

Solving non-monotone equilibrium problems via a DIRECT-type approach

Stefano Lucidi · Mauro Passacantando · Francesco
Rinaldi

Received: date / Accepted: date

Abstract A global optimization approach for solving non-monotone equilibrium problems (EPs) is proposed. The class of (regularized) gap functions is used to reformulate any EP as a constrained global optimization program and some bounds on the Lipschitz constant of such functions are provided. The proposed global optimization approach is a combination of an improved version of the DIRECT algorithm, which exploits local bounds of the Lipschitz constant of the objective function, with local minimizations. Unlike most existing solution methods for EPs, no monotonicity-type condition is assumed in this paper. Preliminary numerical results on several classes of EPs show the effectiveness of the approach.

Keywords Equilibrium problem · Gap function · Global optimization · DIRECT algorithm

Mathematics Subject Classification (2010) 90C26 · 90C33 · 90C56

1 Introduction

Given a bifunction $f : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ and a closed convex set $\mathcal{C} \subseteq \mathbb{R}^n$, we consider the following equilibrium problem:

$$\text{find } x^* \in \mathcal{C} \text{ such that } f(x^*, y) \geq 0, \text{ for any } y \in \mathcal{C}. \quad (\text{EP})$$

This framework is a general mathematical model which includes several problems such as scalar and vector optimization, variational inequality (VI), complementarity, saddle point, Nash equilibrium problems in non-cooperative games and inverse optimization [3, 6].

Several classes of iterative methods to solve EPs have been proposed in the literature: fixed point approaches [32, 35, 36], extragradient methods [17, 25, 26, 40], descent algorithms [8, 11, 22, 31], proximal point methods [7, 21, 34]. All these approaches need, directly or indirectly, some

This paper has been published in Journal of Global Optimization, vol. 83 (4), 2022, pp. 699–725, see: <https://link.springer.com/article/10.1007/s10898-021-01121-z>.

S. Lucidi

Department of Computer, Control and Management Engineering “Antonio Ruberti”, Sapienza University of Rome
via Ariosto, 25, 00185 Roma, Italy
E-mail: lucidi@diag.uniroma1.it

M. Passacantando

Department of Computer Science, University of Pisa
Largo B. Pontecorvo 3, 56127 Pisa, Italy
E-mail: mauro.passacantando@unipi.it

F. Rinaldi

Department of Mathematics “Tullio Levi-Civita”, University of Padova
via Trieste, 63, 35121 Padova, Italy
E-mail: rinaldi@math.unipd.it

monotonicity-type assumption on the bifunction f (e.g. strong or weak monotonicity, pseudo-monotonicity, ∇ -monotonicity, etc.) in order to guarantee the convergence to a solution of (EP). On the other hand, it is well known that, without any need of monotonicity-type assumptions on f , (EP) can be reformulated as an equivalent global optimization problem via merit functions [38]. This fact suggests to use global optimization approaches to solve non-monotone EPs. Global optimization techniques have been considered in the literature only for two special classes of EPs: linear complementarity problems [1,39] and VI problems (a branch and bound method was proposed in [29] and a meta-heuristic algorithm in [30]).

In this paper, we propose a DIRECT-type global optimization approach for solving general EPs, without assuming any monotonicity-type condition on f . In particular, we first reformulate (EP) as a global optimization problem via the well-known gap functions. We analyze the Lipschitz continuity of gap functions and give simple estimates of the Lipschitz constant for some special classes of EPs. Then, we combine the improved version of the DIRECT algorithm developed in [12], which exploits local bounds of the Lipschitz constant of the objective function with local searches to find a global minimum point of the gap function, i.e., a solution of (EP). Finally, we show the effectiveness of our approach with some preliminary numerical experiments on instances coming from the literature and randomly generated instances.

We remark that a possible alternative approach to solve non-monotone EPs relies on the use of simplicial algorithms for fixed point problems. In fact, it is well known that (EP) can be equivalently reformulated as a fixed point problem and the class of simplicial algorithms for fixed point problems does not require any monotonicity assumption to get convergence. In the seminal paper [41], Scarf proposed a constructive proof of the Brouwer's fixed point theorem on the unit simplex. This algorithm, based on a simplicial decomposition of the unit simplex and on integer labeling techniques, generates a sequence of adjacent simplices and stops after a finite number of iterations with a completely labeled simplex yielding an approximate fixed point (see also [23, 42, 43]). Later, improved simplicial algorithms, based on homotopy functions that deform trivial fixed point problems into the original one, allowed for successively finer approximations of fixed points (see, e.g., [9, 10, 14, 15, 24, 44, 45]).

The rest of the paper is organized as follows. In Section 2, we recall some preliminary notions for (EP) and the main properties of gap functions for (EP). In Section 3, we provide some general results on the Lipschitz continuity of gap functions and give explicit bounds of the Lipschitz constant for four classes of problems: affine VIs, VIs with trigonometric terms, affine EPs and EPs with trigonometric terms. Section 4 presents the DIRECT-type global optimization approach and recalls the convergence properties of both the standard version of the DIRECT algorithm and its improved version proposed in [12]. Section 5 reports the results of some preliminary numerical tests and shows that the improved version of DIRECT is more efficient than its standard version on most of the considered instances.

Section 6 shows a numerical comparison between the proposed DIRECT-type approach and a widely used local solver for mixed complementarity problems and variational inequalities. Section 7 shows a numerical comparison between the proposed DIRECT-type approach and two simplicial algorithms for fixed point problems. Conclusions are finally drawn in Section 8.

Throughout the paper we will assume that the feasible set \mathcal{C} is compact, the bifunction f is continuous, $f(x, \cdot)$ is convex and $f(x, x) = 0$ for any $x \in \mathcal{C}$. It is well known that under these assumptions the existence of at least one solution of (EP) is guaranteed (see, e.g., [16]).

2 Preliminary background

First, we briefly recall how some optimization related mathematical models can be formulated in the general format (EP) (see [3,4] for more details).

- *Scalar and vector optimization problems*: finding a global minimizer of a function $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ over a closed set $\mathcal{C} \subseteq \mathbb{R}^n$ is equivalent to solve (EP) with $f(x, y) = \psi(y) - \psi(x)$.
Given m scalar functions $\psi_i : \mathbb{R}^n \rightarrow \mathbb{R}$, a weak Pareto global minimum of the vector function

- $\psi = (\psi_1, \dots, \psi_m)$ over a closed set $\mathcal{C} \subseteq \mathbb{R}^n$ is any $x^* \in \mathcal{C}$ such that there is no $y \in \mathcal{C}$ such that $\psi_i(y) < \psi_i(x^*)$ for any $i = 1, \dots, m$. Finding a weak Pareto global minimum is equivalent to solve (EP) with $f(x, y) = \max_{i=1, \dots, m} [\psi_i(y) - \psi_i(x)]$.
- *Variational inequality and complementarity problems*: given a closed convex set $\mathcal{C} \subseteq \mathbb{R}^n$ and a map $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$, the variational inequality problem consists in finding $x^* \in \mathcal{C}$ such that $\langle F(x^*), y - x^* \rangle \geq 0$ holds for any $y \in \mathcal{C}$. When \mathcal{C} is a closed convex cone of \mathbb{R}^n , the variational inequality problem reduces to a complementarity problem. Solving both problems is equivalent to solve (EP) with $f(x, y) = \langle F(x), y - x \rangle$.
 - *Fixed point problems*: given a closed set $\mathcal{C} \subseteq \mathbb{R}^n$, a fixed point of a map $F : \mathcal{C} \rightarrow \mathcal{C}$ is any $x^* \in \mathcal{C}$ such that $x^* = F(x^*)$. Finding a fixed point of F is equivalent to solve (EP) with $f(x, y) = \langle x - F(x), y - x \rangle$.
 - *Saddle point problems*: given two closed sets $C_1 \subseteq \mathbb{R}^{n_1}$ and $C_2 \subseteq \mathbb{R}^{n_2}$, a saddle point of a function $F : C_1 \times C_2 \rightarrow \mathbb{R}$ is any $x^* = (x_1^*, x_2^*) \in C_1 \times C_2$ such that $F(x_1^*, y_2) \leq F(x_1^*, x_2^*) \leq F(y_1, x_2^*)$ holds for any $y = (y_1, y_2) \in C_1 \times C_2$. Finding a saddle point of F is equivalent to solve (EP) with $\mathcal{C} = C_1 \times C_2$ and $f((x_1, x_2), (y_1, y_2)) = F(y_1, x_2) - F(x_1, y_2)$.
 - *Nash equilibrium problems*: in a noncooperative game with N players, each player i has a set of strategies $C_i \subseteq \mathbb{R}^{n_i}$ and aims at minimizing a cost function $c_i : \mathcal{C} \rightarrow \mathbb{R}$ where $\mathcal{C} = C_1 \times \dots \times C_N$. A Nash equilibrium is any $x^* \in \mathcal{C}$ such that $f_i(x^*) \leq f_i(x^*(y_i))$ holds for any $y_i \in C_i$ and $i = 1, \dots, N$, where $x^*(y_i)$ denotes the vector obtained from x^* by replacing x_i^* with y_i . Finding a Nash equilibrium is equivalent to solve (EP) with f equal to the so-called Nikaido-Isoda bifunction [37], i.e., $f(x, y) = \sum_{i=1}^N [f_i(x(y_i)) - f_i(x)]$.
 - *Inverse optimization problems*: given a closed set $\mathcal{C} \subseteq \mathbb{R}^n$, m functions $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ and p functions $g_j : \mathbb{R}^n \rightarrow \mathbb{R}$, the inverse optimization problem asks to determine a parameter $\lambda^* \in \mathbb{R}_+^m$ such that at least one optimal solution x^* of the problem $\min \{ \sum_{i=1}^m \lambda_i^* f_i(x) : x \in \mathcal{C} \}$ satisfies the constraints $g_j(x^*) \leq 0$ for all $j = 1, \dots, p$. This problem is equivalent to a Nash equilibrium problem with three players: the first player controls the x variables and aims at solving $\min \{ \sum_{i=1}^m \lambda_i f_i(x) : x \in \mathcal{C} \}$, the second player controls the auxiliary variables y and aims at solving $\max \{ \sum_{j=1}^p g_j(x) y_j : y \geq 0 \}$, while the third player simply chooses a vector $\lambda \in \mathbb{R}_+^m$ (or minimizes a constant objective function over \mathbb{R}_+^m). Therefore, also this inverse optimization problem can be formulated in the (EP) format.

We now recall some monotonicity-type definitions for a bifunction $f : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$.

Definition 1 Given a closed convex set $\mathcal{C} \subseteq \mathbb{R}^n$, f is called *monotone* on \mathcal{C} if

$$f(x, y) + f(y, x) \leq 0, \quad \forall x, y \in \mathcal{C};$$

f is called *strongly monotone* on \mathcal{C} if there exists $\tau > 0$ such that $f(x, y) + f(y, x) \leq -\tau \|y - x\|^2$ holds for any $x, y \in \mathcal{C}$;

f is called *pseudomonotone* on \mathcal{C} if the implication

$$f(x, y) \geq 0 \implies f(y, x) \leq 0$$

holds for any $x, y \in \mathcal{C}$;

f is called *weakly monotone* on \mathcal{C} if there exists $\tau > 0$ such that $f(x, y) + f(y, x) \leq \tau \|y - x\|^2$ holds for any $x, y \in \mathcal{C}$;

f is called *∇ -monotone* on \mathcal{C} if $\langle \nabla_x f(x, y) + \nabla_y f(x, y), y - x \rangle \geq 0$ holds for any $x, y \in \mathcal{C}$;

f is called *strictly ∇ -monotone* on \mathcal{C} if $\langle \nabla_x f(x, y) + \nabla_y f(x, y), y - x \rangle > 0$ holds for any $x, y \in \mathcal{C}$ with $x \neq y$;

f is called *strongly ∇ -monotone* on \mathcal{C} if there exists $\tau > 0$ such that $\langle \nabla_x f(x, y) + \nabla_y f(x, y), y - x \rangle \geq \tau \|y - x\|^2$ holds for any $x, y \in \mathcal{C}$.

The monotonicity is a key assumption to guarantee the convergence of extragradient, proximal point and Tikhonov-Browder regularization methods; the strong monotonicity is exploited in those

algorithms based on the reformulation of (EP) as a fixed point problem; pseudomonotonicity is mainly used in extragradient algorithms as well as in fixed point, combined relaxation methods, convex feasibility and proximal point methods; weak monotonicity has been used in proximal point algorithms; both strict and strong ∇ -monotonicity have been widely exploited to devise specific descent methods for (EP). A detailed description of the above monotonicity conditions and a comprehensive analysis of their relationships can be found in [5].

Merit functions allow reformulating (EP) as a global optimization problem, whose optimal value is known *a priori*. Several classes of merit functions for EPs have been introduced in the literature in the last two decades [38]. In this paper, we focus on the class of (regularized) gap functions that is an extension to EPs of the class introduced by Fukushima for VIs [18].

Theorem 1 [31] *For any $\alpha \geq 0$ the gap function*

$$\varphi_\alpha(x) := \max_{y \in \mathcal{C}} \left[-f(x, y) - \frac{\alpha}{2} \|y - x\|^2 \right] \quad (1)$$

has the following properties:

- a) $\varphi_\alpha(x) \geq 0$ for any $x \in \mathcal{C}$;
- b) x^* solves (EP) if and only if $x^* \in \mathcal{C}$ and $\varphi_\alpha(x^*) = 0$;
- c) If $\alpha > 0$ and f is continuously differentiable on $\mathbb{R}^n \times \mathbb{R}^n$, then φ_α is continuously differentiable on \mathbb{R}^n with

$$\nabla \varphi_\alpha(x) = -\nabla_1 f(x, y_\alpha(x)) - \alpha(x - y_\alpha(x)), \quad (2)$$

where $\nabla_1 f(x, y)$ denotes the gradient of $f(\cdot, y)$ at x and $y_\alpha(x)$ is the unique maximizer of problem in (1).

Therefore, the solutions of (EP) coincide with the global minimum points of the optimization problem

$$\begin{aligned} \min \varphi_\alpha(x) \\ x \in \mathcal{C}, \end{aligned} \quad (3)$$

whose global minimum value is zero. We remark that evaluating the gap function φ_α at some point x consists in maximizing a concave (when $\alpha = 0$) or strongly concave (when $\alpha > 0$) function over the set \mathcal{C} . Moreover, the regularization term $\|y - x\|^2$ can be replaced by a more general bifunction satisfying suitable conditions (see [31]).

Several descent methods based on the gap function φ_α have been developed in the literature for solving EPs (see, e.g., [8, 11, 31]). However, their convergence to a solution of (EP) is guaranteed provided that some monotonicity-type condition on the bifunction f is assumed. In this paper, we propose a global optimization approach for solving problem (3) that is not based on any monotonicity-type condition on f . More specifically, we consider a DIRECT-type method (see, e.g., [20]) with local searches. DIRECT (DIvide RECTangle) is a partitioning strategy that samples points in the domain and uses only objective function evaluations to decide what to do next. The boosted version we use here, called \bar{L} -DIRECT and first proposed in [12], exploits overestimates of the Lipschitz constant related to the objective function to improve the way the subsets to be further partitioned are selected. As we will see in the next section, this choice is well-suited to our problem. Indeed, when our problem has some specific structure, an overestimate of the Lipschitz constant for the function φ_α can be easily calculated.

In the rest of the paper, we will consider the class of EPs where the bifunction

$$f(x, y) = \langle F(x, y), y - x \rangle$$

for some map $F : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$. This class of EPs includes two important particular cases: (i) VIs, where the map F only depends on the variable x and (ii) affine EPs, where $F(x, y) = Px + Qy + r$ for some $P, Q \in \mathbb{R}^{n \times n}$ and $r \in \mathbb{R}^n$.

3 Lipschitz continuity of gap functions

In this section, we provide some general results on the Lipschitz continuity of gap function φ_α and show some simple estimates of its Lipschitz constant for four special classes of EPs. The knowledge of the Lipschitz constant of φ_α will be exploited by the global optimization approach described in Section 4 for solving problem (3).

Theorem 2 *Suppose that $B \subseteq \mathbb{R}^n$ is compact, F is continuous on $\mathbb{R}^n \times \mathbb{R}^n$ and $F(\cdot, y)$ is Lipschitz continuous on B , uniformly with respect to y , with constant L_F . Then, for any $\alpha \geq 0$ the function φ_α is Lipschitz continuous on B with constant*

$$L_1 + L_2 L_F + \alpha L_2,$$

where

$$L_1 = \max_{x \in B, y \in C} \|F(x, y)\|, \quad L_2 = \max_{x \in B, y \in C} \|x - y\|. \quad (4)$$

Proof If $x, y \in B$, then the following chain of equalities and inequalities holds:

$$\begin{aligned} \varphi_\alpha(x) - \varphi_\alpha(y) &= \max_{z \in C} \left[\langle F(x, z), x - z \rangle - \frac{\alpha}{2} \|x - z\|^2 \right] - \max_{z \in C} \left[\langle F(y, z), y - z \rangle - \frac{\alpha}{2} \|y - z\|^2 \right] \\ &\leq \max_{z \in C} \left[\langle F(x, z), x - z \rangle - \langle F(y, z), y - z \rangle - \frac{\alpha}{2} \|x - z\|^2 + \frac{\alpha}{2} \|y - z\|^2 \right] \\ &= \max_{z \in C} \left[\langle F(x, z) - F(y, z), x - z \rangle + \langle F(y, z), x - y \rangle + \frac{\alpha}{2} \langle y - x, y - z + x - z \rangle \right] \\ &\leq \max_{z \in C} \left[\|F(x, z) - F(y, z)\| \|x - z\| + \|F(y, z)\| \|x - y\| \right. \\ &\quad \left. + \frac{\alpha}{2} \|y - x\| (\|y - z\| + \|x - z\|) \right] \\ &\leq L_F \|x - y\| (\max_{z \in C} \|x - z\|) + L_1 \|x - y\| \\ &\quad + \frac{\alpha}{2} \|y - x\| \left[\max_{z \in C} \|y - z\| + \max_{z \in C} \|x - z\| \right] \\ &\leq (L_1 + L_2 L_F + \alpha L_2) \|x - y\|, \end{aligned}$$

where the second inequality follows from the Cauchy-Schwarz inequality, the third one from the Lipschitz continuity of F and the last one from the definition of L_2 . \square

Remark 1 Theorem 2 is a generalization of Lemma 2.1 proved in [29], which provides an estimate of the Lipschitz constant of the gap function φ_0 for a VI with Lipschitz continuous operator. In fact, when (EP) reduces to a VI, the regularization parameter $\alpha = 0$ and the set $B = C$, then the value of the Lipschitz constant given in Theorem 2 coincides with that given in [29, Lemma 2.1].

A further estimate of the Lipschitz constant of φ_α , with $\alpha > 0$, can be obtained provided that the map F is smooth.

Theorem 3 *Suppose that $B \subseteq \mathbb{R}^n$ is a convex compact set and F is continuously differentiable on $\mathbb{R}^n \times \mathbb{R}^n$. Then, for any $\alpha > 0$ the function φ_α is Lipschitz continuous on B with constant*

$$L_1 + L_2 L_3(\alpha),$$

where L_1 and L_2 are defined in (4) and

$$L_3(\alpha) = \max_{x \in B, y \in C} \|\alpha I - \nabla_1 F(x, y)\|, \quad (5)$$

where $\nabla_1 F(x, y)$ denotes the Jacobian matrix of $F(\cdot, y)$ at x .

Proof Theorem 1 guarantees that φ_α is continuously differentiable on \mathbb{R}^n with

$$\nabla\varphi_\alpha(x) = F(x, y_\alpha(x)) + [\alpha I - \nabla_1 F(x, y_\alpha(x))^T](y_\alpha(x) - x), \quad x \in \mathbb{R}^n,$$

where

$$y_\alpha(x) = \arg \max_{y \in \mathcal{C}} \left[\langle F(x, y), x - y \rangle - \frac{\alpha}{2} \|y - x\|^2 \right].$$

Let $u, v \in B$. The mean value theorem guarantees that there exists $\xi \in (0, 1)$ such that

$$\varphi_\alpha(u) - \varphi_\alpha(v) = \langle \nabla\varphi_\alpha(z), u - v \rangle,$$

where $z := \xi u + (1 - \xi)v \in B$. Therefore, we get

$$\begin{aligned} |\varphi_\alpha(u) - \varphi_\alpha(v)| &\leq \|\nabla\varphi_\alpha(z)\| \|u - v\| \\ &\leq [\|F(z, y_\alpha(z))\| + \|\alpha I - \nabla_1 F(z, y_\alpha(z))^T\| \|y_\alpha(z) - z\|] \|u - v\| \\ &\leq [L_1 + L_2 L_3(\alpha)] \|u - v\|. \end{aligned}$$

□

In the special case of a VI defined by a smooth map, a third estimate of the Lipschitz constant of φ_α can be proved.

Theorem 4 *Suppose that (EP) is a VI, i.e., $f(x, y) = \langle F(x), y - x \rangle$ for some continuously differentiable map $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$. If $B \subseteq \mathbb{R}^n$ is a convex compact set such that $B \subseteq \mathcal{C}$, then for any $\alpha > 0$ the function φ_α is Lipschitz continuous on B with constant*

$$L_1 + \alpha^{-1} L_1 L_3(\alpha),$$

where L_1 and $L_3(\alpha)$, defined in (4) and (5) respectively, in this special case are equal to

$$L_1 = \max_{x \in B} \|F(x)\|, \quad L_3(\alpha) = \max_{x \in B} \|\alpha I - \nabla F(x)\|.$$

Proof Theorem 1 guarantees that φ_α is continuously differentiable and

$$\nabla\varphi_\alpha(x) = F(x) + [\alpha I - \nabla F(x)^T](y_\alpha(x) - x), \quad x \in \mathbb{R}^n,$$

with

$$y_\alpha(x) = P_{\mathcal{C}}(x - \alpha^{-1}F(x)),$$

where $P_{\mathcal{C}}$ denotes the Euclidean projection on the set \mathcal{C} . If $u, v \in B$, then the mean value theorem implies

$$\varphi_\alpha(u) - \varphi_\alpha(v) = \langle \nabla\varphi_\alpha(z), u - v \rangle,$$

where $z := \xi u + (1 - \xi)v$ for some $\xi \in (0, 1)$. Therefore, we get

$$\begin{aligned} |\varphi_\alpha(u) - \varphi_\alpha(v)| &\leq \|\nabla\varphi_\alpha(z)\| \|u - v\| \\ &\leq [\|F(z)\| + \|\alpha I - \nabla F(z)^T\| \|y_\alpha(z) - z\|] \|u - v\| \\ &= [\|F(z)\| + \|\alpha I - \nabla F(z)^T\| \|P_{\mathcal{C}}(z - \alpha^{-1}F(z)) - P_{\mathcal{C}}(z)\|] \|u - v\| \\ &\leq [\|F(z)\| + \|\alpha I - \nabla F(z)^T\| \|z - \alpha^{-1}F(z) - z\|] \|u - v\| \\ &\leq [L_1 + \alpha^{-1} L_1 L_3(\alpha)] \|u - v\|, \end{aligned}$$

where the third inequality holds since the projection map $P_{\mathcal{C}}$ is nonexpansive, i.e., $\|P_{\mathcal{C}}(x) - P_{\mathcal{C}}(y)\| \leq \|x - y\|$ holds for any $x, y \in \mathbb{R}^n$. □

In the rest of this section we analyze the Lipschitz constant of φ_α for some special classes of EPs.

3.1 Affine VIs defined on a box

Suppose that (EP) is a VI defined by an affine operator $F(x) = Px + r$, for some $P \in \mathbb{R}^{n \times n}$ and $r \in \mathbb{R}^n$, over a box $\mathcal{C} = [l, u]$, where $l, u \in \mathbb{R}^n$. Consider a box $B = [a, b]$, where $a, b \in \mathbb{R}^n$, such that $B \subseteq \mathcal{C}$, i.e., $l \leq a \leq b \leq u$. Then, Theorem 2, Theorem 3 and Theorem 4 guarantee that φ_0 is Lipschitz continuous on B with constant

$$L_1 + L_2 L_F, \quad (6)$$

while, for any $\alpha > 0$, φ_α is Lipschitz continuous on B with constant

$$\min \{L_1 + L_2 L_F + \alpha L_2, \quad L_1 + L_2 L_3(\alpha), \quad L_1 + \alpha^{-1} L_1 L_3(\alpha)\}. \quad (7)$$

We now show that the exact values (or upper bound) of the constants involved in the above formulas can be easily computed.

Estimate of L_1 . The exact value of L_1 is

$$L_1 = \max_{x \in B} \|Px + r\| = \max_{x \in \text{vert}(B)} \|Px + r\|,$$

where $\text{vert}(B)$ denotes the set of vertices of B . Such a evaluation can be computationally expensive since the vertices of B are exponentially many with respect to the number of variables. However, the following upper bounds for L_1 can be easily computed. If we denote by P^+ the Moore-Penrose pseudoinverse matrix of P , then we get

$$\begin{aligned} L_1 &= \max_{x \in B} \|Px + r\| \\ &= \max_{a \leq x \leq b} \|P(x + P^+r) + (I - PP^+)r\| \\ &\leq \|(I - PP^+)r\| + \max_{a \leq x \leq b} \|P(x + P^+r)\| \\ &\leq \|(I - PP^+)r\| + \|P\| \max_{a \leq x \leq b} \|x + P^+r\| \\ &= \|(I - PP^+)r\| + \|P\| \|c(a, b)\| \\ &:= L'_1, \end{aligned}$$

where the i -th component of the vector $c(a, b)$ is defined as $c_i(a, b) = \max\{|(a + P^+r)_i|, |(b + P^+r)_i|\}$. Moreover, the following simple upper bounds hold:

$$L_1 = \max_{a \leq x \leq b} \|P(x - a) + Pa + r\| \leq \|Pa + r\| + \|P\| \|b - a\| := L''_1,$$

$$L_1 = \max_{a \leq x \leq b} \|P(x - b) + Pb + r\| \leq \|Pb + r\| + \|P\| \|b - a\| := L'''_1.$$

Therefore, we have

$$L_1 \leq \tilde{L}_1(P, r, a, b), \quad \text{where} \quad \tilde{L}_1(P, r, a, b) := \min\{L'_1, L''_1, L'''_1\}. \quad (8)$$

Remark 2 In [29] the following upper bound for L_1 is given:

$$L_1 \leq \|P\| \|c(a, b)\|.$$

We remark that this inequality is not true in general, as the following counterexample shows. Let $n = 2$,

$$P = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad r = \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad \text{with } v \neq 0, \quad a = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Then, it is easy to check that $\|P\| = \sqrt{2}$ holds and the pseudoinverse of P is

$$P^+ = \begin{pmatrix} 1/2 & 0 \\ 1/2 & 0 \end{pmatrix},$$

hence $c_i(a, b) = \max\{|a_i|, |b_i|\} = 1$ for $i = 1, 2$. Therefore, $\|P\| \|c(a, b)\| = 2$. On the other hand,

$$L_1 = \max_{x \in B} \|Px + r\| = \max_{0 \leq x \leq 1} \|(x_1 + x_2, v)\| = \|(2, v)\| = \sqrt{4 + v^2} > 2 = \|P\| \|c(a, b)\|.$$

Estimate of L_2 . The exact value of $L_2 = \max_{x \in B, y \in C} \|x - y\|$ can be computed by solving n independent optimization problems of the form

$$\max_{\substack{a_i \leq x_i \leq b_i \\ l_i \leq y_i \leq u_i}} (x_i - y_i)^2 = \max\{(u_i - a_i)^2, (l_i - b_i)^2\},$$

for $i = 1, \dots, n$. Therefore, we have

$$L_2 = \sqrt{\sum_{i=1}^n \max\{(u_i - a_i)^2, (l_i - b_i)^2\}}. \quad (9)$$

Estimates of $L_3(\alpha)$ and L_F . It is easy to check that $L_3(\alpha) = \|\alpha I - P\|$ and $L_F = \|P\|$.

3.2 Nonlinear VIs with trigonometric terms defined on a box

Suppose that (EP) is a VI defined over a box $\mathcal{C} = [l, u]$, with an operator which is the sum of an affine map and a trigonometric map, i.e.,

$$F(x) = Px + r + T(x),$$

where $T_i(x) = w_i \sin(v_i x_i)$, for $i = 1, \dots, n$, $P \in \mathbb{R}^{n \times n}$ and $r, v, w \in \mathbb{R}^n$ with $v, w > 0$. Consider a box $B = [a, b] \subseteq \mathcal{C}$, i.e., $l \leq a \leq b \leq u$. Then, Theorem 2, Theorem 3 and Theorem 4 imply that φ_0 is Lipschitz continuous on B with constant (6), while φ_α , for any $\alpha > 0$, is Lipschitz continuous on B with constant (7).

Estimate of L_1 . An upper bound for L_1 can be computed as follows:

$$\begin{aligned} L_1 &= \max_{x \in B} \|Px + r + T(x)\| \\ &\leq \max_{x \in B} \|Px + r\| + \max_{x \in B} \|T(x)\| \\ &\leq \tilde{L}_1(P, r, a, b) + \|w\|, \end{aligned}$$

where $\tilde{L}_1(P, r, a, b)$ is defined in (8).

Estimate of L_2 . Since L_2 only depends on the B and \mathcal{C} , its exact value is given by (9).

Estimate of $L_3(\alpha)$. The following upper bound can be obtained:

$$L_3(\alpha) = \max_{x \in B} \|\alpha I - P - \nabla T(x)\| \leq \|\alpha I - P\| + \max_{x \in B} \|\nabla T(x)\|.$$

The Jacobian matrix $\nabla T(x)$ is diagonal with

$$[\nabla T(x)]_{ii} = w_i v_i \cos(v_i x_i), \quad i = 1, \dots, n,$$

hence, for any $x \in B$ we get

$$\|\nabla T(x)\| = \max_{1 \leq i \leq n} \{|w_i v_i \cos(v_i x_i)|\} = \max_{1 \leq i \leq n} \{w_i v_i |\cos(v_i x_i)|\} \leq \max_{1 \leq i \leq n} \{w_i v_i\}.$$

Therefore, we have

$$L_3(\alpha) \leq \|\alpha I - P\| + \max_{1 \leq i \leq n} \{w_i v_i\}. \quad (10)$$

Estimate of L_F . The Lipschitz constant of F can be estimated as follows:

$$\begin{aligned}
\|F(x) - F(z)\| &= \|P(x - z) + T(x) - T(z)\| \\
&\leq \|P\|\|x - z\| + \|T(x) - T(z)\| \\
&= \|P\|\|x - z\| + \sqrt{\sum_{i=1}^n w_i^2 [\sin(v_i x_i) - \sin(v_i z_i)]^2} \\
&\leq \|P\|\|x - z\| + \sqrt{\sum_{i=1}^n w_i^2 v_i^2 (x_i - z_i)^2} \\
&\leq \|P\|\|x - z\| + \sqrt{\left[\max_{1 \leq i \leq n} \{w_i v_i\}\right]^2 \sum_{i=1}^n (x_i - z_i)^2} \\
&= \|P\|\|x - z\| + \max_{1 \leq i \leq n} \{w_i v_i\} \|x - z\| \\
&= (\|P\| + \max_{1 \leq i \leq n} \{w_i v_i\}) \|x - z\|,
\end{aligned}$$

where the second inequality holds because the sine function is Lipschitz continuous with constant 1. Therefore, we have

$$L_F \leq \|P\| + \max_{1 \leq i \leq n} \{w_i v_i\}. \quad (11)$$

3.3 Affine EPs defined on a box

Suppose that (EP) is defined by an affine operator $F(x, y) = Px + Qy + r$, for some $P, Q \in \mathbb{R}^{n \times n}$ and $r \in \mathbb{R}^n$, over a box $\mathcal{C} = [l, u]$, where $l, u \in \mathbb{R}^n$. Consider a box $B = [a, b]$. Then, Theorem 2 and Theorem 3 imply that φ_0 is Lipschitz continuous on B with constant (6), while, for any $\alpha > 0$, φ_α is Lipschitz continuous on B with constant

$$\{L_1 + L_2 L_F + \alpha L_2, \quad L_1 + L_2 L_3(\alpha)\}. \quad (12)$$

Estimate of L_1 . The following bound can be easily obtained:

$$\begin{aligned}
L_1 &= \max_{x \in B, y \in \mathcal{C}} \|Px + Qy + r\| \\
&\leq \max_{x \in B, y \in \mathcal{C}} (\|Px\| + \|Qy + r\|) \\
&= \max_{x \in B} \|Px\| + \max_{y \in \mathcal{C}} \|Qy + r\| \\
&\leq \tilde{L}_1(P, 0, a, b) + \tilde{L}_1(Q, r, l, u) := M_1.
\end{aligned} \quad (13)$$

Similarly to the previous bound, we get

$$\begin{aligned}
L_1 &= \max_{x \in B, y \in \mathcal{C}} \|Px + Qy + r\| \\
&\leq \max_{x \in B, y \in \mathcal{C}} (\|Px + r\| + \|Qy\|) \\
&= \max_{x \in B} \|Px + r\| + \max_{y \in \mathcal{C}} \|Qy\| \\
&= \tilde{L}_1(P, r, a, b) + \tilde{L}_1(Q, 0, l, u) := M_2.
\end{aligned} \quad (14)$$

Finally, we have

$$\begin{aligned}
L_1 &= \max_{x \in B, y \in \mathcal{C}} \|Px + Qy + r\| \\
&\leq \max_{x \in B, y \in \mathcal{C}} (\|Px + r/2\| + \|Qy + r/2\|) \\
&= \max_{x \in B} \|Px + r/2\| + \max_{y \in \mathcal{C}} \|Qy + r/2\|
\end{aligned}$$

$$= \tilde{L}_1(P, r/2, a, b) + \tilde{L}_1(Q, r/2, l, u) := M_3, \quad (15)$$

thus $L_1 \leq \min\{M_1, M_2, M_3\}$.

Estimates of L_2 , $L_3(\alpha)$ and L_F . It is easy to check that L_2 is given by (9), $L_3(\alpha) = \|\alpha I - P\|$ and $L_F = \|P\|$.

3.4 Nonlinear EPs with trigonometric terms defined on a box

Suppose that (EP) is defined over a box $\mathcal{C} = [l, u]$, with a nonlinear operator given by the sum of an affine map and a trigonometric map, i.e.,

$$F(x, y) = Px + Qy + r + T(x),$$

where $P, Q \in \mathbb{R}^{n \times n}$, $r \in \mathbb{R}^n$, $T_i(x) = w_i \sin(v_i x_i)$, for $i = 1, \dots, n$, and $v, w \in \mathbb{R}^n$ with $v, w > 0$. Consider a box $B = [a, b]$. Then, Theorem 2 and Theorem 3 imply that φ_0 is Lipschitz continuous on B with constant (6), while, for any $\alpha > 0$, φ_α is Lipschitz continuous on B with constant (12).

Estimate of L_1 . The following bound can be easily obtained:

$$\begin{aligned} L_1 &= \max_{x \in B, y \in \mathcal{C}} \|Px + Qy + r + T(x)\| \\ &\leq \max_{x \in B, y \in \mathcal{C}} \|Px + Qy + r\| + \max_{x \in B} \|T(x)\| \\ &\leq \min\{M_1, M_2, M_3\} + \|w\|, \end{aligned}$$

where M_1, M_2, M_3 are defined in (13)–(15).

Estimates of L_2 , $L_3(\alpha)$ and L_F . It is easy to check that L_2 is given by (9), while the upper bounds for $L_3(\alpha)$ and L_F are given by (10) and (11), respectively.

4 The DIRECT-type algorithms

We now describe a DIRECT-type approach to globally solve the optimization problem (3) that is equivalent to (EP). In this section, we assume that the feasible region \mathcal{C} is a box, i.e., $\mathcal{C} = \{x \in \mathbb{R}^n : l \leq x \leq u\}$.

More specifically, we focus on partition based algorithms, a class of methods with both interesting theoretical properties and efficient computational behavior, and explain why those algorithms represent a good option when dealing with non-monotone EPs. We start by giving some useful details.

Partition based methods produce a sequence of finer and finer partitions $\{\mathcal{H}_k\}$ of the feasible set \mathcal{C} . At each iteration k , the k -th partition is described by:

$$\mathcal{H}_k = \{\mathcal{C}^i : i \in I_k\},$$

where

$$\mathcal{C}^i = \{x \in \mathbb{R}^n : l^i \leq x \leq u^i\}, \quad l^i, u^i \in [l, u], \quad x^i = (l^i + u^i)/2.$$

We further notice that $\mathcal{C} = \cup\{\mathcal{C}^i : i \in I_k\}$ and $\text{Int}(\mathcal{C}^i) \cap \text{Int}(\mathcal{C}^j) = \emptyset$ for all i, j . Then the next partition \mathcal{H}_{k+1} is obtained by selecting and by further partitioning every element of a “particular” subset $\{\mathcal{C}^i : i \in I_k^*\} \subseteq \mathcal{H}_k$, where $I_k^* \subset I_k$. A partition based algorithm is characterized by the rules used to generate the subset of indices I_k^* , and by the strategies applied to further partition the subsets $\{\mathcal{C}^i : i \in I_k^*\}$.

In [29], the authors consider non-monotone VIs and use a Branch and Bound method similar to the one described in [19] to tackle the considered global optimization problems.

Instead, as previously pointed out, we solve non-monotone EPs by means of an algorithm derived from the well-known DIRECT method (see, e.g., [20]). This approach, called \bar{L} -DIRECT and

first proposed in [12], differs from the standard version of **DIRECT** in the way the set of indices I_k^* are defined. In the standard version of **DIRECT**, I_k^* consists of the indices related to the subsets satisfying the definition reported below:

Definition 2 Given a partition $\mathcal{H}_k = \{\mathcal{C}^i : i \in I_k\}$ of \mathcal{C} and a scalar $\varepsilon > 0$, a subset \mathcal{C}^h is *potentially optimal with respect to the function φ_α* if a constant \bar{L}^h exists such that:

$$\begin{aligned} \varphi_\alpha(x^h) - \frac{\bar{L}^h}{2} \|u^h - l^h\| &\leq \varphi_\alpha(x^i) - \frac{\bar{L}^h}{2} \|u^i - l^i\|, & \forall i \in I_k, \\ \varphi_\alpha(x^h) - \frac{\bar{L}^h}{2} \|u^h - l^h\| &\leq \varphi_{\min} - \epsilon |\varphi_{\min}|, \end{aligned}$$

where

$$\varphi_{\min} = \min_{i \in I_k} \varphi_\alpha(x^i). \quad (16)$$

In the \bar{L} -DIRECT algorithm, I_k^* is given by the indices related to those subsets satisfying:

Definition 3 Given a partition $\mathcal{H}_k = \{\mathcal{C}^i : i \in I_k\}$ of \mathcal{C} , a scalar $\varepsilon > 0$, a scalar $\eta > 0$ and a scalar $\bar{L} > 0$, a subset \mathcal{C}^h is \bar{L} -*potentially optimal with respect to the function φ_α* if one of the following conditions is satisfied:

i) A constant $\tilde{L}^h \in (0, \bar{L})$ exists such that:

$$\varphi_\alpha(x^h) - \frac{\tilde{L}^h}{2} \|u^h - l^h\| \leq \varphi_\alpha(x^i) - \frac{\tilde{L}^h}{2} \|u^i - l^i\|, \quad \forall i \in I_k, \quad (17)$$

$$\varphi_\alpha(x^h) - \frac{\tilde{L}^h}{2} \|u^h - l^h\| \leq \varphi_{\min} - \epsilon \max\{|\varphi_{\min}|, \eta\}, \quad (18)$$

where φ_{\min} is given by (16);

ii) The following inequality holds:

$$\varphi_\alpha(x^h) - \frac{\bar{L}}{2} \|u^h - l^h\| \leq \varphi_\alpha(x^i) - \frac{\bar{L}}{2} \|u^i - l^i\|, \quad \forall i \in I_k. \quad (19)$$

Remark 3 The difference between the two is that an overestimate \bar{L} of the Lipschitz constant is used in Definition 3. This fact obviously enhances the way \bar{L} -DIRECT selects the subsets to be partitioned.

Remark 4 As $\bar{L} \rightarrow \infty$, Definition 3 tends to Definition 2 and, hence, the strategy proposed in [12] becomes the one proposed in [20].

We refer to [20] and [12] for detailed descriptions and discussions of the **DIRECT** algorithm and the \bar{L} -DIRECT algorithm. Similarly to any partition-based method, the asymptotic behavior shown by the **DIRECT** and the \bar{L} -DIRECT algorithms is characterized by the partition sequences they produce. Those sequences can be represented equivalently by infinite sequences of nested subsets $\{\mathcal{C}^{i_k}\}$, defined as follows:

Given a set \mathcal{C}^{i_k} at the iteration k , its predecessor $\mathcal{C}^{i_{k-1}}$ is the unique set belonging to the previous partition $\mathcal{H}_{k-1} = \{\mathcal{C}^i : i \in I_{k-1}\}$ such that $\mathcal{C}^{i_k} \subseteq \mathcal{C}^{i_{k-1}}$.

Then, the analysis of theoretical properties of **DIRECT** algorithm and \bar{L} -DIRECT algorithm can be performed by studying the properties of the produced sequences $\{\mathcal{C}^{i_k}\}$. The partitioning strategy used by the **DIRECT** algorithm and the \bar{L} -DIRECT algorithm guarantees (regardless of the particular choice of set I_k^*) that the produced sequences $\{\mathcal{C}^{i_k}\}$ satisfy one of the following properties (see [27]):

- Property 1: an index \bar{k} exists such that $\mathcal{C}^{i_{\bar{k}}} = \mathcal{C}^{i_k}$ for all $k \geq \bar{k}$;
- Property 2: $\bigcap_{k=0}^{\infty} \mathcal{C}^{i_k} = \{\bar{x}\}$, where $\bar{x} \in \mathcal{C}$.

Then the so-called *everywhere dense convergence* can be stated by the following proposition.

Proposition 1 *DIRECT algorithm has the following properties:*

- i) All the sequences of sets $\{\mathcal{C}^{i_k}\}$ produced satisfy Property 2;
- ii) For every $\tilde{x} \in \mathcal{C}$, the DIRECT algorithm produces a sequence of sets $\{\mathcal{C}^{i_k}\}$ satisfying Property 2 and such that

$$\bigcap_{k=0}^{\infty} \mathcal{C}^{i_k} = \{\tilde{x}\}.$$

The properties of the \bar{L} -DIRECT algorithm also depend on the choice of the scalar \bar{L} included in the definition of \bar{L} -potentially optimal subsets. In particular, the following assumption can be introduced.

Assumption 1. *For every global minimum point x^* of problem (3), there exists an index \bar{k} (possibly depending on x^*) such that, if $\mathcal{C}^{j_{\bar{k}}} \in \{\mathcal{C}^i : i \in I_{\bar{k}}\}$ is the subset satisfying $x^* \in \mathcal{C}^{j_{\bar{k}}}$, then*

$$\bar{L} < L,$$

where L is the local Lipschitz constant of the function φ_α over the subset $\mathcal{C}^{j_{\bar{k}}}$.

Now it is possible to state the following result.

Proposition 2 *If Assumption 1 holds, then \bar{L} -DIRECT algorithm has the following properties:*

- i) Every sequence of sets $\{\mathcal{C}^{i_k}\}$ produced by the algorithm which satisfies Property 2 is such that

$$\bigcap_{k=0}^{\infty} \mathcal{C}^{i_k} = \{x^*\},$$

where x^* is a global minimum of problem (3);

- ii) For every global minimum x^* of problem (3), the algorithm produces a sequence of sets $\{\mathcal{C}^{i_k}\}$ satisfying Property 2 and

$$\bigcap_{k=0}^{\infty} \mathcal{C}^{i_k} = \{x^*\};$$

- iii) Let \bar{k} be the index introduced in Assumption 1. Then, for all $k \geq \bar{k}$, the following inequality holds

$$\varphi_\alpha(x^{h_k}) - \varphi_\alpha^* \leq \frac{\bar{L}}{2} \|u^{h_k} - l^{h_k}\|, \quad (20)$$

where the index h_k is given by:

$$\varphi_\alpha(x^{h_k}) - \frac{\bar{L}}{2} \|u^{h_k} - l^{h_k}\| = \min_{i \in I_k} \left\{ \varphi_\alpha(x^i) - \frac{\bar{L}}{2} \|u^i - l^i\| \right\}.$$

Points i) and ii) of the previous proposition guarantee that, as the number of iterations increases, \bar{L} -DIRECT generates points that are more and more clustered around the global minima of problem (3). Point iii) gives a practical stopping criterion for the algorithm. The right-hand side of (20) indeed provides an optimality gap.

Remark 5 Proposition 2 highlights the main difference between the Branch and Bound algorithm used in [29] and the \bar{L} -DIRECT algorithm. In order to guarantee convergence to a global minimum of the Branch and Bound, an overestimate for the Lipschitz constant of φ_α over the whole feasible set \mathcal{C} is needed from the beginning. On the other hand, convergence of the \bar{L} -DIRECT algorithm can be guaranteed by an overestimate of the local Lipschitz constant of φ_α over the subset $\mathcal{C}^{j_{\bar{k}}}$ (keep in mind that this local constant is usually much smaller than the global one). Furthermore, this overestimate is needed only for sufficiently large values of the indices k . Hence, the information obtained from the function values calculated in the first iterations of the algorithm can be exploited to get an overestimate of the required local Lipschitz constant.

Proposition 1 and Proposition 2 imply the following corollary.

Corollary 1 *The DIRECT Algorithm and, if Assumption 1 holds, also the \bar{L} -DIRECT algorithm satisfy the following property:*

For every global minimum point x^ of problem (3) and for every neighborhood $\mathcal{B}(x^*)$ of x^* , an index k exists such that both the algorithms produce a point x^{i_k} satisfying*

$$x^{i_k} \in \mathcal{B}(x^*).$$

The previous result points out that the two DIRECT-based methods can be efficiently combined with local searches within a multistart strategy.

5 Numerical results

In this section, we describe our numerical experience. The goal is twofold: on the one side, we would like to see how DIRECT strategies behave on this class of problems; on the other side, we would like to understand the importance of embedding the Lipschitz constant estimates in those algorithmic schemes. We thus consider two different algorithms in the experiments:

- DIRECT: the standard version of the method with local searches;
- \bar{L} -DIRECT: the modified version with Lipschitz constant estimates and local searches.

In both cases we used the SDBOX algorithm [28] to perform the local search. All algorithms were implemented in Matlab and tests were performed with Matlab v2019b.

We first considered 5 affine VI problems coming from the literature. Then, we considered randomly generated instances for four different classes of problems: affine VIs, nonlinear VIs with trigonometric terms, affine EPs and nonlinear EPs with trigonometric terms. In the analysis of those randomly generated instances we used performance and data profiles [33] with a gate parameter $\tau = 10^{-3}$. Specifically, let S be a set of algorithms and P a set of problems. For each $s \in S$ and $p \in P$, let $t_{p,s}$ be the number of function evaluations required by algorithm s on problem p to satisfy the condition

$$f(x_k) \leq f_L + \tau(f(x_0) - f_L) \quad (21)$$

where $0 < \tau < 1$ and f_L is the best objective function value achieved by any solver on problem p . Then, performance and data profiles of solver s are the following functions

$$\rho_s(\alpha) = \frac{1}{|P|} \left| \left\{ p \in P : \frac{t_{p,s}}{\min\{t_{p,s'} : s' \in S\}} \leq \alpha \right\} \right|,$$

$$d_s(\kappa) = \frac{1}{|P|} |\{p \in P : t_{p,s} \leq \kappa(n_p + 1)\}|$$

where n_p is the dimension of problem p . In our plots we hence show the values α and κ on the x axis, and $\rho_s(\alpha)$ and $d_s(\kappa)$ on the y axis, respectively for performance and data profiles. It is important to notice that the limiting value of $\rho_s(\alpha)$ as $\alpha \rightarrow \infty$ is the percentage of problems that can be solved with the available budget of function evaluations. Hence we have $d_s(\hat{\kappa}) = \lim_{\alpha \rightarrow \infty} \rho_s(\alpha)$, where $\hat{\kappa}$ is the maximum number of simplex gradients performed with the available budget of function evaluations. This indicates how reliable a solver is for the given tolerance τ (see [33] for further details). We can thus use performance and data profiles to have some insights on the relative performances of the compared algorithms. More specifically, we need to pay attention to how steeply and how high those profiles rise, to respectively understand how efficient and reliable a solver is.

In all the experiments, we considered the gap function φ_α defined in (1) with $\alpha = 1$. The detailed results are reported in the next subsections.

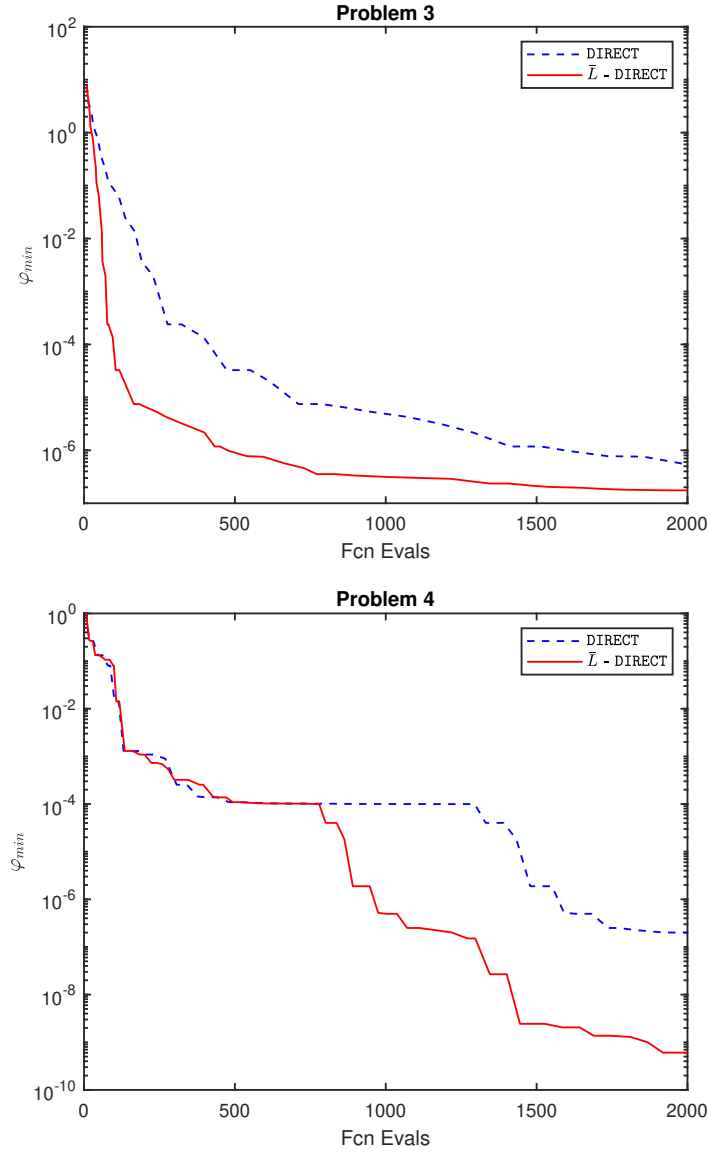


Fig. 1 Comparison between DIRECT and \bar{L} -DIRECT on Problems 3 and 4 from [29].

5.1 Results on VI problems from the literature

We here show results on the Problems 2–6 from paper [29]. In order to consider only affine VIs, we dropped the absolute value in the operator $F(x)$ of Problems 4 and 5. In Table 1, we report, for each problem, the number of function evaluations needed by the two algorithms to reach a certain gap value (we chose 10^{-1} , 10^{-3} , 10^{-5}). As we can easily see, the number of function evaluations is usually smaller for \bar{L} -DIRECT (we report in red the cases where \bar{L} -DIRECT needs a higher number of evaluations). In Figure 1, we further report the plots related to the gap reduction with respect to the number of function evaluations used for Problems 3 and 4. We indicate with φ_{min} the gap value (reported on the y axis) and with **Fcn Evals** the number of function evaluations (reported on the x axis). As we can see, the use of the Lipschitz constant estimate significantly speeds up the algorithm.

Table 1 Comparison between DIRECT and \bar{L} -DIRECT on VI Problems from [29] (# of f.e. to reach a given gap)

Problem	n	DIRECT			\bar{L} -DIRECT		
		10^{-1}	10^{-3}	10^{-5}	10^{-1}	10^{-3}	10^{-5}
2	3	125	431	977	97	313	749
3	4	115	277	711	49	77	165
4	3	79	269	1479	97	223	891
5	5	383	1913	1969	365	1961	2000
6	10	1987	1987	1987	875	1341	1799

5.2 Results on randomly generated affine VIs

We now describe in depth the results obtained on randomly generated affine VI problems. We generated 100 instances with 5 variables. For each instance, the affine operator $F(x) = Px + r$ was randomly built by choosing a matrix P with uniformly distributed random numbers in the interval $[0, 3]$ and a vector r with uniformly distributed random numbers in the interval $[-2, 2]$. The box constraints $\{x \in \mathbb{R}^n : l \leq x \leq u\}$ were generated by considering two vectors l and u with uniformly distributed random numbers in the interval $[-2, 0]$ and $[1, 3]$, respectively. We gave a budget of 600 function evaluations to the considered algorithms (500 for the DIRECT strategies and 100 for the local search). Performance and data profiles are reported in Figure 2. The performance profile plot shows that the \bar{L} -DIRECT (red line) is both much more efficient than DIRECT (blue line), since it gives better performance and satisfies the stopping condition with a smaller number of function evaluations for the 70% of the instances, and more reliable (indeed, the percentage of problems that can be solved with the available budget of function evaluations is higher). If we observe the data profiles, we can further see that \bar{L} -DIRECT solves a higher percentage of problems no matter what the budget used is.

5.3 Results on randomly generated nonlinear VIs with trigonometric terms

In this subsection we report the results obtained on randomly generated VI problems with trigonometric terms. We generated 100 instances with 5 variables in this case as well. For each instance, the operator $F(x) = Px + r + T(x)$, where $T_i(x) = w_i \sin(v_i x_i)$, for $i = 1, \dots, n$, was randomly built by choosing a matrix P with uniformly distributed random numbers in the interval $[0, 3]$ and vectors w, v and r with uniformly distributed random numbers in the interval $(0, 4]$, $(0, 2]$, and $[-2, 2]$, respectively. The box constraints $\{x \in \mathbb{R}^n : l \leq x \leq u\}$ were generated by considering two vectors l and u with uniformly distributed random numbers in the interval $[-2, 0]$ and $[1, 3]$, respectively. We used the same budget of function evaluation given for affine VIs. Performance and data profiles are reported in Figure 3. It is easy to see, by taking a look at the performance profile plot, that the \bar{L} -DIRECT (red line) is again more efficient than DIRECT (blue line), since it gives better performance and satisfies the stopping condition with a smaller number of function evaluations for about the 75% of the instances, and is also more reliable. Data profiles show that \bar{L} -DIRECT solves a higher number of instances no matter what the budget used is.

5.4 Results on randomly generated affine EPs

Here, we report the results obtained on random instances related to affine equilibrium problems. We generated 100 instances with 5 variables in this case as well. For each instance, the operator $F(x) = Px + Qy + r$ was randomly built by choosing a matrix P with uniformly distributed random numbers in the interval $[0, 3]$, a symmetric positive semidefinite matrix Q with uniformly distributed random numbers in the interval $[0, 5]$, and r with uniformly distributed random numbers in the interval $[-2, 2]$. The box constraints $\{x \in \mathbb{R}^n : l \leq x \leq u\}$ were generated by

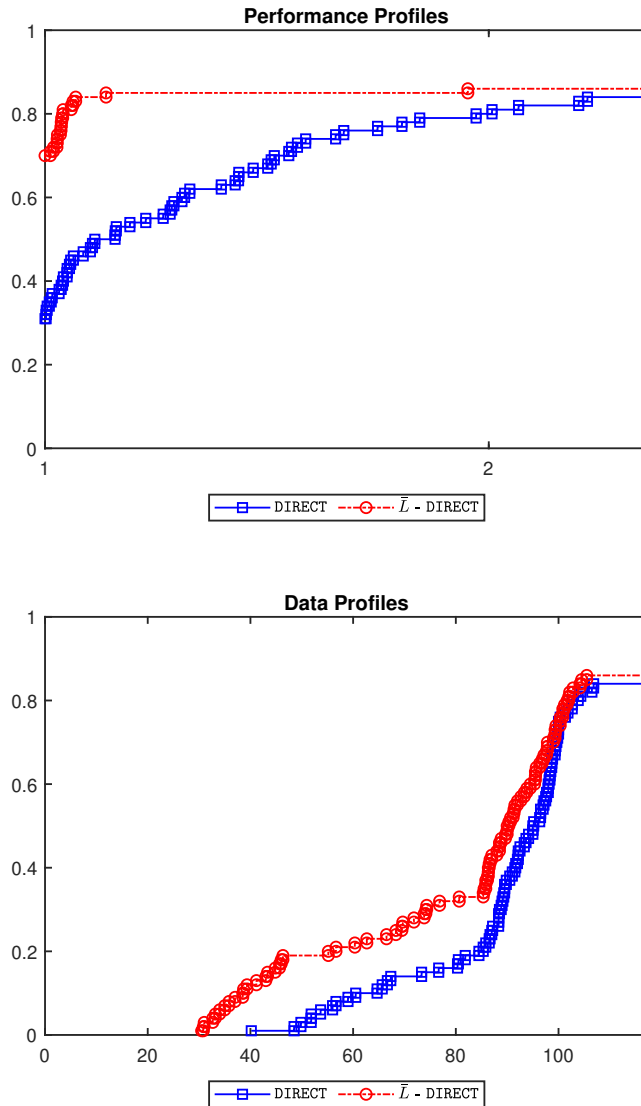


Fig. 2 Performance and data profiles for randomly generated affine VI problems.

considering two vectors l and u with uniformly distributed random numbers in the interval $[-2, 0]$ and $[1, 3]$, respectively. We used the same budget of function evaluations given for the other problems. Performance and data profiles are reported in Figure 4. It is easy to see, by taking a look at the performance profile plot, that the \bar{L} -DIRECT (red line) is again more efficient than DIRECT (blue line), since it gives better performance and satisfies the stopping condition with a smaller number of function evaluations for almost 80% of the instances, and is also more reliable. Data profiles show again that \bar{L} -DIRECT solves a higher number of instances no matter what the budget used is.

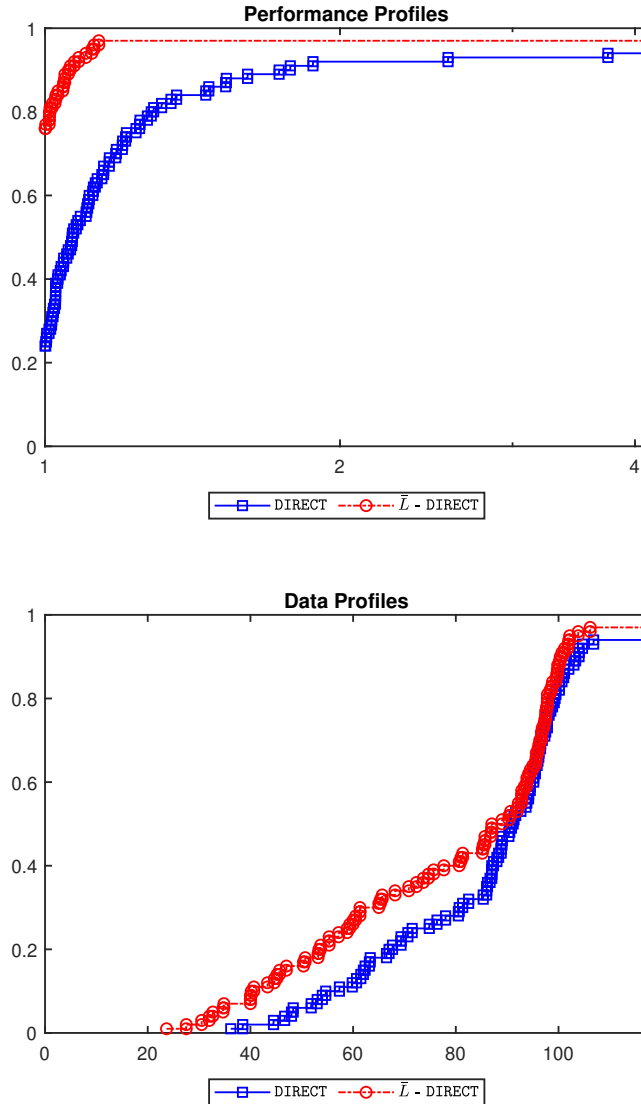


Fig. 3 Performance and data profiles for randomly generated nonlinear VI problems with trigonometric terms.

5.5 Results on randomly generated nonlinear EPs with trigonometric terms

Here, we report the results obtained on random instances related to nonlinear equilibrium problems with trigonometric terms. We generated 100 instances with 5 variables in this case as well. For each instance, the operator $F(x) = Px + Qy + r + T(x)$, where $T_i(x) = w_i \sin(v_i x_i)$, for $i = 1, \dots, n$, was randomly built by choosing a matrix P with uniformly distributed random numbers in the interval $[0, 3]$, a symmetric positive semidefinite matrix Q with uniformly distributed random numbers in the interval $[0, 5]$, vectors w, v and r with uniformly distributed random numbers in the interval $(0, 4]$, $(0, 2]$, and $[-2, 2]$, respectively. The box constraints $\{x \in \mathbb{R}^n : l \leq x \leq u\}$ were generated by considering two vectors l and u with uniformly distributed random numbers in the interval $[-2, 0]$ and $[1, 3]$, respectively. We used the same budget of function evaluations given for the other problems. Performance and data profiles are reported in Figure 5. It is easy to see, by taking a look at the performance profile plot, that the \bar{L} -DIRECT (red line) is more efficient than DIRECT

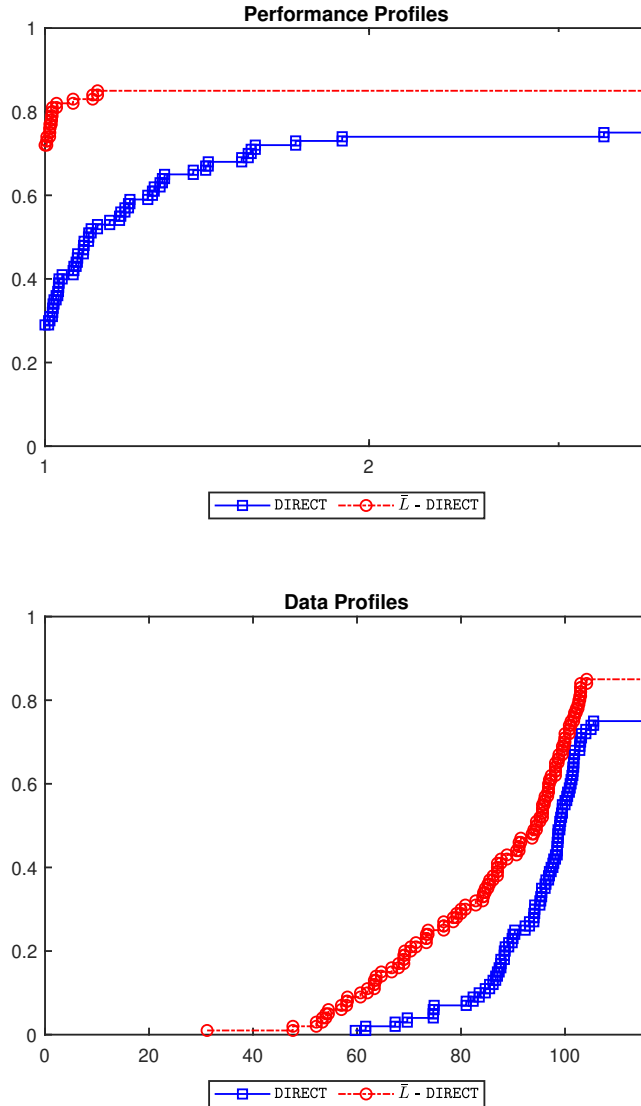


Fig. 4 Performance and data profiles for randomly generated affine EPs.

(blue line), since it gives better performance and satisfies the stopping condition with a smaller number of function evaluations for more than 80% of the instances, and is also more reliable. Data profiles show that \bar{L} -DIRECT solves a higher number of instances for all the budgets.

6 Preliminary comparison with a local solver

In this section, we compare \bar{L} -DIRECT with PATH [13], a widely used local solver for mixed complementarity problems and variational inequalities. The goal in this case is understanding if the proposed approach can really make a difference when dealing with problems that do not satisfy any monotonicity-type assumption. We hence consider once again random VIs with trigonometric terms. The only difference we have with respect to the instances described before is in the vectors w and v , whose components were respectively chosen in the interval $(0, 40]$ and $(0, 20]$ for this

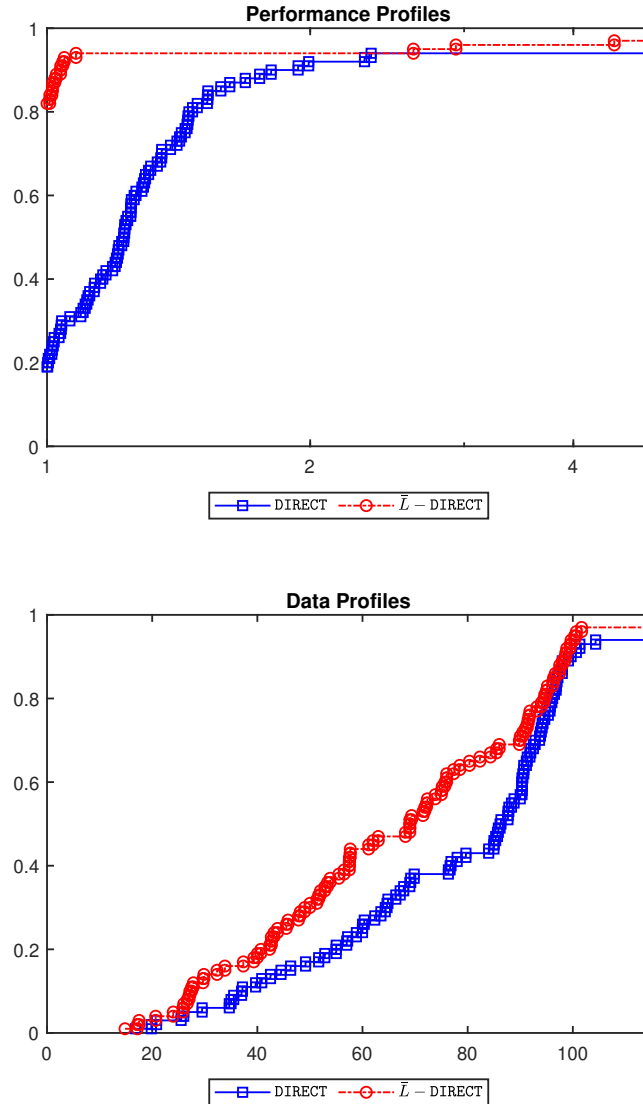


Fig. 5 Performance and data profiles for randomly generated nonlinear EPs with trigonometric terms.

experiment. This should make the problem more challenging for the solvers. In the experiments, we ran `PATH` with its default parameters. Since `PATH` uses first-order information while carrying out the optimization process, we decided to give a budget of 1000 function evaluations to \bar{L} -DIRECT (500 for the `DIRECT` strategy and 500 for the local search). Performance and data profiles are reported in Figure 6. It is easy to see, by taking a look at the performance profile plot, that `PATH` (blue line) is slightly better than \bar{L} -DIRECT (red line) when $\alpha = 0$, but things quickly change as the α value increases. The performance profile related to \bar{L} -DIRECT is indeed always above the one related to `PATH` for values of α larger than 6, and the gap gets larger as α grows. Furthermore, the \bar{L} -DIRECT is more reliable than `PATH`, since it satisfies the stopping condition for more than 90% of the instances, while `PATH` gets the condition satisfied for less than 60% of the instances in the end. Data profiles show that \bar{L} -DIRECT generally solves a higher number of instances. In this case, \bar{L} -DIRECT indeed satisfies the stopping condition for more than 90% of the instances within a budget κ slightly larger than 100 (i.e., just 600 function evaluations), while `PATH` needs to get a

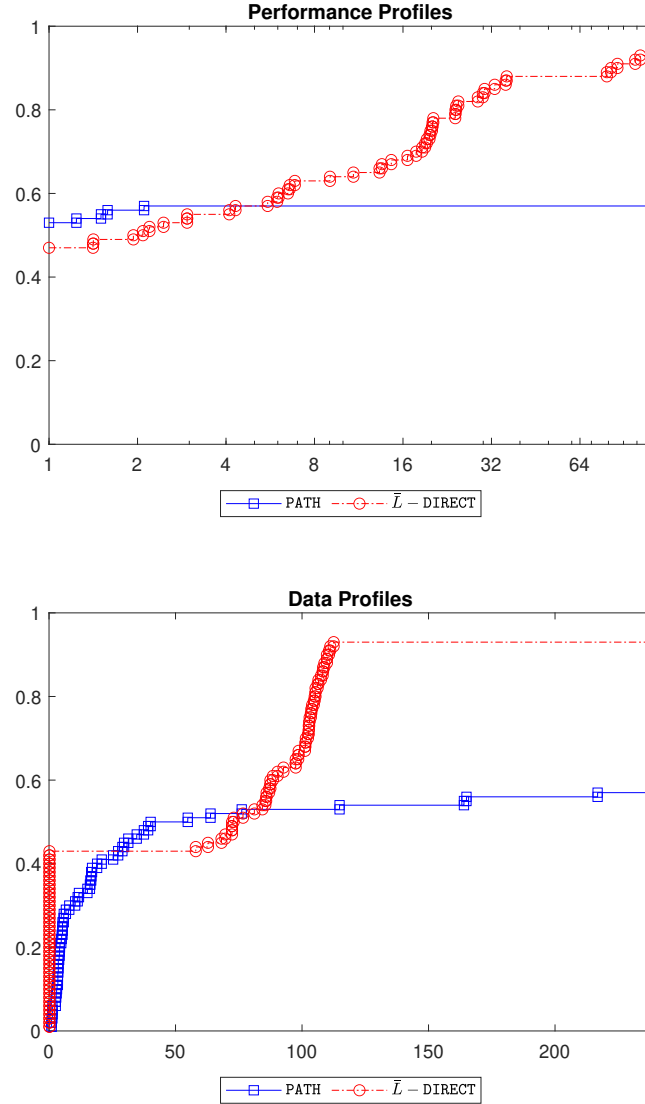


Fig. 6 Comparison with PATH: performance and data profiles ($n = 5$).

value of κ larger than 200 (that is 1200 function and Jacobian evaluations) to satisfy the stopping condition for roughly 60% of the instances. The results clearly show that \bar{L} -DIRECT solves a much larger number of problems as soon as the budget of function evaluations is large enough. As we might expect, PATH is efficient (thanks to the use of first-order information), but not as reliable as \bar{L} -DIRECT. It is interesting to notice that PATH, due to the non-monotonicity of the problems, struggles to find a solution, and actually fails to find it in some cases. We can thus conclude that \bar{L} -DIRECT is a good option when we have no monotonicity-type assumptions on the problem we want to solve.

7 Preliminary comparison with simplicial methods for fixed point problems

In this section, we compare our \bar{L} -DIRECT with two different simplicial methods for fixed point problems from the literature, in order to solve nonlinear equilibrium problems with box constraints $l \leq x \leq u$. More specifically, we considered the seminal Scarf algorithm based on triangulation and labeling techniques (see [41, 42, 43]) and a prototype of the homotopy simplicial algorithms which are referred to as sandwich or restart methods (see Algorithm 1 in [2]).

Since the Scarf algorithm finds an approximation of a fixed point of a continuous mapping of the unit simplex into itself, we considered the following further fixed point reformulation of (EP). Let S be the unit simplex in \mathbb{R}^{n+1} , i.e.,

$$S = \left\{ (x_0, x_1, \dots, x_n) \in \mathbb{R}_+^{n+1} : \sum_{i=0}^n x_i = 1 \right\},$$

and $G : S \rightarrow S$ the mapping defined as follows:

$$G(x) = A_1^{-1}(A_2^{-1}(y_\alpha(A_2(P_{[0,1]^n}(A_1(x))))))),$$

where A_1 maps S into the simplex $\text{conv}\{0, ne_1, \dots, ne_n\} \subset \mathbb{R}^n$ containing the unit box $[0, 1]^n$, where e_1, \dots, e_n are the unit vectors, and is defined as

$$A_1(x_0, x_1, \dots, x_n) = (nx_1, \dots, nx_n),$$

$P_{[0,1]^n}$ is the Euclidean projection on the unit box $[0, 1]^n$, A_2 maps the unit box $[0, 1]^n$ into the box $[l, u]$ and is defined as

$$A_2(x_1, \dots, x_n) = ((u_1 - l_1)x_1 + l_1, \dots, (u_n - l_n)x_n + l_n),$$

and $y_\alpha(x)$ is the unique maximizer of problem in (1). Since the solutions of (EP) coincide with the fixed points of the mapping y_α , it is easy to check that x^* solves (EP) if and only if $A_1^{-1}(A_2^{-1}(x^*))$ is a fixed point of the mapping G . Hence, we applied the Scarf algorithm to find an approximated fixed point of the mapping G . We considered the simplicial decomposition of S given by points $(k_0/D, \dots, k_n/D)$, where D is a large positive integer and k_0, \dots, k_n are non-negative integers summing D . We started the algorithm with $D = 10n$ from the barycenter of the face $x_0 = 0$ of S , as suggested by the Kuhn's method [23]. When a completely labeled simplex (with barycenter x) was found, then the gap function was evaluated at $A_2(A_1(x))$: if its value was below the given tolerance, then the algorithm stopped, otherwise we restarted the algorithm with a new simplicial decomposition given by $D' = 5D$.

The considered prototype of the homotopy simplicial algorithms was applied to find an approximated fixed point of the mapping y_α from \mathbb{R}^n into the box $[l, u]$, i.e., an approximated solution of (EP). We chose the starting point equal to the center $(l + u)/2$ of the box and the grid size $\delta = 0.9$. Similarly to the Scarf algorithm, the gap function was evaluated at the barycenter of the completely labeled simplex found by the algorithm: if its value was below the given tolerance, then the algorithm stopped, otherwise the grid size δ was replaced by $\delta/2$.

Differently from those classic simplicial methods for fixed point problems, our \bar{L} -DIRECT approach tries to efficiently explore the feasible set by exploiting the structure of the problem. In particular, it uses the Lipschitz constant of the objective function to generate points that are clustered around the global minima of the function as the iterations go on (see Proposition 2).

We considered the random instances related to nonlinear equilibrium problems with trigonometric terms described in Section 5.5 and generated two different sets of 100 instances with $n = 5$ and $n = 10$. We gave to all the considered algorithms a budget of 600 function evaluations for the first set of instances and 2400 for the second one. Performance and data profiles are reported in Figure 7 and Figure 8. We can see, by taking a look at the performance profile plots, that the \bar{L} -DIRECT (red line) is more efficient than the competitors, since it gives better performance and satisfies the stopping condition with a smaller number of function evaluations for almost 60% of

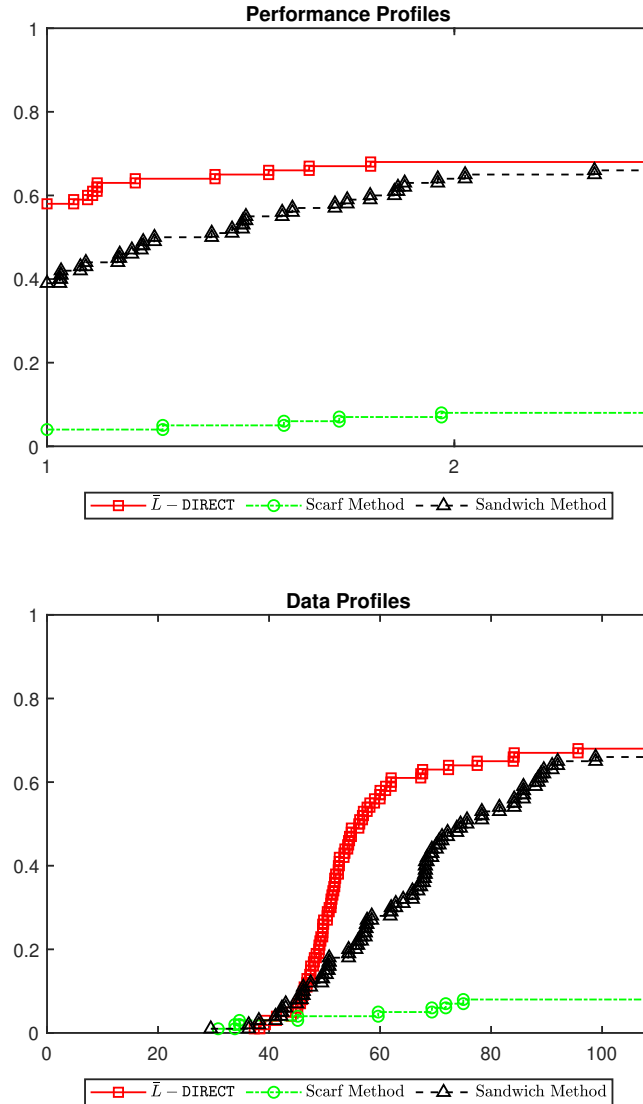


Fig. 7 Comparison with triangulation and simplicial methods: performance and data profiles ($n = 5$).

the instances for $n = 5$, and about 80% for $n = 10$. We can also see that our method is more reliable in both cases. Data profiles show that \bar{L} -DIRECT solves a higher number of instances for budgets larger than 40 simplex gradients with $n = 5$ and larger than 100 with $n = 10$.

8 Conclusions

In this paper, we propose a global optimization approach for solving general EPs without assuming any monotonicity-type condition on f . This approach is based on two phases: (i) reformulate an EP as a global optimization problem via gap functions; (ii) use an improved version of the DIRECT algorithm, which exploits local bounds of the Lipschitz constant of the objective function, combined with local searches to solve the considered global optimization problem. Moreover, we provide some general results on Lipschitz continuity of gap functions and, for some special classes of EPs,

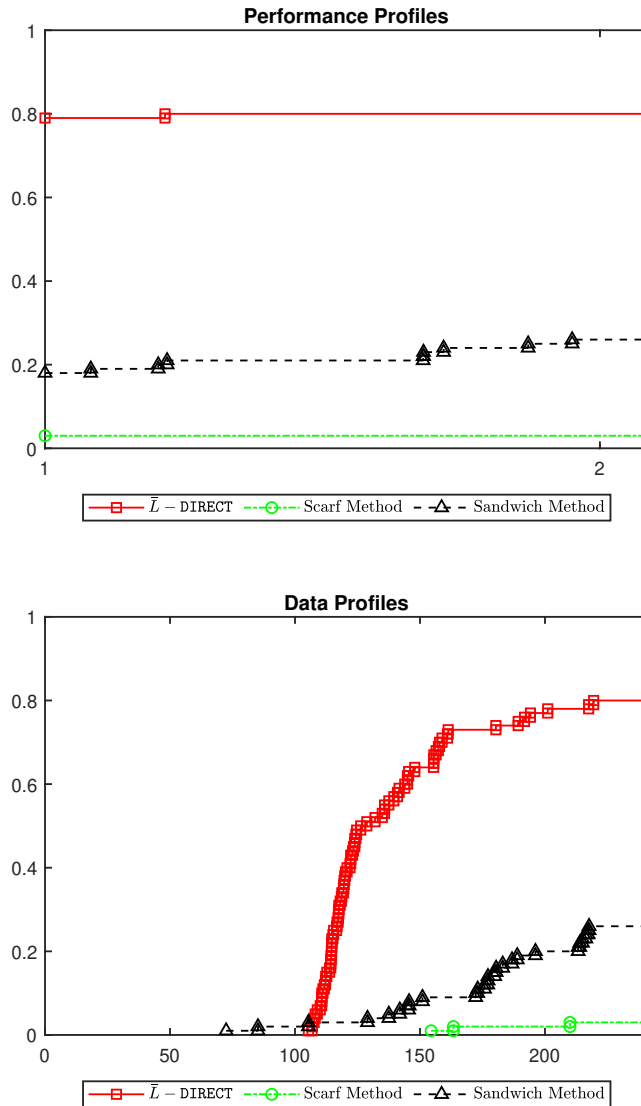


Fig. 8 Comparison with triangulation and simplicial methods: performance and data profiles ($n = 10$).

show simple estimates of their Lipschitz constants that can be exploited in the improved DIRECT algorithm. Preliminary numerical experiments on a set of instances from the literature and sets of randomly generated instances show the effectiveness of our approach for solving non-monotone EPs.

Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgements The work of M. Passacantando and F. Rinaldi has been partially supported by the Italian Government project PRIN2015B5F27W “Nonlinear and Combinatorial Aspects of Complex Networks”. M. Passacantando is member of the Gruppo Nazionale per l’Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA - National Group for Mathematical Analysis, Probability and their Applications) of the Istituto Nazionale di Alta Matematica (INdAM - National Institute of Higher Mathematics).

References

1. F.A. Al-Khayyal. An implicit enumeration procedure for the general linear complementarity problem. *Math. Programming Stud.*, (31):1–20, 1987.
2. E. Allgower and K. Georg. Simplicial and continuation methods for approximating fixed points and solutions to systems of equations. *SIAM Review*, 22(1):28–85, 1980.
3. G. Bigi, M. Castellani, M. Pappalardo, and M. Passacantando. Existence and solution methods for equilibria. *European Journal of Operational Research*, 227(1):1–11, 2013.
4. G. Bigi, M. Castellani, M. Pappalardo, and M. Passacantando. *Nonlinear Programming Techniques for Equilibria*. Springer, Cham, 2019.
5. G. Bigi and M. Passacantando. Twelve monotonicity conditions arising from algorithms for equilibrium problems. *Optimization Methods and Software*, 30(2):323–337, 2015.
6. E. Blum and W. Oettli. From optimization and variational inequalities to equilibrium problems. *Math. student*, 63:123–145, 1994.
7. R. Burachik and G. Kassay. On a generalized proximal point method for solving equilibrium problems in Banach spaces. *Nonlinear Analysis: Theory, Methods & Applications*, 75(18):6456–6464, 2012.
8. O. Chadli, I.V. Konnov, and J.C. Yao. Descent methods for equilibrium problems in a Banach space. *Computers & Mathematics with Applications*, 48(3-4):609–616, 2004.
9. Y. Dai, G. van der Laan, A. J. J. Talman, and Y. Yamamoto. A simplicial algorithm for the nonlinear stationary point problem on an unbounded polyhedron. *SIAM Journal on Optimization*, 1(2):151–165, 1991.
10. C. Dang. *Triangulations and simplicial methods*. Springer, Berlin, 1995.
11. D. Di Lorenzo, M. Passacantando, and M. Sciandrone. A convergent inexact solution method for equilibrium problems. *Optimization Methods and Software*, 29(5):979–991, 2014.
12. G. Di Pillo, G. Liuzzi, S. Lucidi, V. Piccialli, and F. Rinaldi. A DIRECT-type approach for derivative-free constrained global optimization. *Computational Optimization and Applications*, 65(2):361–397, 2016.
13. S. P. Dirkse and M. C. Ferris. The path solver: a nonmonotone stabilization scheme for mixed complementarity problems. *Optimization methods and software*, 5(2):123–156, 1995.
14. B.C. Eaves. Homotopies for computation of fixed points. *Mathematical Programming*, 3(1):1–22, 1972.
15. B.C. Eaves and R. Saigal. Homotopies for computation of fixed points on unbounded regions. *Mathematical Programming*, 3(1):225–237, 1972.
16. K. Fan. A minimax inequality and applications. *Inequalities*, 3:103–113, 1972.
17. S.D. Flåm and A.S. Antipin. Equilibrium programming using proximal-like algorithms. *Mathematical Programming*, 78(1):29–41, 1996.
18. M. Fukushima. Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality. *Mathematical Programming*, 53:99–110, 1992.
19. R. Horst, P.M. Pardalos, and N. Van Thoai. *Introduction to Global Optimization*. Introduction to Global Optimization. Springer US, 2000.
20. D.R. Jones, C.D. Perttunen, and B.E. Stuckman. Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Applications*, 79(1):157–181, 1993.
21. I.V. Konnov. Application of the proximal point method to nonmonotone equilibrium problems. *Journal of Optimization Theory and Applications*, 119(2):317–333, 2003.
22. I.V. Konnov and O.V. Pinyagina. D-gap functions for a class of equilibrium problems in Banach spaces. *Computational Methods in Applied Mathematics Comput. Methods Appl. Math.*, 3(2):274–286, 2003.
23. H.W. Kuhn. Simplicial approximation of fixed points. *Proceedings of the National Academy of Sciences*, 61(4):1238–1242, 1968.
24. H.W. Kuhn and J.G. MacKinnon. Sandwich method for finding fixed points. *Journal of Optimization Theory and Applications*, 17(3):189–204, 1975.
25. N. Langenberg. Interior proximal methods for equilibrium programming: part I. *Optimization*, 62(9):1247–1266, 2013.
26. N. Langenberg. Interior proximal methods for equilibrium programming: part II. *Optimization*, 62(12):1603–1625, 2013.
27. G. Liuzzi, S. Lucidi, and V. Piccialli. Partitioning techniques for global optimization. In V. De Simone, D. Di Serafino, and G. Toraldo, editors, *Recent Advances in Nonlinear Optimization and Equilibrium Problems: A Tribute to Marco D’Apuzzo*, volume 27 of *Quaderni di Matematica, Dipartimento di Matematica, Seconda Università di Napoli*, pages 253–272. Aracne Editrice S.r.l., 2012.
28. S. Lucidi and M. Sciandrone. A derivative-free algorithm for bound constrained optimization. *Computational Optimization and Applications*, 21(2):119–142, 2002.
29. M. Majig, B. Barsbold, R. Enkhbat, and M. Fukushima. A global optimization approach for solving non-monotone variational inequality problems. *Optimization*, 58(7):871–881, 2009.
30. M. Majig, A.R. Hedar, and M. Fukushima. Hybrid evolutionary algorithm for solving general variational inequality problems. *Journal of Global Optimization*, 38(4):637–651, 2007.
31. G. Mastroeni. Gap functions for equilibrium problems. *Journal of Global Optimization*, 27(4):411–426, 2003.
32. G. Mastroeni. On auxiliary principle for equilibrium problems. In *Equilibrium problems and variational models*, pages 289–298. Springer, 2003.
33. J.J. Moré and S.M. Wild. Benchmarking derivative-free optimization algorithms. *SIAM Journal on Optimization*, 20(1):172–191, 2009 .
34. A. Moudafi. Proximal point algorithm extended to equilibrium problems. *Journal of Natural Geometry*, 15(1-2):91–100, 1999.

35. L.D. Muu and T.D. Quoc. Regularization algorithms for solving monotone Ky Fan inequalities with application to a Nash-Cournot equilibrium model. *Journal of Optimization Theory and Applications*, 142(1):185–204, 2009.
36. T.T.V. Nguyen, J.-J. Strodiot, and V.H. Nguyen. A bundle method for solving equilibrium problems. *Mathematical Programming*, 116(1-2):529–552, 2009.
37. H. Nikaido and K. Isoda. Note on noncooperative convex games. *Pacific Journal of Mathematics*, 5:807–815, 1955.
38. M. Pappalardo, G. Mastroeni, and M. Passacantando. Merit functions: a bridge between optimization and equilibria. *Annals of Operations Research*, 240(1):271–299, 2016.
39. P.M. Pardalos and J.B. Rosen. Global optimization approach to the linear complementarity problem. *SIAM Journal on Scientific and Statistical Computing*, 9(2):341–353, 1988.
40. T.D. Quoc, L.D. Muu, and V.H. Nguyen. Extragradient algorithms extended to equilibrium problems. *Optimization*, 57(6):749–776, 2008.
41. H. Scarf. The approximation of fixed points of a continuous mapping. *SIAM Journal on Applied Mathematics*, 15(5):1328–1343, 1967.
42. H. Scarf. The computation of equilibrium prices: An exposition. volume 2 of *Handbook of Mathematical Economics*, pages 1007 – 1061. Elsevier, 1982.
43. H. Scarf and T. Hansen. *The computation of economic equilibria*. Yale University Press, New Haven, 1973.
44. M.J. Todd. *The computation of fixed points and applications*. Lecture notes in economics and mathematical systems ; 124. Springer-Verlag, 1976.
45. G. van der Laan and A.J.J. Talman. A restart algorithm for computing fixed points without an extra dimension. *Mathematical Programming*, 17(1):74–84, 1979.