POLITECNICO DI TORINO
Repository ISTITUZIONALE

A Variation on a Random Coordinate Minimization Method for Constrained Polynomial Optimization

Original
A Variation on a Random Coordinate Minimization Method for Constrained Polynomial Optimization / Calafiore, GIUSEPPE CARLO; Possieri, Corrado. - ELETTRONICO. - (2018). (Intervento presentato al convegno IEEE Conference on Decision and Control tenutosi a Miami, FL nel Dec. 17-19, 2018).

Availability:
This version is available at: 11583/2715553 since: 2018-10-23T10:51:39Z

Publisher:
IEEE

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

20 May 2024

A Variation on a Random Coordinate Minimization Method for Constrained Polynomial Optimization

Giuseppe C. Calafiore, Corrado Possieri

***Abstract*— In this paper, an algorithm is proposed for solving constrained and unconstrained polynomial minimization problems. The algorithm is a variation on random coordinate descent, in which transverse steps are seldom taken. Differently from other methods available in the literature, the proposed technique is guaranteed to converge in probability to the global solution of the minimization problem, even when the objective polynomial is nonconvex. The theoretical results are corroborated by a complexity analysis and by numerical tests that validate its efficiency.**

I. Problem statement and introduction

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. $\mathbb{B}$, $\mathbb{B}^o$ and $\mathbb{S}$ denote the closed and open unit balls and the unit sphere in the Euclidean norm, 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 $\mathcal{A}$, denoted $\varrho \in \mathcal{PD}(\mathcal{A})$, if $\varrho(\mathbf{x}) = 0 \iff \mathbf{x} \in \mathcal{A}$. Given a set $\mathcal{T} \subset \mathbb{R}^n$, let $\mathbb{I}_{\mathcal{T}}(\cdot)$ be the *indicator function of* $\mathcal{T}$, i.e., $\mathbb{I}_{\mathcal{T}}(\mathbf{x}) = 1$, if $\mathbf{x} \in \mathcal{T}$, or $\mathbb{I}_{\mathcal{T}}(\mathbf{x}) = 0$, if $\mathbf{x} \notin \mathcal{T}$. 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 $\lim_{r\to+\infty} \alpha(r) = +\infty$.

Let $\mathbf{x} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top$, with $n \in \mathbb{N}$, be a vector of variables. A *monomial* in $\mathbf{x}$ is a product of the form $\mathbf{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, where $\alpha_i \in \mathbb{N}$, $i = 1, \ldots, n$. Given $\boldsymbol{\alpha}$, let $|\boldsymbol{\alpha}| \doteq \sum_{i=1}^n \alpha_i$. A *polynomial* p in x is a finite, $\mathbb{R}$-linear combination of monomials, $p = \sum_{\boldsymbol{\alpha}\in\mathcal{E}} c_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}}$, where $\mathcal{E} \subset \mathbb{N}^n$ is a finite set and $c_{\boldsymbol{\alpha}} \in \mathbb{R}$, $\forall \boldsymbol{\alpha} \in \mathcal{E}$; the *total degree of* p is $\max\{|\boldsymbol{\alpha}|,\ \boldsymbol{\alpha} \in \mathcal{E}\}$. The *ring* of all the polynomials in $\mathbf{x}$ with coefficients in $\mathbb{R}$ is $\mathbb{R}[\mathbf{x}]$.

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 sequence $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$ such that $\mathbf{x}^k \in \Omega$ for all $k \in \mathbb{N}$ and $\lim_{k\to+\infty} \|\mathbf{x}^k\|_2 = +\infty$, it holds that $\lim_{k\to\infty} f(\mathbf{x}^k) = +\infty$.

B. Problem statement

Given $f(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$, we consider the following minimization problem:

$$\left| \begin{array}{ll} \min & f(\mathbf{x}), \\ \text{with} & \mathbf{x} \in \Omega, \end{array} \right. \tag{1}$$

where $\Omega \subseteq \mathbb{R}^n$ is a convex, closed, full-dimensional set, i.e., $\int_\Omega 1 \, \mathrm{d}\mathbf{x} \neq 0$. The objective of this paper is to design a procedure for computing a solution to the minimization problem (1), i.e., find $f^\star \in \mathbb{R}$ and $\mathbf{x}^\star \in \Omega$ such that

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

Since the polynomial f is not convex on Ω in general, the minimization problem (1) is generally NP-hard, even for very special instances (see [1] for a survey on the computational complexity of the minimization problem (1) over some simple constraint sets). A lot of research effort has been indeed carried out for designing algorithms able to determine a (sub)optimal solution to the minimization problem (1).

One of the most direct methods to determine a solution to the minimization problem (1) is to consider it as a nonlinear programming problem, which can be addressed by using the Karush-Kuhn-Tucker necessary conditions for optimality [2], [3], [4], [5]. However, such techniques are usually not tailored for polynomial problems, thus leading to performances that may vary largely case-by-case [6].

If $\Omega = \mathbb{R}^n$ and the polynomial $f(\mathbf{x})$ is bounded below, several techniques specifically tailored for polynomial problems can be employed to determine the solution to the minimization problem (1), see, e.g., [7]. Since, for unconstrained minimization problems, the minimum of f is attained at a critical point $\mathbf{x}^*$ satisfying

$$\tfrac{\partial}{\partial x_1} f(\mathbf{x}^*) = 0, \ldots, \tfrac{\partial}{\partial x_n} f(\mathbf{x}^*) = 0, \tag{2}$$

methods have been developed for determining the set of all the solutions to the system of equalities (2). Remarkable examples are the tools given in [8], [9], [10], based on the computation of Gröbner bases and eigenvalues, the techniques given in [11], [12], which exploit the concept of A-discriminant, the rational univariate representation given in [13], and the numerical homotopy continuation methods given in [14], [15]. Similar techniques have been used in [16], where it is shown that a solution to the minimization problem (1) can be determined by solving an auxiliary parametric minimization problem that always admits a solution and taking limits.

An entirely different method to determine a solution to the minimization problem (1) when $\Omega = \mathbb{R}^n$ is to determine the largest $\lambda \in \mathbb{R}$ such that the polynomial $f(\mathbf{x}) - \lambda$ is a sum of squares (SOS) in $\mathbb{R}[\mathbf{x}]$, see [17], [18], [19]. Such a problem constitutes a relaxation of the minimization problem (1) and can be solved in polynomial time through semidefinite programming (SDP) [20], [21], providing a lower bound λ on $f^\star$. It is worth noticing that a similar reasoning can be used for determining a solution to the minimization problem (1) if the set Ω is compact and not necessarily convex, by using the results given in [22], [23]. SDP has also be proven successful to solve the minimization problem (1) when $f(\mathbf{x})$

is quadratic, i.e., $f(\mathbf{x}) = \mathbf{x}^\top F\mathbf{x}$, for some $F \in \mathbb{R}^{n\times n}$, see [24], [25], [26], [27], [28].

On the other hand, if Ω is a bounded, convex subset of $\mathbb{R}^n$, the tools given in [29], [30], [31] can be used to determine an approximate solution to the minimization problem (1) within polynomial time, with an assured worst-case performance ratio. In particular, under the assumption that Ω is a convex, compact set with non-empty interior, it is shown in [29] that the minimization problem (1) can be solved by a polynomial-time approximation algorithm with relative approximation ratio $(d+1)!(2d)^{-2d}(n+1)^{-\frac{d-2}{2}}(t+1)^{-\frac{d}{2}}$, where d is the total degree of the polynomial $f \in \mathbb{R}[x_1,\dots,x_n]$ and $t \in \mathbb{R}_{>0}$ is such that $\Omega \subset t\mathbb{B}$.

Among the popular methods that have recently gained increasing interest in large-scale optimization, due to its intrinsic simplicity, is the coordinate minimization method in its various variants [32], [33], [34]. For instance, the *Gauss-Seidel* method is based on updating an estimate $\hat{\mathbf{x}} = [\ \hat{x}_1 \quad \cdots \quad \hat{x}_n\ ]^\top$ of the solution $\mathbf{x}^\star$ to the minimization problem (1) according to the following iterations

$$\hat{\mathbf{x}}_{-i}^{k+1} = \hat{\mathbf{x}}_{-i}^{k}, \tag{3a}$$

$$\hat{x}_i^{k+1} \in \arg\min_{x_i \in \Omega_i^k} f(x_i, \mathbf{x}_{-i}), \tag{3b}$$

where $\hat{\mathbf{x}}_{-i} = [\ \hat{x}_1 \quad \cdots \quad \hat{x}_{i-1} \quad \hat{x}_{i+1} \quad \cdots \quad \hat{x}_n\ ]^\top$ and

$$\Omega_i^k \doteq \{x_i \in \mathbb{R} : [\ \hat{x}_1^k \quad \cdots \quad x_i \quad \cdots \quad \hat{x}_n^k\ ]^\top \in \Omega\}.$$

In its classical implementation, the Gauss-Seidel method updates the entries of $\hat{\mathbf{x}}$ cyclically (i.e., $i = \mathrm{mod}(k,n)+1$), starting from an initial point $\hat{\mathbf{x}}^0 \in \Omega$, and produces a sequence $\{\hat{\mathbf{x}}^k\}_{k\in\mathbb{N}}$, where $\hat{\mathbf{x}}^k = [\ \hat{x}_1^k \quad \cdots \quad \hat{x}_n^k\ ]^\top$.

Convergence results of such a method have been given for both the constrained and the unconstrained case under suitable (pseudo)convexity assumptions [35], [36], [37], [38], [39], [40], [41], [42], even in certain non-differentiable cases [43]. However, for non-convex problems, the classical Gauss-Seidel method need not converge to a critical point. A well-known example is given in [44], where it is shown that the limit points of the sequence $\{\hat{\mathbf{x}}^k\}_{k\in\mathbb{N}}$ need not be critical points of the corresponding minimization problem.

Stochastic versions of the method also exist, in which the coordinate to be updated is chosen each time at random, see [45], [46], [47]. Such methods have been proven succesfull to obtain an η-accurate solution to the minimization problem (1) with probability at least $1-\sigma$, with η and σ being arbitrary number in $\mathbb{R}_{>0}$, provided that the function f to be minimized can be rewritten as the sum of a smooth convex and a nonsmooth convex block-separable function.

In the next section we introduce a variation on the plain random coordinate minimization scheme, which allows the algorithm to seldom take minimization directions that are different from the coordinate axes. Lower and upper bounds on the convergence probability of such an algorithm are given in Section III. Details about the implementation and the computational complexity of this method are given in Section IV. The practical efficiency of the technique is demonstrated in Section V via numerical tests and comparisons with SOS-based methods.

II. A variation on the random coordinate minimization method

The algorithm we propose is described in words as follows: given $p \in [0,1]$, for any current solution estimate $\hat{\mathbf{x}}^k \in \Omega$ at iteration k, with probability p we pick a coordinate direction $i \in \{1,\dots,n\}$ uniformly at random, and with probability $1-p$ we pick a random direction v uniformly on the surface $\mathbb{S}$ of the unit Euclidean ball $\mathbb{B}$ in $\mathbb{R}^n$; we then set $\mathbf{s}^k$ equal to the i-th standard unit vector $\mathbf{e}_i \in \mathbb{S}$ in the first case, or set $\mathbf{s}^k = v$ in the second case. We then update the solution estimate according to the following rule

$$\hat{\mathbf{x}}^{k+1} \in \hat{\mathbf{x}}^k + \lambda^k \mathbf{s}^k, \tag{4a}$$

where

$$\lambda^k = \arg\min_{\lambda \in \mathcal{I}_k} f(\hat{\mathbf{x}}^k + \lambda \mathbf{s}^k), \tag{4b}$$

and $\mathcal{I}_k$ is the (possibly unbounded) interval $\{\lambda \in \mathbb{R} : \hat{\mathbf{x}}^k + \lambda\mathbf{s}^k \in \Omega\}$. Clearly, for $p=1$, the above method is a standard random coordinate minimization method, while for $p=0$ it becomes a random search method with exact line search.

The following theorem guarantees, under some mild assumptions, that the solutions of the inclusion (4) asymptotically converge in probability to the solution of the minimization problem (1).

Theorem 1. *Let $\mathcal{A} \subseteq \mathbb{R}^n$ be the set of all optimal solutions of problem (1), i.e., $\mathcal{A} \doteq \{x^\star \in \Omega :$ such that $f(x) \geq f(x^\star) \forall x \in \Omega\}$. Assume that $\mathcal{A}$ is nonempty, and that either the set Ω is compact or $f \in ru(\Omega)$.*

If $p < 1$ and 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 in probability from Ω, i.e., for each $\varepsilon \in \mathbb{R}_{>0}$, $\sigma \in \mathbb{R}_{>0}$, and $\hat{\mathbf{x}}^0 \in \Omega$, letting $\{\hat{\mathbf{x}}^k\}_{k\in\mathbb{N}}$ be any solution to the stochastic inclusion (4), there exists $K \in \mathbb{N}$ such that $\mathbb{P}(\hat{\mathbf{x}}^i \in \mathcal{A}+\varepsilon\mathbb{B}^o,\ \forall i \in \mathbb{N},\ i \geqslant K) \geqslant 1-\sigma$.

Proof. Note that the dynamics of the stochastic difference inclusion (4) can be rewritten as

$$\hat{\mathbf{x}}^{k+1} \in G(\hat{\mathbf{x}}^k, \mathbf{s}^k), \tag{5}$$

where $G : \Omega \times \mathbb{S} \rightrightarrows \Omega$, $G(\mathbf{x},\mathbf{s}) \doteq \{\mathbf{y} \in \Omega : \exists \lambda^\star \in \mathcal{I}$ such that $\mathbf{y} = \mathbf{x}+\lambda^\star\mathbf{s}$ and $f(\mathbf{x}+\lambda^\star\mathbf{s}) \leqslant f(\mathbf{x}+\lambda\,\mathbf{s}),\ \forall \lambda \in \mathcal{I}\}$, $\mathcal{I} \doteq \{\lambda \in \mathbb{R} : \hat{\mathbf{x}}+\lambda\mathbf{s} \in \Omega\}$, and $\{\mathbf{s}^k\}_{k\in\mathbb{N}}$ is a sequence of independent, identically distributed random variables defined from the probability space $(\Psi,\mathcal{F},\mathbb{P})$. Namely, for each $k \in \mathbb{N}$, the random variable $\mathbf{s}^k : \Psi \to \mathbb{S}$ is such that the probability measure $\mu(F) = \mathbb{P}(\psi \in \Psi : \mathbf{s}^k(\psi) \in F)$ is well defined for each F in the Borel σ-field on $\mathbb{S}$. In particular, for each $k \in \mathbb{N}$, $\mathbf{s}^k \sim (1-p)\mathrm{Uni}(\mathbb{S}) + \frac{p}{n}\sum_{i=1}^n \delta(\mathbf{e}_i)$, where $\delta(\cdot)$ denotes the Kronecker delta.

Thus, let $f^\star = f(\mathbf{x}^\star)$ where $\mathbf{x}^\star$ is any point in $\mathcal{A}$. In order to establish the statement, it is firstly proved that the set

$$\mathcal{L}_c \doteq \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) - f^\star \leqslant c\} \cap \Omega \tag{6}$$

is compact for each $c \in \mathbb{R}_{\geqslant 0}$. If Ω is compact, then $\mathcal{L}_c$ is compact since it is the intersection of a closed and a compact set. On the other hand, if Ω is not compact, but $f \in \text{ru}(\Omega)$, then, given $c \in \mathbb{R}_{\geqslant 0}$, there exists $d \in \mathbb{R}_{\geqslant 0}$ such that $f(\mathbf{x}) - f^\star > c$ for each $\mathbf{x} \in \Omega \setminus d\mathbb{B}$. Thus, $\mathcal{L}_c$ is compact for all $c \in \mathbb{R}_{\geqslant 0}$. In particular, the set $\mathcal{A} = \mathcal{L}_0$ is compact.

Secondly, it is shown that $f(\hat{\mathbf{x}}^{k+1}) \leqslant f(\hat{\mathbf{x}}^k)$ for all $k \in \mathbb{N}$. In fact, assume, by contradiction, that $f(\hat{\mathbf{x}}^{k+1}) > f(\hat{\mathbf{x}}^k)$. This implies that there exists $\mathbf{s} \in \mathbb{S}$ such that $f(\hat{\mathbf{x}}^k + \mathbf{s} \arg\min_{\lambda \in \mathcal{I}^k} f(\hat{\mathbf{x}}^k + \lambda \mathbf{s})) > f(\hat{\mathbf{x}}^k)$ leading to a contradiction by the definition of the $\arg\min(\cdot)$ function. Furthermore, if $\hat{\mathbf{x}}^0 \in \Omega$, then, by construction, one has that $\hat{\mathbf{x}}^k \in \Omega$ for all $k \in \mathbb{N}$. Therefore, letting $f_0 = f(\hat{\mathbf{x}}^0)$, one has that $\hat{\mathbf{x}}^k \in \mathcal{L}_{f_0}$ for all $k \in \mathbb{N}$, i.e., the set-valued mapping $G : \mathbb{R}^n \times \mathbb{S} \rightrightarrows \mathbb{R}^n$ given in (5) is locally bounded. Moreover, since the set Ω is either compact or $f \in \text{ru}(\Omega)$, by the same reasoning given above about the compactness of the sets $\mathcal{L}_c$, the mapping $\mathbf{s} \mapsto \text{graph}(G(\cdot, \mathbf{s})) \doteq \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^n : y \in G(\mathbf{x}, \mathbf{s})\}$ has closed values. Hence, measurability of such a mapping follows by Example 5.22 and Exercise 14.9 of [48]. Therefore, the set-valued mapping $G : \mathbb{R}^n \times \mathbb{S} \rightrightarrows \mathbb{R}^n$ satisfies Standing Assumption 1 of [49], [50], [51], which guarantees existence of random solutions of (4).

Thus, let $V : \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$ 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 $f \in \text{ru}(\Omega)$. Since $V(\mathbf{x}) = 0$ if and only if $\mathbf{x} \in \mathcal{A}$, the set $\mathcal{A}$ is compact, and V is radially unbounded, by Lemma 4.3 of [52], there exist class $\mathcal{K}_\infty$ functions $\underline{\alpha}$ and $\overline{\alpha}$ such that $\underline{\alpha}(\|\mathbf{x}\|_{\mathcal{A}}) \leqslant V(\mathbf{x}) \leqslant \overline{\alpha}(\|\mathbf{x}\|_{\mathcal{A}})$. By the reasoning given above about the monotonic behavior of $f(\hat{\mathbf{x}}^k)$, it results that $\sup_{\mathbf{g} \in G(\mathbf{x},\mathbf{s})} V(\mathbf{g}) \leqslant V(\mathbf{x})$ for all $(\mathbf{x}, \mathbf{s}) \in \Omega \times \mathbb{S}$. Furthermore, for each $\mathbf{x} \notin \mathcal{A}$, letting $\upsilon = \frac{f(\mathbf{x}) - f^\star}{2}$, $\upsilon \in \mathbb{R}_{>0}$, the set $\mathcal{L}_\upsilon$ has nonzero measure since, by assumption, the set $(\mathcal{A} + \nu\mathbb{B}) \cap \Omega$ has nonzero measure for all $\nu \in (0, \nu^\star)$, and f is continuous. Thus, for each $\mathbf{x} \notin \mathcal{A}$, there exists a measurable selection $\mathcal{S}$ of $\mathbb{S}$ such that $\sup_{\mathbf{g} \in G(\mathbf{x},\mathbf{s})} V(\mathbf{g}) < V(\mathbf{x})$ for all $\mathbf{s} \in \mathcal{S}$ (see Fig. 1 for a graphical representation of such a selection).

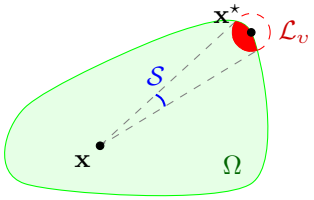


Fig. 1: Measurable selection $\mathcal{S}$ of $\mathbb{S}$.

Therefore, there exists $\varrho \in \mathcal{PD}(\mathcal{A})$ such that

$$\int_{\mathbb{S}} \sup_{\mathbf{g} \in G(\mathbf{x},\mathbf{s})} V(\mathbf{g})\, \mu(\text{d}\mathbf{s}) = \frac{(1-p)\Gamma(\frac{n}{2})}{2\pi^{n/2}} \int_{\mathbb{S}} \sup_{\mathbf{g} \in G(\mathbf{x},\mathbf{s})} V(\mathbf{g})\, \text{d}\mathbf{s} + \frac{p}{n} \sum_{i=1}^{n} \sup_{\mathbf{g} \in G(\mathbf{x},\mathbf{e}_i)} V(\mathbf{g}) \leqslant V(\mathbf{x}) - \varrho(\mathbf{x}),$$

where $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$, for all $\mathbf{x} \in \Omega$. Thus, by a trivial extension of Theorem 1 of [53], since if $\hat{\mathbf{x}}^0 \in \Omega$ then $\hat{\mathbf{x}}^k \in \Omega$ for all $k \in \mathbb{N}$, the set $\mathcal{A}$ if asymptotically stable in probability from Ω. □

It is worth noticing that, in order to establish Theorem 1, no assumption is needed on the convexity of the polynomial f. The following two remarks and two examples discuss the assumptions made in Theorem 1.

Remark 1. If Ω is not compact and $f \notin \text{ru}(\Omega)$, then the set-valued mapping $G(\mathbf{x}, \mathbf{s})$ given in (5) need not be locally bounded for all $(\mathbf{x}, \mathbf{s}) \in \Omega \times \mathbb{S}$, thus leading to undesirable behaviors, as, e.g., unbounded random solutions (see the following example).

Example 1. Define the polynomial $f(\mathbf{x}) = x_2^2 x_1^2 + x_1^2 - 2x_2 x_1 + 1$, let $\Omega = \mathbb{R}^2$, and consider the minimization problem (1). By [54], the polynomial f can be rewritten as $f(\mathbf{x}) = x_1^2 + (1 - x_1 x_2)^2$. Thus, letting $\mathbf{y}^k = \begin{bmatrix} \frac{1}{k} & k \end{bmatrix}^\top$, one has that $\lim_{k \to +\infty} \|\mathbf{y}^k\| = +\infty$, but $\lim_{k \to +\infty} f(\mathbf{y}^k) = 0$, i.e., $f \notin \text{ru}(\Omega)$. Note that, since $n = 2$, the random variable $\mathbf{s}$ can be parametrized as $\mathbf{s}(\vartheta) = \begin{bmatrix} \cos(\vartheta) & \sin(\vartheta) \end{bmatrix}^\top$, where ϑ is another random variable, $\vartheta \sim (1-p)\text{Uni}([0, 2\pi]) + \frac{p}{2}\delta(0) + \frac{p}{2}\delta(\frac{\pi}{2})$, where $\delta(\cdot)$ denotes the Kronecker delta. The following Fig. 2 depicts the map $\vartheta \mapsto \arg\min_\lambda f(\lambda\, \mathbf{s}(\vartheta))$.

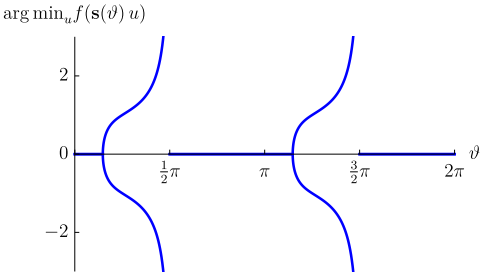


Fig. 2: Graph of $\vartheta \mapsto \arg\min_\lambda f(\lambda\, \mathbf{s}(\theta))$.

As shown by such a figure, the map G given in (5) is unbounded, thus possibly leading to large computational errors when using the proposed method to solve the minimization problem (1) (which, in this case, does not have a solution).

Remark 2. The assumption about measurability of the set $(\mathcal{A} + \nu\mathbb{B}) \cap \Omega$ essentially requires that there exist sufficiently many points in Ω that are close to the optimal set $\mathcal{A}$. Such an assumption is not satisfied if the set Ω is not full dimensional. The following example shows that, in such a case, the solutions of the stochastic difference inclusion (4) need not converge to a solution to the minimization problem (1).

Example 2. Consider the following minimization problem

$$\left| \begin{array}{ll} \min & f(\mathbf{x}), \\ \text{with} & x_1 + x_2 + x_3 = 0, \end{array} \right. \tag{7}$$

where $f = x_1^2 + 6x_2x_1 - 2x_3x_1 + 2x_1 + 10x_2^2 + 6x_3^2 + 12x_2 - 10x_2x_3 - 14x_3 + 10$. By letting $x_1 = -x_2 - x_3$ and by using the "complete the squares" procedure given in [55], one obtains that f constrained to $\Omega = \{\mathbf{x} \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 0\}$ can be rewritten as $f(\mathbf{x})_{|\Omega} = 10(1 + \frac{1}{2}x_2 - \frac{4}{5}x_3)^2 + \frac{5}{2}(x_2 - \frac{4}{5}x_3)^2 + x_3^2$. Therefore, the minimization problem (7) admits a solution, $f^\star = \frac{25}{9}$ and $\mathbf{x}^\star = [\ -\frac{13}{9} \quad \frac{1}{3} \quad \frac{10}{9}\ ]^\top$, and, for each sequence $\{\mathbf{y}^k\}_{k\in\mathbb{N}}$ such that $\mathbf{y}^k \in \Omega$ for all $k \in \mathbb{N}$ and $\lim_{k\to+\infty} \|\mathbf{y}^k\| = +\infty$, one has that $\lim_{k\to+\infty} f(\mathbf{y}^k) = +\infty$, i.e., $f \in \text{ru}(\Omega)$. However, since $(\mathcal{A} + \nu\mathbb{B}) \cap \Omega$ has null measure, the stochastic difference inclusion (4) does not converge to the solution to the minimization problem (4). Namely, given $\mathbf{x}^0 \in \Omega$, $\mathbb{P}(\mathbf{x}^{k+1} \neq \mathbf{x}^k) = 0$ for all $k \in \mathbb{N}$, since $\mathbf{x}^{k+1} \neq \mathbf{x}^k$ only if $\mathbf{s}^k \in \Omega$ and $\mathbb{P}(\mathbf{s}^k \in \Omega) = 0$.

As shown in Remarks 1, 2 and in Examples 1, 2, the assumptions of Theorem 1 are the minimal set of conditions that have to be satisfied in order to guarantee well–behaved convergence of the stochastic variation of the Gauss-Seidel method given in (4). In particular, as shown in Example 1, if the function $f(\mathbf{x})$ is not radially unbounded on Ω, then the solutions to the stochastic difference inclusion (4) need not be bounded. On the other hand, if the set Ω has null measure, then, by Remark 2, the proposed method fails to converge in the form given in (4). Furthermore, the proof of Theorem 1 establishes also numerical stability of the proposed method, at least in the unconstrained case. In fact, by [56], the existence of a smooth Lyapunov function for the stochastic difference inclusion (4) establishes (semiglobal, practical) robustness with respect to small constant perturbations.

III. Bounds on the convergence probability of the proposed method

The tools given in [53], [51] can be used to determine bounds on the convergence time of the proposed method.

Let the assumptions of Theorem 1 hold and let $G : \mathbb{R}^n \times \mathbb{S} \rightrightarrows \mathbb{R}^n$ be the set-valued mapping given in (5). Thus, given $p \in \mathbb{R}_{\geqslant 0}$, $p < 1$, and an open set $\mathcal{O} \subset \mathbb{R}^n$ for each $\mathbf{x} \in \Omega$, define $\ell_{\subset\mathcal{O}}(0, \mathbf{x}) \doteq 1$, $\ell_{\cap\mathcal{O}}(0, \mathbf{x}) \doteq 0$, and, for each $(\mathbf{x}, k) \in \Omega \times \mathbb{N}$, define $\ell_{\subset\mathcal{O}}(k+1, \mathbf{x})$ and $\ell_{\cap\mathcal{O}}(k+1, \mathbf{x})$ as

$$\begin{aligned} \ell_{\subset\mathcal{O}}(k+1, \mathbf{x}) \doteq{}& \tfrac{(1-p)\Gamma(\frac{n}{2})}{2\pi^{n/2}} \int_{\mathbb{S}} \min_{\mathbf{g}\in G(\mathbf{x},\mathbf{s})} \mathbb{I}_{\mathcal{O}}(\mathbf{g}) \ell_{\subset\mathcal{O}}(k, \mathbf{x})\, \mathrm{d}\mathbf{s} \\ &+ \tfrac{p}{n} \sum_{i=1}^{n} \min_{\mathbf{g}\in G(\mathbf{x},\mathbf{e}_i)} \mathbb{I}_{\mathcal{O}}(\mathbf{g}) \ell_{\subset\mathcal{O}}(k, \mathbf{x}), \end{aligned} \tag{8a}$$

$$\begin{aligned} \ell_{\cap\mathcal{O}}(k+1, \mathbf{x}) \doteq{}& \tfrac{(1-p)\Gamma(\frac{n}{2})}{2\pi^{n/2}} \int_{\mathbb{S}} \min_{\mathbf{g}\in G(\mathbf{x},\mathbf{s})} \mathbb{J}_{\mathcal{O}}(k, \mathbf{g})\, \mathrm{d}\mathbf{s} \\ &+ \tfrac{p}{n} \sum_{i=1}^{n} \min_{\mathbf{g}\in G(\mathbf{x},\mathbf{e}_i)} \mathbb{J}_{\mathcal{O}}(k, \mathbf{g}), \end{aligned} \tag{8b}$$

where $\mathbb{J}_{\mathcal{O}}(k, \mathbf{g}) \doteq \{\mathbb{I}_{\mathcal{O}}(\mathbf{g}), \mathbb{I}_{\mathbb{R}^n\setminus\mathcal{O}}(\mathbf{g}) \ell_{\cap\mathcal{O}}(k, \mathbf{g})\}$. The following theorem shows how $\ell_{\subset\mathcal{O}}$ and $\ell_{\cap\mathcal{O}}$ can be used to bound the convergence probability of the proposed method

Theorem 2. *Let the assumptions of Theorem 1 hold. Given $\eta \in \mathbb{R}_{>0}$, let $\mathcal{O}_1 = \mathbb{R}^n \setminus \{\mathbf{x} \in \Omega : f(\mathbf{x}) - f^\star \leqslant \eta\}$ and let $\mathcal{O}_2 = \mathcal{A} + \varepsilon\mathbb{B}^o$, where $\varepsilon \in \mathbb{R}_{>0}$ is such that $\sup_{\mathbf{x}\in(\mathcal{A}+\varepsilon\mathbb{B}^o)} f(\mathbf{x}) \leqslant f^\star + \eta$. Thus, letting $\mathcal{S}(\mathbf{x}^0)$ be the set of all the solutions to the stochastic difference inclusion (4) starting at $\mathbf{x}^0$, one has that, for each $K \in \mathbb{N}$,*

$$\begin{aligned} \ell_{\cap\mathcal{O}_2}(K, \mathbf{x}^0) &\leqslant \mathbb{P}(f(\mathbf{x}^i) - f^\star \leqslant \eta,\ \forall i \geqslant K) \\ &\leqslant 1 - \ell_{\subset\mathcal{O}_1}(K, \mathbf{x}^0), \end{aligned} \tag{9}$$

for all random solutions $\{\mathbf{x}^k\}_{k\in\mathbb{N}} \in \mathcal{S}(\mathbf{x}^0)$.

Proof. If the assumptions of Theorem 1 are met, then the stochastic difference inclusion (5) satisfies Standing Assumption 1 of [51] and the set $\mathcal{L}_\eta$ defined in (6) is compact. Thus, the set $\mathcal{O}_1$ is open, whence implying, by [51, Lem. 1], that the function $\ell_{\subset\mathcal{O}_1} : \mathbb{N} \times \Omega \to [0, 1]$ is well-defied and

$$\ell_{\subset\mathcal{O}_1}(k, \mathbf{x}^0) = \inf_{\{\mathbf{z}^k\}_{k\in\mathbb{N}}\in\mathcal{S}(\mathbf{x}^0)} \mathbb{E}\left[\prod_{i=1}^{k} \mathbb{I}_{\mathcal{O}_1}(\mathbf{z}^k)\right],$$

i.e., $\ell_{\subset\mathcal{O}_1}(k, \mathbf{x}^0)$ constitute a lower bound over all the random solutions from $\mathbf{x}^0 \in \Omega$ for the probability of staying in the set $\mathcal{O}_1$ for k time steps. Therefore, since $\mathbf{x}^0 \in \Omega$ implies that $\mathbf{x}^k \in \Omega$, $k \in \mathbb{N}$, for each $\{\mathbf{x}^k\}_{k\in\mathbb{N}} \in \mathcal{S}(\mathbf{x}^0)$, $1 - \ell_{\subset\mathcal{O}_1}(K, \mathbf{x}^0)$ constitutes an upper bound over all the solutions from $\mathbf{x}^0$ for the probability of reaching the set $\mathcal{L}_\eta$. Thus, since the set $\mathcal{L}_\eta$ is positively invariant with respect to the stochastic inclusion (4) (see Theorem 1), we have that

$$\begin{aligned} &\mathbb{P}(f(\mathbf{x}^i) - f^\star \leqslant \eta,\ \forall i \geqslant K) \\ =&\mathbb{P}(f(\mathbf{x}^i) - f^\star \leqslant \eta,\ \forall i \geqslant K+1 \wedge \mathbf{x}^K \in \mathcal{L}_\eta) \\ =&\mathbb{P}(f(\mathbf{x}^i) - f^\star \leqslant \eta,\ \forall i \geqslant K+1 \mid \mathbf{x}^K \in \mathcal{L}_\eta)\mathbb{P}(\mathbf{x}^K \in \mathcal{L}_\eta) \\ =&\mathbb{P}(\mathbf{x}^K \in \mathcal{L}_\eta), \end{aligned}$$

and that

$$\begin{aligned} &\mathbb{P}(\exists i \leqslant K \text{ s.t. } f(\mathbf{x}^i) - f^\star \leqslant \eta) \\ =&\mathbb{P}(\exists i \leqslant K \text{ s.t. } f(\mathbf{x}^i) - f^\star \leqslant \eta \mid \mathbf{x}^K \in \mathcal{L}_\eta)\mathbb{P}(\mathbf{x}^K \in \mathcal{L}_\eta) \\ &+ \mathbb{P}(\exists i \leqslant K \text{ s.t. } f(\mathbf{x}^i) - f^\star \leqslant \eta \mid \mathbf{x}^K \notin \mathcal{L}_\eta)\mathbb{P}(\mathbf{x}^K \notin \mathcal{L}_\eta) \\ =&\mathbb{P}(\exists i \leqslant K \text{ s.t. } f(\mathbf{x}^i) - f^\star \leqslant \eta \mid \mathbf{x}^K \in \mathcal{L}_\eta)\mathbb{P}(\mathbf{x}^K \in \mathcal{L}_\eta) \\ =&\mathbb{P}(\mathbf{x}^K \in \mathcal{L}_\eta). \end{aligned}$$

Therefore, it results that $\mathbb{P}(f(\mathbf{x}^i) - f^\star \leqslant \eta,\ \forall i \geqslant K) = \mathbb{P}(\mathbf{x}^k \in \mathcal{L}_\eta) = \mathbb{P}(\exists i \in \{1, \dots, K\} \text{ such that } f(\mathbf{x}^i) - f^\star \leqslant \eta) \leqslant 1 - \ell_{\subset\mathcal{O}_1}(K, \mathbf{x}^0)$, for all $\{\mathbf{x}^k\}_{k\in\mathbb{N}} \in \mathcal{S}(\mathbf{x}^0)$.

On the other hand, we have that $f(\mathbf{x}) - f^\star \leqslant \eta$ for all $\mathbf{x} \in \mathcal{O}_2$. By [51, Lem. 2], since the stochastic difference inclusion (5) satisfies Standing Assumption 1 of [51], the function $\ell_{\subset\mathcal{O}_1} : \mathbb{N} \times \Omega \to [0, 1]$ is well-defined and

$$\ell_{\cap\mathcal{O}_2}(k, \mathbf{x}^0) = \inf_{\{\mathbf{z}^k\}_{k\in\mathbb{N}}\in\mathcal{S}(\mathbf{x}^0)} \mathbb{E}\left[\max_{i\in\{1\dots,k\}} \mathbb{I}_{\mathcal{O}_2}(\mathbf{z}^k)\right],$$

i.e., $\ell_{\cap\mathcal{O}_2}(k, \mathbf{x}^0)$ constitute a lower bound over all the random solutions from $\mathbf{x}^0 \in \Omega$ for the probability of reaching the set $\mathcal{O}_2$ in (at most) k time steps. Thus, since $f(\mathbf{x}^{k+1}) \leqslant f(\mathbf{x}^k)$ and hence $\mathbb{P}(f(\mathbf{x}^i) - f^\star \leqslant \eta,\ \forall i \geqslant K \mid \exists j \in \{1, \dots, K\} \text{ such that } \mathbf{x}^j \in \mathcal{O}_2) = 1$, it results that $\mathbb{P}(f(\mathbf{x}^i) - f^\star \leqslant \eta,\ \forall i \geqslant K) \geqslant \mathbb{P}(f(\mathbf{x}^i) - f^\star \leqslant \eta,\ \forall i \geqslant K \wedge \exists j \in \{1, \dots, K\} \text{ such that } \mathbf{x}^j \in \mathcal{O}_2) =$

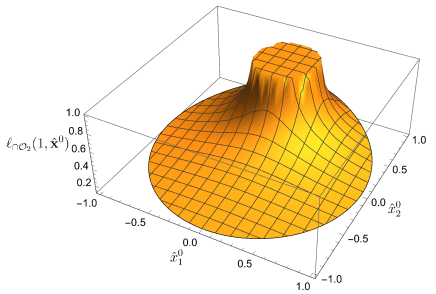
$\mathbb{P}(\exists j \in \{1,\dots,K\}$ such that $\mathbf{x}^j \in \mathcal{O}_2) \geqslant \ell_{\cap\mathcal{O}_2}(K,\mathbf{x}^0)$, for all random solutions $\{\mathbf{x}^k\}_{k\in\mathbb{N}} \in \mathcal{S}(\mathbf{x}^0)$. □

By Theorem 2, the functions $\ell_{\cap\mathcal{O}}$ and $\ell_{\subset\mathcal{O}}$ defined in (8) can be used to determine upper and lower bounds on the probability of obtaining an η-accurate solution to the minimization problem (1) within k time steps, starting from the initial condition $\mathbf{x}^0$. Note that, for each $\varepsilon,\eta \in \mathbb{R}_{>0}$ and all $\mathbf{x}^0 \in \Omega$, by Theorem 1, it results that $\lim_{k\to\infty} \ell_{\cap\mathcal{O}_2}(k,\mathbf{x}^0) = 1$ and $\lim_{k\to\infty} \ell_{\subset\mathcal{O}_1}(k,\mathbf{x}^0) = 0$ (see also [50], [53]).

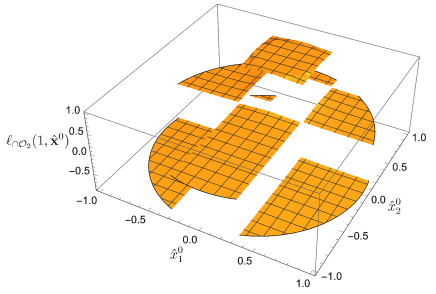
Example 3. Consider the minimization problem (1) with $\Omega = \mathbb{R}^2$ and $f = x_1^2 + x_2^2 - x_2 + 1$. Since $f \in \text{ru}(\mathbb{R}^2)$ and $\Omega = \mathbb{R}^2$, the assumptions of Theorem 1 are met, whence the function $\ell_{\cap\mathcal{O}}$ given in (8) is well-defined. By the same reasoning used in Example 1, for the considered minimization problem, the stochastic system (5) reads as

$$x_1^{k+1} = \tfrac{1}{4}(\sin(2\theta^k) - 2x_2^k\sin(2\theta^k) - 2x_1^k\cos(2\theta^k) + 2x_1^k),$$
$$x_2^{k+1} = \tfrac{1}{2}(\sin^2(\theta^k) - x_1^k\sin(2\theta^k) + 2x_2^k\cos^2(\theta^k)),$$

where $\{\theta^k\}_{k\in\mathbb{N}}$ is a sequence of i.i.d. random variables, $\theta^k \sim (1-p)\text{Uni}([0,2\pi]) + \frac{p}{2}\delta(0) + \frac{p}{2}\delta(\frac{\pi}{2})$, for each $k \in \mathbb{N}$. Assume that the objective is to characterize the convergence probabilities of obtaining a 0.1-accurate solution to the minimization problem (1), within 1 step, from the initial conditions $\hat{\mathbf{x}}^0 \in \mathbb{B}$. It can be easily derived that $\varepsilon = 0.1$ is such that $\sup_{\mathbf{x}\in(\mathcal{A}+\varepsilon\mathbb{B}^o)} f(\mathbf{x}) \leqslant f^\star + 0.1$. Thus, let $\mathcal{O}_2 = \mathcal{A} + 0.1\,\mathbb{B}^o$. Figure 3 depicts the values assumed by the function $\ell_{\cap\mathcal{O}_2}$ for two different values of p.



(a) $p = 0$.



(b) $p = 0.9$.

Fig. 3: Value of the function $\ell_{\cap\mathcal{O}_2}(1,\hat{\mathbf{x}}^0)$ for $\hat{\mathbf{x}}^0 \in \mathbb{B}$.

As shown by such a figure, the value chosen for p affects the convergence probabilities of the proposed method. As a matter of fact, by Theorem 2, $\mathbb{P}(f(\mathbf{x}^i) - f^\star \leqslant \eta, \forall i \geqslant K) \geqslant \ell_{\cap\mathcal{O}_2}(K,\mathbf{x}^0)$, thus, the probability of obtaining a 0.1-accurate solution to the minimization problem (1), within 1 step, starting from $\hat{\mathbf{x}}^0$, is lower bounded by $\ell_{\cap\mathcal{O}_2}(1,\mathbf{x}^0)$. Thus, since $\ell_{\cap\mathcal{O}_2}(K,\mathbf{x}^0)$ is highly affected by the chosen value of p, the convergence probability of the proposed algorithm strongly depends on the choice of p. In particular, if p is close to 1, then the points $\hat{\mathbf{x}}^0$ in $\mathcal{O}_2 + \text{Span}(\mathbf{e}_i)$ (for some $i \in \{1,\dots,n\}$) are more likely to converge to $\mathcal{O}_2$ (see Figure 3a). On the other hand, if p is close to 0, given $\hat{\mathbf{x}}^0$, the convergence probability just depends on the measure of the set $\mathcal{S} \subset \mathbb{S}$ that is such that $G(\hat{\mathbf{x}}^0,\mathbf{s}) \subset \mathcal{O}_2$ for all $\mathbf{s} \in \mathcal{S}$. Therefore, when implementing the proposed algorithm, the parameter p has to be chosen as a trade-off between computational efficiency (see Section IV) and convergence probability (see Theorem 2).

IV. Implementation notes and complexity

In this section, we provide some tools to implement the proposed variation of the random coordinate minimization scheme to solve the minimization problem (1). Firstly, note that given a monomial $\mathbf{x}^{\boldsymbol{\alpha}} \in \mathbb{R}[\mathbf{x}]$, it can be equivalently rewritten as $\mathbf{x}^{\boldsymbol{\alpha}} = \mathbf{x}_{-i}^{\boldsymbol{\alpha}_{-i}} x_i^{\alpha_i}$, for each $i \in \{1,\dots,n\}$. Similarly, given $f \in \mathbb{R}[\mathbf{x}]$, the *$i$-th coordinate-wise polynomial* of f at $\mathbf{x}_{-i}$ is the univariate polynomial in x_i with coefficients in $\mathbb{R}[\mathbf{x}_{-i}]$ that is obtained by considering all values in $\mathbf{x}_{-i}$ being fixed, and only the i-th variate x_i as variable, i.e.,

$$f_i(x_i) = \sum_{\boldsymbol{\alpha}\in\mathcal{E}} c_\alpha \mathbf{x}^{\boldsymbol{\alpha}} = \sum_{\boldsymbol{\alpha}\in\mathcal{E}} c_\alpha \mathbf{x}_{-i}^{\boldsymbol{\alpha}_{-i}} x_i^{\alpha_i} = \sum_{\boldsymbol{\alpha}\in\mathcal{E}} \tilde{c}_\alpha(\mathbf{x}_{-i}) x_i^{\alpha_i}$$

where $\tilde{c}_\alpha(\mathbf{x}_{-i}) \doteq c_\alpha \mathbf{x}_{-i}^{\boldsymbol{\alpha}_{-i}} \in \mathbb{R}[\mathbf{x}_{-i}]$. The i-th coordinate-wise polynomial f_i is extremely useful when we update the estimate of the solution to the minimization problem (1) with $\mathbf{s}^k = \mathbf{e}_i$ (note that $\mathbb{P}(\mathbf{s}^k = \mathbf{e}_i) = \frac{p}{n}$). In fact, in such a case, letting $f_i^k(x_i) = \sum_{\boldsymbol{\alpha}\in\mathcal{E}} \tilde{c}_\alpha(\hat{\mathbf{x}}_{-i}^k) x_i^{\alpha_i}$, $f_i^k \in \mathbb{R}[x_i]$, and letting $\mathcal{X}_i^k = \{x_i \in \mathbb{R} : f_i^k(x_i) \leqslant f_i^k(\hat{x}_i^{k-1})\}$,

$$\underline{x}_i^k = \min\{x_i \in \Omega_i^k : x_i \geqslant \min \mathcal{X}_i^k\},$$
$$\overline{x}_i^k = \max\{x_i \in \Omega_i^k : x_i \leqslant \max \mathcal{X}_i^k\},$$

which can be computed by using the bisection method [57], the value of $\hat{x}_i^{k+1}$ can be determined by solving the following (univariate) minimization problem

$$\left|\begin{array}{ll} \min & f_i^k(x_i), \\ \text{with} & x_i \in [\underline{x}_i^k, \overline{x}_i^k], \end{array}\right. \tag{10}$$

and, letting $x_i^\star$ be the set of all the solutions to the minimization problem (10), setting $\hat{x}_i^{k+1} \in x_i^\star$. Note that the minimization problem (10) can be solved by computing the roots $r_1^k,\dots,r_h^k \in [\underline{x}_i^k, \overline{x}_i^k]$ of $\frac{d}{dx_i} f_i^k(x_i)$, and letting

$$x_i^\star = \underset{r\in\{r_1^k,\dots,r_h^k,\underline{x}_i^k,\overline{x}_i^k\}}{\arg\min} f_i^k(r). \tag{11}$$

On the other hand, assume that at the k-th time step, $\mathbf{s}^k$ is picked in $\mathbb{S} \setminus \{\mathbf{e}_1,\dots,\mathbf{e}_n\}$. Let $\tilde{f}^k(\lambda) \doteq f(\hat{\mathbf{x}}^k + \lambda\mathbf{s}^k) = \sum_{\boldsymbol{\alpha}\in\mathcal{E}} c_\alpha(\mathbf{x}^k + \lambda\mathbf{s}^k)^{\boldsymbol{\alpha}}$. To obtain an explicit polynomial representation of $\tilde{f}^k(\lambda)$, we consider the j-th monomial

$$m_{\boldsymbol{\alpha}}(\lambda) \doteq (\mathbf{x}^k + \lambda\mathbf{s}^k)^{\boldsymbol{\alpha}} = \prod_{i=1}^n (x_i^k + \lambda s_i^k)^{\alpha_i}.$$

Each term in the product above can be written as

$$(x_i^k + \lambda s_i^k)^{\alpha_i} = \sum_{j=0}^{\alpha_i} \binom{\alpha_i}{j} (x_i^k)^{\alpha_i - j} (s_i^k)^j \cdot \lambda^j = \sum_{j=0}^{\alpha_i} \zeta_{i,j}^{\alpha_i}(x_i^k, s_i^k)\lambda^j,$$

where

$$\zeta_{i,j}^{\alpha_i}(x_i^k, s_i^k) \doteq \binom{\alpha_i}{j} (x_i^k)^{\alpha_i - j} (s_i^k)^j.$$

Therefore, by defining the row vector $\boldsymbol{\zeta}_i^{\alpha_i}(x_i^k, s_i^k) \doteq [\ \zeta_{i,0}^{\alpha_i}(x_i^k, s_i^k) \ \cdots \ \zeta_{i,\alpha_i}^{\alpha_i}(x_i^k, s_i^k)\ ]$, the product in the expression for $m_{\boldsymbol{\alpha}}(\lambda)$ can be readily obtained by taking the convolution of the $\boldsymbol{\zeta}_i^{\alpha_i}$'s, i.e., letting $\boldsymbol{\omega}_{\boldsymbol{\alpha}}(x_i^k, s_i^k) = [\ \omega_{\boldsymbol{\alpha},0}(x_i^k, s_i^k) \ \cdots \ \omega_{\boldsymbol{\alpha},|\boldsymbol{\alpha}|}(x_i^k, s_i^k)\ ]$, $\omega_{\boldsymbol{\alpha}}(x_i^k, s_i^k) \doteq \boldsymbol{\zeta}_1^{\alpha_1}(x_i^k, s_i^k) * \boldsymbol{\zeta}_2^{\alpha_2}(x_i^k, s_i^k) * \cdots * \boldsymbol{\zeta}_n^{\alpha_n}(x_i^k, s_i^k)$, where $*$ denotes the convolution operator, we have that $m_{\boldsymbol{\alpha}}(\lambda) = \sum_{i=0}^{|\boldsymbol{\alpha}|} \omega_{\boldsymbol{\alpha},i}(x_i^k, s_i^k)\lambda^i$, and hence

$$\tilde{f}^k(\lambda) = \sum_{\boldsymbol{\alpha}\in\mathcal{E}} \sum_{i=0}^{|\boldsymbol{\alpha}|} c_{\boldsymbol{\alpha}} \omega_{\boldsymbol{\alpha},i}(x_i^k, s_i^k)\lambda^i.$$

Once that the polynomial $\tilde{f}^k(\lambda)$ has been determined, let $\Lambda^k = \{\lambda \in \mathbb{R} : \tilde{f}^k(\lambda) \leqslant f(\hat{\mathbf{x}}^{k-1})\}$, and

$$\underline{\lambda}^k = \min\{\lambda \in \mathcal{I}_k : \lambda \geqslant \min \Lambda^k\},$$
$$\overline{\lambda}^k = \max\{\lambda \in \mathcal{I}_k : \lambda \leqslant \max \Lambda^k\},$$

which can be computed by using the bisection method. Thus, the next value of the estimate of the solution to the minimization problem (1) can be obtained by solving the (univariate) minimization problem

$$\left| \begin{array}{ll} \min & \tilde{f}^k(\lambda), \\ \text{with} & \lambda \in [\underline{\lambda}^k, \overline{\lambda}^k], \end{array} \right. \tag{12}$$

and, letting $\lambda^\star$ be the set of all the solutions to such a minimization problem, setting $\hat{\mathbf{x}}^{k+1} \in \hat{\mathbf{x}}^k + \lambda^\star \mathbf{s}^k$. Note that the minimization problem (12) can be solved by computing the roots $\lambda_1^k, \ldots, \lambda_h^k \in [\underline{\lambda}^k, \overline{\lambda}^k]$ of $\frac{d}{d\lambda}\tilde{f}^k(\lambda)$, and letting

$$\lambda^\star = \operatorname*{arg\,min}_{\lambda \in \{\lambda_1^k, \ldots, \lambda_h^k, \underline{\lambda}^k, \overline{\lambda}^k\}} \tilde{f}^k(\lambda). \tag{13}$$

Letting $d \in \mathbb{R}_{>0}$ be the total degree of the polynomial f to be minimized, in order to determine the value of the monomials $\hat{\mathbf{x}}_{-i}^{\boldsymbol{\alpha}_{-i}}$, one has to carry out up to $n^2\, d$ elementary operations (i.e., additions or multiplications) [58]. Thus, the i-th coordinate-wise polynomial of f at $\hat{\mathbf{x}}_{-i}^k$ can be determined by carrying out up to $d\,(d+1)\,n\binom{d+n}{d+1}$ since there are $\sum_{r=0}^{d}\binom{n+r-1}{r} = \frac{d+1}{n}\binom{n+d}{d+1}$ vectors $\boldsymbol{\alpha} \in \mathbb{N}^n$ such that $|\boldsymbol{\alpha}| \leqslant d$. On the other hand, in order to determine each of the vectors $\boldsymbol{\zeta}_i^{\alpha_i}(x_i^k, s_i^k)$ and $\boldsymbol{\omega}_{\boldsymbol{\alpha}}(x_i^k, s_i^k)$ one has to carry out up to $2d^3(d+1)$ and $\frac{5}{2}n\,(d+1)(d\,n+d+2)$ elementary operations, respectively. Thus, the polynomial $\tilde{f}^k(\lambda)$ can be determined by carrying out up to $\frac{1}{2}(d+1)^2(4d^3 + 5d\,n + 5d + 10)\binom{d+n}{d+1}$ elementary operations. The extremal points of the set $\mathcal{X}_i^k$ (respectively, Λ^k) can be determined by computing the roots of the polynomial $f_i^k(x_i) - f_i^k(\hat{x}_i^{k-1})$ (respectively, $\tilde{f}^k(\lambda) - f(\hat{\mathbf{x}}^{k-1})$), thus carrying out up to d^6 elementary operations [59]. Hence, in order to determine $\underline{x}_i^k, \overline{x}_i^k$ (respectively, $\underline{\lambda}^k, \overline{\lambda}^k$) by using the bisection method, one has to carry out up to $2\,M\log_2(\frac{2t}{\epsilon})$ elementary operations, where M is the number of elementary operations required to determine whether $\mathbf{x} \in \Omega$, t is such that $\{\mathbf{x} \in \Omega : f(\mathbf{x}) \leqslant f(\hat{\mathbf{x}}^0)\} \subset t\mathbb{B}$, and ϵ is the desired precision. Finally, in oder to determine $x_i^\star$ (respectively, $\lambda^\star$) one has to carry out $(d-1)^6$ elementary operations to determine the roots of $\frac{d}{dx_i}f_i^k(x_i)$ (respectively, $\frac{d}{d\lambda}\tilde{f}^k(\lambda)$) and up to $2(d-1)(d+1)$ elementary operations to evaluate $f_i^k(x_i)$ (respectively, $\tilde{f}^k(\lambda)$) over $\{r_1^k, \ldots, r_h^k, \underline{x}_i^k, \overline{x}_i^k\}$ (respectively, $\{\lambda_1^k, \ldots, \lambda_h^k, \underline{\lambda}^k, \overline{\lambda}^k\}$). Therefore, for each $(d, n) \in \mathbb{N} \times \mathbb{N}$, at each iteration of (4), is more computationally efficient to update the estimate of the solution to problem (1) by using $\mathbf{s}^k \in \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. This improved efficiency motivates the interest in letting $p > 0$ (but, $p < 1$) when implementing the proposed variation of the random coordinate descent method.

V. Numerical tests

In this section, we demonstrate the practical efficiency of the proposed minimization algorithm through some numerical tests. Randomly generated experiments have been carried out to compare the procedure given in (4), with the `MATLAB` toolbox `SOSTOOLS` [60] interfaced with the external solver `SeDuMi` [61]. All the experiments have been carried out on a laptop with an Intel Core i5 CPU (2.4 GHz) and an 8 GB, 1600 MHz, DDR3 RAM.

In each experiment, it has been assumed that $\Omega = \mathbb{R}^n$, and the polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ to be minimized has been generated by firstly determining a polynomial $\bar{f} \in \mathbb{R}[x_1, \ldots, x_n]$ of total degree $2d-1$, with coefficients being random integers uniformly distributed in the interval $[-100, 100]$ and, secondly, letting $f = \bar{f} + x_1^{2d} + \cdots + x_2^{2d}$. By [16], with such a choice, the minimization problem (1) always admits a solution (i.e., $\mathcal{A} \neq \emptyset$), $f \in \mathrm{ru}(\mathbb{R}^n)$, and the set $\mathcal{A} + \nu\mathbb{B}$ has nonzero measure for each $\nu \in \mathbb{R}_{>0}$. Thus, by Theorem 1, the solutions of the stochastic difference inclusion (4) converge to the solution of the minimization problem (1). Note that similar polynomials have been used in [7] to validate SOS-based optimization methods.

Since, by Theorems 1 and 2, the proposed method converges asymptotically in probability to the solution to the minimization problem (1) (possibly, not in finite time), the iteration given in (4) has been interrupted as soon as $\|\hat{\mathbf{x}}^{k+1} - \hat{\mathbf{x}}^k\| = |\lambda^k|$ is smaller than a given tolerance (set to 10^{-3} in the tests, whose results are reported hereafter) for at least an assigned number of time steps (set to 10 in the tests, whose results are reported hereafter).

In each test, the same set of 100 polynomials, randomly generated as detailed above, has been used as input to either

(a) the method given in (4) with $p = 0.5$, with the stopping criterion detailed above and $\hat{\mathbf{x}}^0 = \mathbf{0}$ (implemented in `Mathematica` [62]);
(b) the `MATLAB` toolbox `SOSTOOLS`, interfaced with the solver `SeDuMi`, using the function `findbound` to find

a global lower bound for the input polynomial. It is worth noticing that the polynomials in the considered class have a global minima that generally have large negative values, of the order of -100^{2d} [7]. This ill-conditioning of the problem possibly leads to numerical problems for the interior-point algorithm used by SOSTOOLS. Therefore, in order to carry out a fair comparison, as suggested in [7], the function findbound has been applied to the polynomial $100^{-2d} f$ rather than to f, and the corresponding lower bound λ (which is such that $100^{-2d} f - \lambda$ is an SOS) has been multiplied by 100^{2d}.

A total of 25 tests (corresponding to 2500 polynomials) have been considered. Each entry of the sub-tables of Table I corresponds to the same pair $(n, d) \in \mathbb{N} \times \mathbb{N}$ (i.e., to the same 100 input polynomials generated randomly as detailed above) and reports the average execution time either of the method given in (4) with $p = 0.5$ (Sub-table Ia) or of the MATLAB toolbox SOSTOOLS (Sub-table Ib).

TABLE I: Average execution times (in seconds).

		$2d$				
		2	4	6	8	10
n	1	0.005955	0.008497	0.011883	0.024923	0.022796
	2	0.014629	0.031447	0.053511	0.083747	0.123574
	3	0.022149	0.071583	0.174957	0.425468	0.998607
	4	0.042049	0.138056	0.469544	1.87048	5.47279
	5	0.061493	0.234072	1.29932	6.78617	25.4617

(a) Method given in (4) with $p = 0.5$.

		$2d$				
		2	4	6	8	10
n	1	0.52157	0.59757	0.66152	0.68767	0.7788
	2	0.64789	0.80812	1.1724	1.7391	2.5219
	3	0.68604	1.2432	2.5541	5.2407	9.7755
	4	0.69486	1.873	5.5549	16.761	35.732
	5	0.77382	2.6632	11.133	48.493	216.19

(b) MATLAB toolbox SOSTOOLS.

In order to further corroborate the effectiveness of the proposed technique, Table II reports the average number of iterations K performed by the method given in (4) to determine a solution to the minimization problem (1), whereas Table III reports the percentage of tests in which $f(\hat{\mathbf{x}}^K) < \lambda$, where λ is the largest integer such that $f(\mathbf{x}) - \lambda$ is an SOS (which has been determined by using the MATLAB toolbox SOSTOOLS through the function findbound).

As shown by Sub-Table Ia, the proposed variation of the random coordinate descent method is able to determine efficiently (i.e., within reasonable computational times), an estimate of the solution to the minimization problem (1). As shown by Table II, the number of iterations that such an algorithm has to carry out to converge to a solution to the minimization problem (1) is increasing with the number of variables n and with the total degree d of the polynomial to be minimized. This is essentially due to the fact that, letting $G : \mathbb{R}^n \times \mathbb{S} \rightrightarrows \mathbb{R}^n$ be the set-valued mapping given in (5), for each $\mathbf{x} \in \mathbb{R}^n$, the measure of the set $\mathcal{S} \subset \mathbb{S}$ such that $G(\mathbf{x}, \mathbf{s}) \subset \mathcal{A} + \varepsilon \mathbb{B}^o$ for all $\mathbf{s} \in \mathcal{S}$ decreases as n and d increase. However, as shown by Table II, the number of iterations K to be carried out in order to determine a solution to the minimization problem (1) through the proposed algorithm grows mildly with n and d (the data reported in Table II are nicely fitted by the expression $K = 6.094\, e^{0.5192\, n + 0.2197\, d}$), thus highlighting the fact that the proposed technique can be employed to solve minimization problems involving a large number of variables and with the objective function being a polynomial of large total degree (note that, for $n = 5$ and $2d = 10$, each polynomial f has been obtained by the weighted sum of 2007 monomials).

TABLE II: Average number of iterations to determine a solution to the minimization problem (1).

		$2d$				
		2	4	6	8	10
n	1	10	10	10	10	10
	2	17.9	29.1	33.7	36.2	37
	3	29.9	58.3	70.2	79.4	89.8
	4	53.9	93.4	108	134	144
	5	74.8	128	161	201	235

TABLE III: Percentage of tests in which $f(\hat{\mathbf{x}}^K) < \lambda$.

		$2d$				
		2	4	6	8	10
n	1	29	100	99	92	91
	2	41	99	100	100	100
	3	45	97	100	100	100
	4	46	99	100	100	100
	5	62	98	100	100	100

In all the tests that have been carried out, we obtained that either $f(\hat{\mathbf{x}}^K) < \lambda$ or $|f(\hat{\mathbf{x}}^K) - \lambda| \leqslant \max\{|f(\hat{\mathbf{x}}^K)|, \lambda|\}$, showing that the proposed minimization algorithm is able to determine a good approximate of the solution to the minimization problem (1). By comparing Sub-tables Ia and Ib, it can be noticed that the execution time of the proposed algorithm is smaller than the one of SOS-based methods, especially if both the number of variables n and the total degree $2d$ of the polynomial to be minimized have large values. Furthermore, as shown by Table III, if n and $2d$ have large values (namely, $n \geqslant 3$ and $2d \geqslant 6$), the proposed minimization technique provides a better estimate of the solution to the minimization problem (1).

REFERENCES

[1] 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.

[2] W. Karush, "Minima of functions of several variables with inequalities as side conditions," Master's thesis, University of Chicago, 1939.

[3] H. W. Kuhn and A. W. Tucker, "Nonlinear programming," in *Proc. Berkeley Symp. Math. Stat. Probab.*, (Berkeley, CA), Univ. California Press, 1951.

[4] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *J. ACM*, vol. 42, no. 6, pp. 1115–1145, 1995.

[5] L. Qi, "Extrema of a real polynomial," *J. Global Optim.*, vol. 30, no. 4, pp. 405–433, 2004.

[6] Z. Li, S. He, and S. Zhang, *Approximation methods for polynomial optimization: Models, Algorithms, and Applications*. New York: Springer Sci. Bus. Media, 2012.

[7] 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.

[8] D. A. Cox, J. Little, and D. O'Shea, *Using algebraic geometry*. New York: Springer Sci. Bus. Media, 2006.

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

[10] L. Menini, C. Possieri, and A. Tornambe, "Dead-beat regulation of mechanical juggling systems," *Asian J. Control*, vol. 20, no. 2, pp. 1–11, 2018.

[11] K. Hägglöf, P. O. Lindberg, and L. Svensson, "Computing global minima to polynomial optimization problems using Gröbner bases," *J. Global Optim.*, vol. 7, no. 2, pp. 115–125, 1995.

[12] I. M. Gelfand, M. Kapranov, and A. Zelevinsky, *Discriminants, resultants, and multidimensional determinants*. New York: Springer Sci. Bus. Media, 2008.

[13] F. Rouillier, "Solving zero-dimensional systems through the rational univariate representation," *Appl. Algebra Eng. Commun. Comput.*, vol. 9, no. 5, pp. 433–461, 1999.

[14] T.-Y. Li, "Numerical solution of multivariate polynomial systems by homotopy continuation methods," *Acta Numer.*, vol. 6, pp. 399–436, 1997.

[15] J. Verschelde, "Polynomial homotopies for dense, sparse and determinantal systems." arXiv:math/9907060, 1999.

[16] B. Hanzon and D. Jibetean, "Global minimization of a multivariate polynomial using matrix methods," *J. Global Optim.*, vol. 27, no. 1, pp. 1–23, 2003.

[17] N. Z. Shor, "Class of global minimum bounds of polynomial functions," *Cybern. Syst. Anal.*, vol. 23, no. 6, pp. 731–734, 1987.

[18] N. Z. Shor and P. I. Stetsyuk, "The use of a modification of the r-algorithm for finding the global minimum of polynomial functions," *Cybern. Syst. Anal.*, vol. 33, p. 482497, 1997.

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

[20] Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

[21] H. Wolkowicz, R. Saigal, and L. Vandenberghe, *Handbook of semidefinite programming: theory, algorithms, and applications*, vol. 27. New York: Springer Sci. Bus. Media, 2012.

[22] M. Putinar, "Positive polynomials on compact semi-algebraic sets," *Indiana Univ. Math. J.*, vol. 42, no. 3, pp. 969–984, 1993.

[23] T. Jacobi and A. Prestel, "Distinguished representations of strictly positive polynomials," *J. Reihe Angew. Math.*, vol. 532, p. 223, 2001.

[24] Y. Nesterov, "Semidefinite relaxation and nonconvex quadratic optimization," *Optim. Meth. Softw.*, vol. 9, no. 1-3, pp. 141–160, 1998.

[25] Y. Ye, "Approximating quadratic programming with bound and quadratic constraints," *Math. Grog.*, vol. 84, no. 2, pp. 219–226, 1999.

[26] A. Nemirovski, C. Roos, and T. Terlaky, "On maximization of quadratic form over intersection of ellipsoids with common center," *Math. Program.*, vol. 86, no. 3, pp. 463–473, 1999.

[27] M. Charikar and A. Wirth, "Maximizing quadratic programs: Extending Grothendieck's inequality," in *Proc. 45th Ann. IEEE Symp. Found. Comp. Sci.*, pp. 54–60, IEEE, 2004.

[28] N. Alon and A. Naor, "Approximating the cut-norm via Grothendieck's inequality," *SIAM J. Comp.*, vol. 35, no. 4, pp. 787–803, 2006.

[29] S. He, Z. Li, and S. Zhang, "General constrained polynomial optimization: an approximation approach," Tech. Rep. SEEM2009-05, Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong, 2009.

[30] S. He, Z. Li, and S. Zhang, "Approximation algorithms for discrete polynomial optimization," *J. Oper. Res. Soc. China*, vol. 1, no. 1, pp. 3–36, 2013.

[31] X. Zhang, C. Ling, and L. Qi, "Semidefinite relaxation bounds for bi-quadratic optimization problems with quadratic constraints," *J. Global Optim.*, vol. 49, no. 2, pp. 293–311, 2011.

[32] H. Rosenbrock, "An automatic method for finding the greatest or least value of a function," *The Computer Journal*, vol. 3, no. 3, pp. 175–184, 1960.

[33] G. Zoutendijk, "Nonlinear programming, computational methods," in *Integer and nonlinear programming* (J. Abadie, ed.), pp. 37–86, North-Holland, 1970.

[34] M. Sassano, D. Carnevale, and A. Astolfi, "Extremum seeking-like observer for nonlinear systems," *IFAC Proc. Vol.*, vol. 44, no. 1, pp. 1849–1854, 2011.

[35] N. Zadeh, "A note on the cyclic coordinate ascent method," *Management Sci.*, vol. 16, no. 9, pp. 642–644, 1970.

[36] D. G. Luenberger and Y. Ye, *Linear and nonlinear programming*. Springer, 1984.

[37] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: numerical methods*, vol. 23. Englewood Cliffs, NJ: Prentice hall, 1989.

[38] L. Grippo and M. Sciandrone, "Globally convergent block-coordinate techniques for unconstrained optimization," *Optim. Methods Software*, vol. 10, no. 4, pp. 587–637, 1999.

[39] A. Auslender, "Asymptotic properties of the Fenchel dual functional and applications to decomposition problems," *J. Optim. Theory Appl.*, vol. 73, no. 3, pp. 427–449, 1992.

[40] Z.-Q. Luo and P. Tseng, "On the convergence of the coordinate descent method for convex differentiable minimization," *J. Optim. Theory Appl.*, vol. 72, no. 1, pp. 7–35, 1992.

[41] M. Patriksson, "Decomposition methods for differentiable optimization problems over cartesian product sets," *Comput. Optim. Appl.*, vol. 9, no. 1, pp. 5–42, 1998.

[42] L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear Gauss–Seidel method under convex constraints," *Oper. Res. Lett.*, vol. 26, no. 3, pp. 127–136, 2000.

[43] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *J. Optim. Theory Appl.*, vol. 109, no. 3, pp. 475–494, 2001.

[44] M. J. Powell, "On search directions for minimization algorithms," *Math. Program.*, vol. 4, no. 1, pp. 193–201, 1973.

[45] Y. Nesterov, "Efficiency of coordinate descent methods on huge-scale optimization problems," *SIAM J. Optim.*, vol. 22, no. 2, pp. 341–362, 2012.

[46] I. Necoara and D. Clipici, "Efficient parallel coordinate descent algorithm for convex optimization problems with separable constraints: Application to distributed mpc," *J. Process Control*, vol. 23, no. 3, pp. 243 – 253, 2013.

[47] P. Richtárik and M. Takáč, "Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function," *Math. Program.*, vol. 144, no. 1-2, pp. 1–38, 2014.

[48] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*. New York: Springer Sci. Bus. Media, 2009.

[49] A. Subbaraman and A. R. Teel, "A converse Lyapunov theorem for strong global recurrence," *Automatica*, vol. 49, no. 10, pp. 2963–2974, 2013.

[50] 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.

[51] 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.

[52] H. K. Khalil, *Noninear Systems*. Prentice-Hall, New Jersey, 1996.

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

[54] 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.

[55] 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.

[56] 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.

[57] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge Univ. Press, 2004.
[58] W. Rheinboldt, C. Mesztenyi, and J. Fitzgerald, "On the evaluation of multivariate polynomials and their derivatives," *BIT Numer. Math.*, vol. 17, no. 4, pp. 437–450, 1977.
[59] 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.
[60] A. Papachristodoulou, J. Anderson, G. Valmorbida, S. Prajna, P. Seiler, and P. Parrilo, "SOSTOOLS version 3.00 sum of squares optimization toolbox for MATLAB." arXiv:1310.4716, 2013.
[61] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Methods Software*, vol. 11, no. 1-4, pp. 625–653, 1999.
[62] Wolfram Research, Inc., "Mathematica, Version 11.2." Champaign, IL, 2017.