AN ALGORITHM FOR SEMI-INFINITE POLYNOMIAL OPTIMIZATION

J.B. LASSERRE

Abstract. We consider the semi-infinite optimization problem:

$$f^* := \min_{\mathbf{x}\in\mathbf{X}} \{f(\mathbf{x}) : g(\mathbf{x},\mathbf{y}) \leq 0, \forall \mathbf{y}\in\mathbf{Y}_{\mathbf{x}}\},$$

where f, g are polynomials and $\mathbf{X} \subset \mathbb{R}^n$ as well as $\mathbf{Y}_{\mathbf{x}} \subset \mathbb{R}^p$, $\mathbf{x} \in \mathbf{X}$, are compact basic semi-algebraic sets. To approximate f^* we proceed in two steps. First, we use the "joint+marginal" approach of the author [8] to approximate from above the function $\mathbf{x} \mapsto \Phi(\mathbf{x}) = \sup\{g(\mathbf{x},\mathbf{y}) : \mathbf{y} \in \mathbf{Y}_{\mathbf{x}}\}$ by a polynomial $\Phi_d \geq \Phi$, of degree at most $2d$, with the strong property that Φ_d converges to Φ for the L_1-norm, as $d \to \infty$ (and in particular, almost uniformly for some subsequence (d_ℓ), $\ell \in \mathbb{N}$). Then we solve the polynomial optimization problem $f^*_d = \min_{\mathbf{x}\in\mathbf{X}}\{f(\mathbf{x}) : \Phi_d(\mathbf{x}) \leq 0\}$ via a (by now standard) hierarchy of semidefinite relaxations. It turns out that the optimal value $f^*_d \geq f^*$ converges to f^* as $d \to \infty$. In practice we let d be fixed, small, and relax the constraint $\Phi_d \leq 0$ to $\Phi_d(\mathbf{x}) \leq \epsilon$ with $\epsilon > 0$, allowing to change ϵ dynamically.

1. Introduction

Consider the semi-infinite optimization problem:

$$\mathbf{P}: \qquad f^* := \min_{\mathbf{x}\in\mathbf{X}} \{f(\mathbf{x}) : g(\mathbf{x},\mathbf{y}) \leq 0, \forall \mathbf{y}\in\mathbf{Y}_{\mathbf{x}}\}, \tag{1.1}$$

where $\mathbf{X} \subset \mathbb{R}^n$, $\mathbf{Y}_{\mathbf{x}} \subset \mathbb{R}^p$ for every $\mathbf{x} \in \mathbf{X}$, and some functions $f\colon \mathbb{R}^n \to \mathbb{R}$, $g\colon \mathbb{R}^n \times \mathbb{R}^p :\to \mathbb{R}$.

Problem $\mathbf{P}$ is called a *semi-infinite* optimization problem because of the infinitely many constraints $g(\mathbf{x},\mathbf{y}) \leq 0$ for all $\mathbf{y} \in \mathbf{Y}_{\mathbf{x}}$ (for each fixed $\mathbf{x} \in \mathbf{X}$). It has many applications and particularly in robust control.

In full generality $\mathbf{P}$ is a very hard problem and most methods aiming at computing (or at least approximating) f^* use discretization to overcome the difficult semi-infinite constraint $g(\mathbf{x},\mathbf{y}) \leq 0$ for all $\mathbf{y} \in \mathbf{Y}_{\mathbf{x}}$. Namely, in typical approaches where $\mathbf{Y}_{\mathbf{x}} \equiv \mathbf{Y}$ for all $\mathbf{x} \in \mathbf{X}$ (i.e. no dependence on $\mathbf{x}$), the set $\mathbf{Y} \subset \mathbb{R}^p$ is discretized on a finite grid and if the resulting nonlinear programming problems are

Key words and phrases. Polynomial optimization; min-max optimization; robust optimization; semidefinite relaxations.

This work was performed during a visit in November 2010 at CRM (Centre de Recerca Matematica), a Mathematics center from UAB (Universitat Autnoma de Barcelona), and the author wishes to gratefully acknowledge financial support from CRM.

solved to global optimality, then convergence to a global optimum of the semi-infinite problem occurs as the grid size vanishes (see e.g. the discussion and the many references in [9]). Alternatively, in [9] the authors provide lower bounds on f^* by discretizing $\mathbf{K}$ and upper bounds via convex relaxations of the inner problem $\max_{\mathbf{y}\in\mathbf{Y}}\{g(\mathbf{x},\mathbf{y})\} \leq 0$. In [10] the authors also use a discretization scheme of $\mathbf{K}$ but now combined with a hierarchy of *sum of squares* convex relaxations for solving to global optimality.

Contribution. We restrict ourselves to problem $\mathbf{P}$ where:

- f, g are polynomials, and
- $\mathbf{X} \subset \mathbb{R}^n$ and $\mathbf{Y}_\mathbf{x} \subset \mathbb{R}^p$, $\mathbf{x} \in \mathbf{X}$, are compact basic semi-algebraic sets.

Then in this context we provide a numerical scheme whose novelty with respect to previous works is to avoid discretization of the set $\mathbf{Y}_\mathbf{x}$. Instead we use the "joint+marginal" methodology for parametric polynomial optimization developed by the author in [8], to provide a sequence of polynomials $(\Phi_d) \subset \mathbb{R}[\mathbf{x}]$ (with degree $2d$, $d \in \mathbb{N}$) that approximate from above the function $\Phi(\mathbf{x}) := \max_\mathbf{y} \{g(\mathbf{x},\mathbf{y}) \,:\, \mathbf{y} \in \mathbf{Y}_\mathbf{x}\}$, and with the strong property that if $d \to \infty$ then $\Phi_d \to \Phi$ in the L_1-norm. (In particular, $\Phi_{d_\ell} \to \Phi$ almost uniformly on $\mathbf{X}$ for some subsequence (d_ℓ), $\ell \in \mathbb{N}$.) Then, ideally, one could solve the nested sequence of polynomial optimization problems:

$$\mathbf{P}_d: \quad f^*_d = \min\{f(\mathbf{x}) \,:\, \Phi_d(\mathbf{x}) \,\leq\, 0\,\}, \qquad d = 1, 2, \ldots \tag{1.2}$$

because, as we show in the paper, the resulting sequence (f^*_d) is monotone non decreasing and converges to the desired value f^* as $d \to \infty$. Moreover, for fixed d, one may approximate (and often solve exactly) (1.2) by solving a hierarchy of semidefinite relaxations, as defined in [5]. However, as the size of these semidefinite relaxations increases very fast with d, in practice one rather let d be fixed, small, and relax the constraint $\Phi_d(\mathbf{x}) \leq 0$ to $\Phi_d(\mathbf{x}) \leq \epsilon$ for some scalar $\epsilon > 0$ that one may adjust dynamically during the algorithm.

2. Notation, definitions and preliminary results

Let $\mathbb{R}[\mathbf{x}]$ (resp. $\mathbb{R}[\mathbf{x},\mathbf{y}]$) denote the ring of real polynomials in the variables $\mathbf{x} = (x_1, \ldots, x_n)$ (resp. $\mathbf{x}$ and $\mathbf{y} = (y_1, \ldots, y_p)$), whereas $\Sigma[\mathbf{x}]$ (resp. $\Sigma[\mathbf{x},\mathbf{y}]$) denote its subset of sums of squares.

Let $\mathbb{R}[\mathbf{y}]_k \subset \mathbb{R}[\mathbf{y}]$ denote the vector space of real polynomials of degree at most k. For every $\alpha \in \mathbb{N}^n$ the notation $\mathbf{x}^\alpha$ stands for the monomial $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ and for every $i \in \mathbb{N}$, let $\mathbb{N}^p_d := \{\beta \in \mathbb{N}^n : \sum_j \beta_j \leq d\}$ whose cardinal is $s(d) = \binom{n+d}{n}$. A polynomial $f \in \mathbb{R}[\mathbf{x}]$ is written

$$\mathbf{x} \mapsto f(\mathbf{x}) \,=\, \sum_{\alpha\in\mathbb{N}^n} f_\alpha\, \mathbf{x}^\alpha,$$

and f can be identified with its vector of coefficients $\mathbf{f} = (f_\alpha)$ in the canonical basis. For a real symmetric matrix $\mathbf{A}$ the notation $\mathbf{A} \succeq 0$ stands for $\mathbf{A}$ is positive semidefinite.

A real sequence $\mathbf{y} = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, has a *representing measure* if there exists some finite Borel measure μ on $\mathbb{R}^n$ such that

$$y_\alpha = \int_{\mathbb{R}^n} \mathbf{x}^\alpha \, d\mu(\mathbf{x}), \qquad \forall\, \alpha \in \mathbb{N}^n.$$

Moment matrix. The *moment* matrix associated with a sequence $\mathbf{y} = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, is the real symmetric matrix $\mathbf{M}_d(\mathbf{y})$ with rows and columns indexed by $\mathbb{N}^n_d$, and whose entry (α, β) is just $y_{\alpha+\beta}$, for every $\alpha, \beta \in \mathbb{N}^n_d$. If $\mathbf{y}$ has a representing measure μ then $\mathbf{M}_d(\mathbf{y}) \succeq 0$ because

$$\langle \mathbf{f}, \mathbf{M}_d(\mathbf{y})\mathbf{f} \rangle \;=\; \int f^2 \, d\mu \;\geq 0, \qquad \forall\, \mathbf{f} \,\in \mathbb{R}^{s(d)}.$$

Localizing matrix. With $\mathbf{y}$ as above and $g \in \mathbb{R}[\mathbf{x}]$ (with $g(\mathbf{x}) = \sum_\gamma g_\gamma \mathbf{x}^\gamma$), the *localizing* matrix associated with $\mathbf{y}$ and g is the real symmetric matrix $\mathbf{M}_d(g\,\mathbf{y})$ with rows and columns indexed by $\mathbb{N}^n_d$, and whose entry (α, β) is just $\sum_\gamma g_\gamma y_{\alpha+\beta+\gamma}$, for every $\alpha, \beta \in \mathbb{N}^n_d$. If $\mathbf{y}$ has a representing measure μ whose support is contained in the set $\{\mathbf{x} \,:\, g(\mathbf{x}) \geq 0\}$ then $\mathbf{M}_d(g\,\mathbf{y}) \succeq 0$ because

$$\langle \mathbf{f}, \mathbf{M}_d(g\,\mathbf{y})\mathbf{f} \rangle \;=\; \int f^2 \, g \, d\mu(\mathbf{x}) \;\geq 0, \qquad \forall\, \mathbf{f} \,\in \mathbb{R}^{s(d)}.$$

Definition 2.1 (Archimedean property)**.** A set of polynomials $q_j \in \mathbb{R}[\mathbf{x}]$, $j = 0, \ldots, p$ (with $q_0 = 1$), satisfy the archimedean property if the quadratic polynomial $\mathbf{x} \mapsto M - \|\mathbf{x}\|^2$ can be written in the form:

$$M - \|\mathbf{x}\|^2 = \sum_{j=0}^{m} \sigma_j(\mathbf{x})\, q_j(\mathbf{x}),$$

for some sums of squares polynomials $(\sigma_j) \subset \Sigma[\mathbf{x}]$.

Let $\mathbf{D} := \{\mathbf{x} \in \mathbb{R}^n \,:\, q_j(\mathbf{x}) \geq 0, \, j = 1, \ldots, p\}$, and given a polynomial $h \in \mathbb{R}[\mathbf{x}]$, consider the hierarchy of semidefinite programs:

$$\text{(2.1)} \qquad \left\{ \begin{array}{rll} \rho_\ell = \max\limits_{\mathbf{z}} & L_{\mathbf{z}}(h) & \\ \text{s.t.} & \mathbf{M}_\ell(\mathbf{z}), \mathbf{M}_{\ell - v_j}(q_j\,\mathbf{z}) & \succeq 0, \quad j = 1, \ldots, p, \end{array} \right.$$

where $\mathbf{z} = (z_\alpha)$, $\alpha \in \mathbb{N}^n_{2\ell}$, and $v_j = \lceil (\deg q_j)/2 \rceil$, $j = 1, \ldots, p$.

Theorem 2.2 ([5, 7])**.** *Let a family of polynomials $(q_j) \subset \mathbb{R}[\mathbf{x}]$ satisfy the Archimedean property. Then as $\ell \to \infty$, $\rho_\ell \uparrow h^* = \min_{\mathbf{x}} \{h(\mathbf{x}) \,:\, \mathbf{x} \in \mathbf{D}\}$. Moreover, if $\mathbf{z}^*$ is an optimal solution of (2.1) and*

$$\text{(2.2)} \qquad \operatorname{rank} \mathbf{M}_\ell(\mathbf{z}^*) \;=\; \operatorname{rank} \mathbf{M}_{\ell - v}(\mathbf{z}^*) \; (=: r)$$

then $\rho_\ell = h^*$ and one may extract r global minimizers $\mathbf{x}^*_k \in \mathbf{D}$, $k = 1, \ldots, r$.

The size of the semidefinite program (2.1) grows as $O(\ell^n)$ with ℓ and so becomes rapidly prohibitive. Therefore, and even though practice reveals that convergence is fast and often finite, so far, the above methodology is limited to small to medium size problems. However for larger size problems with sparsity in the data and/or symmetries, *adhoc* and tractable versions of (2.1) exist. See for instance the sparse version of (2.1) proposed in [11], and whose convergence was proved in [6] when the sparsity pattern satifies the so-called *running intersection property.*

3. Main result

Let $p_s \in \mathbb{R}[\mathbf{x}]$, $s = 1, \ldots, sx$, and $h_j \in \mathbb{R}[\mathbf{x}, \mathbf{y}]$, $j = 1, \ldots, m$, be given polynomials and let $\mathbf{X} \subset \mathbb{R}^n$ be the basic semi-algebraic set

$$\mathbf{X} := \{\mathbf{x} \in \mathbb{R}^n \,:\, p_s(\mathbf{x}) \geq 0, \quad s = 1, \ldots, sx\}.$$

Next, for every $\mathbf{x} \in \mathbb{R}^n$, let $\mathbf{Y}_\mathbf{x} \subset \mathbb{R}^p$ be the basic semi-algebraic set described by:

$$\mathbf{Y}_\mathbf{x} \,=\, \{\,\mathbf{y} \in \mathbb{R}^p \,:\, h_j(\mathbf{x}, \mathbf{y}) \,\geq\, 0, \quad j = 1, \ldots, m\,\}, \tag{3.1}$$

and with $\mathbf{B} \supseteq \mathbf{X}$, let $\mathbf{K} \subset \mathbb{R}^n \times \mathbb{R}^p$ be the set

$$\mathbf{K} \,:=\, \{(\mathbf{x}, \mathbf{y}) \,\in\, \mathbb{R}^{n+p} \,:\, \mathbf{x} \in \mathbf{B}; \quad h_j(\mathbf{x}, \mathbf{y}) \geq 0, \quad j = 1, \ldots, m\}. \tag{3.2}$$

Observe that problem $\mathbf{P}$ in (1.1) is equivalent to:

$$\mathbf{P}: \qquad f^* = \min_{\mathbf{x} \in \mathbf{X}} \{\, f(\mathbf{x}) \,:\, \Phi(\mathbf{x}) \,\leq\, 0\,\} \tag{3.3}$$

$$\text{where} \quad \Phi(\mathbf{x}) = \max_{\mathbf{y}} \{g(\mathbf{x}, \mathbf{y}) \,:\, \mathbf{y} \in \mathbf{Y}_\mathbf{x}\,\}, \quad \mathbf{x} \in \mathbf{B}. \tag{3.4}$$

Lemma 3.1. *Let $\mathbf{K} \subset \mathbb{R}^{n+p}$ in (3.2) be compact and assume that for every $\mathbf{x} \in \mathbf{B} \subset \mathbb{R}^n$, the set $\mathbf{Y}_\mathbf{x}$ defined in (3.1) is nonempty. Then Φ is upper semicontinuous (u.s.c.) on $\mathbf{B}$. Moreover, if there is some compact set $\mathbf{Y} \subset \mathbb{R}^p$ such that $\mathbf{Y}_\mathbf{x} = \mathbf{Y}$ for every $\mathbf{x} \in \mathbf{B}$, then Φ is continuous on $\mathbf{B}$.*

Proof. Let $\mathbf{x}_0 \in \mathbf{B}$ be fixed, arbitrary, and let $(\mathbf{x}_k)_{k\in\mathbb{N}} \subset \mathbf{B}$ be a sequence that converges to $\mathbf{x}_0$ and such that

$$\limsup_{\mathbf{x} \to \mathbf{x}_0} \Phi(\mathbf{x}) \,=\, \lim_{k \to \infty} \Phi(\mathbf{x}_k).$$

As $\mathbf{K}$ is compact then so is $\mathbf{Y}_\mathbf{x}$ for every $\mathbf{x} \in \mathbf{B}$. Therefore, as $\mathbf{Y}_\mathbf{x} \neq \emptyset$ for all $\mathbf{x} \in \mathbf{B}$ and f is continuous, there exists an optimal solution $\mathbf{y}_k^* \in \mathbf{Y}_{\mathbf{x}_k}$ for every k. By compactness there exist a subsequence (k_ℓ) and $\mathbf{y}^* \in \mathbb{R}^p$ such that $(\mathbf{x}_{k_\ell}, \mathbf{y}_{k_\ell}^*) \to (\mathbf{x}_0, \mathbf{y}^*) \in \mathbf{K}$, as $\ell \to \infty$. Hence

$$\begin{aligned}
\limsup_{\mathbf{x} \to \mathbf{x}_0} \Phi(\mathbf{x}) &= \lim_{k \to \infty} \Phi(\mathbf{x}_k) \\
&= \lim_{k \to \infty} f(\mathbf{x}_k, \mathbf{y}_k^*) \,=\, \lim_{\ell \to \infty} f(\mathbf{x}_{k_\ell}, \mathbf{y}_{k_\ell}^*) \\
&= f(\mathbf{x}_0, \mathbf{y}^*) \,\leq\, \Phi(\mathbf{x}_0),
\end{aligned}$$

which proves that Φ is u.s.c. at $\mathbf{x}_0$. As $\mathbf{x}_0 \in \mathbf{B}$ was arbitrary, Φ is u.s.c. on $\mathbf{B}$.

Next, assume that there is some compact set $\mathbf{Y} \subset \mathbb{R}^p$ such that $\mathbf{Y}_\mathbf{x} = \mathbf{Y}$ for every $\mathbf{x} \in \mathbf{B}$. Let $\mathbf{x}_0 \in \mathbf{B}$ be fixed arbitrary with $\Phi(\mathbf{x}_0) = f(\mathbf{x}_0, \mathbf{y}_0^*)$ for some $\mathbf{y}_0^* \in \mathbf{Y}$. Let $(\mathbf{x}_n) \subset \mathbf{B}$, $n \in \mathbb{N}$, be a sequence such that $\mathbf{x}_n \to \mathbf{x}_0$ as $n \to \infty$, and $\Phi(\mathbf{x}_0) \geq \liminf_{\mathbf{x} \to \mathbf{x}_0} \Phi(\mathbf{x}) = \lim_{n\to\infty} \Phi(\mathbf{x}_n)$. Again, let $\mathbf{y}_n^* \in \mathbf{Y}$ be such that $\Phi(\mathbf{x}_n) = f(\mathbf{x}_n, \mathbf{y}_n^*)$, $n \in \mathbb{N}$. By compactness, consider an arbitrary converging subsequence $(n_\ell) \subset \mathbb{N}$, i.e., such that $(\mathbf{x}_{n_\ell}, \mathbf{y}_{n_\ell}^*) \to (\mathbf{x}_0, \mathbf{y}^*) \in \mathbf{K}$ as $\ell \to \infty$, for some $\mathbf{y}^* \in \mathbf{Y}$. Suppose that $\Phi(\mathbf{x}_0)\,(= f(\mathbf{x}_0, \mathbf{y}_0^*)) > f(\mathbf{x}_0, \mathbf{y}^*)$, say $\Phi(\mathbf{x}_0) > f(\mathbf{x}_0, \mathbf{y}^*) + \delta$ for some $\delta > 0$. By continuity of f, $f(\mathbf{x}_{n_\ell}, \mathbf{y}_{n_\ell}^*) < f(\mathbf{x}_0, \mathbf{y}^*) + \delta/2$ for every $\ell > \ell_1$ (for some ℓ_1). But again, by continuity, $|f(\mathbf{x}_{n_\ell}, \mathbf{y}_0^*) - f(\mathbf{x}_0, \mathbf{y}_0^*)| < \delta/3$ whenever $\ell > \ell_2$ (for some ℓ_2). And so we obtain the contradiction

$$\begin{aligned} \Phi(\mathbf{x}_{n_\ell}) &\geq f(\mathbf{x}_{n_\ell}, \mathbf{y}_0^*) \;>\; \Phi(\mathbf{x}_0) - \delta/3 \\ \Phi(\mathbf{x}_{n_\ell}) &= f(\mathbf{x}_{n_\ell}, \mathbf{y}_{n_\ell}^*) \;<\; \Phi(\mathbf{x}_0) - \delta/2, \end{aligned}$$

whenever $\ell > \max[\ell_1, \ell_2]$. Therefore, $f(\mathbf{x}_0, \mathbf{y}_0^*) = f(\mathbf{x}_0, \mathbf{y}^*)$ and so,

$$f(\mathbf{x}_0, \mathbf{y}_0^*) \;=\; \Phi(\mathbf{x}_0) \;=\; f(\mathbf{x}_0, \mathbf{y}^*) \;=\; \lim_{\ell\to\infty} \Phi(\mathbf{x}_{n_\ell}) \;=\; \liminf_{\mathbf{x}\to\mathbf{x}_0} \Phi(\mathbf{x}_0) \leq \Phi(\mathbf{x}_0),$$

which combined with Φ being u.s.c., yields that Φ is continuous at $\mathbf{x}_0$. $\square$

We next explain how to

- approximate the function $\mathbf{x} \mapsto \Phi(\mathbf{x})$ on $\mathbf{B}$ by a polynomial, and
- evaluate (or at least approximate) $\Phi(\mathbf{x})$ for some given $\mathbf{x} \in \mathbf{B}$, to check whether $\Phi(\mathbf{x}) \leq 0$.

Indeed, these are the two main ingredients of the algorithm that we present later.

3.1. **Certificate of $\Phi(\mathbf{x}) \leq 0$.** For every $\mathbf{x} \in \mathbf{X}$ and $j = 1, \ldots, m$, let $h_j^\mathbf{x} \in \mathbb{R}[\mathbf{y}]$ be the polynomial $\mathbf{y} \mapsto h_j^\mathbf{x}(\mathbf{y}) := h_j(\mathbf{x}, \mathbf{y})$, $j = 1, \ldots, m$, and consider the hierarchy of semidefinite programs:

$$(3.5)\qquad \mathbf{Q}_\ell(\mathbf{x}):\quad \left\{ \begin{array}{rll} \rho_\ell(\mathbf{x}) = \displaystyle\max_{\mathbf{z}} & L_\mathbf{z}(g_\mathbf{x}) & \\ \text{s.t.} & \mathbf{M}_\ell(\mathbf{z}), \mathbf{M}_{\ell-v_j}(h_j^\mathbf{x}\,\mathbf{z}) \;\succeq 0, & j = 1, \ldots, m, \end{array} \right.$$

where $\mathbf{z} = (z_\alpha)$, $\alpha \in \mathbb{N}^n_{2\ell}$, and $v_j = \lceil (\deg h_j^\mathbf{x})/2 \rceil$, $j = 1, \ldots, m$. Obviously one has $\rho_\ell(\mathbf{x}) \geq \Phi(\mathbf{x})$ for every ℓ, and

Corollary 3.2. *Let $\mathbf{x} \in \mathbf{X}$ and assume that the polynomials $(h_j^\mathbf{x}) \subset \mathbb{R}[\mathbf{y}]$ satisfy the Archimidean property. Then:*

(a) *As $\ell \to \infty$, $\rho_\ell(\mathbf{x}) \downarrow \Phi(\mathbf{x}) = \max\{g(\mathbf{x}, \mathbf{y}) \,:\, \mathbf{y} \in \mathbf{Y}_\mathbf{x}\}$. In particular, if $\rho_\ell(\mathbf{x}) \leq 0$ for some ℓ, then $\Phi(\mathbf{x}) \leq 0$.*

(b) *Moreover, if $\mathbf{z}^*$ is an optimal solution of (3.5) that satisfies*

$$\operatorname{rank} \mathbf{M}_\ell(\mathbf{z}^*) \;=\; \operatorname{rank} \mathbf{M}_{\ell-v}(\mathbf{z}^*)\; (=: r),$$

(where $v := \max_j [v_j]$), then $\rho_\ell(\mathbf{x}) = \Phi(\mathbf{x})$ and there are r global maximizers $\mathbf{y}(k) \in \mathbf{Y}_\mathbf{x}$, $k = 1, \ldots, r$.

Corollary 3.2 is a direct consequence of Theorem 2.2.

3.2. **Approximating the function Φ.** Let $\mathbf{B} \supseteq \mathbf{X}$ be a *simple* set like e.g., a simplex, a box or an ellipsoid and let μ be the finite Borel probability measure uniformly distributed on $\mathbf{B}$. Therefore, the vector $\gamma = (\gamma_\alpha)$, $\alpha \in \mathbb{N}^n$, of moments of μ, i.e.,

$$\gamma_\alpha := \int_{\mathbf{B}} \mathbf{x}^\alpha \, d\mu(\mathbf{x}), \qquad \alpha \in \mathbb{N}^n,$$

can be computed easily. For instance, in the sequel we assume that $\mathbf{B} = [-1,1]^n$, and let $\theta_i \in \mathbb{R}[\mathbf{x},\mathbf{y}]$ be the polynomial $(\mathbf{x},\mathbf{y}) \mapsto \theta_i(\mathbf{x},\mathbf{y}) := 1 - x_i^2$, $i = 1,\ldots,n$.

Observe that the function Φ is defined in (3.4) via a *parametric* polynomial optimization problem (with $\mathbf{x}$ being the parameter vector). Therefore, following [8], let $r_j = \lceil (\deg h_j)/2 \rceil$, $j = 1,\ldots,m$, and consider the semidefinite relaxation:

$$(3.6) \qquad \left\{ \begin{array}{rll} \rho_d = \displaystyle\max_{\mathbf{z}} & L_{\mathbf{z}}(g) & \\ \text{s.t.} & \mathbf{M}_d(\mathbf{z}), \mathbf{M}_{d-r_j}(h_j\,\mathbf{z}) & \succeq 0, \quad j = 1,\ldots,m \\ & \mathbf{M}_{d-1}(\theta_i\,\mathbf{z}) & \succeq 0, \quad i = 1,\ldots,n \\ & L_{\mathbf{z}}(\mathbf{x}^\alpha) & = \gamma_\alpha, \quad \alpha \in \mathbb{N}^n_{2d}, \end{array} \right.$$

where $\mathbf{z} = (z_{\alpha\beta})$, $(\alpha,\beta) \in \mathbb{N}^{n+p}_{2d}$. Writing $g_0 \equiv 1$, the dual of the semidefinite program reads

$$(3.7) \qquad \left\{ \begin{array}{rl} \rho_d^* = \displaystyle\min_{q,\sigma_j} & \displaystyle\int_{\mathbf{B}} q(\mathbf{x})\, d\mu(\mathbf{x}) \\ \text{s.t.} & q(\mathbf{x}) - g(\mathbf{x},\mathbf{y}) = \displaystyle\sum_{j=0}^m \sigma_j(\mathbf{x},\mathbf{y}) h_j(\mathbf{y}) \\ & q \in \mathbb{R}[\mathbf{x}]_{2d},\ \sigma_j \in \Sigma[\mathbf{x},\mathbf{y}] \\ & \deg \sigma_j\, h_j \leq 2d, \quad j = 0,\ldots,m. \end{array} \right.$$

It turns out that any optimal solution of the semidefinite program (3.7) permits to approximate Φ in a strong sense.

Theorem 3.3 ([8]). *Let $\mathbf{K} \subset \mathbb{R}^{n+p}$ in (3.2) be compact. Assume that the polynomials $h_j, \theta_i \in \mathbb{R}[\mathbf{x},\mathbf{y}]$ satisfy the Archimedean property and assume that for every $\mathbf{x} \in \mathbf{B}$, the set $\mathbf{Y}_{\mathbf{x}}$ defined in (3.1) is nonempty. Let $\Phi_d \in \mathbb{R}[\mathbf{x}]_{2d}$ be an optimal solution of (3.7). Then:*

(a) *$\Phi_d \geq \Phi$ and as $d \to \infty$,*

$$(3.8) \qquad \int_{\mathbf{B}} (\Phi_d(\mathbf{x}) - \Phi(\mathbf{x}))\, d\mu(\mathbf{x}) \;=\; \int_{\mathbf{B}} |\, \Phi_d(\mathbf{x}) - \Phi(\mathbf{x})\,|\, d\mu(\mathbf{x}) \;\to\; 0,$$

that is, $\Phi_d \to \Phi$ for the $L_1(\mathbf{B},\mu)$-norm[1].

(b) *There is a subsequence (d_ℓ), $\ell \in \mathbb{N}$, such that $\Phi_{d_\ell} \to \Phi$, μ-almost uniformly*[2] *in $\mathbf{B}$, as $\ell \to \infty$.*

[1] $L_1(\mathbf{B},\mu)$ is the Banach space of μ-integrable functions on $\mathbf{B}$, with norm $\|f\| = \int_{\mathbf{B}} |f| d\mu$.

[2] If one fixes $\epsilon > 0$ arbitrary then there is some $A \in \mathcal{B}(\mathbf{B})$ such that $\mu(A) < \epsilon$ and $\Phi_{d_\ell} \to \Phi$ uniformly on $\mathbf{B} \setminus A$, as $\ell \to \infty$.

The proof of (a) can be found in [8], whereas (b) follows from (a) and [1, Theorem 2.5.3].

3.3. **An algorithm.** The idea behind the algorithm is to approximate $\mathbf{P}$ in (1.1) with the *polynomial* optimization problem: $(\mathbf{P}_d^\epsilon)$:

$$\mathbf{P}_d^\epsilon: \quad f_d^\epsilon = \min_{\mathbf{x}\in\mathbf{X}} \{ f(\mathbf{x}) \,:\, \Phi_d(\mathbf{x}) \leq \epsilon \}, \qquad d = 1, 2, \ldots \tag{3.9}$$

with $d \in \mathbb{N}, \epsilon > 0$ fixed, and Φ_d as in Theorem 3.3, for every $d = 1, \ldots$.

Obviously, for $\epsilon = 0$ one has $f_d^0 \geq f^*$ for all d because by definition $\Phi_d \geq \Phi$ for every $d \in \mathbb{N}$. However it may happen that $\mathbf{P}_d^0$ has no solution. Also, if $\mathbf{x}^*$ is an optimal solution of $\mathbf{P}$ and $\Phi_d(\mathbf{x}^*) < 0$, it may happen that $\Phi_d(\mathbf{x}^*) > 0$ if d is not large enough. This is why one needs to relax the constraint $\Phi \leq 0$ to $\Phi_d \leq \epsilon$ for some $\epsilon > 0$. However, in view of Theorem 3.3, one expects that $f_d^\epsilon \approx f^*$ provided that d and ϵ are sufficiently large and small, respectively. And indeed:

Theorem 3.4. *Assume that $\mathbf{X}$ is the closure of an open set. Let $\epsilon \geq 0$ be fixed, arbitrary and with f_d^ϵ be as in (3.9), let $\mathbf{x}_d^\epsilon \in \mathbf{X}$ be any optimal solution of (3.9) (including the case where $\epsilon = 0$), and let*

$$\tilde{f}_d^\epsilon := \min\{f_\ell^\epsilon : \ell = 1, \ldots, d\} = f(\mathbf{x}_{\ell(d)}^\epsilon) \quad \text{for some } \ell(d) \in \{1, \ldots, d\}.$$

(a) *If $\epsilon > 0$ there exists $d_\epsilon \in \mathbb{N}$ such that for every $d \geq d_\epsilon$, $f(\mathbf{x}_{\ell(d)}^\epsilon) < f^* + \epsilon$.*

(b) *If there is an optimal solution $\mathbf{x}^* \in \mathbf{X}$ of (1.1) such that $\Phi(\mathbf{x}^*) < 0$, then there exists $d_0 \in \mathbb{N}$ such that for every $d \geq d_0$, $f^* \leq f(\mathbf{x}_{\ell(d)}^0) < f^* + \epsilon$.*

Proof. (a) With $\epsilon > 0$ fixed, arbitrary, let $\mathbf{x}_\epsilon^* \in \mathbf{X}$ be such that $\Phi(\mathbf{x}_\epsilon^*) \leq 0$ and $f(\mathbf{x}_\epsilon^*) < f^* + \epsilon/2$. We may assume that $\mathbf{x}_\epsilon^*$ is not on the boundary of $\mathbf{X}$. Let $O_\epsilon^1 := \{\mathbf{x} \in \mathbf{X} : \Phi(\mathbf{x}) < \epsilon/2\}$ which is an open set because Φ is u.s.c. (by Lemma 3.1), and so $\mu(O_\epsilon^1) > 0$. Next, as f is continuous, there exists $\rho_0 > 0$ such that $f < f^* + \epsilon$ whenever $\mathbf{x} \in O_\epsilon^2 := \{\mathbf{x} \in \mathbf{X} : \|\mathbf{x} - \mathbf{x}_\epsilon^*\| < \rho_0\}$. Observe that $\rho := \mu(O_\epsilon^1 \cap O_\epsilon^2) > 0$ because $\mathbf{x}_\epsilon^* \in O_\epsilon^1 \cap O_\epsilon^2$. Next, by Theorem 3.3(b), there is a subsequence (d_ℓ), $\ell \in \mathbb{N}$, such that $\Phi_{d_\ell} \to \Phi$, μ-almost uniformly on $\mathbf{B}$. Hence, there is some Borel set $A_\epsilon \subset \mathbf{B}$, and integer $\ell_\epsilon \in \mathbb{N}$, such that $\mu(A_\epsilon) < \rho/2$ and $\sup_{\mathbf{x}\in\mathbf{X}\setminus A_\epsilon} |\Phi(\mathbf{x}) - \Phi_{d_\ell}(\mathbf{x})| < \epsilon/2$ for all $\ell \geq \ell_\epsilon$. In particular, as $\mu(A_\epsilon) < \rho/2 < \mu(O_\epsilon^1 \cap O_\epsilon^2)$, the set $\Delta_\epsilon := (O_\epsilon^1 \cap O_\epsilon^2) \setminus A_\epsilon$ has positive μ-measure. Therefore, $f(\mathbf{x}) < f^* + \epsilon$ and $\Phi_{d_\ell}(\mathbf{x}) < \epsilon$ whenever $\ell \geq \ell_\epsilon$ and $\mathbf{x} \in \Delta_\epsilon$, which in turn implies $f_{d_\ell}^\epsilon < f^* + \epsilon$, and consequently, $\tilde{f}_d^\epsilon = f(\mathbf{x}_{\ell(d)}^\epsilon) < f^* + \epsilon$, the desired result.

(b) Let $\epsilon' := -\Phi(\mathbf{x}^*)$, and let $O_{\epsilon'}^1 := \{\mathbf{x} \in \mathbf{X} : \Phi(\mathbf{x}) < -\epsilon'/2\}$ which is a nonempty open set because it contains $\mathbf{x}^*$. Let $O_{\epsilon'}^2$ be as O_ϵ^2 in the proof of (a), but now with $\mathbf{x}_{\epsilon'}^* = \mathbf{x}^* \in \mathbf{X}$. Both $O_{\epsilon'}^1$ and $O_{\epsilon'}^2$ are nonempty open sets because they contain $\mathbf{x}^*$. The rest of the proof is like for the proof of (a), but noticing that now for every $\mathbf{x} \in \Delta_{\epsilon'}$ one has $\Phi_{d_\ell}(\mathbf{x}) < -\epsilon'/2 + \epsilon'/2 = 0$, and so $\mathbf{x}$ is feasible for (3.9) with $\epsilon = 0$. Next, by feasiblity $f(\mathbf{x}) \geq f^*$ since the resulting feasible set in (3.9) is smaller than that of (1.2) because $\Phi_d \geq \Phi$, for all d. And so $f^* \leq f(\mathbf{x}) < f^* + \epsilon$ whenever $\mathbf{x} \in \Delta_\epsilon$, and $\ell \geq \ell_\epsilon$, from which (b) follows. $\square$

Theorem 3.4 provides a rationale behind the algorithm that we present below. In solving (3.9) with d sufficiently large and small ϵ (or even $\epsilon = 0$), f_d^ϵ would provide a good approximation of f^*. But in principle, computing the global optimum f_d^ϵ is still a difficult problem. However, $\mathbf{P}_d^\epsilon$ is a polynomial optimization problem. Therefore, by Theorem 2.2, if the polynomials $(p_s) \subset \mathbb{R}[\mathbf{x}]$ that define $\mathbf{X}$ satisfy the Archimedean property (see Definition 2.1) we can approximate f_d^ϵ from below, as closely as desired, by a monotone sequence (f_{dt}^ϵ), $t \in \mathbb{N}$, obtained by solving the hierarchy of semidefinite relaxations (2.1), which here read:

$$(3.10) \qquad \left\{ \begin{array}{rll} f_{dt}^\epsilon = \displaystyle\min_{\mathbf{z}} & L_{\mathbf{z}}(f) & \\ \text{s.t.} & \mathbf{M}_t(\mathbf{z}), \mathbf{M}_{t-d}(\epsilon - \Phi_d\, \mathbf{z}) & \succeq 0 \\ & \mathbf{M}_{t-t_s}(p_s\, \mathbf{z}) & \succeq 0, \quad s = 1, \ldots, sx, \end{array} \right.$$

where $t_s = \lceil (\deg p_s)/2 \rceil$, $s = 1, \ldots, sx$.

Corollary 3.5. *Assume that the polynomials $(p_s) \subset \mathbb{R}[\mathbf{x}]$ satisfy the Archimedean property. Then $f_{dt}^* \uparrow f_d^*$ as $t \to \infty$. Moreover, if $\mathbf{z}^*$ is an optimal solution of (3.10) and*

$$(3.11) \qquad \operatorname{rank} \mathbf{M}_t(\mathbf{z}^*) \;=\; \operatorname{rank} \mathbf{M}_{t-t_0}(\mathbf{z}^*) \; (=: r)$$

(where $t_0 := \max[d, \max_s[t_s]]$) then $f_{dt}^* = f_d^*$ and one may extract r global minimizers $\mathbf{x}_d^*(k) \in \mathbf{X}$, $k = 1, \ldots, r$. That is, for every $k = 1, \ldots, r$, $f(\mathbf{x}_d^*(k)) = f_d^*$ and $\Phi_d(\mathbf{x}_d^*(k)) \leq \epsilon$.

However, given a minimizer $\mathbf{x}_d^* \in \mathbf{X}$, if one the one hand $\Phi_d(\mathbf{x}_d^*) \leq \epsilon$, on the other hand it may not satisfy $\Phi(\mathbf{x}_d^*) \leq 0$. (Recall that checking whether $\Phi(\mathbf{x}_d^*) \leq 0$ can be done via solving the hierarchy of relaxations $\mathbf{Q}_\ell(\mathbf{x})$ in (3.5) with $\mathbf{x} := \mathbf{x}_d^*$.) If this happens then one solves again (3.10) for a smaller value of ϵ, etc., until one obtains some $\mathbf{x}_d^* \in \mathbf{X}$ with $\Phi(\mathbf{x}_d^*) \leq 0$.

Finally, and as already mentioned, if d is relatively large, the size of semidefinite relaxations (3.10) to compute f_{dt}^ϵ in (3.10) becomes too large for practical implementation (as one must have $t \geq d$). So in practice one let d be fixed at a small value, typically the smallest possible value of d, i.e., 1 (Φ_d is quadratic) or 2 (Φ_d is quartic)), and one updates ϵ as indicated above. So the resulting algorithm reads:

Algorithm.
Input: $\ell, d, k^* \in \mathbb{N}$, $\epsilon_0 > 0$ (e.g. $\epsilon_0 := 10^{-1}$), $d \in \mathbb{N}$, $\tilde{\mathbf{x}} := \star$, $f(\star) = +\infty$.
Output: $f(\mathbf{x}_d^*)$ with $\mathbf{x}_d^* \in \mathbf{X}$ and $\Phi(\mathbf{x}_d^*) \leq 0$.
Step 1: Set $k = 1$ and $\epsilon(k) = 1$.
Step 2: While $k \leq k^*$, solve $\mathbf{P}_d^{\epsilon(k)}$ in (3.9) $\to \mathbf{x}_k^* \in \mathbf{X}$.
Step 3: Solve $\mathbf{Q}_\ell(\mathbf{x}_k^*)$ in (3.5) $\to \rho_\ell(\mathbf{x}_k^*)$.

If $-\epsilon_0 \leq \rho_\ell(\mathbf{x}_k^*) \leq 0$ set $\mathbf{x}_d^* := \mathbf{x}_k^*$ and STOP.
If $\rho_\ell(\mathbf{x}_k^*) < -\epsilon_0$ then:

- if $f(\tilde{\mathbf{x}}) > f(\mathbf{x}_k^*)$ then set $\tilde{\mathbf{x}} := \mathbf{x}_k^*$. If $k = k^*$ then $\mathbf{x}_k^* := \tilde{\mathbf{x}}$.
- set $\epsilon(k+1) := 2\epsilon(k)$, $k := k+1$ and go to Step 2.

If $\rho_\ell(\mathbf{x}_k^*) > 0$ set $\epsilon(k+1) := \epsilon(k)/2$, $k := k+1$ and go to Step 2.

Observe that in Step 2 of the above algorithm, one assumes that by solving $\mathbf{P}_d^{\epsilon(k)}$ one obtains $\mathbf{x}_k^* \in \mathbf{X}$.

3.4. **Numerical experiments.** We have taken Examples $2, 7, 9$, K, M, N, all from Bhattacharjee et al. [9, Appendix A] and whose data are polynomials, except for problem L. For the latter problem, the non-polynomial function $\mathbf{x} \mapsto \min[0, (x_1 - x_2)]$ is semi-algebraic and can be generated by introducing an additional variable x_3, with the polynomial constraints:

$$x_3^2 = (x_1 - x_2)^2; \qquad x_3 \geq 0.$$

Indeed, $2\min[0, (x_1 - x_2)] = x_1 - x_2 - x_3$.

Although these examples are quite small, they are still non trivial (and even difficult) to solve, and we wanted to test the above methodology with small relaxation order d. In fact we have even considered the smallest possible d, i.e., $d = 1$ (Φ_d is quadratic). Results in Table 1 are quite good since by using the semidefinite relaxation of minimal order "d" one obtains an optimal value f_d^* quite close to f^*, at the price of updating ϵ several times.

Next, for Problem L, if we now increase d to $d = 2$, we improve the optimal value which becomes $f_d^* = 0.3849$ with $\epsilon = 2.2$. However, for Problem M, increasing d does not improve the optimal value.

	best known value	f_d^*	final value of ϵ
problem 2	0.194	0.198	1.895
problem 7	1.0	1.41	5
problem 9	-12.0	-14.47*	0
problem K	-3.0	-3.0	3.037
problem L	0.3431	0.435	2.295
problem M	1.0	2.25	2.592
problem N	0.0	10^{-8}	0

TABLE 1. Examples of [2, Table 6.1] with minimal d

4. Conclusion

We have presented an algorithm for semi-infinite (global) polynomial optimization whose novelty with respect to previous works is to *not* rely on a discretization scheme. Instead, it uses a polynomial approximation Φ_d of the function Φ, obtained by solving some semidefinite relaxation attached to the "joint+marginal" approach developed in [8] for parametric optimization, which guarantees (strong) convergence $\Phi_d \to \Phi$ in L_1-norm. Then for fixed d, one has to solve a polynomial optimization problem, which can be done by solving an appropriate hierarchy of semidefinite relaxations. Of course, and especially in view of the present status of semidefinite solvers, so far the present methodology is limited to small to medium size problems, unless sparsity in the data and/or symmetries are taken into account appropriately, as described in e.g [6, 11]. Preliminary results on non trivial (but small size) examples are encouraging.

References

[1] R.B. Ash. *Real Analysis and Probability.* Academic Press Inc., Boston (1972).

[2] B. Bhattacharjee, W.H. Green Jr., P.I. Barton. Interval mthods for semi-infinite programs, Comput. Optim. and Appl. **30**, 63–93 (2005).

[3] A. Ben-Tal, S. Boyd, A. Nemirovski. Extending Scope of Robust Optimization: Comprehensive Robust Counterparts of Uncertain Problems, Math. Program. Sér. B **107**, 63–89 (2006).

[4] D. Henrion, J.B. Lasserre, J. Lofberg. GloptiPoly 3: moments, optimization and semidefinite programming, Optim. Methods and Softwares **24**, 761–779 (2009).

[5] J.B. Lasserre. Global optimization with polynomials and the problem of moments. SIAM J. Optim. **11**, 796–817 (2001).

[6] J.B. Lasserre. Convergent SDP-relaxations in polynomial optimization with sparsity, SIAM J. Optim. **17**, 822–843 (2006).

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

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

[9] A. Mitsos, P. Lemonidis, Cha Kun Lee, P.I. Barton. Relaxation-based bounds for semi-infinite programs, SIAM J. Optim. **19**, 77–113 (2008).

[10] P. Parpas, B. Rustem. An algorithm for the global optimization of a class of continuous minimax problems, J. Optim. Theor. Appl. **141**, 461–473 (2009).

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

J.B. Lasserre
LAAS-CNRS
7 Avenue du Colonel Roche
31077 Toulouse Cédex 4
France.
E-mail address: lasserre@laas.fr