TRACTABLE APPROXIMATIONS OF SETS DEFINED WITH QUANTIFIERS

JEAN B. LASSERRE

Abstract. Given a compact basic semi-algebraic set $\mathbf{K} \subset \mathbb{R}^n \times \mathbb{R}^m$, a simple set $\mathbf{B}$ (box or ellipsoid), and some semi-algebraic function f, we consider sets defined with quantifiers, of the form

$$\begin{aligned}\mathbf{R}_f &:= \{\mathbf{x} \in \mathbf{B} \quad : \quad f(\mathbf{x},\mathbf{y}) \leq 0 \text{ for all } \mathbf{y} \text{ such that } (\mathbf{x},\mathbf{y}) \in \mathbf{K}\}\\ \mathbf{D}_f &:= \{\mathbf{x} \in \mathbf{B} \quad : \quad f(\mathbf{x},\mathbf{y}) \leq 0 \text{ for some } \mathbf{y} \text{ such that } (\mathbf{x},\mathbf{y}) \in \mathbf{K}\}.\end{aligned}$$

The former set $\mathbf{R}_f$ is particularly useful to qualify "robust" decisions $\mathbf{x}$ versus noise parameter $\mathbf{y}$ (e.g. in robust optimization) whereas the latter set $\mathbf{D}_f$ (a projection) is useful in design optimization when one does not want to work with its lifted representation $\{(\mathbf{x},\mathbf{y}) \in \mathbf{K} : f(\mathbf{x},\mathbf{y}) \geq 0\}$. We provide a systematic procedure to obtain a sequence of explicit inner (resp. outer) approximations that converge to $\mathbf{R}_f$ (resp. $\mathbf{D}_f$) in a strong sense. An additional feature is that each approximation is the sublevel set of a single polynomial whose vector of coefficients is an optimal solution of a semidefinite program. Several extensions are also proposed, and in particular, approximations for sets of the form

$$\mathbf{R}_F := \{\mathbf{x} \in \mathbf{B} \,:\, (\mathbf{x},\mathbf{y}) \in F \text{ for all } \mathbf{y} \text{ such that } (\mathbf{x},\mathbf{y}) \in \mathbf{K}\}$$

where F is some other basic-semi algebraic set, and also sets defined with two quantifiers.

1. Introduction

Consider two sets of variables $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ coupled with a constraint $(\mathbf{x},\mathbf{y}) \in \mathbf{K}$, where $\mathbf{K} \subset \mathbb{R}^n \times \mathbb{R}^m$ is some compact basic semi-algebraic set defined by:

$$\mathbf{K} := \{(\mathbf{x},\mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^m \,:\, \mathbf{x} \in \mathbf{B}; \quad g_j(\mathbf{x},\mathbf{y}) \geq 0,\ j = 1,\dots,s\} \tag{1.1}$$

for some polynomials g_j, $j = 1,\dots,s$, and let $\mathbf{B} \subset \mathbb{R}^n$ be a simple set (e.g. some box or ellipsoid).

With $f : \mathbf{K} \to \mathbb{R}$ a given semi-algebraic function on $\mathbf{K}$ (that is, its graph $\{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in \mathbf{K}\}$ is a semi-algebraic set), and

$$\mathbf{K}_{\mathbf{x}} := \{\mathbf{y} \in \mathbb{R}^m \,:\, (\mathbf{x},\mathbf{y}) \in \mathbf{K}\}, \tag{1.2}$$

consider the two sets:

$$\mathbf{R}_f := \{\, \mathbf{x} \in \mathbf{B} \,:\, f(\mathbf{x},\mathbf{y}) \leq 0 \text{ for all } \mathbf{y} \in \mathbf{K}_{\mathbf{x}} \,\}, \tag{1.3}$$

This work was partially supported by a grant from the Gaspar Monge Program on Optimization and Operations Research (PGMO) of the "Fondation mathématique Jacques Hadamard" (France). It was completed in 2013 while the author was a Visiting Fellow at the Isaac Newton Institute (Cambridge, UK) of which financial support is also gratefully acknowledged.

and

$$\mathbf{D}_f := \{ \mathbf{x} \in \mathbf{B} : f(\mathbf{x},\mathbf{y}) \leq 0 \text{ for some } \mathbf{y} \in \mathbf{K}_{\mathbf{x}} \}. \tag{1.4}$$

Both sets $\mathbf{R}_f$ and $\mathbf{D}_f$ which include a *quantifier* in their definition, are semi-algebraic and are interpreted as *robust* sets of variables $\mathbf{x}$ with respect to the other set of variables $\mathbf{y}$, and to some performance criterion f.

Indeed, in the first case (1.3) one may think of "$\mathbf{x}$" as *decision* variables which should be *robust* with respect to some *noise* (or perturbation) $\mathbf{y}$ in the sense that no matter what the admissible level of noise $\mathbf{y} \in \mathbf{K}_{\mathbf{x}}$ is, the constraint $f(\mathbf{x},\mathbf{y}) \leq 0$ is satisfied whenever $\mathbf{x} \in \mathbf{R}_f$. For instance, such sets $\mathbf{R}_f$ are fundamental in robust control.

On the other hand, in the second case (1.4) the vector $\mathbf{x}$ should be interpreted as *design* variables (or parameters), and the set $\mathbf{K}_{\mathbf{x}}$ defines a set of admissible decisions $\mathbf{y} \in \mathbf{K}_{\mathbf{x}}$ within the framework of design $\mathbf{x}$. And so $\mathbf{D}_f$ is the set of *robust* design parameters $\mathbf{x}$, in the sense that for every value of the design parameter $\mathbf{x} \in \mathbf{D}_f$, there is at least one admissible decision $\mathbf{y} \in \mathbf{K}_{\mathbf{x}}$ with performance level $f(\mathbf{x},\mathbf{y}) \leq 0$. Notice that $\mathbf{D}_{-f} \supseteq \overline{\mathbf{B} \setminus \mathbf{R}_f}$, and in a sense robust optimization is dual to design optimization.

The semi-algebraic function f as well as the set $\mathbf{K}$ can be fairly complicated and therefore in general both sets $\mathbf{R}_f$ and $\mathbf{D}_f$ are non convex so that their exact description can be fairly complicated as well! Needless to say that robust optimization problems with constraints of the form $\mathbf{x} \in \mathbf{R}_f$, are very difficult solve. In principle when $\mathbf{K}$ is a basic semi-algebraic set[1], quantifier elimination is possible via algebraic techniques; see e.g. Bochnak et al. [2]. However, in practice quantifier elimination is very costly and untractable.

On the other hand, design optimization problems with a constraint of the form $\mathbf{x} \in \mathbf{D}_f$ can be formulated directly in the lifted space of variables $(\mathbf{x},\mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ (i.e. by adding the constraints $f(\mathbf{x},\mathbf{y}) \leq 0$; $(\mathbf{x},\mathbf{y}) \in \mathbf{K}$) and so with no approximation. But sometimes one may be interested in getting a description of the set $\mathbf{D}_f$ itself in $\mathbb{R}^n$ because its "shape" is hidden in the lifted $(\mathbf{x},\mathbf{y})$-description, or because optimizing over $\mathbf{K} \cap \{(\mathbf{x},\mathbf{y}) : f(\mathbf{x},\mathbf{y}) \leq 0\}$ may not be practical. However, if the projection of a basic semi-algebraic set (like e.g. $\mathbf{D}_f$) is semi-algebraic, it is not necessarily *basic* semi-algebraic and could be a complicated union of several basic semi-algebraic sets (hence not very useful in practice). So in this case one wishes to obtain a relatively simple and tractable description of $\mathbf{D}_f$.

So a less ambitious but more practical goal is to obtain *tractable* approximations of such sets $\mathbf{R}_f$ (or $\mathbf{D}_f$). Then such approximations can be used for various purposes, optimization being only one potential application.

Contribution. In this paper we provide a hierarchy $(\mathbf{R}_f^k)$ (resp. $(\mathbf{D}_f^k)$), $k \in \mathbb{N}$, of *inner* approximations for $\mathbf{R}_f$ (resp. *outer* approximations for $\mathbf{D}_f$). These two hierarchies have three essential characteristic features:

[1]A basic semi-algebraic set is the intersection $\cap_{j=1}^m \{\mathbf{x} : g_j(\mathbf{x}) \geq 0\}$ of super level sets of finitely many polynomials $(g_j) \subset \mathbb{R}[\mathbf{x}]$.

- (a) Each set $\mathbf{R}_f^k \subset \mathbb{R}^n$ (resp. $\mathbf{D}_f^k$), $k \in \mathbb{N}$, has a very simple description in terms of the sublevel set $\{\mathbf{x} : \overline{p_k}(\mathbf{x}) \leq 0\}$ (resp. $\{\mathbf{x} : \underline{p_k}(\mathbf{x}) \leq 0\}$) associated with a *single* polynomial $\overline{p_k}$ (resp. $\underline{p_k}$).

- (b) Both hierarchies $(\mathbf{R}_f^k)$ and $(\mathbf{D}_f^k)$, $k \in \mathbb{N}$, converge in a strong sense since we prove that (under some conditions) $\mathrm{vol}\,(\mathbf{R}_f \setminus \mathbf{R}_f^k) \to 0$ (resp. $\mathrm{vol}\,(\mathbf{D}_f \setminus \mathbf{D}_f^k) \to 0$) as $k \to \infty$ (and where "vol$(\cdot)$" denotes the Lebesgue volume). In other words, for k sufficiently large, the inner approximations $\mathbf{R}_f^k$ (resp. outer approximations $\mathbf{D}_f^k$) coincide with $\mathbf{R}_f$ (resp. $\mathbf{D}_f$) up to a set of very small Lebesgue measure.

- (c) Computing the vector of coefficients of the above polynomial $\overline{p_k}$ (resp. $\underline{p_k}$) reduces to solving a semidefinite program whose size is parametrized by k.

Hence for instance, the constraint $\overline{p_k}(\mathbf{x}) \leq 0$ (resp. $\underline{p_k}(\mathbf{x}) \leq 0$) can be used in any robust (resp. design) polynomial optimization problem on $\mathbf{B}$, as a substitute for $\mathbf{x} \in \mathbf{R}_f$ (resp. $\mathbf{x} \in \mathbf{D}_f$), thereby eliminating the variables $\mathbf{y}$. One then obtains a standard polynomial optimization problem $\mathbf{P}$ for which one may apply the hierarchy of semidefinite relaxations defined in [9] to obtain a sequence of lower bounds (and sometimes an optimal solution if the size of the resulting is moderate or if some sparsity pattern can be used for larger size problems). For mored details, the interested reader is refereed to [9] (and Waki et al [12] for semidefinite relaxations that use a sparsity pattern). But the sets $\mathbf{R}_f^k$ can also be used in other applications to provide a certificate for robustness as membership in $\mathbf{R}_f^k$ is easy to check and the approximation is from inside.

We first obtain inner (resp. outer) approximations of $\mathbf{R}_f$ (resp. $\mathbf{D}_f$) when f is a *polynomial*. To do so we extensively use a previous result of the author [8] which allows to approximate in a strong sense the optimal value of a parametric optimization problem. We then extend the methodology to the case where f is a semi-algebraic function on $\mathbf{K}$, whose graph Ψ_f is explicitly described by a basic semi-algebraic set[2]. This methodology had been already used in Henrion and Lasserre [3] to provide (convergent) inner approximations for the particular case of a set defined by matrix polynomial inequalities. The present contribution can be viewed as an extension of [3] to the more general framework (1.3)-(1.4) and with f semi-algebraic.

Finally, we also provide several extensions, and in particular, we consider:

- The case where one also enforces the computed inner or outer approximations to be a *convex* set. This can be interesting for optimization purposes but of course, in this case convergence as in (b) is lost.

- The case where $f(\mathbf{x}, \mathbf{y}) \leq 0$ is now replaced with a polynomial matrix inequality $\mathbf{F}(\mathbf{x}, \mathbf{y}) \preceq 0$, i.e., $\mathbf{F}(\cdot, \cdot)$ is a real symmetric $m \times m$ matrix such that $\mathbf{F}_{ij} \in \mathbb{R}[\mathbf{x}, \mathbf{y}]$ for each entry (i, j). A converging hierarchy of inner approximations for $\mathbf{R}_f$ has been already provided in [3].

- The case where $\mathbf{R}_f$ is now replaced with the set $\mathbf{R}_F$ defined by:

$$\mathbf{R}_F = \{\mathbf{x} \in \mathbf{B} : (\mathbf{x}, \mathbf{y}) \in F \text{ for all } \mathbf{y} \text{ such that } (\mathbf{x}, \mathbf{y}) \in \mathbf{K}\},$$

where F is some basic-semi-algebraic set. And a similarly extension is also possible for sets $\mathbf{D}_f$ defined accordingly.

[2] That is, $\Psi_f = \{((\mathbf{x}, \mathbf{y}), f(\mathbf{x}, \mathbf{y}))) : (\mathbf{x}, \mathbf{y}) \in \mathbf{K}\} = \{(\mathbf{x}, \mathbf{y}, v_r) : (\mathbf{x}, \mathbf{y}) \in \mathbf{K};\ h_\ell(\mathbf{x}, \mathbf{y}, \mathbf{v}) \geq 0,\ \ell = 1, \ldots, s\}$, for some polynomials $h_\ell \in \mathbb{R}[\mathbf{x}, \mathbf{y}, \mathbf{v}]$, and where $\mathbf{v} \in \mathbb{R}^r$.

- The case where we now how have two quantifiers, so that for instance,

$$\mathbf{R}_f \,=\, \{\mathbf{x}\in\mathbf{B}_\mathbf{x}\,:\,\exists\,\mathbf{y}\in\mathbf{B}_\mathbf{y}\text{ s.t. } f(\mathbf{x},\mathbf{y},\mathbf{u})\leq 0,\,\forall \mathbf{u}:\,(\mathbf{x},\mathbf{y},\mathbf{u})\in\mathbf{K}\},$$

for some boxes $\mathbf{B}_\mathbf{x}\subset\mathbb{R}^n$, $\mathbf{B}_\mathbf{y}\subset\mathbb{R}^m$, and some compact set $\mathbf{K}\subset\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^s$.

2. Notation and definitions

Let $\mathbb{R}[\mathbf{x}]$ denote the ring or real polynomials in the variables $\mathbf{x}=(x_1,\ldots,x_n)$, and let $\mathbb{R}[\mathbf{x}]_d$ be the vector space of real polynomials of degree at most d. Similarly, let $\Sigma[\mathbf{x}]\subset\mathbb{R}[\mathbf{x}]$ denote the convex cone of real polynomials that are sums of squares (SOS) of polynomials, and $\Sigma[\mathbf{x}]_d\subset\Sigma[\mathbf{x}]$ its subcone of SOS polynomials of degree at most $2d$. Denote by $\mathcal{S}^m$ the space of $m\times m$ real symmetric matrices. For a given matrix $\mathbf{A}\in\mathcal{S}^m$, the notation $\mathbf{A}\succeq 0$ (resp. $\mathbf{A}\succ 0$) means that $\mathbf{A}$ is positive semidefinite (resp. positive definite), i.e., all its eigenvalues are real and nonnegative (resp. positive).

Moment matrix. With $\mathbf{z}=(z_\alpha)$ being a sequence indexed in the canonical basis $(\mathbf{x}^\alpha)$ of $\mathbb{R}[\mathbf{x}]$, let $L_\mathbf{z}:\mathbb{R}[\mathbf{x}]\to\mathbb{R}$ be the so-called Riesz functional defined by:

$$f\quad(=\sum_\alpha f_\alpha\,\mathbf{x}^\alpha)\quad\mapsto\quad L_\mathbf{z}(f)\,=\,\sum_\alpha f_\alpha\,z_\alpha,$$

and let $\mathbf{M}_d(\mathbf{z})$ be the symmetric matrix with rows and columns indexed in the canonical basis $(\mathbf{x}^\alpha)$, and defined by:

$$\mathbf{M}_d(\mathbf{z})(\alpha,\beta)\,:=\,L_\mathbf{z}(\mathbf{x}^{\alpha+\beta})\,=\,z_{\alpha+\beta},\quad \alpha,\beta\in\mathbb{N}^n_d \tag{2.1}$$

with $\mathbb{N}^n_d:=\{\alpha\in\mathbb{N}^n\,:\,|\alpha|\,(=\sum_i\alpha_i)\leq d\}$.

If $\mathbf{z}$ has a representing measure μ, i.e., if $z_\alpha=\int\mathbf{x}^\alpha d\mu$ for every $\alpha\in\mathbb{N}^n$, then

$$\langle\mathbf{f},\mathbf{M}_d(\mathbf{z})\mathbf{f}\rangle\,=\,\int f(\mathbf{x})^2\,d\mu(\mathbf{x})\,\geq\,0,\qquad\forall\,f\in\mathbb{R}[\mathbf{x}]_d,$$

and so $\mathbf{M}_d(\mathbf{z})\succeq 0$. In particular, if μ has a density h with respect to the Lebesgue measure, positive on some open set B, then $\mathbf{M}_d(\mathbf{z})\succ 0$ because

$$0\,=\,\langle\mathbf{f},\mathbf{M}_d(\mathbf{z})\mathbf{f}\rangle\,\geq\,\int_B f(\mathbf{x})^2\,h(\mathbf{x})d\mathbf{x}\quad\Rightarrow\quad f=0.$$

Localizing matrix. Similarly, with $\mathbf{z}=(z_\alpha)$ and $g\in\mathbb{R}[\mathbf{x}]$ written

$$\mathbf{x}\mapsto g(\mathbf{x})\,=\,\sum_{\gamma\in\mathbb{N}^n} g_\gamma\,\mathbf{x}^\gamma,$$

let $\mathbf{M}_d(g\,\mathbf{y})$ be the symmetric matrix with rows and columns indexed in the canonical basis $(\mathbf{x}^\alpha)$, and defined by:

$$\mathbf{M}_d(g\,\mathbf{z})(\alpha,\beta)\,:=\,L_\mathbf{z}\left(g(\mathbf{x})\,\mathbf{x}^{\alpha+\beta}\right)\,=\,\sum_\gamma g_\gamma\,z_{\alpha+\beta+\gamma},\qquad\forall\,\alpha,\beta\in\mathbb{N}^n_d. \tag{2.2}$$

If $\mathbf{z}$ has a representing measure μ, then $\langle\mathbf{f},\mathbf{M}_d(g\,\mathbf{z})\mathbf{f}\rangle=\int f^2 g d\mu$, and so if μ is supported on the set $\{\mathbf{x}\,:\,g(\mathbf{x})\geq 0\}$, then $\mathbf{M}_d(g\,\mathbf{z})\succeq 0$ for all $d=0,1,\ldots$ because

$$\langle\mathbf{f},\mathbf{M}_d(g\,\mathbf{z})\mathbf{f}\rangle\,=\,\int f(\mathbf{x})^2 g(\mathbf{x})\,d\mu(\mathbf{x})\,\geq\,0,\qquad\forall\,f\in\mathbb{R}[\mathbf{x}]_d. \tag{2.3}$$

In particular, if μ is the Lebesgue measure and g is positive on some open set B, then $\mathbf{M}_d(g\,\mathbf{z}) \succ 0$ because

$$0 \,=\, \langle \mathbf{f}, \mathbf{M}_d(g\,\mathbf{z})\mathbf{f}\rangle \,\geq\, \int_B f(\mathbf{x})^2\, g(\mathbf{x})d\mathbf{x} \quad \Rightarrow \quad f=0.$$

3. Main result

Let $\mathbf{K}$ be the basic semi-algebraic set defined in (1.1) for some polynomials $g_j \subset \mathbb{R}[\mathbf{x},\mathbf{y}]$, $j=1,\ldots,s$, and with simple set (box or ellipsoid) $\mathbf{B}\subset\mathbb{R}^n$.

Denote by $L_1(\mathbf{B})$ the Lebesgue space of measurable functions $h:\mathbf{B}\to\mathbb{R}$ that are integrable with respect to the Lebesgue measure on $\mathbf{B}$, i.e., such that $\int_{\mathbf{B}}|h|d\mathbf{x}<\infty$.

Given $f\in\mathbb{R}[\mathbf{x},\mathbf{y}]$, consider the mappings $\overline{J_f}:\mathbf{B}\to\mathbb{R}\cup\{-\infty\}$ and $\underline{J_f}:\mathbf{B}\to\mathbb{R}\cup\{+\infty\}$, respectively defined by:

$$\mathbf{x}\mapsto\overline{J_f}(\mathbf{x}) \quad := \quad \sup_{\mathbf{y}}\{f(\mathbf{x},\mathbf{y})\,:\,\mathbf{y}\in\mathbf{K}_{\mathbf{x}}\},\qquad \mathbf{x}\,\in\,\mathbf{B}. \tag{3.1}$$

$$\mathbf{x}\mapsto\underline{J_f}(\mathbf{x}) \quad := \quad \inf_{\mathbf{y}}\{f(\mathbf{x},\mathbf{y})\,:\,\mathbf{y}\in\mathbf{K}_{\mathbf{x}}\},\qquad \mathbf{x}\,\in\,\mathbf{B}. \tag{3.2}$$

The function $\overline{J_f}$ (resp. $\underline{J_f}$) is upper (resp. lower) semi-continuous. We will need the following intermediate result.

Theorem 3.1. *Let $\mathbf{K}\subset\mathbb{R}^n\times\mathbb{R}^m$ be compact. If $\mathbf{K}_{\mathbf{x}}\neq\emptyset$ for every $\mathbf{x}\in\mathbf{B}$, there exists a sequence of polynomials $(\overline{p_k})\subset\mathbb{R}[\mathbf{x}]$ (resp. $(\underline{p_k})\subset\mathbb{R}[\mathbf{x}]$), $k\in\mathbb{N}$, such that $\overline{p_k}(\mathbf{x})\geq f(\mathbf{x},\mathbf{y})$ (resp. $\underline{p_k}(\mathbf{x})\leq f(\mathbf{x},\mathbf{y})$) for all $\mathbf{y}\in\mathbf{K}_{\mathbf{x}}$, $\mathbf{x}\in\mathbf{B}$, and such that*

$$\lim_{k\to\infty}\int_{\mathbf{B}}|\overline{p_k}(\mathbf{x})-\overline{J_f}(\mathbf{x})|\,d\mathbf{x} \quad = \quad 0 \quad \text{[Convergence in } L_1(\mathbf{B})\text{]} \tag{3.3}$$

$$\lim_{k\to\infty}\int_{\mathbf{B}}|\underline{J_f}(\mathbf{x})-\underline{p_k}(\mathbf{x})\,|\,d\mathbf{x} \quad = \quad 0 \quad \text{[Convergence in } L_1(\mathbf{B})\text{]} \tag{3.4}$$

Proof. To prove (3.3) observe that $\overline{J_f}$ being bounded and upper semi-continuous on $\mathbf{B}$, there exists a nonincreasing sequence (f_k), $k\in\mathbb{N}$, of bounded continuous functions $f_k:\mathbf{B}\to\mathbb{R}$ such that $f_k(\mathbf{x})\downarrow\overline{J_f}(\mathbf{x})$ for all $\mathbf{x}\in\mathbf{B}$, as $k\to\infty$. Moreover, as $k\to\infty$, by the Monotone Convergence Theorem:

$$\int_{\mathbf{B}} f_k(\mathbf{x})\,d\mathbf{x}\;\to\;\int_{\mathbf{B}}\overline{J_f}(\mathbf{x})\,d\mathbf{x}\qquad\text{as } k\to\infty,$$

and so

$$\int_{\mathbf{B}}|f_k(\mathbf{x})-\overline{J_f}(\mathbf{x})|\,d\mathbf{x}\,=\,\int_{\mathbf{B}}(f_k(\mathbf{x})-\overline{J_f}(\mathbf{x}))\,d\mathbf{x}\;\to\;0\qquad\text{as } k\to\infty,$$

that is, $f_k\to\overline{J_f}$ for the $L_1(\mathbf{B})$-norm. Next, by the Stone-Weierstrass theorem, for every $k\in\mathbb{N}$, there exists $p_k\in\mathbb{R}[\mathbf{x}]$ such that $\sup_{\mathbf{x}\in\mathbf{B}}|p_k-f_k|<(2k)^{-1}$ and so $\overline{p_k}:=p_k+k^{-1}\geq f_k\geq\overline{J_f}$ on $\mathbf{B}$. In addition,

$$\begin{aligned}\lim_{k\to\infty}\int_{\mathbf{B}}|\overline{p_k}(\mathbf{x})-\overline{J_f}(\mathbf{x})|\,d\mathbf{x} \quad &= \quad \lim_{k\to\infty}\int_{\mathbf{B}}\underbrace{|\overline{p_k}(\mathbf{x})-f_k(\mathbf{x})|}_{\leq k^{-1}}+|f_k(\mathbf{x})-\overline{J_f}(\mathbf{x})|\,d\mathbf{x}\\ &\leq \quad \lim_{k\to\infty}\left(k^{-1}\mathrm{vol}(\mathbf{B})+\int_{\mathbf{B}}|f_k(\mathbf{x})-\overline{J_f}(\mathbf{x})|\,d\mathbf{x}\right)\\ &\leq \quad \lim_{k\to\infty}\int_{\mathbf{B}}|f_k(\mathbf{x})-\overline{J_f}(\mathbf{x})|\,d\mathbf{x}\,=\,0\end{aligned}$$

□

3.1. **Robust optimization.** Let $\mathbf{R}_f$ be as in (1.3) and let $\mathbf{B} \subset \mathbb{R}^n$ be the set in Theorem 3.1, assumed to have nonempty interior.

Theorem 3.2. *Let $\mathbf{K} \subset \mathbb{R}^n \times \mathbb{R}^m$ in (1.1) be compact and $\mathbf{K}_{\mathbf{x}} \neq \emptyset$ for every $\mathbf{x} \in \mathbf{B}$. Assume that $\{\mathbf{x} \in \mathbf{B} : \overline{J_f}(\mathbf{x}) = 0\}$ has Lebesgue measure zero.*

Let $\mathbf{R}_f^k := \{\mathbf{x} \in \mathbf{B} : \overline{p_k}(\mathbf{x}) \leq 0\}$, where $\overline{p_k} \in \mathbb{R}[\mathbf{x}]$ is as in Theorem 3.1. Then $\mathbf{R}_f^k \subset \mathbf{R}_f$ for every k, and

$$\text{vol}\left(\mathbf{R}_f \setminus \mathbf{R}_f^k\right) \to 0 \quad \textit{as } k \to \infty. \tag{3.5}$$

Proof. By Theorem 3.1 $\overline{p_k} \to \overline{J_f}$ in $L_1(\mathbf{B})$. Therefore by [1, Theorem 2.5.1], $\overline{g_k}$ converges to $\overline{J_f}$ in measure, that is, for every $\epsilon > 0$,

$$\lim_{k\to\infty} \text{vol}\left(\{\, \mathbf{x} : |\overline{p_k}(\mathbf{x}) - \overline{J_f}(\mathbf{x})| \geq \epsilon\}\right) = 0. \tag{3.6}$$

Next, as $\overline{J_f}$ is upper semi-continuous on $\mathbf{B}$, the set $\{\mathbf{x} : \overline{J_f}(\mathbf{x}) < 0\}$ is open and as the set $\{\mathbf{x} \in \mathbf{B} : \overline{J_f}(\mathbf{x}) = 0\}$ has Lebesgue measure zero,

$$\begin{aligned}
\text{vol}(\mathbf{R}_f) = \text{vol}\left(\{\mathbf{x} \in \mathbf{B} : \overline{J_f}(\mathbf{x}) < 0\}\right) &= \text{vol}\left(\bigcup_{\ell=1}^{\infty} \{\mathbf{x} \in \mathbf{B} : \overline{J_f}(\mathbf{x}) \leq -1/\ell\}\right) \\
&= \lim_{\ell\to\infty} \text{vol}\left(\{\mathbf{x} \in \mathbf{B} : \overline{J_f}(\mathbf{x}) \leq -1/\ell\}\right) \\
&= \lim_{\ell\to\infty} \text{vol}\left(\mathbf{R}_f(\ell)\right),. \qquad (3.7)
\end{aligned}$$

where $\mathbf{R}_f(\ell) := \{\mathbf{x} \in \mathbf{B} : \overline{J_f}(\mathbf{x}) \leq -1/\ell\}$. Next, $\mathbf{R}_f(\ell) \subseteq \mathbf{R}_f$ for every $\ell \geq 1$, and

$$\text{vol}\left(\mathbf{R}_f(\ell)\right) = \text{vol}\left(\mathbf{R}_f(\ell) \cap \{\mathbf{x} : \overline{p_k}(\mathbf{x}) > 0\}\right) + \text{vol}\left(\mathbf{R}_f(\ell) \cap \{\mathbf{x} : \overline{p_k}(\mathbf{x}) \leq 0\}\right).$$

Observe that by (3.6), $\text{vol}\left(\mathbf{R}_f(\ell) \cap \{\mathbf{x} : \overline{p_k}(\mathbf{x}) > 0\}\right) \to 0$ as $k \to \infty$. Therefore,

$$\begin{aligned}
\text{vol}(\mathbf{R}_f(\ell)) &= \lim_{k\to\infty} \text{vol}\left(\mathbf{R}_f(\ell) \cap \underbrace{\{\mathbf{x} : \overline{p_k}(\mathbf{x}) \leq 0\}}_{=\mathbf{R}_f^k}\right) \qquad (3.8) \\
&\leq \lim_{k\to\infty} \text{vol}\left(\mathbf{R}_f^k\right) \leq \text{vol}\left(\mathbf{R}_f\right).
\end{aligned}$$

As $\mathbf{R}_f^k \subset \mathbf{R}_f$ for all k, letting $\ell \to \infty$ and using (3.7) yields the desired result. □

3.2. **Design optimization.** Let $\mathbf{D}_f$ be as in (1.4) and let $\mathbf{B} \subset \mathbb{R}^n$ be the set in Theorem 3.1, assumed to have nonempty interior.

Corollary 3.3. *Let $\mathbf{K} \subset \mathbb{R}^n \times \mathbb{R}^m$ in (1.1) be compact and $\mathbf{K}_{\mathbf{x}} \neq \emptyset$ for every $\mathbf{x} \in \mathbf{B}$. Assume that $\{\mathbf{x} \in \mathbf{B} : \underline{J_f}(\mathbf{x}) = 0\}$ has Lebesgue measure zero.*

Let $\mathbf{D}_f^k := \{\mathbf{x} \in \mathbf{B} : \underline{p_k}(\mathbf{x}) \leq 0\}$, where $\underline{p_k} \in \mathbb{R}[\mathbf{x}]$ is as in Theorem 3.1. Then $\mathbf{D}_f^k \supset \mathbf{D}_f$ for every k, and

$$\text{vol}(\mathbf{D}_f \setminus \mathbf{D}_f^k) \to 0 \quad \textit{as } k \to \infty. \tag{3.9}$$

Proof. The proof uses same arguments as in the proof of Theorem 3.2 with $-f$ in lieu of f. Indeed, $\mathbf{D}_f = \mathbf{B} \setminus \Delta_f$ with

$$\begin{aligned}\Delta_f &:= \{\mathbf{x} \in \mathbf{B} \,:\, -f(\mathbf{x},\mathbf{y}) \,<\, 0 \text{ for all } \mathbf{y} \in \mathbf{K}_\mathbf{x}\}\\ &= \{\mathbf{x} \in \mathbf{B} \,:\, \sup_\mathbf{y}\{-f(\mathbf{x},\mathbf{y}) \,:\, \mathbf{y} \in \mathbf{K}_\mathbf{x}\} \,<\, 0\}\\ &= \{\mathbf{x} \in \mathbf{B} \,:\, \overline{J_{-f}}(\mathbf{x}) \,<\, 0\} \,=\, \{\mathbf{x} \in \mathbf{B} \,:\, \underline{J_f}(\mathbf{x}) \,>\, 0\}\end{aligned}$$

and since $\{\mathbf{x} \in \mathbf{B} \,:\, \underline{J_f}(\mathbf{x}) \,=\, 0\}$ has Lebesgue measure zero,

$$\mathrm{vol}(\Delta_f) = \mathrm{vol}\left(\{\mathbf{x} \in \mathbf{B} \,:\, \overline{J_{-f}}(\mathbf{x}) \,\leq\, 0\}\right) \,=\, \mathrm{vol}\left(\{\mathbf{x} \in \mathbf{B} \,:\, \underline{J_f}(\mathbf{x}) \,\geq\, 0\}\right).$$

Hence by Theorem 3.2 applied to $-f$,

$$\lim_{k\to\infty} \mathrm{vol}\left(\{\mathbf{x} \in \mathbf{B} \,:\, \underline{p_k}(\mathbf{x}) \geq 0\}\right) \,=\, \mathrm{vol}\,(\Delta_f)\,,$$

which in turn implies the desired result

$$\begin{aligned}\lim_{k\to\infty} \mathrm{vol}\left(\{\mathbf{x} \in \mathbf{B} \,:\, \underline{p_k}(\mathbf{x}) \leq 0\}\right) &= \lim_{k\to\infty} \mathrm{vol}\left(\{\mathbf{x} \in \mathbf{B} \,:\, \underline{p_k}(\mathbf{x}) < 0\}\right)\\ &= \mathrm{vol}\,(\mathbf{B} \setminus \Delta_f) \,=\, \mathrm{vol}\,(\mathbf{D}_f)\,.\end{aligned}$$

because $\mathrm{vol}\left(\{\mathbf{x} \in \mathbf{B} : \underline{p_k}(\mathbf{x}) = 0\}\right) = 0$ for every k. □

Hence for robust polynomial optimization problems where one wishes to optimize over the set $\mathbf{R}_f$, one may reinforce the complicated (and untractable) constraint $\mathbf{x} \in \mathbf{R}_f$ by instead considering the inner approximation obtained with the two much simpler constraints $\mathbf{x} \in \mathbf{B}$ and $\overline{p_k}(\mathbf{x}) \geq 0$. Similarly, for design problems where one wishes to work with $\mathbf{D}_f$ and not its lifted representation $\{f(\mathbf{x},\mathbf{y}) \geq 0;\, (\mathbf{x},\mathbf{y}) \in \mathbf{K}\}$, one may instead use the outer approximation $\{\mathbf{x} \,:\, \underline{p_k}(\mathbf{x}) \leq 0\}$. In both cases, if k is sufficiently large then the resulting conservatism (resp. laxism) introduced by these respective approximations is negligible.

3.3. **Practical computation.** In this section we follow [8] and show how to compute a sequence of polynomials $(\overline{p_k}) \subset \mathbb{R}[\mathbf{x}]$, $k \in \mathbb{N}$, as defined in Theorem 3.2. (As expected, a similar procedure also applies to compute a sequence of polynomials $(\underline{p_k}) \subset \mathbb{R}[\mathbf{x}]$, $k \in \mathbb{N}$, as defined in Corollary 3.3.)

With $\mathbf{K} \subset \mathbb{R}^n \times \mathbb{R}^m$ as in (1.1) and compact, we assume that we know some $M > 0$ such that $M - \|\mathbf{y}\|^2 \geq 0$ whenever $(\mathbf{x},\mathbf{y}) \in \mathbf{K}$. Next, and possibly after re-scaling of the g_j's, we may and will set $M = 1$, $\mathbf{B} = [-1,1]^n$. Next, let

$$(3.10) \qquad \gamma_\alpha \,:=\, \int_\mathbf{B} \mathbf{x}^\alpha \, d\lambda(\mathbf{x}) \,=:=\, \mathrm{vol}(\mathbf{B})^{-1} \int_\mathbf{B} \mathbf{x}^\alpha \, d\mathbf{x}, \qquad \alpha \in \mathbb{N}^n,$$

be the moments of the (scaled) Lebesgue measure λ on $\mathbf{B}$, which are easy to compute. Moreover, letting $g_{s+1}(\mathbf{y}) := 1 - \|\mathbf{y}\|^2$, and $x_i \mapsto \theta_i(\mathbf{x}) := 1 - x_i^2$, $i = 1,\ldots,n$, for convenience we redefine $\mathbf{K} \subset \mathbb{R}^n \times \mathbb{R}^m$ to be the basic semi-algebraic set

$$(3.11) \quad \mathbf{K} \,=\, \{\,(\mathbf{x},\mathbf{y}) \,:\, g_j(\mathbf{x},\mathbf{y}) \,\geq\, 0, \quad j = 1,\ldots,s+1;\, \theta_i(\mathbf{x}) \geq 0,\, i = 1,\ldots,n\}.$$

With $v_j := \lceil \deg(g_j)/2 \rceil$, $j = 0, \ldots, m$, and for fixed $k \geq \max_j [v_j]$, consider the following optimization problem

$$\begin{array}{rll} \rho_k = \displaystyle\min_{p,\sigma_j,\psi_i} & \displaystyle\int_{\mathbf{B}} p(\mathbf{x})\, d\lambda(\mathbf{x}) & \\ \text{s.t.} & p(\mathbf{x}) - f(\mathbf{x},\mathbf{y}) = \displaystyle\sum_{j=0}^{s+1} \sigma_j(\mathbf{x},\mathbf{y})\, g_j(\mathbf{x},\mathbf{y}) + \sum_{i=1}^{n} \psi_i(\mathbf{x},\mathbf{y})\, \theta_i(\mathbf{x}) & \\ & p \in \mathbb{R}[\mathbf{x}]_{2k};\ \sigma_j \in \Sigma_{k-v_j}[\mathbf{x},\mathbf{y}],\ j = 0, \ldots, s+1 & \\ & \psi_i \in \Sigma_{k-1}[\mathbf{x},\mathbf{y}],\ i = 1, \ldots, n. & \end{array} \tag{3.12}$$

The above optimization problem (3.12) is a semidefinite program. Indeed :
- The criterion $\int_{\mathbf{B}} p(\mathbf{x})\, d\mathbf{x}$ is linear in the coefficients $\mathbf{p} = (p_\alpha)$, $\alpha \in \mathbb{N}^n_{2k}$, of the unknown polynomial $p \in \mathbb{R}[\mathbf{x}]_k$. In fact,

$$\int_{\mathbf{B}} p(\mathbf{x})\, d\mathbf{x} \;=\; \sum_{\alpha \in \mathbb{N}^n_{2k}} p_\alpha \underbrace{\int_{\mathbf{B}} \mathbf{x}^\alpha\, d\lambda(\mathbf{x})}_{\gamma_\alpha} \;=\; \sum_{\alpha \in \mathbb{N}^n_{2k}} p_\alpha\, \gamma_\alpha.$$

- The constraint

$$p(\mathbf{x}) - f(\mathbf{x},\mathbf{y}) = \sum_{j=0}^{s+1} \sigma_j(\mathbf{x},\mathbf{y})\, g_j(\mathbf{x},\mathbf{y}) + \sum_{i=1}^{n} \psi_i(\mathbf{x},\mathbf{y})\, \theta_i(\mathbf{x}),$$

with $p \in \mathbb{R}[\mathbf{x}]_{2k}$; $\sigma_j \in \Sigma_{k-v_j}[\mathbf{x},\mathbf{y}]$, $j = 0, \ldots, s$, and $\psi_i \in \Sigma_{k-1}[\mathbf{x},\mathbf{y}]$, $k = 1, \ldots, n$, reduces to

- linear equality constraints between the coefficients of the polynomials p, σ_j and ψ_i, to satisfy the identity, and
- Linear Matrix Inequality (LMI) constraints to ensure that σ_j and ψ_i are all SOS polynomials of degree bounded by $2(k - v_j)$ and $2(k-1)$ respectively.

The dual of the semidefinite program (3.12) reads:

$$\begin{array}{rll} \rho^*_k = \displaystyle\min_{\mathbf{z}} & L_{\mathbf{z}}(f) & \\ \text{s.t.} & \mathbf{M}_{k-v_j}(g_j\, \mathbf{z}) \succeq 0, & j = 0, \ldots, s+1 \\ & \mathbf{M}_{k-1}(\theta_i\, \mathbf{z}) \succeq 0, & i = 1, \ldots, n \\ & L_{\mathbf{z}}(\mathbf{x}^\alpha) \,=\, \gamma_\alpha, & \alpha \in \mathbb{N}^n_{2k}, \end{array} \tag{3.13}$$

where $\mathbf{z} = (z_{\alpha\beta})$, $(\alpha, \beta) \in \mathbb{N}^{n+m}_{2k}$, and $L_{\mathbf{z}} : \mathbb{R}[\mathbf{x},\mathbf{y}] \to \mathbb{R}$ is the Riesz functional introduced in §2. Similarly, $\mathbf{M}_k(g_j\, \mathbf{z})$ (resp. $\mathbf{M}_k(\theta_i\, \mathbf{z})$) is the localizing matrix associated with the sequence $\mathbf{z}$ and the polynomial g_j (resp. θ_i), also introduced in §2.

Next we extend [8, Theorem 3.5] and prove that both (3.12) and its dual (3.13) have an optimal solution whenever $\mathbf{K}$ has nonempty interior.

Theorem 3.4. *Let $\mathbf{K}$ be as in (3.11) with nonempty interior, and assume that $\mathbf{K}_{\mathbf{x}} \neq \emptyset$ for every $\mathbf{x} \in \mathbf{B}$.*

Then there is no duality gap between the semidefinite program (3.12) and its dual (3.13). Moreover (3.12) (resp. (3.13)) has an optimal solution $p^*_k \in \mathbb{R}[\mathbf{x}]_{2k}$ (resp. $\mathbf{z}^* = (z^*_{\alpha\beta})$, $(\alpha, \beta) \in \mathbb{N}^{n+m}_{2k}$). Moreover,

$$\lim_{k\to\infty} \int_{\mathbf{B}} |p^*_k(\mathbf{x}) - \overline{J_f}(\mathbf{x})|\, d\mathbf{x} \,=\, 0 \quad \textit{[Convergence in } L_1(\mathbf{B})\textit{]}. \tag{3.14}$$

Proof. As $\mathbf{K}$ has a nonempty interior it contains an open set $O\subset\mathbb{R}^n\times\mathbb{R}^m$. Let $O_{\mathbf{x}}\subset\mathbf{B}$ be the projection of O onto $\mathbf{B}$, so that its ($\mathbb{R}^m$) Lebesgue volume is positive. Let μ be the finite Borel measure on $\mathbf{K}$ defined by

$$\mu(A\times B)\,:=\,\int_A \phi(B\,|\,\mathbf{x})\,d\lambda(\mathbf{x}),\qquad A\in\mathcal{B}(\mathbb{R}^n),\;B\in\mathcal{B}(\mathbb{R}^m),$$

where for every $\mathbf{x}\in O_{\mathbf{x}}$, $\phi(d\mathbf{y}\,|\,\mathbf{x})$ is the probability measure on $\mathbb{R}^m$, supported on $\mathbf{K}_{\mathbf{x}}$, and defined by:

$$\phi(B\,|\,\mathbf{x})=\text{vol}\,(\mathbf{K}_{\mathbf{x}}\cap B)/\text{vol}(\mathbf{K}_{\mathbf{x}}),\qquad \forall B\in\mathcal{B}(\mathbb{R}^m)$$

(and where here vol$(\cdot)$ denotes the Lebesgue volume in $\mathbb{R}^m$). And on $\mathbf{B}\setminus O_{\mathbf{x}}$, the probability $\phi(d\mathbf{y}\,|\,\mathbf{x})$ is an arbitrary probability measure on $\mathbf{K}_{\mathbf{x}}$.

Let $\mathbf{z}=(z_{\alpha\beta})$, $(\alpha,\beta)\in\mathbb{N}^{n+m}_{2k}$, be the moments of μ. As $\mathbf{K}\supset O$, $\mathbf{M}_{k-v_j}(g_j\,\mathbf{z})\succ 0$ (resp. $\mathbf{M}_{k-1}(\theta_i\,\mathbf{z})\succ 0$) for $j=0,\ldots,s{+}1$ (resp. for $i=1,\ldots,n$). Indeed otherwise suppose that $\mathbf{M}_{k-v_j}(g_j\,\mathbf{z})\mathbf{u}=0$ for some non trivial vector $\mathbf{u}$. Then one obtains the contradiction

$$0\,=\,\langle\mathbf{u},\mathbf{M}_{d-v_j}(g_j\,\mathbf{z})\mathbf{u}\rangle\,=\,\int u(\mathbf{x})^2g_j(\mathbf{x})\,\mu(d\mathbf{x})\,>\,\int_O u(\mathbf{x})^2g_j(\mathbf{x})\,\mu(d\mathbf{x})\,>\,0.$$

Moreover, by construction of μ, its marginal on $\mathbf{B}$ is the (scaled) Lebesgue measure λ on $\mathbf{B}$ and so

$$L_{\mathbf{z}}(\mathbf{x}^\alpha)\,=\,\int_{\mathbf{B}}\mathbf{x}^\alpha\,d\lambda(\mathbf{x})\,=\,\gamma_\alpha,\qquad \alpha\in\mathbb{N}^n_{2k}.$$

In other words, $\mathbf{z}$ is a strictly feasible solution of (3.13), i.e., Slater's condition holds for the semidefinite program (3.13). By a now standard result in convex optimization, this implies that $\rho_k=\rho_k^*$, and (3.12) has an optimal solution if ρ_k is finite. So it remains to show that indeed ρ_k is finite and (3.13) is solvable.

Observe that from the constraint $\mathbf{M}_{k-1}(g_{s+1}\,\mathbf{z})\succeq 0$, and $\mathbf{M}_{k-1}(\theta_i\,\mathbf{z})\succeq 0$, $i=1,\ldots,n$, we deduce that any feasible solution $\mathbf{z}$ of (3.13) satisfies:

$$L_{\mathbf{z}}(y_\ell^{2k})\,\le\,1,\quad\forall\ell=1,\ldots,m;\qquad L_{\mathbf{z}}(x_i^{2k})\,\le\,1,\quad\forall i=1,\ldots,n.$$

Moreover, we also have $L_{\mathbf{z}}(1)=\gamma_0$, and so by [7, Lemma 4.3, p. 111] this implies $|z_{\alpha\beta}|\,\le\,\max[\gamma_0,1]$ for all $(\alpha,\beta)\in\mathbb{N}^{n+m}_{2k}$. Therefore, the feasible set is compact as closed and bounded, which in turn implies that (3.13) has an optimal solution $\mathbf{z}^*$. And as Slater's condition holds for (3.13) the dual (3.12) also has an optimal solution. Finally (3.14) follows from [8, Theorem 3.5] □

Remark 3.5. In fact, in Theorem 3.2 on may impose the sequence $(\overline{p_k})\subset\mathbb{R}[\mathbf{x}]$, $k\in\mathbb{N}$, to be monotone, i.e., such that $\overline{J_f}\,\le\,\overline{p_k}\,\le\,\overline{p_{k-1}}$ on $\mathbf{B}$, for all $k\ge 2$. And similarly for Corollary 3.3. For the practical computation of such a monotone sequence, in the semidefinite program (3.12) it suffices to include the additional constraint (or positivity certificate)

$$p^*_{k-1}(\mathbf{x})-p(\mathbf{x})\,=\,\sum_{i=0}^n\phi_i(\mathbf{x})\,\theta_i(\mathbf{x}),\quad \phi_0\in\Sigma[\mathbf{x}]_k,\,\phi_i\in\Sigma[\mathbf{x}]_{k-1},\,i\ge 1,$$

where $\theta_0=1$ and $p^*_{k-1}\in\mathbb{R}[\mathbf{x}]_{k-1}$ is the optimal solution computed at the previous step $k-1$. In this case the inner approximations $(\mathbf{R}^k_f)$, $k\in\mathbb{N}$, form a nested sequence since $\mathbf{R}^k_f\subseteq\mathbf{R}^{k+1}_f$ for all k. Similarly the outer approximations $(\mathbf{D}^k_f)$, $k\in\mathbb{N}$, also form a nested sequence since $\mathbf{D}^{k+1}_f\subseteq\mathbf{D}^k_f$ for all k.

4. Extensions

4.1. **Semi-algebraic functions.** Suppose for instance that given $q_1, q_2 \in \mathbb{R}[\mathbf{x},\mathbf{y}]$, one wants to characterize the set

$$\{\mathbf{x}\in\mathbf{B}:\quad \min\,[q_1(\mathbf{x},\mathbf{y}),q_2(\mathbf{x},\mathbf{y})]\,\leq\,0\quad\text{for all }\mathbf{y}\in\mathbf{K}_{\mathbf{x}}\},$$

where $\mathbf{K}_{\mathbf{x}}$ has been defined in (1.2), i.e., the set $\mathbf{R}_f$ associated with the semi-algebraic function $(\mathbf{x},\mathbf{y}) \mapsto f(\mathbf{x},\mathbf{y}) = \min[q_1(\mathbf{x},\mathbf{y}),q_2(\mathbf{x},\mathbf{y})]$. If f would be the semi-algebraic function $\max[q_1(\mathbf{x},\mathbf{y}),q_2(\mathbf{x},\mathbf{y})]$, characterizing $\mathbf{R}_f$ would reduce to the polynomial case (with some easy adjustments). But for $f = \min[q_1,q_2]$ this characterization is not so easy, and in fact is significantly more complicated. However, even though f is not a polynomial any more, we shall next see that the above methodology also works for semi-algebraic functions, a much larger class than the class of polynomials. Of course there is no free lunch and the resulting computational burden increases because one needs additional lifting variables to represent the semi-algebraic function.

With $\mathbf{S}\subset\mathbb{R}^n$ being semi-algebraic, recall that $f:\mathbf{S}\to\mathbb{R}$ is a semi-algebraic function if its graph $\{(\mathbf{x},f(\mathbf{x})):\mathbf{x}\in\mathbf{S}\}$ is a semi-algebraic set. And in fact, the graph of every semi-algebraic function is the projection of some *basic* semi-algebraic set in a lifted space. For more details the interested reader is referred to e.g. Lasserre and Putinar [10, p. 418].

So with $\mathbf{K}\subset\mathbb{R}^n\times\mathbb{R}^n$ as in (1.1), let $f:\mathbf{K}\to\mathbb{R}$ be a semi-algebraic function whose graph $\Psi_f = ((\mathbf{x},\mathbf{y}),f(\mathbf{x},\mathbf{y}))$ is the projection $\{(\mathbf{x},\mathbf{y},v_r)\in\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}\}$ of the basic semi-algebraic set $\widehat{\mathbf{K}}\subset\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^r$ defined by:

$$\widehat{\mathbf{K}}\,:=\,\{(\mathbf{x},\mathbf{y},\mathbf{v})\,:\,(\mathbf{x},\mathbf{y})\in\mathbf{K};\quad h_\ell(\mathbf{x},\mathbf{y},\mathbf{v})\geq 0,\quad \ell=1,\ldots,N\},\tag{4.1}$$

for some polynomials $(h_\ell)\subset\mathbb{R}[\mathbf{x},\mathbf{y},\mathbf{v}]$. That is,

$$\Psi_f\;=\;\{(\mathbf{x},\mathbf{y},v_r)\,:\,(\mathbf{x},\mathbf{y},\mathbf{v})\,\in\,\widehat{\mathbf{K}}\}.$$

For every $\mathbf{x}\in\mathbf{B}$, the set $\mathbf{K}_{\mathbf{x}}$ in (1.2) can be rewritten in the equivalent form:

$$\widehat{\mathbf{K}}_{\mathbf{x}}\,:=\,\{\mathbf{y}\,\in\,\mathbb{R}^m\,:\,(\mathbf{x},\mathbf{y},\mathbf{v})\,\in\,\widehat{\mathbf{K}}\,\},\qquad \mathbf{x}\in\mathbf{B}.\tag{4.2}$$

Then the functions $\overline{J_f}$ and $\underline{J_f}$ defined in (3.1)-(3.2) become

$$\mathbf{x}\mapsto\overline{J_f}(\mathbf{x})\quad:=\quad\sup_{\mathbf{y}}\{v_r\,:\,\mathbf{y}\in\widehat{\mathbf{K}}_{\mathbf{x}}\,\},\qquad \mathbf{x}\,\in\,\mathbf{B}.\tag{4.3}$$

$$\mathbf{x}\mapsto\underline{J_f}(\mathbf{x})\quad:=\quad\inf_{\mathbf{y}}\{v_r\,:\,\mathbf{y}\in\widehat{\mathbf{K}}_{\mathbf{x}}\,\},\qquad \mathbf{x}\,\in\,\mathbf{B}.\tag{4.4}$$

And Theorem 3.1 now reads:

Theorem 4.1. *Let $\widehat{\mathbf{K}}$ as in (4.1) be compact. If $\mathbf{K}_{\mathbf{x}}\neq\emptyset$ for every $\mathbf{x}\in\mathbf{B}$, there exists a sequence of polynomials $(\overline{p_k})\subset\mathbb{R}[\mathbf{x}]$ (resp. $(\underline{p_k})\subset\mathbb{R}[\mathbf{x}]$), $k\in\mathbb{N}$, such that $\overline{p_k}(\mathbf{x})\geq f(\mathbf{x},\mathbf{y})$ (resp. $\underline{p_k}(\mathbf{x})\leq f(\mathbf{x},\mathbf{y})$) for all $\mathbf{y}\in\mathbf{K}_{\mathbf{x}}$, $\mathbf{x}\in\mathbf{B}$, and such that*

$$\lim_{k\to\infty}\int_{\mathbf{B}}|\,\overline{p_k}(\mathbf{x})-\overline{J_f}(\mathbf{x})|\,d\mathbf{x}\quad=\quad 0\quad\text{[Convergence in }L_1(\mathbf{B})\text{]}\tag{4.5}$$

$$\lim_{k\to\infty}\int_{\mathbf{B}}|\underline{J_f}(\mathbf{x})-\underline{p_k}(\mathbf{x})\,|\,d\mathbf{x}\quad=\quad 0\quad\text{[Convergence in }L_1(\mathbf{B})\text{]}\tag{4.6}$$

Proof. Let λ be the Lebesgue measure on the box $\mathbf{B} \subset \mathbb{R}^n$, scaled to be a probability measure, and let $\mathbf{M}(\widehat{\mathbf{K}})$ be the set of finite Borel measures on $\widehat{\mathbf{K}} \subset \mathbb{R}^{n+m+r}$, and consider the following infinite-dimensional linear programs $\mathbf{P}$:

$$\rho := \sup_{\mu \in M(\widehat{\mathbf{K}})} \left\{ \int_{\widehat{\mathbf{K}}} v_r \, d\mu(\mathbf{x}, \mathbf{y}, \mathbf{v}) \, : \, \pi \, \mu = \lambda \right\}, \tag{4.7}$$

and its dual $\mathbf{P}^*$:

$$\rho^* := \inf_{h \in C(\mathbf{B})} \left\{ \int_{\mathbf{B}} h \, d\lambda \, : \, p(\mathbf{x}) - v_r \geq 0, \quad \forall (\mathbf{x}, \mathbf{y}, \mathbf{v}) \in \widehat{\mathbf{K}} \right\}, \tag{4.8}$$

where $C(\mathbf{B})$ is the Banach space of continuous functions on $\mathbf{B}$, equipped with the sup-norm.

Recall that $v_r = f(\mathbf{x}, \mathbf{y})$ whenever $(\mathbf{x}, \mathbf{y}, \mathbf{v}) \in \widehat{\mathbf{K}}$, and so for every feasible solution μ of (4.7) one has

$$\int_{\widehat{\mathbf{K}}} v_r \, d\mu(\mathbf{x}, \mathbf{y}, \mathbf{v}) \, \leq \, \int_{\widehat{\mathbf{K}}} \overline{J_f}(\mathbf{x}) \, d\mu(\mathbf{x}, \mathbf{y}, \mathbf{v}) \, = \, \int_{\mathbf{B}} \overline{J_f}(\mathbf{x}) \, d\lambda(\mathbf{x}),$$

and so $\rho \leq \int_{\mathbf{B}} \overline{J_f} d\lambda$.

As f is continuous on $\mathbf{K}$ and $\mathbf{K}_{\mathbf{x}} \neq \emptyset$ for every $\mathbf{x} \in \mathbf{B}$, $\overline{J_f}(\mathbf{x}) = f(\mathbf{x}, \mathbf{y}^*(\mathbf{x}))$ for some $\mathbf{y}^*(\mathbf{x}) \in \mathbf{K}_{\mathbf{x}}$, and for all $\mathbf{x} \in \mathbf{B}$. So for every $\mathbf{x} \in \mathbf{B}$, let $\Delta^*(\mathbf{x}) := \{(\mathbf{y}, \mathbf{v}) \in \mathbf{K}_{\mathbf{x}} \times \mathbb{R}^r \, : \, v_r = f(\mathbf{x}, \mathbf{y}^*(\mathbf{x})) = \overline{J_f}(\mathbf{x})\}$, which is nonempty. Define the Borel measure $\mu^* \in M(\widehat{\mathbf{K}})$ as $\varphi(d(\mathbf{y}, \mathbf{v})|\mathbf{x})\lambda(d\mathbf{x})$, where $\varphi(\cdot|\mathbf{x})$ is a stochastic kernel (or conditional distribution given $\mathbf{x} \in \mathbf{B}$) uniformly supported on $\Delta^*(\mathbf{x})$, for every $\mathbf{x} \in \mathbf{B}$. Then by construction $\pi\mu^* = \lambda$ and

$$\int_{\widehat{\mathbf{K}}} v_r \, d\mu^*(\mathbf{x}, \mathbf{y}, \mathbf{v}) \, = \, \int_{\mathbf{B}} \left(\int_{\Delta^*(\mathbf{x})} v_r \varphi(d(\mathbf{y}, \mathbf{v})|\mathbf{x}) \right) d\lambda(\mathbf{x}) \, = \, \int_{\mathbf{B}} \overline{J_f}(\mathbf{x}) \, d\lambda(\mathbf{x}) \, \leq \, \rho,$$

where the last inequality is because μ^* is admissible for $\mathbf{P}$. And so $\rho = \int_{\mathbf{B}} \overline{J_f} d\lambda$.

Next, proceeding exactly as for the proof of [8, Lemma 2.5], one may show that there is no duality gap between $\mathbf{P}$ and its dual $\mathbf{P}^*$. So consider a minimizing sequence $(h_k) \subset C(\mathbf{B})$, $k \in \mathbb{N}$, with $\lim_{k \to \infty} \int_{\mathbf{B}} h_k d\lambda = \rho$, and $h_k(\mathbf{x}) \geq v_r$ for all $(\mathbf{x}, \mathbf{y}, \mathbf{v}) \in \widehat{\mathbf{K}}$, or equivalently, $h_k \geq \overline{J_f}$ on $\mathbf{B}$.

As $\mathbf{B}$ is compact, by the Stone-Weierstrass theorem, for each k, let $p_k \in \mathbb{R}[\mathbf{x}]$ be such that $\sup \{|h_k(\mathbf{x}) - p_k(\mathbf{x})| : \mathbf{x} \in \mathbf{B}\} < 1/k$. Then $\overline{p_k}(\mathbf{x}) := 1/k + p_k(\mathbf{x}) \geq h_k(\mathbf{x}) \geq \overline{J_f}(\mathbf{x})$ for all $\mathbf{x} \in \mathbf{B}$, and so

$$\lim_{k \to \infty} \int_{\mathbf{B}} \overline{p_k} \, d\lambda \, = \, \lim_{k \to \infty} \int_{\mathbf{B}} h_k \, d\lambda \, = \, \int_{\mathbf{B}} \overline{J_f} \, d\lambda = \rho.$$

And so one obtains the desired $L_1(\mathbf{B})$-convergence, $\int_{\mathbf{B}} |\overline{p_k} - \overline{J_f}| \, d\lambda \to 0$ as $k \to \infty$, i.e., (3.3) holds. And (3.4) is obtained in a similar fashion. □

And so Theorem 3.2, Corollary 3.3, Theorem 3.4 and Theorem **??** all apply, where the semidefinite program (3.12) now reads:

$$
\begin{array}{rl}
\rho_k = \displaystyle\min_{p,\sigma_j,\psi_i,\chi_\ell} & \displaystyle\int_{\mathbf{B}} p(\mathbf{x})\,d\mathbf{x} \\
\text{s.t.} & p(\mathbf{x}) - v_r = \displaystyle\sum_{j=0}^{s+1} \sigma_j(\mathbf{x},\mathbf{y},\mathbf{v})\,g_j(\mathbf{x},\mathbf{y}) \\
 & \quad + \displaystyle\sum_{i=1}^{n} \psi_i(\mathbf{x},\mathbf{y},\mathbf{v})\,\theta_i(\mathbf{x}) + \sum_{\ell=1}^{N} \chi_\ell(\mathbf{x},\mathbf{y},\mathbf{v})\,h_\ell(\mathbf{x},\mathbf{y},\mathbf{v}) \\
 & p \in \mathbb{R}[\mathbf{x}]_{2k};\ \sigma_j \in \Sigma_{k-v_j}[\mathbf{x},\mathbf{y},\mathbf{v}],\ j = 0,\ldots,s+1 \\
 & \psi_i \in \Sigma_{k-1}[\mathbf{x},\mathbf{y},\mathbf{v}],\ i = 1,\ldots,n. \\
 & \chi_i \in \Sigma_{k-q_\ell}[\mathbf{x},\mathbf{y},\mathbf{v}],\ \ell = 1,\ldots,N.
\end{array}
\tag{4.9}
$$

where $q_\ell = \lceil \deg(h_\ell)/2 \rceil$, $\ell = 1,\ldots,N$.

Example 1. For instance suppose that $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is the semi-algebraic function $(\mathbf{x},\mathbf{y}) \mapsto f(\mathbf{x},\mathbf{y}) := \min[q_1(\mathbf{x},\mathbf{y}), q_2(\mathbf{x},\mathbf{y})]$. Then using $a \wedge b = \frac{1}{2}(a + b - |a-b|)$ and $|a-b| = \theta \geq 0$ with $\theta^2 = (a-b)^2$,

$$
\begin{aligned}
\widehat{\mathbf{K}} = \{(\mathbf{x},\mathbf{y},\mathbf{v}) \ : \ & (\mathbf{x},\mathbf{y}) \in \mathbf{K};\ v_1^2 = (q_1(\mathbf{x},\mathbf{y}) - q_2(\mathbf{x},\mathbf{y}))^2; \quad v_1 \geq 0; \\
& 2v_2 = q_1(\mathbf{x},\mathbf{y}) + q_2(\mathbf{x},\mathbf{y}) - v_1\},
\end{aligned}
$$

and

$$\Psi_f = \{((\mathbf{x},\mathbf{y}), f(\mathbf{x},\mathbf{y}))\} = \{(\mathbf{x},\mathbf{y},v_2) : (\mathbf{x},\mathbf{y},v_1,v_2) \in \widehat{\mathbf{K}}\}.$$

4.2. **Convex inner approximations.** It is worth mentioning that enforcing convexity of inner approximations of $\mathbf{R}_f$ is easy. But of course there is some additional computational cost and the convergence in Theorem **??** is lost in general.

To enforce convexity of the level set $\{\mathbf{x} \in \mathbf{B} : p_k^*(\mathbf{x}) \leq 0\}$ it suffices to require that p_k^* is convex on $\mathbf{B}$, i.e., adding the constraint

$$\langle \mathbf{u}, \nabla^2 p_k^*(\mathbf{x})\,\mathbf{u}\rangle \geq 0, \qquad \forall (\mathbf{x},\mathbf{u}) \in \mathbf{B} \times \mathbf{U},$$

where $\mathbf{U} := \{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u}\|^2 \leq 1\}$. The latter constraint can in turn be enforced by the Putinar positivity certificate

$$\langle \mathbf{u}, \nabla^2 p_k^*(\mathbf{x})\,\mathbf{u}\rangle = \sum_{i=0}^{n} \omega_i(\mathbf{x},\mathbf{u})\,\theta_i(\mathbf{x}) + \omega_{n+1}(\mathbf{x},\mathbf{u})\,\theta_{n+1}(\mathbf{x},\mathbf{u}), \tag{4.10}$$

for some SOS polynomials $(\omega_i) \subset \Sigma[\mathbf{x},\mathbf{u}]$ (and where $\theta_{n+1}(\mathbf{x},\mathbf{u}) = 1 - \|\mathbf{u}\|^2$).

Then (4.10) can be included in the semidefinite program (3.12) with $\omega_0 \in \Sigma[\mathbf{x},\mathbf{u}]_k$, and $\omega_i \in \Sigma[\mathbf{x},\mathbf{u}]_{k-1}$, $i = 1,\ldots n+1$. However, now $\mathbf{z} = (z_{\alpha,\gamma,\beta})$, $(\alpha,\beta,\gamma) \in \mathbb{N}^{2n+m}$, and so solving the resulting semidefinite program is more demanding.

4.3. **Polynomial matrix inequalities.** Let $\mathbf{A}_\alpha \in \mathcal{S}_m$, $\alpha \in \mathbb{N}^n_d$, be real symmetric matrices and let $\mathbf{B} \subset \mathbb{R}^n$ be a given box. Consider the set

$$\mathbf{S} := \{\mathbf{x} \in \mathbf{B} \ : \ \mathbf{A}(\mathbf{x}) \succeq 0\}, \tag{4.11}$$

where $\mathbf{A} \in \mathbb{R}[\mathbf{x}]^{m\times m}$ is the matrix polynomial

$$\mathbf{x} \mapsto \mathbf{A}(\mathbf{x}) := \sum_{\alpha \in \mathbb{N}^n_d} \mathbf{x}^\alpha\,\mathbf{A}_\alpha.$$

If $\mathbf{A}(\mathbf{x})$ is linear in $\mathbf{x}$ then $\mathbf{S}$ is convex and (4.11) is an LMI description of $\mathbf{S}$ which is very nice as it can be used efficiently in semidefinite programming.

In the general case the description (4.11) of $\mathbf{S}$ is called a Polynomial Matrix Inequality (PMI) and cannot be used as efficiently as in the convex case. Indeed $\mathbf{S}$ is a basic semi-algebraic set with an alternative description in terms of the box constraint $\mathbf{x} \in \mathbf{B}$ and m additional polynomial inequality constraints (including the constraint $\det(\mathbf{A}(\mathbf{x})) \geq 0$). However, this latter description may not be very appropriate either because the degree of polynomials involved in that description is potentially as large as d^m which precludes from its use for practical computation (e.g., for optimization purposes).

On the other hand, for polynomial optimization problems with a PMI constraint $\mathbf{A}(\mathbf{x}) \succeq 0$, one may still define an appropriate and *ad hoc* hierarchy of semidefinite relaxations, as described in Hol and Scherer [5, 6], and Henrion and Lasserre [4]. But even if more economical than the hierarchy using the former description of $\mathbf{S}$ with m (high degree) polynomials, this latter approach may not still be ideal. In particular it is not clear how to detect (and then take benefit of) some possible structured sparsity to reduce the computational cost.

So in the general case and when d^m is not small, one may be interested in a description of $\mathbf{S}$ simpler than the PMI (4.11) so that it can used more efficiently.

Let $\mathbf{Y} := \{\mathbf{y} \in \mathbb{R}^m : \|\mathbf{y}\|^2 = 1\}$ denote the unit sphere of $\mathbb{R}^m$. Then with $(\mathbf{x}, \mathbf{y}) \mapsto f(\mathbf{x}, \mathbf{y}) := -\langle \mathbf{y}, \mathbf{A}(\mathbf{x})\mathbf{y} \rangle$, the set $\mathbf{S}$ has the alternative and equivalent description

$$\mathbf{S} = \{\mathbf{x} \in \mathbf{B} : f(\mathbf{x}, \mathbf{y}) \leq 0, \quad \forall \mathbf{y} \in \mathbf{Y}\} =: \mathbf{R}_f, \tag{4.12}$$

which involves the quantifier "$\forall$". Therefore the machinery developed in §3 can be applied to define the hierarchy of inner approximations $\mathbf{R}_f^k \subset \mathbf{S}$ in Theorem 3.2, where for each k, $\mathbf{R}_f^k = \{\mathbf{x} \in \mathbf{B} : \overline{p_k}(\mathbf{x}) \leq 0\}$ for some polynomial $\overline{p_k}$ of degree k. Observe that if $\mathbf{x} \mapsto \mathbf{A}(\mathbf{x})$ is not a constant matrix, then with

$$\mathbf{x} \mapsto \overline{J_f}(\mathbf{x}) := \sup_{\mathbf{y}} \{f(\mathbf{x}, \mathbf{y}) : \mathbf{y} \in \mathbf{Y}\}, \qquad \mathbf{x} \in \mathbf{B},$$

the set $\{\mathbf{x} : \overline{J_f}(\mathbf{x}) = 0\}$ has Lebesgue measure zero because $\overline{J_f}(\mathbf{x})$ is the largest eigenvalue of $-\mathbf{A}(\mathbf{x})$. Hence by Theorem 3.2

$$\mathrm{vol}\left(\mathbf{R}_f^k\right) \to \mathrm{vol}(\mathbf{S}), \qquad \text{as } k \to \infty.$$

Notice that computing $\overline{p_k}$ has required to introduce the m additional variables $\mathbf{y}$ but the degree of f is not larger than $d+2$ if d is the maximum degree of the entries.

Importantly for computational purposes, if the polynomial $\mathbf{x} \mapsto f(\mathbf{x}, \mathbf{y})$ has some structured sparsity[3] then $(\mathbf{x}, \mathbf{y}) \mapsto f(\mathbf{x}, \mathbf{y})$ inherits the same structured sparsity (but with now $\cup_{j=1}^p (\mathbf{x}_j, \mathbf{y})$ in lieu of $\cup_{j=1}^p \mathbf{x}_j$). And so in particular, for computing $\overline{p_k}$ one may use the sparse version of the hierarchy of semidefinite relaxations introduced in Waki et al. [12] which permits to handle problems with a significantly large number of variables.

[3]That is, $f(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^p f_j(\mathbf{x}_j, \mathbf{y})$ where $\mathbf{x}_j = (x_{i_1}, \ldots, x_{i_j})$, with $\{i_1, \ldots, i_j\} \subset \{1, \ldots, n\}$.

Example 2. The following illustrative example is taken from Henrion and Lasserre [3]. With $n = 2$, let $\mathbf{B} \subset \mathbb{R}^2$ be the unit disk $\{\mathbf{x} : \|\mathbf{x}\|^2 \leq 1\}$, and let

$$\mathbf{A}(\mathbf{x}) := \left[\begin{array}{cc} 1 - 16x_1x_2 & x_1 \\ x_1 & 1 - x_1^2 - x_2^2 \end{array} \right]; \quad \mathbf{S} := \{\mathbf{x} \in \mathbf{B} : \mathbf{A}(\mathbf{x}) \succeq 0\}.$$

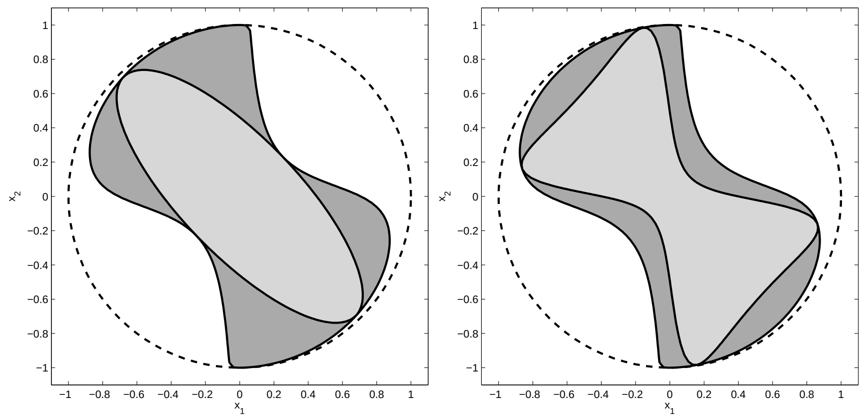

FIGURE 1. Example 2: $\mathbf{R}_f^1$ (left) and $\mathbf{R}_f^2$ (right) inner approximations (light gray) of $\mathbf{S}$ (dark gray) embedded in unit disk $\mathbf{B}$ (dashed)

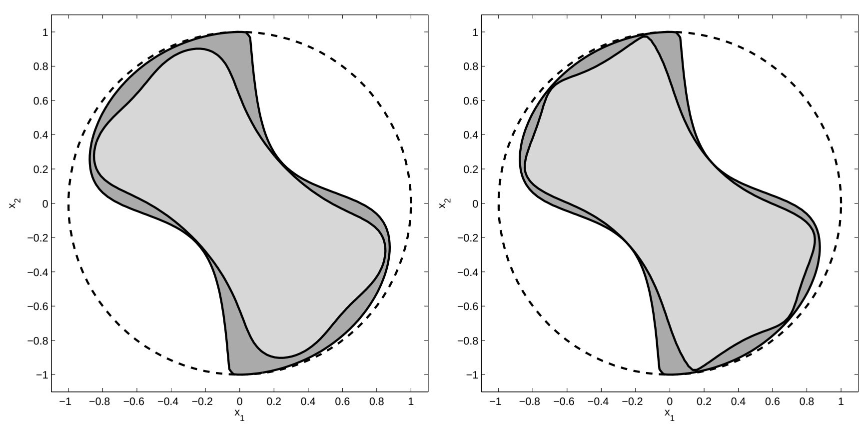

FIGURE 2. Example 2: $\mathbf{R}_f^3$ (left) and $\mathbf{R}_f^4$ (right) inner approximations (light gray) of $\mathbf{S}$ (dark gray) embedded in unit disk $\mathbf{B}$ (dashed).

In Figure 1 is displayed $\mathbf{S}$ and the degree two $\mathbf{R}_f^1$ and four $\mathbf{R}_f^2$ inner approximations of $\mathbf{S}$, whereas in Figure 2 are displayed the $\mathbf{R}_f^3$ and $\mathbf{R}_f^4$ inner approximations of $\mathbf{S}$. One may see that with $k = 4$, $\mathbf{R}_f^4$ is already a quite good approximation of $\mathbf{S}$.

4.4. **Several functions f.** We now consider sets of the form

$$\mathbf{R}_F := \{\mathbf{x} \in \mathbf{B} \,:\, (\mathbf{x},\mathbf{y}) \in F \text{ for all } \mathbf{y} \text{ such that } (\mathbf{x},\mathbf{y}) \in \mathbf{K}\}$$

where $F \subset \mathbb{R}^n \times \mathbb{R}^m$ is a basic-semi algebraic set defined by

$$F \,:=\, \{(\mathbf{x},\mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^m \,:\, f_\ell(\mathbf{x},\mathbf{y}) \,\leq\, 0, \quad \forall \ell = 1,\ldots,q\},$$

for some polynomials $(f_\ell) \subset \mathbb{R}[\mathbf{x},\mathbf{y}]$, $\ell = 1,\ldots,q$. Of course it is a particular case of the previous section with the semi-algebraic function $f = f_1 \vee f_2 \cdots \vee f_q$, but in this case a simpler approach is possible.

Let $\overline{p_{k\ell}} \in \mathbb{R}[\mathbf{x}]$ be the polynomial in Theorem 3.1 associated with $\overline{J_{f_\ell}}$, $\ell = 1,\ldots,q$, and let the set $\mathbf{R}_F^k$ be defined by

$$\mathbf{R}_F^k \,:=\, \{\mathbf{x} \in \mathbb{R}^n \,:\, \overline{p_{k\ell}}(\mathbf{x}) \,\leq\, 0, \quad \ell = 1,\ldots,q\} \,=\, \bigcap_{\ell=1}^{q} \mathbf{R}_{f_\ell}^k,$$

where for each $\ell = 1,\ldots,q$, the set $\mathbf{R}_{f_\ell}^k$ is defined in the obvious manner.

The sets $(\mathbf{R}_F^k) \subset \mathbf{R}_F$, $k \in \mathbb{N}$, provide a sequence of inner approximations of $\mathbf{R}_F$ with the nice property that

$$\lim_{k\to\infty} \mathrm{vol}\left(\mathbf{R}_F^k\right) \,=\, \mathrm{vol}\left(\mathbf{R}_F\right),$$

whenever the set

$$\{\mathbf{x} \in \mathbf{B} : \sup_{\mathbf{y}} \{\max_{\ell} f_\ell(\mathbf{x},\mathbf{y}) \,:\, \mathbf{y} \in \mathbf{K}_{\mathbf{x}}\} = 0\}$$

has Lebesgue measure zero.

4.5. **Sets defined with two quantifiers.** Consider three types of variables $(\mathbf{x},\mathbf{y},\mathbf{u}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^s$, a box $\mathbf{B}_{\mathbf{x}} \subset \mathbb{R}^n$, a box $\mathbf{B}_{\mathbf{y}} \subset \mathbb{R}^m$, and a set $\mathbf{K} \subset \mathbf{B}_{\mathbf{x}} \times \mathbf{B}_{\mathbf{y}} \times \mathbf{U}$. It is assumed that for each $(\mathbf{x},\mathbf{y}) \in \mathbf{B}_{\mathbf{xy}}$ $(= \mathbf{B}_{\mathbf{x}} \times \mathbf{B}_{\mathbf{y}})$,

$$\mathbf{K}_{\mathbf{xy}} \,:=\, \{\mathbf{u} \in \mathbf{U} \,:\, (\mathbf{x},\mathbf{y},\mathbf{u}) \,\in\, \mathbf{K}\} \,\neq\, \emptyset.$$

Sets with $\forall$, $\exists$. Consider a set $\mathbf{D}'_f$ of the form

$$\mathbf{D}'_f \,:=\, \{\mathbf{x} \in \mathbf{B}_{\mathbf{x}} \,:\, \exists\, \mathbf{y} \in \mathbf{B}_{\mathbf{y}} \text{ such that } f(\mathbf{x},\mathbf{y},\mathbf{u}) \leq 0 \text{ for all } \mathbf{u} \in \mathbf{K}_{\mathbf{xy}}\}. \tag{4.13}$$

Such a set is not easy to handle, in particular for optimizing over it. So it is highly desirable to approximate as closely as possible such a set $\mathbf{D}'_f$ with a set having a much simpler description, and in particular a description with *no* quantifier. We propose to use the methodology of §3 to provide such approximations.

One proceeds as follows. First define the function $\overline{J_f} : \mathbf{B}_{\mathbf{xy}} \to \mathbb{R}$ by:

$$\overline{J_f}(\mathbf{x},\mathbf{y}) \,:=\, \sup_{\mathbf{u}} \{f(\mathbf{x},\mathbf{y},\mathbf{u}) \,:\, (\mathbf{x},\mathbf{y},\mathbf{u}) \in \mathbf{K}\},$$

so that $\mathbf{D}'_f = \{\mathbf{x} \in \mathbf{B}_{\mathbf{x}} : \exists\, \mathbf{y} \in \mathbf{B}_{\mathbf{y}} \text{ such that } \overline{J_f}(\mathbf{x},\mathbf{y}) \leq 0\}$.

By Theorem 3.2 (adapted to the present context), the function $\overline{J_f}$ can be approximated by some polynomial $\overline{p_k} \in \mathbb{R}[\mathbf{x},\mathbf{y}]$ such that $\overline{p_k}(\mathbf{x},\mathbf{y}) \geq \overline{J_f}(\mathbf{x},\mathbf{y})$ for all $(\mathbf{x},\mathbf{y}) \in \mathbf{B}_{\mathbf{xy}}$. Then as $\overline{p_k} \geq \overline{J_f}$ on $\mathbf{B}_{\mathbf{xy}}$ for all k, one has the inclusions

$$\mathbf{D}'_f \supset \mathbf{D}_f^k := \{\mathbf{x} \in \mathbf{B}_{\mathbf{x}} \,:\, \exists\, \mathbf{y} \in \mathbf{B}_{\mathbf{y}} \text{ such that } \overline{p_k}(\mathbf{x},\mathbf{y}) \leq 0\}, \quad \forall k.$$

With no loss of generality one may and will assume that for every $k \geq 1$,

$$\overline{J_f}(\mathbf{x},\mathbf{y}) \,\leq\, \overline{p_k}(\mathbf{x},\mathbf{y}) \,\leq\, \overline{p_{k-1}}(\mathbf{x},\mathbf{y}), \qquad \forall (\mathbf{x},\mathbf{y}) \in \mathbf{B}_{\mathbf{xy}}.$$

(See Remark 3.5.) Hence $\mathbf{D}'_f \supset \mathbf{D}^k_f \supset \mathbf{D}^{k-1}_f$ for all $k \geq 2$, i.e., the $(\mathbf{D}^k_f)$, $k \in \mathbb{N}$, is a monotone non decreasing nested sequence.

Lemma 4.2. *Let $\mathbf{D}'_f$ be as in (4.13) and assume that the set*

$$\{(\mathbf{x},\mathbf{y}) \in \mathbf{B}_{\mathbf{xy}} : \overline{J_f}(\mathbf{x},\mathbf{y}) = 0\}$$

has Lebesgue measure zero. Then

$$\lim_{k\to\infty} \mathrm{vol}\left(\{(\mathbf{x},\mathbf{y}) \in \mathbf{B}_{\mathbf{xy}} : \overline{p_k}(\mathbf{x},\mathbf{y}) \leq 0\}\right) = \mathrm{vol}\left(\{(\mathbf{x},\mathbf{y}) \in \mathbf{B}_{\mathbf{xy}} : \overline{J_f}(\mathbf{x},\mathbf{y}) \leq 0\}\right)$$

Therefore we have obtained a convergent hierarchy of inner approximations

$$\mathbf{H}^k_{\mathbf{xy},f} \,:=\, \{(\mathbf{x},\mathbf{y}) \in \mathbf{B}_{\mathbf{xy}} : \overline{p_k}(\mathbf{x},\mathbf{y}) \leq 0\}, \quad k \in \mathbb{N},$$

of the lifted representation $\mathbf{H}_{\mathbf{xy},f} := \{(\mathbf{x},\mathbf{y}) \in \mathbf{B}_{\mathbf{xy}} : \overline{J_f}(\mathbf{x},\mathbf{y}) \leq 0\}$ of $\mathbf{D}'_f$. Each set $\mathbf{H}^k_{\mathbf{xy},f}$ has a very simple description in terms of the sublevel set of some polynomial and with no quantifier.

But one may even obtain approximations of $\mathbf{D}'_f$ itself rather than approximations of its lifted representation $\mathbf{H}_{\mathbf{xy},f}$. Following §3 again, for each $k \in \mathbb{N}$, one may construct outer approximations

$$\mathbf{D}^{k\ell}_f \,:=\, \{\mathbf{x} \in \mathbf{B}_{\mathbf{x}} \,:\, \underline{p_{k\ell}}(\mathbf{x}) \leq 0\,\} \,\supset\, \mathbf{D}^k_f, \quad \forall \ell,$$

for some polynomial $\underline{p_{k\ell}} \in \mathbb{R}[\mathbf{x}]$ such that $\underline{p_{k\ell}}(\mathbf{x}) \leq \overline{p_k}(\mathbf{x},\mathbf{y})$ for all $(\mathbf{x},\mathbf{y}) \in \mathbf{B}_{\mathbf{x}} \times \mathbf{B}_{\mathbf{y}}$, and all k, ℓ. Moreover, for every k, if $\mathrm{vol}\,(\{\mathbf{x} \in \mathbf{B}_{\mathbf{x}} : \overline{p_k}(\mathbf{x},\mathbf{y}) = 0\}) = 0$ then

$$\lim_{\ell\to\infty} \mathrm{vol}\left(\mathbf{D}^{k\ell}_f\right) \,=\, \mathrm{vol}\left(\mathbf{D}^k_f\right).$$

But in general we will not have: $\lim\limits_{k\to\infty} \mathrm{vol}\left(\mathbf{D}^k_f\right) \,=\, \mathrm{vol}\left(\mathbf{D}'_f\right)$ as $k \to \infty$.

Sets with $\exists$, $\forall$. Consider now a set $\mathbf{R}'_f$ of the form

$$\mathbf{R}'_f \,:=\, \{\mathbf{x} \in \mathbf{B}_{\mathbf{x}} \,:\, \forall\, \mathbf{y} \in \mathbf{B}_{\mathbf{y}},\, \exists\, \mathbf{u} \in \mathbf{K}_{\mathbf{xy}} \text{ such that } f(\mathbf{x},\mathbf{y},\mathbf{u}) \leq 0\}. \tag{4.14}$$

As for $\mathbf{D}'_f$, such a set is not easy to handle, in particular for optimizing over it. So again it is highly desirable to approximate as closely as possible such a set $\mathbf{R}'_f$ with a set having a much simpler description, and in particular a description with *no* quantifier. So proceeding In a similar fashion as before, first define the function $\underline{J_f} : \mathbf{B}_{\mathbf{xy}} \to \mathbb{R}$ by:

$$\underline{J_f}(\mathbf{x},\mathbf{y}) \,:=\, \inf_{\mathbf{u}}\, \{f(\mathbf{x},\mathbf{y},\mathbf{u}) \,:\, (\mathbf{x},\mathbf{y},\mathbf{u}) \in \mathbf{K}\},$$

so that $\mathbf{R}'_f = \{\mathbf{x} \in \mathbf{B}_{\mathbf{x}} : \ \underline{J_f}(\mathbf{x},\mathbf{y}) \leq 0 \text{ for all } \mathbf{y} \in \mathbf{B}_{\mathbf{y}}\}$.

By Corollary 3.3 (adapted to the present context), the function $\underline{J_f}$ can be approximated by some polynomial $\underline{p_k} \in \mathbb{R}[\mathbf{x},\mathbf{y}]$ such that $\underline{p_k}(\mathbf{x},\mathbf{y}) \leq \underline{\overline{J_f}}(\mathbf{x},\mathbf{y})$ for all $(\mathbf{x},\mathbf{y}) \in \mathbf{B}_{\mathbf{xy}}$. Then as $\underline{p_k} \leq \overline{J_f}$ on $\mathbf{B}_{\mathbf{xy}}$ for all k, one has the inclusions

$$\mathbf{R}'_f \subset \mathbf{R}^k_f := \{\mathbf{x} \in \mathbf{B}_{\mathbf{x}} \,:\, \underline{p_k}(\mathbf{x},\mathbf{y}) \leq 0 \text{ for all } \mathbf{y} \in \mathbf{B}_{\mathbf{y}}\}, \quad \forall k.$$

And so we have the following analogue of Lemma 4.2

Lemma 4.3. *Let $\mathbf{R}'_f$ be as in (4.14) and assume that the set*

$$\{(\mathbf{x},\mathbf{y})\in\mathbf{B}_{\mathbf{xy}}:\underline{J_f}(\mathbf{x},\mathbf{y})=0\}$$

has Lebesgue measure zero. Then

$$\lim_{k\to\infty}\mathrm{vol}\left(\{(\mathbf{x},\mathbf{y})\in\mathbf{B}_{\mathbf{xy}}:\underline{p_k}(\mathbf{x},\mathbf{y})\leq 0\}\right)=\mathrm{vol}\left(\{(\mathbf{x},\mathbf{y})\in\mathbf{B}_{\mathbf{xy}}:\underline{J_f}(\mathbf{x},\mathbf{y})\leq 0\}\right)$$

Therefore we have obtained a convergent hierarchy of outer approximations

$$\Delta^k_{\mathbf{xy},f}\;:=\;\{(\mathbf{x},\mathbf{y})\in\mathbf{B}_{\mathbf{xy}}:\underline{p_k}(\mathbf{x},\mathbf{y})\leq 0\},\quad k\in\mathbb{N},$$

of the lifted representation $\Delta_{\mathbf{xy},f}:=\{(\mathbf{x},\mathbf{y})\in\mathbf{B}_{\mathbf{xy}}:\underline{J_f}(\mathbf{x},\mathbf{y})\leq 0\}$ of $\mathbf{R}'_f$. Again, each set $\Delta^k_{\mathbf{xy},f}$ has a very simple description in terms of the sublevel set of some polynomial and with no quantifier.

But one may even obtain approximations of $\mathbf{R}'_f$ itself rather than approximations of its lifted representation $\Delta_{\mathbf{xy},f}$. Following §3 again, for each $k\in\mathbb{N}$, one may construct inner approximations

$$\mathbf{R}^{k\ell}_f\;:=\;\{\mathbf{x}\in\mathbf{B}_{\mathbf{x}}\;:\;\overline{p_{k\ell}}(\mathbf{x})\leq 0\;\}\;\subset\;\mathbf{R}^k_f,\quad\forall\ell,$$

for some polynomial $\overline{p_{k\ell}}\in\mathbb{R}[\mathbf{x}]$ such that $\overline{p_{k\ell}}(\mathbf{x})\geq\underline{p_k}(\mathbf{x},\mathbf{y})$ for all $(\mathbf{x},\mathbf{y})\in\mathbf{B}_{\mathbf{xy}}$, and all k,ℓ. Moreover, for every k, if $\mathrm{vol}\left(\{\mathbf{x}\in\mathbf{B}_{\mathbf{x}}:\underline{p_k}(\mathbf{x},\mathbf{y})=0\}\right)=0$ then

$$\lim_{\ell\to\infty}\mathrm{vol}\left(\mathbf{R}^{k\ell}_f\right)\;=\;\mathrm{vol}\left(\mathbf{R}^k_f\right).$$

But in general we will not have: $\lim_{k\to\infty}\mathrm{vol}\left(\mathbf{R}^k_f\right)\;=\;\mathrm{vol}\left(\mathbf{R}'_f\right)$ as $k\to\infty$.

References

[1] R. Ash. *Real Analysis and Probability*, Academic Press, Boston, USA, 1972.

[2] J. Bochnak, M. Coste and M.-F. Roy. *Real Algebraic Geometry*, Springer Verlag, New York, 1998

[3] D. Henrion and J.B. Lasserre. Inner approximations for polynomial matrix inequalities and robust stability regions, IEEE Trans. Auto. Control **57**, pp. 1456–1467, 2012

[4] D. Henrion and J.B. Lasserre. Convergent relaxations of polynomial matrix inequalities and static output feedback, IEEE Trans. Auto. Control **51**, pp. 192–202, 2006

[5] C. W. J. Hol and C. W. Scherer. Sum of squares relaxations for polynomial semidenite programming. *Proc. Symp. on Mathematical Theory of Networks and Systems* (MTNS), Leuven, Belgium, July 2004.

[6] C. W. J. Hol and C. W. Scherer. A sum of squares approach to xed-order H1 synthesis, in *Positive Polynomials in Control*, D. Henrion, A. Garulli (Eds.), LNCIS, Springer Verlag, Berlin, 2005.

[7] J.B. Lasserre and T. Netzer. SOS approximations of nonnegative polynomials via simple high-degree perturbations, Math. Z. **256**, pp. 99–112, 2007

[8] J.B. Lasserre. A "joint+marginal" approach to parametric polynomial optimization, SIAM J. Optim. **20**, pp. 1995–2022, 2010

[9] J.B. Lasserre. *Moments, Positive Polynomials and Their Applications*, Imperial College Press, London, 2009.

[10] J.B. Lasserre and M. Putinar. Positivity and optimization: Beyond polynomials, in: *Handbook on Semidefinite, Conic and Polynomial Optimization*, M. Anjos and J.B. Lasserre (Eds.), Springer, New York, 2011, pp. 407–434.

[11] M. Putinar. Positive polynomials on compact semi-algebraic sets, Indiana Univ. Math. J. **42**, pp. 969–984, 1993

[12] S. Waki, S. Kim, M. Kojima and M. Maramatsu. Sums of squares and semidefinite programming relaxations for polynomial optimization with structured sparsity, SIAM J. Optim. **17**, pp. 218–242, 2006

LAAS-CNRS and Institute of Mathematics, University of Toulouse, LAAS, 7 avenue du Colonel Roche, 31077 Toulouse Cédex 4, France, Tel: +33561336415
E-mail address: lasserre@laas.fr