Random Coordinate Minimization Method with Eventual Transverse Directions for Constrained Polynomial Optimization

Random Coordinate Minimization Method with Eventual Transverse Directions for Constrained Polynomial Optimization

Giuseppe C. Calafiore, Carlo Novara and Corrado Possieri

***Abstract*— In this paper, we propose a novel algorithm for the solution of polynomial optimization problems. In particular, we show that, under mild assumptions, such problems can be solved by performing a random coordinate-wise minimization and, eventually, when a coordinate-wise minimum has been reached, an univariate minimization along a randomly chosen direction. The theoretical results are corroborated by a numerical example where the given procedure is compared with several other methods able to solve polynomial problems.**

I. Introduction

Several problems arising from system analysis and control design can be recast into optimization problems involving polynomial objective and constraint functions. Some remarkable examples are the computation of the parametric stability margin of a control system affected by uncertainty [1], the estimation of the domain of attraction of an equilibrium of a nonlinear systems [2], the estimation of the fundamental matrix from point correspondences [3], the robust design of control systems that satisfy several specifications [4], and the simultaneous stabilization of multiple plants [5].

Thus, a lot of research effort has been spent to design procedures to address such problems, which may be surprisingly difficult to solve [6]. Among the tools that are available in the literature to solve these (generically non-convex) programming problems [7], convex relaxations are some of the most widely employed. A convex relaxation is essentially a convex problem whose solution provides information about the solution to the original problem. Methods belonging to such a class are sum-of-squares (briefly, SOS) optimization [8], [9], the method of moments [10], [11], and linear-matrix-inequality (breifly, LMI) relaxations [12], [13]. The drawback of these methods is that, although the convex relaxation can be solved with very efficient methods, non-convex constraints are usually substituted with relaxed, but often nonequivalent, constraints that lead to conservative solutions to the problem [14]. As an example, the largest real γ such that $f(\mathbf{x}) - \gamma$ is an SOS is just a lower bound on $\min f(\mathbf{x})$ that need not be tight when dealing with global problems [15], especially if the degree of f and the number of variables involved in the minimization problem is large [16]. Nonetheless, when dealing with positivity on bounded domains [17], [18], the cone of SOS polynomials is densely embedded in the cone of positive polynomials [19], i.e., there is no gap.

The main objective of this paper is to propose a novel algorithm to solve polynomial optimization problems that does not rely on convex relaxations. In particular, we show that, if one performs a random coordinate descent up to a coordinate-wise minimum and, when such a minimum is reached, performs a minimization along a random transverse direction, then one obtains a sequence that asymptotically converges to the global solution of the optimization problem.

The main advantage of this procedure with respect to the global minimization is that, at each of its steps, one has to solve just an univariate problem that can be easily addressed (even exactly [20], [21]) using standard tools [22]. The main difference between the algorithm proposed here and the one given in [23] is that here transverse coordinates are allowed just once that a coordinate-wise minimum is reached, whereas in [23] they are allowed at each iteration. This leads to a substantially different procedure that may lead to faster convergence times, especially in the case that the problem to be solved is convex or coordinate-wise convex [24].

Differently from classical nonlinear programming methods, such as gradient-based algorithms [25], the given algorithm asymptotically tends to the globally optimal solution.

II. Problem statement

A. Notation

Let $\mathbb{Z}$, $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}_{\geqslant 0}$, and $\mathbb{R}_{>0}$ denote the sets of integer, natural, real, nonnegative real, and positive real numbers, respectively. Let $\mathbb{B}$, $\mathbb{B}^o$ and $\mathbb{S}$ denote the closed and open unit balls and the unit sphere in the Euclidean norm, respectively. The symbol $\overline{1,n}$ denotes the set $\{1, 2, \dots, n\}$. The symbol $\mathbf{v}_i$ denotes the i-th coordinate direction in the Euclidean space and $\mathbb{V} \doteq \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$. Given a vector $\mathbf{h} = \begin{bmatrix} h_1 & \cdots & h_m \end{bmatrix}$, the inequality $\mathbf{h} \leqslant 0$ is understood entry-wise, i.e., $h_1 \leqslant 0, \dots, h_m \leqslant 0$. The symbols $\vee$ and $\wedge$ denote the logical `or` and `and` operators, respectively.

Given a compact set $\mathcal{A} \subset \mathbb{R}^n$, $\|\mathbf{x}\|_{\mathcal{A}} \doteq \inf_{\mathbf{y}\in\mathcal{A}} \|\mathbf{x} - \mathbf{y}\|_2$ denotes the ℓ_2 distance between $\mathbf{x} \in \mathbb{R}^n$ and $\mathcal{A}$.

A function $\varrho : \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$ is positive semidefinite with respect to $\mathcal{A} \subset \mathbb{R}^n$, denoted $\varrho \in \mathrm{pd}(\mathcal{A})$, if $\varrho(\mathbf{x}) = 0 \iff \mathbf{x} \in \mathcal{A}$. A continuous function $\alpha : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ is of class $\mathcal{K}_\infty$ if it is strictly increasing, $\alpha(0) = 0$, and $\lim_{r\to+\infty} \alpha(r) = +\infty$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *radially unbounded* on $\Omega \subset \mathbb{R}^n$, denoted $f \in \mathrm{ru}(\Omega)$, if for every $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$ such that $\mathbf{x}^k \in \Omega$ for all $k \in \mathbb{N}$ and $\lim_{k\to+\infty} \|\mathbf{x}^k\|_2 = +\infty$, one has $\lim_{k\to\infty} f(\mathbf{x}^k) = +\infty$.

$\mathrm{Uni}(\mathcal{C})$ denotes the uniform distribution over the compact set $\mathcal{C}$, whereas $\sim$ reads as "has distribution".

Following [26], letting $\mathbf{x} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top$, with $n \in \mathbb{N}$, be a vector of variables, the *ring* of all the polynomials in $\mathbf{x}$ with coefficients in $\mathbb{R}$ is denoted $\mathbb{R}[\mathbf{x}]$.

G. C. Calafiore, C. Novara and C. Possieri are with Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino, 10129 Torino, Italy. G. C. Calafiore is also with IEIIT-CNR Torino, 10129 Torino, Italy. (e-mails: {giuseppe.calafiore, carlo.novara, corrado.possieri}@polito.it)

B. Problem statement

Given $f, h_1, \dots, h_m \in \mathbb{R}[\mathbf{x}]$, consider the following polynomial minimization problem (briefly, PMP):

$$\left| \begin{array}{ll} \min & f(\mathbf{x}), \\ \text{with} & h_i(\mathbf{x}) \leqslant 0, \quad i = 1, \dots, m. \end{array} \right. \tag{1}$$

The main purpose of this paper is to design a procedure for computing a solution to the PMP (1), i.e., find $f^\star \in \mathbb{R}$ and

$$\mathbf{x}^\star \in \Omega \doteq \{\mathbf{x} \in \mathbb{R}^n : \mathbf{h}(\mathbf{x}) \leqslant 0\},$$

where $\mathbf{h} \doteq [\ h_1 \quad \cdots \quad h_m\ ]^\top$, such that

$$f^\star = f(\mathbf{x}^\star) = \min_{\mathbf{x} \in \Omega} f(\mathbf{x}).$$

Since f need not be convex on the *feasible set* Ω, the minimization problem (1) is generally NP-hard, even for simple instances of the constraint set [27].

Due to the fact that a lot of problems arising from the analysis and design of control system can be reduced to a PMP in the form (1), a lot of research effort has been spent to specifically solve it (see [23] for a review of the different methods available in the literature to pursue such a goal).

III. Random coordinate descent with eventual transverse directions

The algorithm that we propose to solve the PMP (1) is described in words as follows: perform a classical random coordinate descent on the PMP (1) while there exists a coordinate direction such that, by performing a step of the coordinate descent along such a direction, the objective function f decreases of at least $\omega \in \mathbb{R}_{>0}$. On the other hand, when such a condition does not hold, pick a random direction $\mathbf{s}^k \in \mathbb{S}$ and perform a transverse descent along such a direction. More formally, we update the estimate of the solution to the PMP (1) according to the following rule

$$\mathbf{x}^{k+1} \in \mathbf{x}^k + \lambda^k\, \mathbf{s}^k, \tag{2a}$$

where

$$\lambda^k = \operatorname*{arg\,min}_{\lambda \in \Omega_{\mathbf{s}^k}(\mathbf{x}^k)} f(\mathbf{x}^k + \lambda\, \mathbf{s}^k), \tag{2b}$$

the set $\Omega_s(\mathbf{x})$ is defined as

$$\Omega_{\mathbf{s}}(\mathbf{x}) = \{\lambda \in \mathbb{R} : h_j(\mathbf{x} + \lambda\, \mathbf{s}) \leqslant 0,\ j = 1, \dots, m\}, \tag{2c}$$

$\mathbf{s}^k \sim \mathrm{Uni}(\mathbb{V})$, if $\mathbf{x}^k \in \mathcal{D}_1$, or $\mathbf{s}^k \sim \mathrm{Uni}(\mathbb{S})$, if $\mathbf{x}^k \in \mathcal{D}_2$, with

$$\begin{aligned} \mathcal{D}_1 \doteq \{&\mathbf{x} \in \mathbb{R}^n : \exists i \in \overline{1, n} \wedge x_i \in \mathbb{R} \text{ s.t.} \\ &f(\mathbf{x}) - f(x_i, \mathbf{x}_{-i}) \geqslant \omega \wedge \mathbf{h}(x_i, \mathbf{x}_{-i}) \leqslant 0\}, \end{aligned} \tag{2d}$$

$$\begin{aligned} \mathcal{D}_2 \doteq \{&\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) - f(x_i, \mathbf{x}_{-i}) \leqslant \omega, \\ &\forall x_i \in \mathbb{R} \text{ s.t. } \mathbf{h}(x_i, \mathbf{x}_{-i}) \leqslant 0,\ i \in \overline{1, n}\}, \end{aligned} \tag{2e}$$

while $\mathbf{x}_{-i} \doteq [\ x_1 \ \cdots \ x_{i-1} \ x_{i+1} \ \cdots \ x_n\ ]^\top$ and $f(x_i, \mathbf{x}_{-i}) \doteq f(x_1 \dots, x_n)$.

Note that, since the set $\mathcal{D}_1$ can be equivalently defined as $\mathcal{D}_1 = \bigcup_{i=1}^n \{\mathbf{x} \in \mathbb{R}^n : \exists x_i \in \mathbb{R} \text{ s.t. } f(\mathbf{x}) - f(x_i, \mathbf{x}_{-i}) \geqslant \omega \wedge \mathbf{h}(x_i, \mathbf{x}_{-i}) \leqslant 0\}$, which is the finite union of closed sets, it is closed [28]. Similarly, since the set $\mathcal{D}_2$ can be equivalently defined as $\mathcal{D}_2 = \bigcap_{i=1}^n \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) - f(x_i, \mathbf{x}_{-i}) \leqslant \omega,\ \forall x_i \in \mathbb{R} \text{ s.t. } \mathbf{h}(x_i, \mathbf{x}_{-i}) \leqslant 0\}$, which is the finite intersection of closed sets, it is closed. Furthermore, by construction $\mathcal{D}_1 \cup \mathcal{D}_2 \supset \Omega$, although they do not constitute a partition of Ω since $\mathcal{D}_1 \cap \mathcal{D}_2 \neq \emptyset$.

Therefore, letting $\mathcal{J}_1 = \mathcal{D}_1 \cap \Omega$ and $\mathcal{J}_2 = \mathcal{D}_2 \cap \Omega$ (which are closed sets since $\mathcal{D}_1$, $\mathcal{D}_2$, and Ω are closed and satisfy $\mathcal{J}_1 \cup \mathcal{J}_2 = \Omega$ by construction) and letting the set-valued mapping $G : \Omega \times \mathbb{S} \rightrightarrows \Omega$ be defined as

$$\begin{aligned} G(\mathbf{x}, \mathbf{s}) \doteq \{&\mathbf{y} \in \Omega : \exists \lambda^\star \in \mathbb{R} \text{ s.t. } \mathbf{y} = \mathbf{x} + \lambda^\star\, \mathbf{s} \\ &\wedge f(\mathbf{y}) \leqslant f(\mathbf{x} + \lambda\, \mathbf{s}),\ \forall \lambda \text{ s.t. } \mathbf{h}(\mathbf{x} + \lambda\, \mathbf{s}) \leqslant 0\}, \end{aligned}$$

the algorithm dynamics given by (2) can be rewritten in the form of the difference inclusion

$$\mathbf{x}^{k+1} \in G(\mathbf{x}^k, \mathbf{s}_1^k), \qquad \mathbf{x}^k \in \mathcal{J}_1, \tag{3a}$$

$$\mathbf{x}^{k+1} \in G(\mathbf{x}^k, \mathbf{s}_2^k), \qquad \mathbf{x}^k \in \mathcal{J}_2, \tag{3b}$$

where $\{\mathbf{s}_1^k\}_{k \in \mathbb{N}}$ and $\{\mathbf{s}_2^k\}_{k \in \mathbb{N}}$ are two sequences of independent, identically distributed (i.i.d.) random variables defined from the probability space $(\Psi, \mathcal{F}, \mathbb{P})$. Namely, the random variables $\mathbf{s}_1^k : \Psi \to \mathbb{V}$ (resp., $\mathbf{s}_2^k : \Psi \to \mathbb{S}$) are such that the probability measure $\mu_1(F) \doteq \mathbb{P}(\psi \in \Psi : \mathbf{s}_1^k(\psi) \in F)$ (resp., $\mu_2(F) \doteq \mathbb{P}(\psi \in \Psi : \mathbf{s}_2^k(\psi) \in F)$) is well defined and independent of k for each F in the Borel σ-field on $\mathbb{V}$ (resp., $\mathbb{S}$). In particular, $\mathbf{s}_1 \sim \mathrm{Uni}(\mathbb{V})$ and $\mathbf{s}_2 \sim \mathrm{Uni}(\mathbb{S})$.

Define the set $\mathcal{A}$ of all the solutions to the PMP (1),

$$\mathcal{A} \doteq \{\mathbf{x} \in \Omega : f(\mathbf{x}) \leqslant f(\mathbf{y}),\ \forall \mathbf{y} \in \Omega\}.$$

The following lemma establishes some regularity properties of the set-valued map $G(\mathbf{x}, \mathbf{s})$.

Lemma 1 ([23])**.** *Assume that $\mathcal{A} \neq \emptyset$ and that either Ω is compact or $f \in \mathrm{ru}(\Omega)$. Then, the set $\mathcal{A}$ is compact, the set-valued mapping $G : \Omega \times \mathbb{S} \rightrightarrows \Omega$ is locally bounded and $\mathbf{s} \mapsto \mathrm{graph}(G(\cdot, \mathbf{s})) \doteq \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^n : \mathbf{y} \in G(\mathbf{x}, \mathbf{s})\}$ is measurable with closed values. Moreover, one has that*

$$G(\Omega, \mathbb{S}) \doteq \bigcup_{\mathbf{x} \in \Omega} \bigcup_{\mathbf{s} \in \mathbb{S}} G(\mathbf{x}, \mathbf{s}) \subset \Omega.$$

Lemma 1 states that the stochastic difference inclusion (3) satisfies the assumptions of [29], [30], [31], thus establishing existence and completeness of maximal random solutions. The set of all such solutions starting at $\mathbf{x}^0 \in \Omega$ is $\mathcal{R}(\mathbf{x}^0)$.

We now recall the definition of *asymptotic stability in probability for* (3). A compact set $\mathcal{A} \subset \mathbb{R}^n$ is *stable in probability for* (3) if for each $\varepsilon \in \mathbb{R}_{>0}$ and $\sigma \in \mathbb{R}_{>0}$ there exists $\delta \in \mathbb{R}_{>0}$ such that

$$\begin{aligned} \mathbf{x}^0 \in (\mathcal{A} + \delta\, \mathbb{B}) \cap \Omega, \{\mathbf{x}^k\}_{k \in \mathbb{N}} \in \mathcal{R}(\mathbf{x}^0) \\ \implies \mathbb{P}(\mathbf{x}^k(\psi) \in \mathcal{A} + \varepsilon\, \mathbb{B}^o,\ k \in \mathbb{N}) \geqslant 1 - \sigma. \end{aligned} \tag{4}$$

On the other hand, the set $\mathcal{A} \subset \mathbb{R}^n$ is *attractive in probability for* (3) if for each $\varepsilon \in \mathbb{R}_{>0}$, $\sigma \in \mathbb{R}_{>0}$ and $\Delta \in \mathbb{R}_{>0}$, there exists $K \in \mathbb{R}_{>0}$ such that

$$\begin{aligned} \mathbf{x}^0 \in (\mathcal{A} + \Delta\, \mathbb{B}) \cap \Omega, \{\mathbf{x}^k\}_{k \in \mathbb{N}} \in \mathcal{R}(\mathbf{x}^0) \\ \implies \mathbb{P}(\mathbf{x}^k(\psi) \in \mathcal{A} + \varepsilon\, \mathbb{B}^o,\ k \geqslant K) \geqslant 1 - \sigma. \end{aligned} \tag{5}$$

The set $\mathcal{A} \subset \mathbb{R}^n$ is *asymptotically stable in probability for* (3) if it is both attractive and stable in probability for (3) [29].

Establishing that the set $\mathcal{A}$ is asymptotically stable in probability for an algorithm attempting at solving an optimization problem is a particularly desirable property. In fact, the condition given in (4) implies that if one already possess an accurate solution $\mathbf{x}^0$ to the optimization problem, then its (pseudo)optimality is not lost by applying the optimization algorithm with $\mathbf{x}^0$ as initial guess. On the other hand, the condition given in (5) implies that, although the initial guess of the optimization algorithm is far from optimality, its state eventually converge to an arbitrarily good approximate of the solution. In the following theorem, we prove that indeed the set $\mathcal{A}$ is asymptotically stable in probability for (3).

Theorem 1. *Let the assumptions of Lemma 1 hold and assume, additionally, that there exists $\nu^\star \in \mathbb{R}_{>0}$ such that the set $(\mathcal{A}+\nu\mathbb{B})\cap\Omega$ has nonzero measure for all $\nu \in (0, \nu^\star)$. Then, the set $\mathcal{A}$ is asymptotically stable for* (3)*.*

Proof. Let $\mathbf{x}^\star$ be any point in $\mathcal{A}$ and let $f^\star = f(\mathbf{x}^\star)$. Since the set $\mathcal{A}$ is nonempty and the polynomial f is continuous $f^\star \in \mathbb{R}$. Thus, let $V : \Omega \to \mathbb{R}_{\geqslant 0}$ be any smooth function such that $V(\mathbf{x}) = f(\mathbf{x}) - f^\star$ for all $\mathbf{x} \in \Omega$ and $\lim_{\|\mathbf{x}\|\to\infty} V(\mathbf{x}) = +\infty$. Such a function exists since either Ω is compact or $f \in \mathrm{ru}(\Omega)$. Since the set $\mathcal{A}$ is compact, $V(\mathbf{x}) = 0$ if and only if $\mathbf{x} \in \mathcal{A}$, and V is radially unbounded, by [32, p. 54], there exist class $\mathcal{K}_\infty$ functions $\underline{\alpha}$ and $\overline{\alpha}$ such that

$$\underline{\alpha}(\|\mathbf{x}\|_{\mathcal{A}}) \leqslant V(\mathbf{x}) \leqslant \overline{\alpha}(\|\mathbf{x}\|_{\mathcal{A}}).$$

Furthermore, by considering that $G(\Omega, \mathbb{S}) \subset \Omega$ and that, by definition of the set-valued mapping $G : \Omega \times \mathbb{S} \rightrightarrows \Omega$, $f(\mathbf{g}) \leqslant f(\mathbf{x})$ for all $g \in G(\mathbf{x}, \mathbf{s})$ and all $\mathbf{s} \in \mathbb{S}$, one has that

$$\sup_{\mathbf{g}\in G(\mathbf{x},\mathbf{s})} V(\mathbf{g}) \leqslant V(\mathbf{x}), \quad \forall (\mathbf{x}, \mathbf{s}) \in \Omega \times \mathbb{S}.$$

Additionally, in view of the definition of the set $\mathcal{D}_1$ given in (2d), one has that, for all $\mathbf{x} \in \mathcal{J}_1$,

$$\frac{1}{n}\sum_{i=1}^{n} \sup_{\mathbf{g}\in G(\mathbf{x},\mathbf{v}_i)} V(\mathbf{g}) \leqslant V(\mathbf{x}) - \frac{\omega}{n}.$$

On the other hand, given $\mathbf{x} \in \mathcal{J}_2 \setminus \mathcal{A}$, let $\upsilon^* = \frac{f(\mathbf{x})-f^\star}{2} \in \mathbb{R}_{>0}$ and let $\upsilon^\diamond = \min\{\upsilon^*, \upsilon^\star\}$. Since $(\mathcal{A} + \nu\mathbb{B}) \cap \Omega$ has nonzero measure for all $\nu \in (0, \nu^\star)$, the set

$$\mathcal{L}_{\upsilon^\diamond} \doteq \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) - f^\star \leqslant \upsilon^\diamond\} \cap \Omega$$

has nonzero measure. Therefore, for each $\mathbf{x} \in \mathcal{J}_2 \setminus \mathcal{A}$, there exists a measurable selection $\mathcal{S}_{\mathbf{x}}$ of $\mathbb{S}$ such that $\sup_{\mathbf{g}\in G(\mathbf{x},\mathbf{s})} V(\mathbf{g}) \leqslant \frac{V(\mathbf{x})}{2}$ for all $\mathbf{s} \in \mathcal{S}_{\mathbf{x}}$. Hence, for all $\mathbf{x} \in \mathcal{J}_2$, there is $\bar{\varrho} \in \mathrm{pd}(\mathcal{A})$ such that

$$\begin{aligned}
\int_{\mathbb{S}} \sup_{\mathbf{g}\in G(\mathbf{x},\mathbf{s})} V(\mathbf{g})\, \mu_1(\mathrm{d}\mathbf{s}_1) &= \Gamma \int_{\mathbb{S}} \sup_{\mathbf{g}\in G(\mathbf{x},\mathbf{s})} V(\mathbf{g})\, \mathrm{d}\mathbf{s}_1 \\
&\leqslant \Gamma \left(\frac{V(\mathbf{x})}{2} \int_{\mathcal{S}} 1\, \mathrm{d}\mathbf{s}_1 + V(\mathbf{x}) \int_{\mathbb{S}\setminus\mathcal{S}} 1\, \mathrm{d}\mathbf{s}_1 \right) \\
&\leqslant V(\mathbf{x}) - \frac{1}{2}\Gamma\, V(\mathbf{x}) \int_{\mathcal{S}_{\mathbf{x}}} 1\, \mathrm{d}\mathbf{s}_1 \\
&\leqslant V(\mathbf{x}) - \bar{\varrho}(\mathbf{x}), \qquad (6)
\end{aligned}$$

where $\Gamma = \frac{1}{2\pi^{n/2}} \int_0^\infty x^{\frac{n}{2}-1} e^{-x} dx$. Therefore, letting

$$\varrho(\mathbf{x}) \doteq \min\left\{\frac{\omega}{n}, \bar{\varrho}(\mathbf{x})\right\},$$

where $\bar{\varrho}(\mathbf{x})$ is any function in $\mathrm{pd}(\mathcal{A})$ such that $\bar{\varrho}(\mathbf{x}) \leqslant \frac{1}{2}\Gamma\, V(\mathbf{x}) \int_{\mathcal{S}_{\mathbf{x}}} 1\, \mathrm{d}\mathbf{s}$, that is a function in $\mathrm{pd}(\mathcal{A})$, it results that

$$\begin{aligned}
\int_{\mathbb{S}} \sup_{\mathbf{g}\in G(\mathbf{x},\mathbf{s})} V(\mathbf{g})\, \mu_1(\mathrm{d}\mathbf{s}_1) &\leqslant V(\mathbf{x}) - \varrho(\mathbf{x}), \quad \mathbf{x} \in \mathcal{J}_1, \\
\int_{\mathbb{S}} \sup_{\mathbf{g}\in G(\mathbf{x},\mathbf{s})} V(\mathbf{g})\, \mu_2(\mathrm{d}\mathbf{s}_2) &\leqslant V(\mathbf{x}) - \varrho(\mathbf{x}), \quad \mathbf{x} \in \mathcal{J}_2.
\end{aligned}$$

Therefore, the assumptions of Corollary 1 of [33] are satisfied by the discrete-time dynamics of the stochastic difference inclusion (3) and hence the set $\mathcal{A}$ is asymptotically stable in probability for (3). □

Note that the hypotheses of Theorem 1 are met if Ω is a regular closed set (i.e., the closure of its interior coincides with the set itself). In the remainder of this section, some further insights on the random solutions of (3) are given.

In the following corollary, we prove that solutions to (3) eventually jump outside the set $\mathcal{J}_1$ with probability 1.

Corollary 1. *Let the assumptions of Theorem 1 hold. Then, for each $\Delta \in \mathbb{R}_{>0}$ and $\sigma \in \mathbb{R}_{>0}$, there is $K \in \mathbb{N}$ such that*

$$\begin{aligned}
&\mathbf{x}^0 \in (\mathcal{A} + \Delta\,\mathbb{B}) \cap \mathcal{J}_1, \{\mathbf{x}^k\}_{k\in\mathbb{N}} \in \mathcal{R}(\mathbf{x}^0) \\
&\quad \implies \mathbb{P}(\exists k \leqslant K \text{ such that } \mathbf{x}^k(\psi) \notin \mathcal{J}_1) \geqslant 1 - \sigma. \qquad (7)
\end{aligned}$$

Proof. Note that, by definition of the set $\mathcal{D}_1$ in (2d), one has that $\mathbf{x} \in \mathcal{J}_1$ if and only if there exists $i \in \overline{1, n}$ and $\mathbf{g} \in G(\mathbf{x}, \mathbf{v}_i)$ such that $f(\mathbf{x}) - f(\mathbf{g}) \geqslant \omega$. Since, by construction, $G(\mathcal{A}, \mathbb{S}) \subset \mathcal{A}$, this implies that there exists $\varepsilon \in \mathbb{R}_{>0}$ such that $(\mathcal{A} + \varepsilon\,\mathbb{B}^o) \cap \mathcal{J}_1 = \emptyset$ (and hence, since $\mathcal{J}_1 \cup \mathcal{J}_2 = \Omega$ and $\mathcal{A} \subset \Omega$, one has that $\mathcal{A}$ is a subset of $\mathcal{J}_2$). This implies that $\mathbb{P}(\exists k \leqslant K$ such that $\mathbf{x}^k(\psi) \notin \mathcal{J}_1) \geqslant \mathbb{P}(\mathbf{x}^k(\psi) \in \mathcal{A} + \varepsilon\,\mathbb{B}^o,\, k \geqslant K)$. Thus, the proof is concluded by the fact that, by Theorem 1, the set $\mathcal{A}$ is asymptotically stable in probability for (3) and hence (5) holds. □

Note that, although Corollary 1 establishes that the solutions of (3) eventually jump out of the set $\mathcal{J}_1$ almost surely, this does not implies that they do not visit the set $\mathcal{J}_1$ again. Indeed, there may exist $\mathbf{x} \in \mathcal{J}_2$ and $\mathbf{s}_2 \in \mathbb{S}$ such that $\mathbf{g} \in \mathcal{J}_1$, for some $\mathbf{g} \in G(\mathbf{x}, s_2)$. This is one of the key facts that ensures asymptotic stability in probability of the set $\mathcal{A}$ for (3). However, as it has been established in Corollary 1, $\mathbf{x} \in \mathcal{A} + \varepsilon\,\mathbb{B}^o$ implies that $\mathbf{x} \in \mathcal{J}_2$ and hence random solutions to (3) are in $\mathcal{J}_2$ when they are close to $\mathcal{A}$.

Finally, it is worth noticing that, if the set $\mathcal{A}$ is known, then results wholly similar to [23, Thm. 2] can be used to find bounds on the number of iterations that have to be carried out to determine an η-accurate solution to the PMP (1). An alternative approach to compute such bounds is to employ the Lyapunov function V given in the proof of Theorem 1. However, since such a function is essentially the objective function f of the PMP (1), such bounds are largely dependent on the optimization problem to be solved.

IV. Implementation notes

In Section III, it has been proved that the set of all the solutions to the PMP (1) is asymptotically stable in probability for the minimization method given in (2). The main objective of this section is to show how such a method can be actually implemented to solve the PMP (1).

By Theorem 1, if the procedure given in (2) is run indefinitely starting at $\mathbf{x}^0 \in \Omega$, then the sequence $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$ converges to $\mathcal{A}$ in probability. In numerical practice, however, such iterations have to be interrupted. In view of the proof of Corollary 1, if $\mathbf{x}$ is close to $\mathcal{A}$, then $\mathbf{x}$ is also in $\mathcal{J}_2$. Hence, in order to interrupt the iterations of the proposed procedure when a sufficiently good estimate of the solution to the PMP (1) has been obtained, we require that,

$$\mathbf{x}^k \in \mathcal{J}_2 \wedge \frac{|f(\mathbf{x}^k) - f(\mathbf{x}^{k-1})|}{|f(\mathbf{x}^k)| + 1} \leqslant \epsilon, \tag{8}$$

for L consecutive steps, where ϵ is a numerical tolerance. In fact, notice that, when $f(\mathbf{x}^k) \gg 1$, one has that

$$\frac{|f(\mathbf{x}^k) - f(\mathbf{x}^{k-1})|}{|f(\mathbf{x}^k)| + 1} \simeq \frac{|f(\mathbf{x}^k) - f(\mathbf{x}^{k-1})|}{\min\{|f(\mathbf{x}^k)|, |f(\mathbf{x}^{k-1})|\}},$$

that is the relative improvement obtained by using an iteration of the proposed method. On the other hand, when $f(\mathbf{x}^k) \ll 1$, one has that

$$\frac{|f(\mathbf{x}^k) - f(\mathbf{x}^{k-1})|}{|f(\mathbf{x}^k)| + 1} \simeq |f(\mathbf{x}^k) - f(\mathbf{x}^{k-1})|,$$

that is the absolute improvement obtained by using an iteration of the proposed method. Therefore, letting K be the step at which the iterations of (2) are interrupted, requiring that (8) holds for $k = K - L + 1, \ldots, K$, essentially corresponds to require that the obtained estimate is in the set of the block-coordinate (pseudo)minima and that the improvement obtained using transverse directions is lower than the given threshold for at least L steps.

Thus, consider Algorithm 1 that implements (2).

It can be easily noticed that the sequence $\{\mathbf{x}^k\}_{k=0}^K$ in Algorithm 1 is a random sample of a solution of the stochastic difference inclusion (3). In order to reduce its computational complexity, such an algorithm has been designed in order to avoid checking at each step whether $\mathbf{x}^k$ is in $\mathcal{J}_1$ or in $\mathcal{J}_2$. In fact, rather than checking if either $\mathbf{x} \in \mathcal{J}_1$ or $\mathbf{x} \in \mathcal{J}_2$, the algorithm iteratively removes from the set $\mathcal{V}$ the directions that are such that no improvement is obtained by updating the solution along such a direction. Hence, when eventually $\mathcal{V}$ is empty, a transverse direction is taken since $\mathbf{x} \in \mathcal{J}_2$.

In order to carry out Algorithm 1, an initial feasible guess $\mathbf{x}^0 \in \Omega$ has to be known. If this is not the case, however, it is possible to find $\mathbf{x}^0 \in \Omega$ by solving the following PMP with an additional variable η,

$$\left| \begin{array}{ll} \min & \eta, \\ \text{with} & h_i(\mathbf{x}) - \eta \leqslant 0, \quad i = 1, \ldots, m. \end{array} \right. \tag{11}$$

Note that $\mathbf{x}^0 = \mathbf{0}$ and $\eta^0 = \max\{h_1(\mathbf{0}), \ldots, h_m(\mathbf{0})\}$ is always in the feasible set of the PMP (11), whence Algorithm 1 can be applied to solve such a problem.

Algorithm 1

Input: the PMP (1), $\omega \in \mathbb{R}_{>0}$, $\mathbf{x}^0 \in \Omega$, a numerical tolerance $\epsilon \in \mathbb{R}_{>0}$, and a positive integer $L \in \mathbb{N}$

Output: estimates $\hat{\mathbf{x}}^\star$ and $\hat{f}^\star$ of $\mathbf{x}^\star \in \mathcal{A}$ and $f^\star$,

$k \leftarrow 0$, $\varkappa \leftarrow 0$, $\mathcal{V} \leftarrow \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$
while $\varkappa \leqslant L$ **do**
  pick $\mathbf{v}_i$ in $\mathcal{V}$ uniformly at random
  define $\breve{f}(\lambda) \doteq f(\mathbf{x} + \lambda\,\mathbf{v}_i)$ and $\breve{\mathbf{h}}(\lambda) \doteq \mathbf{h}(\mathbf{x} + \lambda\,\mathbf{v}_i)$
  let $\lambda^\star$ be a solution to the univariate PMP

$$\left| \begin{array}{ll} \min & \breve{f}(\lambda), \\ \text{with} & \breve{h}_j(\lambda) \leqslant 0, \quad j = 1, \ldots, m \end{array} \right. \tag{9}$$

  $\mathbf{y} \leftarrow \mathbf{x}^k + \lambda^\star\,\mathbf{v}_i$
  if $f(\mathbf{x}^k) - f(\mathbf{y}) \geqslant \omega$ **then**
    $k \leftarrow k + 1$, $\mathbf{x}^k \leftarrow \mathbf{y}$, $\mathcal{V} \leftarrow \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, $\varkappa \leftarrow 0$
  else
    $\mathcal{V} \leftarrow \mathcal{V} \setminus \{\mathbf{v}_i\}$
    if $\mathcal{V} = \emptyset$ **then**
      pick $\mathbf{s}$ in $\mathbb{S}$ uniformly at random
      define $\breve{f}(\lambda) \doteq f(\mathbf{x} + \lambda\,\mathbf{s})$ and $\breve{\mathbf{h}}(\lambda) \doteq \mathbf{h}(\mathbf{x} + \lambda\,\mathbf{s})$
      let $\lambda^\star$ be a solution to the univariate PMP

$$\left| \begin{array}{ll} \min & \breve{f}(\lambda), \\ \text{with} & \breve{h}_j(\lambda) \leqslant 0, \quad j = 1, \ldots, m \end{array} \right. \tag{10}$$

      $k \leftarrow k+1$, $\mathbf{x}^k \leftarrow \mathbf{x}^{k-1} + \lambda^\star\,\mathbf{s}$, $\mathcal{V} \leftarrow \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$
      if $\frac{|f(\mathbf{x}^k) - f(\mathbf{x}^{k-1})|}{|f(\mathbf{x}^k)| + 1} \leqslant \epsilon$ **then**
        $\varkappa \leftarrow \varkappa + 1$
      else
        $\varkappa \leftarrow 0$
return $\hat{\mathbf{x}}^\star = \mathbf{x}^k$ and $\hat{f}^\star = f(\mathbf{x}^k)$

At each step of Algorithm 1, one has to solve an univariate PMP (namely, (9) and (10)). Such problems can be easily solved using the tools given in Section IV.A of [22], i.e., by computing the set of all the real roots of the (univariate) polynomials $\frac{df(\lambda)}{d\lambda}$ and $h_j(\lambda)$, $j = 1, \ldots, m$ [34], [35] and comparing the values attained by $f(\lambda)$ at such points. Note that the PMPs (9) and (10) can be solved even exactly using either symbolic computation or SOS-based tools [15].

V. Numerical example

In this section, we show the effectiveness of the proposed minimization method by using it to solve a PMP that cannot be easily solved by means of other methods. In particular, we show how Algorithm 1 allows us to compute the distance between a polynomial surface and the origin.

Computing the distance between the point $\mathbf{y}$ and the surface $\mathcal{V} \doteq \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) = 0\}$ corresponds to solving

$$\left| \begin{array}{ll} \min & \|\mathbf{x} - \mathbf{y}\|_2^2, \\ \text{with} & h(\mathbf{x}) = 0. \end{array} \right. \tag{12}$$

The following proposition establishes that a solution to the PMP (12) can be determined by solving the following PMP

$$\left| \begin{array}{ll} \min & \|\mathbf{x} - \mathbf{y}\|_2^2, \\ \text{with} & h(\mathbf{x}) \leqslant 0. \end{array} \right. \tag{13}$$

Proposition 1. *Let* $(\mathbf{x}^\star, f^\star) \in \mathbb{R}^n \times \mathbb{R}$ *be a solution to the PMP* (13)*. Then, it is also a solution to the PMP* (12)*.*

Proof. The statement is proved by showing that if $(\mathbf{x}^\star, f^\star)$ is a solution to the PMP (13), then $h(\mathbf{x}^\star) = 0$. Assume, by contradiction, that $h(\mathbf{x}^\star) < 0$. Since $\mathcal{Z} := \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) < 0\}$ is open and $\mathbf{x}^\star \in \mathcal{Z}$, there exists an open neighborhood $\mathcal{Q}$ of $\mathbf{x}^\star$ such that $h(\mathbf{x}) < 0$, $\forall \mathbf{x} \in \mathcal{Q}$. Therefore, there is $\lambda \in (0,1)$ such that, letting $\hat{\mathbf{x}} = \lambda \mathbf{x}^\star + (1-\lambda)\mathbf{y}$, $h(\hat{\mathbf{x}}) \leqslant 0$. Hence, the ℓ_2 distance between $\hat{\mathbf{x}}$ and $\mathbf{y}$ is

$$\|\hat{\mathbf{x}} - \mathbf{y}\|_2^2 = \lambda \|\mathbf{x}^\star - \mathbf{y}\|_2^2 = \lambda f^\star < f^\star,$$

i.e., $(\mathbf{x}^\star, f^\star)$ is not a solution to the PMP (13), thus leading to a contradiction and concluding the proof. □

In view of Proposition 1, let $n = 3$ and consider the following PMP (taken from Section VII.B of [13] and deeply analyzed in [36], where it is shown that a tight lower bound may be difficult to obtain)

$$\left| \begin{array}{ll} \min & \|\mathbf{x}\|_2^2, \\ \text{with} & h(\mathbf{x}) \leqslant 0, \end{array} \right. \tag{14}$$

where $\|\mathbf{x}\|_2^2 = x_1^2 + x_2^2 + x_3^2$ and

$$h(\mathbf{x}) = 10\left(x_3^6 - 3x_1^2x_2^2x_3^2 + x_1^2x_2^4 + x_1^4x_2^2\right) - \|\mathbf{x}\|_2^6 + 1.$$

Note that $\|\mathbf{x}\|_2^2 \in \text{ru}(\mathbb{R}^3)$ and that, for each point $\mathbf{x}$ in the feasible set of the PMP (14) and each $\varepsilon \in \mathbb{R}_{>0}$, the set $(\mathbf{x} + \varepsilon\,\mathbb{B}^o) \cap \{\mathbf{x} \in \mathbb{R}^3 : h(\mathbf{x}) \leqslant 0\}$ has nonzero measure. Furthermore, by using the tools given in [22], it can be derived that the set $\mathcal{A}$ of the optimal solutions to the PMP (14) is nonempty. Therefore, the hypotheses of Theorem 1 are met and hence $\mathcal{A}$ is asymptotically stable in probability for (3), i.e., the proposed method can be used to compute a solution of the PMP (14).

Despite the PMP (14) seems to be simple (it involves just three variables, the objective function is quadratic and convex, and there is one constraint), it cannot be easily solved by using methods taken from the literature. In fact[1]:

- the function `findbound` of the `Matlab` package `SOSTOOLS` [9] returns $\gamma = 0.9820$ as a lower bound for $f^\star$ when invoked with maximum degree equal to 18 (execution time = 162.008 s);
- by choosing the almost right set of monomials, i.e., the ones of degree 2 4, and 12 in x_1, x_2, x_3, the `Matlab` package `SOSTOOLS` returns $\gamma = 0.98395$ as a lower bound for $f^\star$ (execution time = 28.593 s);
- the function `msol` of the `Matlab` toolbox `GloptiPoly` [37] returns 0.9898 as a lower bound for $f^\star$ when invoked with order 10 of the SDP relaxation (execution time = 44.942 s);
- the `Matlab` function `fmincon` [38] is not able to determine a feasible point and returns an error;
- the function `optimize` of the `Matlab` toolbox `Yalmip` [39], interfaced with the nonlinear solver `PENLAB` [40], returns an error;

[1]All the computations reported hereafter have been carried out on a laptop with an Intel i5 CPU (2.4 GHz) and 8 GB, 1600 MHz, DDR3 RAM.

This is essentially due to the fact that the feasible set Ω of problem (14) is unbounded. In fact, the performance of the above toolboxes can be improved by adding bounds on the optimization variables. For instance, by adding an additional unit Euclidean ball constraint, `GloptiPoly` [37] returns 1 as a lower bound for $f^\star$ when invoked with order 4 of the SDP relaxation (execution time = 2.4111 s);

Hence, we applied the proposed technique to the PMP (14) in order to verify its effectiveness. In particular, we used Algorithm 1 with $L = 100$, $\omega = 10^{-3}$, $\varepsilon = 10^{-3}$, and with 1000 different initial conditions $\mathbf{x}^0$, which have been determined by solving the PMP (11). Such a latter problem has been solved through Algorithm 1 with the same parameters as above and with $\mathbf{x}^0 = \mathbf{0}$ and $\eta^0 = h(\mathbf{0})$. The average execution time required to determine both the initial guess $\mathbf{x}^0$ and the solution $\mathbf{x}^\star$ to the PMP (14) via Algorithm 1 is 3.3807 s, which is smaller that the computational times required to solve the problem by using other packages.

Figure 1 depicts 20 outcomes of such an algorithm.

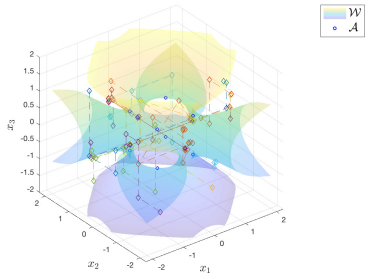

Fig. 1: Sets $\mathcal{W}$, $\mathcal{A}$, and outcomes of Algorithm 1.

As shown by this figure, Algorithm 1 converges toward a solution to the problem by firstly minimizing along the coordinate directions and just eventually taking transverse jumps. Note that different outcomes of Algorithm 1 need not converge to the same point in $\mathcal{A}$, which, in the considered case, is constituted by 12 different optimal points. Indeed, even in the case that $\mathbf{x}^{k_1}$ is already close to a point in $\mathcal{A}$, there may exist $k_2 > k_1$ such that $\mathbf{x}^{k_2}$ is close to another point in $\mathcal{A}$ (see the outcome depicted in orange). Indeed, as shown by Figure 1, if the set $\mathcal{A}$ is not a singleton, then the proposed algorithm does not guarantee point-wise asymptotic stability of $\mathcal{A}$, i.e., every point in $\mathcal{A}$ need not be Lyapunov stable although the overall set $\mathcal{A}$ is asymptotically stable.

To further corroborate our results, Figure 2 depicts $\{f(\mathbf{x}^k)\}_k$, where each $\{\mathbf{x}^k\}_k$ is one of the 1000 sequences computed by Algorithm 1.

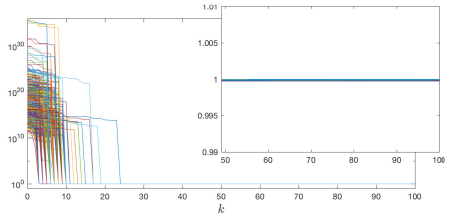

Fig. 2: Values attained by the polynomial $f(\mathbf{x})$ along the sequences computed by Algorithm 1.

As shown by such a figure, in all the 1000 simulations the state hits the set $\{\mathbf{x} \in \Omega : f(\mathbf{x}) \leqslant f^\star + 0.005\}$ in less than 100 iterations, despite the application of Algorithm 1 for the computation of the initial guess $\mathbf{x}^0$ may lead to quite large initial values of the objective function f.

VI. Conclusions

In this paper, it has been shown that several polynomial optimization problems can be solved by performing a random coordinate-wise minimization and, eventually, when a coordinate-wise minimum has been reached, by performing a minimization along a random transverse direction.

In particular, it has been shown that, under mild assumption, such a method converges to the set of the globally optimal solutions to the polynomial minimization problem with probability arbitrarily close to 1, provided that a sufficient number of iterations are carried out. Thus, differently from classical nonlinear optimization methods, such as gradient-based algorithms [25], the given method asymptotically tends to the globally optimal solution.

Robustness of the convergence of the considered algorithm with respect to small perturbations follows by [41]. In fact, the Lyapunov arguments used in the proof of Theorem 1 establish also practical and semi-global robustness of the proposed minimization algorithm with respect to small errors, such as those induced by numerical computations.

Finally, note that, as shown in Section V, the proposed method seems to be very computationally efficient while requiring minimal tuning of its parameters to guarantee convergence.

References

[1] S. P. Bhattacharyya and L. H. Keel, "Robust control: the parametric approach," in *Advances in Control Education*, pp. 49–52, 1995.

[2] R. Genesio, M. Tartaglia, and A. Vicino, "On the estimation of asymptotic stability regions: State of the art and new proposals," *IEEE Trans. Autom. Control*, vol. 30, no. 8, pp. 747–755, 1985.

[3] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *Int. J. Comput. vision*, vol. 27, no. 2, pp. 161–195, 1998.

[4] P. Dorato, "Quantified multivariate polynomial inequalities. the mathematics of practical control design problems," *IEEE Control Syst. Mag.*, vol. 20, no. 5, pp. 48–58, 2000.

[5] J. Garloff, "Application of Bernstein expansion to the solution of control problems," *Reliable Comp.*, vol. 6, no. 3, pp. 303–320, 2000.

[6] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge Univ. Press, 2004.

[7] D. P. Bertsekas, *Nonlinear programming*. Belmont: Athena Sci., 1999.

[8] P. A. Parrilo and B. Sturmfels, "Minimizing polynomial functions," in *Algorithmic and Quantitative Real Algebraic Geometry: DIMACS Series in Discrete Mathematics and Theoretical Computer Science* (S. Basu and L. González-Vega, eds.), pp. 83–99, Am. Math. Soc., 2003.

[9] S. Prajna, A. Papachristodoulou, P. Seiler, and P. A. Parrilo, "SOSTOOLS and its control applications," in *Positive Polynomials in Control* (D. Henrion and A. Garulli, eds.), pp. 273–292, Heidelberg: Springer, 2005.

[10] J. B. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM J. Optim.*, vol. 11, no. 3, pp. 796–817, 2001.

[11] J. B. Lasserre, "An explicit equivalent positive semidefinite program for nonlinear 0-1 programs," *SIAM J. Optim.*, vol. 12, no. 3, pp. 756–769, 2002.

[12] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory*, vol. 15. Siam, 1994.

[13] G. Chesi, A. Tesi, A. Vicino, and R. Genesio, "An LMI approach to constrained optimization with homogeneous forms," *Syst. Contr. Lett*, vol. 42, no. 1, pp. 11–19, 2001.

[14] D. Henrion and J.-B. Lasserre, "Solving nonconvex optimization problems," *IEEE Control Syst. Mag.*, vol. 24, no. 3, pp. 72–83, 2004.

[15] P. A. Parrilo, *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*. PhD thesis, California Institute of Technology, 2000.

[16] G. Blekherman, "There are significantly more nonegative polynomials than sums of squares," *Israel J. Math.*, vol. 153, pp. 355–380, 2006.

[17] D. Henrion and J.-B. Lasserre, "Detecting global optimality and extracting solutions in Gloptipoly," in *Positive polynomials in control* (D. Henrion and A. Garulli, eds.), pp. 293–310, Springer, 2005.

[18] C. Josz and D. Henrion, "Strong duality in Lasserre's hierarchy for polynomial optimization," *Optim. Lett.*, vol. 10, no. 1, pp. 3–10, 2016.

[19] J. B. Lasserre and T. Netzer, "SOS approximations of nonnegative polynomials via simple high degree perturbations," *Math. Z.*, vol. 256, no. 1, pp. 99–112, 2007.

[20] L. Menini and A. Tornambe, "Exact sum of squares decomposition of univariate polynomials," in *54th IEEE Conf. Decis. Control*, pp. 1072–1077, 2015.

[21] L. Menini, C. Possieri, and A. Tornambe, "Algebraic certificates of (semi)definiteness for polynomials over fields containing the rationals," *IEEE Tran. Autom. Control*, vol. 63, no. 1, pp. 158–173, 2018.

[22] L. Menini, C. Possieri, and A. Tornambe, "Algebraic methods for multi-objective optimal design of control feedbacks for linear systems," *IEEE Trans. Autom. Control*, 2019.

[23] G. C. Calafiore and C. Possieri, "A variation on a random coordinate minimization method for constrained polynomial optimization," *IEEE Control Syst. Lett.*, vol. 2, no. 3, pp. 531–536, 2018.

[24] G. C. Calafiore and L. El Ghaoui, *Optimization models*. Cambridge university press, 2014.

[25] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Math. Prog.*, vol. 106, no. 1, pp. 25–57, 2006.

[26] D. A. Cox, J. Little, and D. O'Shea, *Ideals, varieties, and algorithms*. New York: Springer, 2015.

[27] E. De Klerk, "The complexity of optimizing over a simplex, hypercube or sphere: a short survey," *Cent. Eur. J. Oper. Res.*, vol. 16, no. 2, pp. 111–125, 2008.

[28] N. Bourbaki, *General topology*. Springer, 1998.

[29] A. R. Teel, "A Matrosov theorem for adversarial Markov decision processes," *IEEE Trans. Autom. Control*, vol. 58, no. 8, pp. 2142–2148, 2013.

[30] A. R. Teel, J. P. Hespanha, and A. Subbaraman, "A converse Lyapunov theorem and robustness for asymptotic stability in probability," *IEEE Trans. Autom. Control*, vol. 59, no. 9, pp. 2426–2441, 2014.

[31] C. Possieri and A. R. Teel, "A Lyapunov theorem certifying global weak reachability for stochastic difference inclusions with random inputs," *Syst. Control Lett.*, vol. 109, pp. 37–42, 2017.

[32] R. Goebel, R. G. Sanfelice, and A. R. Teel, *Hybrid Dynamical Systems*. Princeton Univ. Press, 2012.

[33] A. R. Teel, "Stochastic hybrid inclusions with diffusive flows," in *53rd IEEE Conf. Decis. Control*, pp. 3071–3076, IEEE, 2014.

[34] G. E. Collins and A. G. Akritas, "Polynomial real root isolation using Descarte's rule of signs," in *Proc. Third ACM Symp. Symb. Algebraic Computation*, pp. 272–275, ACM, 1976.

[35] B. Sturmfels, *Solving systems of polynomial equations*. Am. Math. Soc., 2002.

[36] G. Chesi, A. Garulli, A. Tesi, and A. Vicino, "Solving quadratic distance problems: an LMI-based approach," *IEEE Trans. Autom. Control*, vol. 48, no. 2, pp. 200–212, 2003.

[37] D. Henrion and J.-B. Lasserre, "Gloptipoly: Global optimization over polynomials with Matlab and SeDuMi," *ACM Trans. Math. Software*, vol. 29, no. 2, pp. 165–194, 2003.

[38] R. H. Byrd, J. C. Gilbert, and J. Nocedal, "A trust region method based on interior point techniques for nonlinear programming," *Math. Prog.*, vol. 89, no. 1, pp. 149–185, 2000.

[39] J. Löfberg, "Pre-and post-processing sum-of-squares programs in practice," *IEEE Trans. Autom. Control*, vol. 54, no. 5, pp. 1007–1011, 2009.

[40] J. Fiala, M. Kočvara, and M. Stingl, "PENLAB: A MATLAB solver for nonlinear semidefinite optimization." arXiv:1311.5240, 2013.

[41] S. Grammatico, A. Subbaraman, and A. R. Teel, "Discrete-time stochastic control systems: A continuous Lyapunov function implies robustness to strictly causal perturbations," *Automatica*, vol. 49, no. 10, pp. 2939–2952, 2013.