathbf{x}), f_2(\mathbf{x}))\}$ . The image space  $\mathbb{R}^2$  is partially ordered with the positive orthant  $\mathbb{R}_+^2$ , that is, for every  $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^2$ ,  $\mathbf{y}_1 \geq \mathbf{y}_2$  stands for  $\mathbf{y}_2 - \mathbf{y}_1 \in \mathbb{R}_+^2$ . A point  $\bar{\mathbf{x}}$  is called a *weakly Edgeworth-Pareto optimal point*, when there is no  $\mathbf{x} \in \mathbf{S}$  such that  $f_j(\mathbf{x}) < f_j(\bar{\mathbf{x}})$ ,  $j = 1, 2$ . The Pareto curve is the set of weakly Edgeworth-Pareto optimal points. For more details on multicriteria optimization, we refer the interested reader to [Jah10] and the references therein.

The methodology of [MHL14] consists of reformulating the initial bicriteria optimization problem to use a hierarchy of semidefinite approximations for parametric polynomial optimization problems. Then, one can apply the framework developed in [Las10] and build a hierarchy of semidefinite programs, allowing to approximate as closely as desired the Pareto curve. Here we propose to study outer approximations of the set  $\mathbf{F} = (f_1(\mathbf{S}), f_2(\mathbf{S}))$  since points along the boundary of a tight outer approximation are expected to be close to the Pareto curve.

**Example 3.** Let consider the two-dimensional nonlinear problem proposed in [WCSF01]:  $\min_{\mathbf{x} \in \mathbf{S}} \{(f_1(\mathbf{x}), f_2(\mathbf{x}))\}$ , with  $f_1(\mathbf{x}) := \frac{(x_1+x_2-7.5)^2}{4} + (x_2 - x_1 + 3)^2$ ,  $f_2(\mathbf{x}) := x_1 + x_2^2$  and  $\mathbf{S} := \{\mathbf{x} \in \mathbb{R}^2 : -(x_1 - 2)^3/2 - x_2 + 2.5 \geq 0, -x_1 - x_2 + 8(x_2 - x_1 + 0.65)^2 + 3.85 \geq 0\}$ . Instead of  $f_1$ , we consider  $\tilde{f}_1 := (f_1(\mathbf{x}) - a_1)/(b_1 - a_1)$ , where  $a_1$  and  $b_1$  are given by  $a_1 := \min_{\mathbf{x} \in \mathbf{S}} f_1(\mathbf{x})$  and  $b_1 := f_1(\bar{\mathbf{x}})$  with  $\bar{\mathbf{x}}$  a solution of  $\min_{\mathbf{x} \in \mathbf{S}} f_2(\mathbf{x})$ . Similarly, we consider a scaled criterion  $\tilde{f}_2$  defined from  $f_2$ . A preprocessing step consists in computing lower and upper bounds of the polynomial  $f_1$  (resp.  $f_2$ ) over  $\mathbf{S}$  to define  $\tilde{f} = (\tilde{f}_1, \tilde{f}_2)$ . Doing so, one ensures that  $\tilde{f}(\mathbf{S})$  is a subset of the unit ball  $\mathbf{B}$  and the present methodology applies.

In this case, the Pareto curve is nonconvex and disconnected. As depicted in Figure 5 and Figure 6, it is difficult to obtain precise approximations of the whole image set, in particular for the subset  $\mathbf{F} \cap \mathbf{B}_0$ , with an ellipse  $\mathbf{B}_0 := \{\mathbf{x} \in \mathbb{R}^2 : ((x_1 - 13.7)/1.7)^2 + ((x_2 - 1.8)/0.5)^2 \leq 1\}$ . Figure 7 displays more precise outer approximations of degree 8 (a) and degree 10 (b).

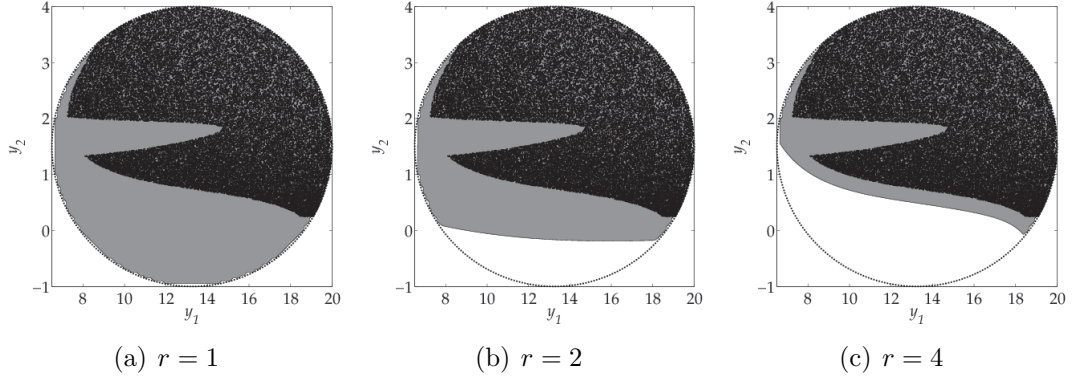


Figure 5: Outer approximations  $\mathbf{F}_r^1$  (light gray) of  $\mathbf{F}$  (black dot samples) for Example 3, for  $r = 1, 2, 4$ .

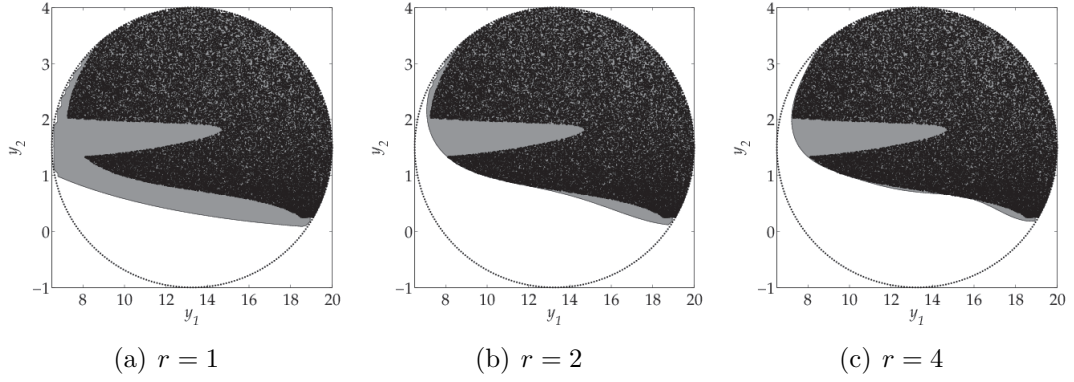


Figure 6: Outer approximations  $\mathbf{F}_r^2$  (light gray) of  $\mathbf{F}$  (black dot samples) for Example 3, for  $r = 1, 2, 4$ .

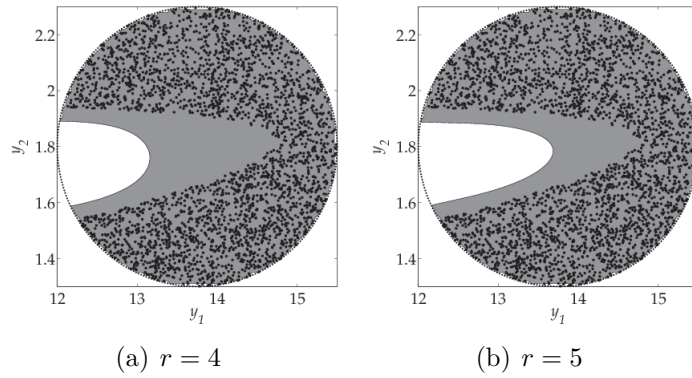


Figure 7: Outer approximations  $\mathbf{F}_r^1$  (light gray) of  $\mathbf{F} \cap \mathbf{B}_0$  (black dot samples) for Example 3, for  $r = 4, 5$ .

## 6.4 Semialgebraic image of semialgebraic sets

Given a semialgebraic set  $\mathbf{S}$  as in (1), Methods 1 and 2 can be extended to approximate the image of  $\mathbf{S}$  under a semialgebraic application  $f = (f_1, \dots, f_m)$ . To do so, we follow [LP10] and introduce lifting variables to represent non-polynomial components involved in  $f_1, \dots, f_m$ , as well as additional polynomial constraints.

Proceeding as in [LP10], for each semialgebraic function  $f_j$ , one introduces additional variables  $\mathbf{x}^j := (x_1^j, \dots, x_{t_j}^j)$  such that the graph  $\{(\mathbf{x}, f_j(\mathbf{x})) : \mathbf{x} \in \mathbf{S}\} = \{(\mathbf{x}, x_{t_j}^j) : (\mathbf{x}, \mathbf{x}^j) \in \hat{\mathbf{S}}_j\}$  for some semialgebraic set  $\hat{\mathbf{S}}_j \subseteq \mathbb{R}^{n+t_j}$ . In the end, one works with the lifted set  $\hat{\mathbf{S}} := \{(\mathbf{x}, \mathbf{x}^1, \dots, \mathbf{x}^m) : (\mathbf{x}, \mathbf{x}^j) \in \hat{\mathbf{S}}_j, j = 1, \dots, m\}$ .

**Example 4.** Here, we consider the image of the two-dimensional unit ball  $\mathbf{S} := \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_2^2 \leq 1\}$  under the semialgebraic application  $f(\mathbf{x}) := (\min(x_1 + x_1x_2, x_1^2), x_2 - x_1^3)/3$ . Remind that  $2\min(a, b) = a + b - |a - b|$ , so that  $2\min(x_1 + x_1x_2, x_1^2) = x_1 + x_1x_2 + x_1^2 - |x_1 + x_1x_2 - x_1^2|$ . To handle the absolute value, we introduce an additional variable  $x_3$  together with the equality constraint  $x_3^2 = (x_1 + x_1x_2)^2$  and the inequality constraint  $x_3 \geq 0$ .

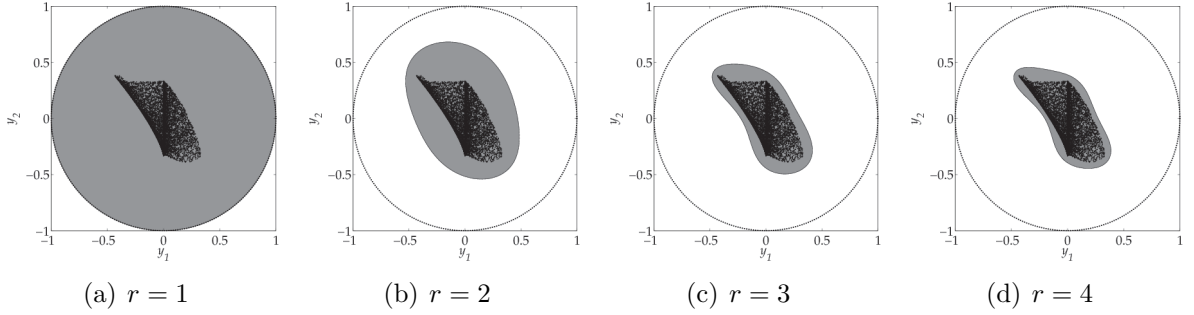


Figure 8: Outer approximations  $\mathbf{F}_r^1$  (light gray) of  $\mathbf{F}$  (black dot samples) for Example 4, for  $r = 1, 2, 3, 4$ .

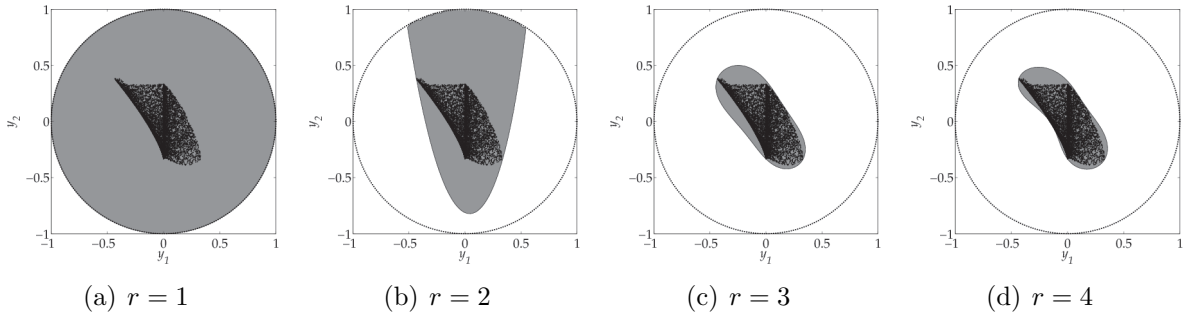


Figure 9: Outer approximations  $\mathbf{F}_r^2$  (light gray) of  $\mathbf{F}$  (black dot samples) for Example 4, for  $r = 1, 2, 3, 4$ .

As for Example 1, we report in Table 2 the data related to the semidefinite problems solved by MOSEK to compute approximations of increasing degrees, while using Method

1, Method 2 and Method 2 with the lifting strategy. Method 2 fails to compute polynomial approximations of degree higher than eight ( $r = 4$ ), the system running out of memory (indicated with the symbol “—”). The lifting strategy described in Section 5.1.2 overcomes this practical limitation.

Table 2: Comparison of timing results for Example 4

relaxation order $r$		1	2	3	4	5
Method 1	vars	66	438	3137	16993	73213
	size	45	226	1008	3387	9075
	time (s)	0.68	0.85	1.16	14.41	147.32
Method 2	vars	715	12243	89695	—	—
	size	295	1957	6283	—	—
	time (s)	0.83	3.29	52.55	—	—
Method 2 with lifting	vars	78	540	3788	20216	87475
	size	51	273	1262	4247	11508
	time (s)	0.68	0.96	1.83	14.44	174.80

## 7 Discussion and conclusion

In this work, we propose two methods to approximate polynomial images of basic compact semialgebraic sets, a numerical approximation alternative to exact computer algebra methods when the latter are too computationally demanding. In its present form, this methodology is applicable to problems of modest size, except if some sparsity can be taken into account, as explained earlier. Therefore, to handle larger size problems, the methodology needs to be adapted. A topic of further investigation is to search for alternative positivity certificates, less demanding than the SOS certificates used in this paper but more efficient than the LP based certificates as defined in [Han88, Vas03]. On the one hand, the latter are appealing since they yield a hierarchy of LP relaxations (as opposed to semidefinite relaxations as in this paper). Moreover, today’s LP solvers can handle huge size LP problems, which is far from being the case for semidefinite solvers. On the other hand, it has been shown in [Las09] that generically finite convergence cannot occur for convex problems, except for the linear case.

## Acknowledgments

This work was partly funded by an award of the Simone and Cino del Duca foundation of Institut de France.

## References

- [AFP00] L. Ambrosio, N. Fusco, D. Pallara. Functions of bounded variation and free discontinuity problems. Oxford University Press, UK, 2000.
- [Ash72] R. B. Ash. Real analysis and probability. Academic Press, New York, 1972.
- [Bar02] A. Barvinok. A course in convexity. American Mathematical Society, Providence, NJ, 2002.
- [Bas14] S. Basu. Algorithms in real algebraic geometry: a survey. [arXiv:1409.1534](#), September 2014.
- [BCR98] J. Bochnak, M. Coste, M.-F. Roy. Real algebraic geometry. Springer, Berlin, 1998.
- [Col74] G. E. Collins. Quantifier elimination for real closed fields by cylindrical algebraic decomposition—preliminary report. SIGSAM Bull., 8(3):80–90, 1974.
- [CK77] B. D. Craven, J. Koliha. Generalization of Farkas’ Theorem. SIAM J. Math. Anal. 8:983–997, 1977.
- [Dah12] J. Dahl. Extending the conic optimizer in MOSEK with semidefinite cones. International Symposium on Mathematical Programming (ISMP), Berlin, Germany, 2012.
- [GJ88] D. Yu. Grigoriev, N. N. Vorobjov Jr. Solving systems of polynomial inequalities in subexponential time. Journal of Symbolic Computation, 5(1–2):37–64, 1988.
- [Han88] D. Handelman. Representing polynomials by positive linear functions on compact convex polyhedra. Pacific Journal of Mathematics, 132(1):35–62, 1988.
- [HD12] H. Hong, M. Safey El Din. Variant quantifier elimination. Journal of Symbolic Computation, 47(7):883–901, 2012.
- [HLS09] D. Henrion, J.-B. Lasserre, and C. Savorgnan. Approximate volume and integration for basic semialgebraic sets. SIAM Review, 51(4):722–743, 2009.
- [Jah10] J. Jahn. Vector optimization: theory, applications, and extensions. Springer, Berlin, 2010.
- [Las06] J.-B. Lasserre. Convergent SDP relaxations in polynomial optimization with sparsity. SIAM Journal on Optimization, 17(3):822–843, 2006.
- [Las09] J.-B. Lasserre. Moments, positive polynomials and their applications. Imperial College Press, UK, 2010.
- [Las10] J.-B. Lasserre. A “joint+marginal” approach to parametric polynomial optimization. SIAM Journal on Optimization, 20(4):1995–2022, 2010.



- [Las13] J.-B. Lasserre. Tractable approximations of sets defined with quantifiers. Isaac Newton Institute for Mathematical Sciences, Cambridge, UK, preprint 13019, 2013.
- [LN07] J.-B. Lasserre, T. Netzer. SOS approximations of nonnegative polynomials via simple high degree perturbations. *Math. Z.*, 256(1):99–112, 2007.
- [LP10] J.-B. Lasserre, M. Putinar. Positivity and optimization for semi-algebraic functions. *SIAM Journal on Optimization*, 20(6):3364–3383, 2010.
- [Löf04] J. Löfberg. Yalmip: a toolbox for modeling and optimization in Matlab. In *IEEE Symposium on CACSD*, Taipei, Taiwan, 2004.
- [MHL14] V. Magron, D. Henrion, J.-B. Lasserre. Approximating pareto curves using semidefinite relaxations. *Operations Research Letters*, 42(6–7):432 – 437, 2014.
- [RF10] H. L. Royden, P. Fitzpatrick. *Real analysis*. 4th edition. Prentice Hall, NJ, 2010.
- [Tar51] A. Tarski. *A decision method for elementary algebra and geometry*. University of California Press, 1951.
- [Trn05] M. Trnovská. Strong duality conditions in semidefinite programming. *Journal of Electrical Engineering (Bratislava)*, 56:1–5, 2005.
- [Vas03] F.-H. Vasilescu. Spectral measures and moment problems. In A. Gheondea et al. (eds.). *Spectral analysis and its applications*. Ion Colojoară anniversary volume. Theta Series in Advanced Mathematics 2, 173-215, 2003.
- [VB94] L. Vandenberghe, S. Boyd. Semidefinite programming. *SAM Review*, 38:49–95, 1994.
- [WCSF01] B. Wilson, D. Cappelleri, T. W. Simpson, M. Frecker. Efficient Pareto frontier exploration using surrogate approximations. *Optimization and Engineering*, 2:31–50, 2001.
- [WKKM06] H. Waki, S. Kim, M. Kojima, M. Muramatsu. Sums of squares and semidefinite programming relaxations for polynomial optimization problems with structured sparsity. *SIAM Journal on Optimization*, 17(1):218–242, 2006.