Optimization Tools for Solving Equilibrium Problems with Nonsmooth Data

Giancarlo Bigi*   Massimo Pappalardo*[1]   Mauro Passacantando*

*Dipartimento di Informatica, Università di Pisa,

Largo B.Pontecorvo 3, 56127 Pisa, Italia

Communicated by Jean-Pierre Crouzeix

---

[1] Corresponding author: tel: +39 050 2212750; fax: +39 050 2212726. *e-mail address:*

massimo.pappalardo@unipi.it.

Optimization Tools for Solving Equilibrium Problems with Nonsmooth Data

Giancarlo Bigi · **Massimo Pappalardo** ·
Mauro Passacantando

Abstract The paper deals with the gap function approach for equilibrium problems with locally Lipschitz data. The gap function inherits the locally Lipschitz continuity of the data. Hence, the connections between its generalized directional derivatives, monotonicity conditions on the equilibrium bifunction and descent properties can be analysed. In turn, this analysis leads to devise two descent methods. Finally, the results of preliminary numerical tests are reported.

G. Bigi · M. Pappalardo · M. Passacantando

Dipartimento di Informatica, Università di Pisa, Largo B.Pontecorvo 3, 56127 Pisa, Italia

E-mail: massimo.pappalardo@unipi.it (corresponding author)

G. Bigi

E-mail: giancarlo.bigi@unipi.it

M. Passacantando

E-mail: mauro.passacantando@unipi.it

Keywords nonsmooth equilibrium problem · gap functions · monotonicity · descent methods

Mathematics Subject Classification (2000) 90C33 · 49M37

1 Introduction

In this paper we consider the so-called equilibrium problem with nonsmooth data in a finite-dimensional setting, following its mathematical format as given in [1,2]. This format was shaped on the well-known Ky Fan's minimax inequality [3] and has attracted increasing attention ever since its introduction. Indeed, it provides a rather general model that includes scalar and vector optimization, inverse optimization, variational inequalities, fixed point, complementarity, saddle points and noncooperative games as particular cases.

Many classes of methods for solving the equilibrium problem have been developed: fixed point and extragradient methods, descent methods, proximal point and Tikhonov-Browder regularization methods (see, for instance, the survey paper [4]). In this paper we focus on algorithms that are based on descent procedures.

Descent techniques exploit the reformulation of the equilibrium problem as an optimization problem through suitable merit functions (see, for instance, [4,5]), which are generally referred to as gap functions. Many descent type algorithms have been developed supposing that the equilibrium bifunction is continuously differentiable [6–16]. This assumption guarantees the differentiability of the gap function; moreover, convergence results require some kind of

monotonicity assumption on the gradients of the equilibrium bifunction. Entering nonsmoothness brings in some difficulties: the differentiability of the gap function is generally lost and monotonicity conditions have to be addressed through generalized derivatives.

When the nonsmooth equilibrium problem takes the shape of a variational inequality, i.e., the equilibrium bifunction is affine in the second argument, with a Lipschitz operator, the analysis of nonsmooth gap functions leads to error bounds [17–19] and to devise solution methods under the strong monotonicity of the operator [20–22]. The algorithms require the explicit knowledge of the modulus of strong monotonicity [21, 22] or the Lipschitz constant [20].

In the general case, some algorithms have been developed just for those particular problems in which the nonsmooth terms of the bifunction are additively separable [23, 24]. Anyhow, the connections between directional derivatives, monotonicity and descent properties given in [25] pave the way to a general framework for descent type methods. In this paper we deepen the analysis of [25] using the generalized directional derivatives of the equilibrium bifunction and we exploit them to devise descent algorithms for the general case.

The paper is structured in the following way. Section 2 recalls the gap function approach, analyses how the local Lipschitz continuity of the equilibrium bifunction is inherited by the gap function and provides an upper estimate of its generalized directional derivative. Section 3 introduces monotonicity conditions on f through generalized directional derivatives and explores their con-

nections with stationarity and descent properties of the gap function. Section 4 exploits the results of the previous sections to devise two different solution methods and to prove their convergence. Finally, Section 5 reports the results of some preliminary numerical tests.

2 Equilibria and Gap Functions

Given a bifunction $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ and a closed convex set $C \subseteq \mathbb{R}^n$, the format of the *equilibrium problem* reads

$$\text{find} \quad x^* \in C \quad \text{s.t.} \quad f(x^*, y) \geq 0, \qquad \forall \, y \in C. \qquad (EP)$$

Throughout the paper, we suppose the function $f(x, \cdot)$ to be convex and $f(x, x) = 0$ for any $x \in \mathbb{R}^n$. Moreover, if C is bounded and $f(\cdot, y)$ is upper semicontinuous, then the solution set is nonempty, closed and bounded (see, for instance [4, Section 2] for further existence results relaxing the two above assumptions).

A function $\varphi : C \to \mathbb{R}$ is said to be a *gap function* for (EP) if it is non-negative on C, and x^* solves (EP) if and only if $x^* \in C$ and $\varphi(x^*) = 0$. Thus, gap functions allow reformulating an equilibrium problem as a global optimization problem, whose optimal value is known a priori. In order to build gap functions with good properties, it is helpful to consider a continuously differentiable auxiliary bifunction $h : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ satisfying the conditions:

- $h(x, y) \geq 0$ for all $x, y \in C$ and $h(z, z) = 0$ for all $z \in C$,
- $h(x, \cdot)$ is strongly convex for all $x \in C$,

  - $\nabla_y h(z, z) = 0$ for all $z \in C$,

  - $\langle \nabla_x h(x, y) + \nabla_y h(x, y), y - x \rangle \geq 0$ for all $x, y \in C$.

A bifunction with the above properties can be obtained taking $h(x, y) = g(y - x)$ for some continuously differentiable and strongly convex function $g : \mathbb{R}^n \to \mathbb{R}_+$ with $g(0) = 0$.

Given any $\alpha > 0$, the auxiliary problem principle (see, for instance, [26]) guarantees that the solution set of (EP) coincides with the solution set of the following problem:

$$\text{find } x^* \in C \text{ s.t. } f(x^*, y) + \alpha h(x^*, y) \geq 0, \qquad \forall \, y \in C. \qquad (EP_\alpha)$$

Therefore, the value function

$$\varphi_\alpha(x) := \max \{ -f(x, y) - \alpha \, h(x, y) \; : \; y \in C \} \qquad (1)$$

is a gap function for (EP). The optimization problem in (1) has a strongly concave objective function, hence it admits a unique solution $y_\alpha(x)$; moreover, x^* solves *(EP)* if and only if $y_\alpha(x^*) = x^*$ (see [6]). Indeed, the inequality

$$\frac{\alpha \, \tau_x}{2} \|x - y_\alpha(x)\|^2 \leq \varphi_\alpha(x)$$

hold for any $x \in C$, and it follows from [27, Theorem 6.1.2] applied to the strongly convex function $f(x, \cdot) + \alpha h(x, \cdot)$, with $\tau_x > 0$ being the modulus of strong convexity of $h(x, \cdot)$.

If f is continuously differentiable, then φ_α is continuously differentiable as well (see [6]), while if f is nonsmooth and continuous, the differentiability of the gap function is generally lost and only continuity is preserved (see [28]).

However, stronger conditions such as the local Lipschitz continuity of φ_α are needed to develop descent methods for (EP). The following assumption on f provides the right tool.

Assumption (A1) Given any bounded set $D \subseteq C$, the function $f(\cdot, y)$ is locally Lipschitz continuous on C uniformly in $y \in D$.

Indeed, in this framework local Lipschitz continuity is preserved.

Theorem 2.1 *If assumption (A1) holds, then φ_α is locally Lipschitz continuous on C.*

Proof Let $x \in C$, $\varepsilon > 0$ and D the intersection of C with the closed ball $B(y_\alpha(x), \varepsilon)$. Assumption (A1) and the convexity of $f(x, \cdot)$ guarantee the continuity of f on $C \times C$, hence the mapping y_α is continuous [28]. Thus, there exists $\delta > 0$ such that $y_\alpha(u) \in D$ for any $u \in B(x, \delta)$. Since h is continuously differentiable and (A1) holds, there exists a constant $L > 0$ such that

$$\varphi_\alpha(u) - \varphi_\alpha(v) = -f(u, y_\alpha(u)) - \alpha \, h(u, y_\alpha(u)) - \varphi_\alpha(v)$$

$$\leq f(v, y_\alpha(u)) - f(u, y_\alpha(u)) + \alpha \, h(v, y_\alpha(u)) - \alpha \, h(u, y_\alpha(u))$$

$$\leq L \, \|u - v\|$$

holds for any $u, v \in B(x, \delta)$. $\square$

Remark 2.1 Assumption (A1) is clearly satisfied if f is continuously differentiable. More generally, it is satisfied whenever f is locally Lipschitz continuous. For instance, if (EP) is a variational inequality problem, i.e., $f(x, y) =$

$\langle F(x), y - x \rangle$, with the operator $F : \mathbb{R}^n \to \mathbb{R}^n$ being locally Lipschitz continuous, then f is locally Lipschitz continuous.

When the gap function φ_α is locally Lipschitz continuous near $x \in C$, its generalized directional derivative

$$\varphi_\alpha^\circ(x; d) := \limsup_{\substack{z \to x \\ t \downarrow 0}} \left[\varphi_\alpha(z + t\, d) - \varphi_\alpha(z) \right] / t$$

at x in any direction $d \in \mathbb{R}^n$ is finite [29]. In the following, $f_x^\circ((x, y); d)$ denotes the generalized direction derivative of the function $f(\cdot, y)$ at x in the direction d and $\partial_x f(x, y)$ the generalized gradient of $f(\cdot, y)$ at x, i.e.,

$$\partial_x f(x, y) := \{ \xi \in \mathbb{R}^n : \ f_x^\circ((x, y); d) \geq \langle \xi, d \rangle, \quad \forall\, d \in \mathbb{R}^n \},$$

while $f_y'((x, y); d)$ the directional derivative of the convex function $f(x, \cdot)$ at y in the direction d. Accordingly, $\nabla_x h(x, y)$ denotes the gradient of the function $h(\cdot, y)$ at x.

An upper estimate of $\varphi_\alpha^\circ(x; d)$, which is based on the generalized directional derivative of f and the gradient of h, can be achieved relying on the following additional assumption.

Assumption (A2) The graph of the set-valued map $(x, y) \mapsto \partial_x(-f)(x, y)$, i.e., the set

$$\{ (x, y, \xi) \in C \times C \times \mathbb{R}^n : \ \xi \in \partial_x(-f)(x, y) \},$$

is closed.

Theorem 2.2 *If (A1) and (A2) hold, then the generalized directional derivative of the gap function φ_α satisfies the inequality*

$$\varphi_\alpha^\circ(x; d) \leq f_x^\circ((x, y_\alpha(x)); -d) + \alpha \langle \nabla_x h(x, y_\alpha(x)), -d \rangle \qquad (2)$$

for any $x \in C$, any direction $d \in \mathbb{R}^n$ and any $\alpha > 0$.

Proof By definition, there exist two sequences $z^k \to x$ and $t_k \downarrow 0$ such that

$$\varphi_\alpha^\circ(x; d) = \lim_{k \to \infty} (\varphi_\alpha(z^k + t_k d) - \varphi_\alpha(z^k))/t_k.$$

Let $u^k = z^k + t_k d$. The mean value theorem [29, Theorem 2.3.7] guarantees

$$
\begin{aligned}
(\varphi_\alpha(u^k) - \varphi_\alpha(z^k))/t_k &\leq [-f(u^k, y_\alpha(u^k)) + f(z^k, y_\alpha(u^k)) + \\
&\quad -\alpha\, h(u^k, y_\alpha(u^k)) + \alpha\, h(z^k, y_\alpha(u^k))]/t_k \\
&= \langle \xi^k, u^k - z^k \rangle / t_k \qquad\qquad (3)\\
&\quad -\alpha \langle \nabla_x h(z^k + \eta_k\, d, y_\alpha(u^k)), u^k - z^k \rangle / t_k \\
&= \langle \xi^k, d \rangle - \alpha \langle \nabla_x h(z^k + \eta_k\, d, y_\alpha(u^k)), d \rangle,
\end{aligned}
$$

where $\xi^k \in \partial_x(-f)(z^k + \theta_k\, d, y_\alpha(u^k))$ and $\theta_k, \eta_k \in (0, t_k)$. Assumption (A1) and the continuity of y_α guarantee the existence of one index $\bar{k}$ and a constant $L > 0$ such that the function $(-f)(\cdot, y_\alpha(u^k))$ is locally Lipschitz continuous at $z^k + \theta_k\, d$ with constant L for any $k \geq \bar{k}$. Therefore, $\|\xi^k\| \leq L$ holds for any $k \geq \bar{k}$ (see [29, Proposition 2.1.2]). Without loss of generality, we can suppose $\xi^k \to \xi$, hence assumption (A2) implies $\xi \in \partial_x(-f)(x, y_\alpha(x))$. Taking the limit

in (3), the chain of inequalities and equalities

$$\varphi_\alpha^\circ(x; d) \leq \langle \xi, d \rangle + \alpha \langle \nabla_x h(x, y_\alpha(x)), -d \rangle$$

$$\leq (-f)_x^\circ((x, y_\alpha(x)); d) + \alpha \langle \nabla_x h(x, y_\alpha(x)), -d \rangle$$

$$= f_x^\circ((x, y_\alpha(x)); -d) + \alpha \langle \nabla_x h(x, y_\alpha(x)), -d \rangle.$$

follows.   □

Assumption (A2) is satisfied when f is continuously differentiable or concave with respect to the first argument and in the particular case of a variational inequality with a locally Lipschitz operator.

Proposition 2.1 *Assumption (A2) is satisfied if any of the following conditions holds:*

a) f is continuously differentiable on $C \times C$;

b) $f(x, y) = \langle F(x), y - x \rangle$, with F locally Lipschitz continuous on C;

c) $f(\cdot, y)$ is concave on an open convex set D with $C \subseteq D$ for any $y \in C$.

Proof a) By definition, the continuity of $\nabla_x f$ coincides with assumption (A2).

b) Let $y \in C$ be given. Lemma 3.1 in [20] guarantees

$$\partial_x(-f)(x, y) = \partial F(x)^T(x - y) + F(x), \tag{4}$$

where $\partial F(x)$ is the generalized Jacobian of F at x (see [29, Definition 2.6.1]). Suppose $(x_k, y_k) \to (x, y)$, $\xi_k \to \xi$ and $\xi_k \in \partial_x(-f)(x_k, y_k)$. Hence, (4) implies $\xi_k = A_k^T(x_k - y_k) + F(x_k)$ for some $A_k \in \partial F(x_k)$. Since ∂F is upper semicontinuous at x [29, Proposition 2.6.2c], there exists $M > 0$ such that

$\|A_k\| \leq M$ for any $k \in \mathbb{N}$. Thus, we can suppose $A_k \to A$ without any loss of generality. Since the graph of ∂F is closed [29, Proposition 2.6.2b], then $A \in \partial F(x)$. Therefore $\xi_k \to \xi = A^T(x-y) + F(x) \in \partial_x(-f)(x,y)$, showing that assumption (A2) holds.

c) It follows immediately from Theorems 23.4 and 24.5 in [30]. $\square$

3 Stationarity and Descent

Though (EP) can be equivalently formulated as the optimization problem

$$\min \left\{ \varphi_\alpha(x) \ : \ x \in C \right\}, \tag{5}$$

there are still some difficulties to overcome in order to exploit descent methods. In fact, (5) is a nonconvex problem since the gap function φ_α is generally nonconvex: descent algorithms provide just local minima while the required solutions are actually the global minima of (5). A standard way to overcome this issue relies on monotonicity conditions that guarantee all the stationary points of (5) to be global minima (see, for instance, [6,7,11]). When the bifunction f is continuously differentiable, these conditions can be addressed through the gradient of f. Precisely, if f is strictly ∇-monotone on C, i.e.,

$$\langle \nabla_x f(x,y) + \nabla_y f(x,y), x-y \rangle < 0 \qquad \forall\, x,y \in C, x \neq y,$$

the stationary points of (5) coincide with its global minima [6, Theorem 2.1] though the convexity of φ_α may be still missing. When f is just locally Lipschitz, this kind of monotonicity condition can be addressed through its generalized directional derivatives in the following way.

Definition 3.1 f is called *strictly generalized ∇-monotone* on C if

$$f_x^\circ((x,y); x-y) + f_y'((x,y); x-y) < 0, \qquad \forall\, x,y \in C, x \neq y. \qquad (6)$$

Proposition 3.1 f *is strictly generalized ∇-monotone on C if any of the following conditions holds:*

a) f *is continuously differentiable on $C \times C$ and strictly ∇-monotone on C;*

b) $f(x,y) = \langle F(x), y-x \rangle$, *with F locally Lipschitz continuous on C and any matrix $A \in \mathbb{R}^{n \times n}$ in the generalized Jacobian $\partial F(x)$ is positive definite for any $x \in C$;*

c) $f(\cdot, y)$ *is strictly concave on an open set D with $C \subseteq D$ for any $y \in C$;*

d) $f(\cdot, y)$ *is concave and $f(x, \cdot)$ strictly convex on an open convex set D with $C \subseteq D$ for any $x, y \in C$.*

Proof a) By definition, strict and strict generalized ∇-monotonicity coincide.

b) Since $f_y'((x,y); x-y) = \langle F(x), x-y \rangle$ and

$$f_x^\circ((x,y); x-y) = \max_{A \in \partial F(x)} \langle A^T(y-x) - F(x), x-y \rangle, \qquad (7)$$

strictly generalized ∇-monotonicity follows from the positive definiteness of A.

c) The strict concavity of $f(\cdot, y)$ implies

$$
\begin{aligned}
f_x^\circ((x,y); x-y) &= (-f)_x^\circ((x,y); y-x) \\
&= (-f)_x'((x,y); y-x) \\
&= -f_x'((x,y); y-x) \\
&< f(x,y),
\end{aligned}
\qquad (8)
$$

while the convexity of $f(x, \cdot)$ guarantees

$$f_y'((x,y); x-y) \leq -f(x,y). \qquad (9)$$

Hence, (6) follows just summing the inequalities in (8) and (9).

d) The same as *c)* just switching the strict inequality from (8) to (9).   □

Under condition (6) the equivalence between stationarity and global optimality is preserved also in the nonsmooth case.

Theorem 3.1 *Suppose (A1) and (A2) hold. If f is strictly generalized ∇-monotone on C, then the following statements hold:*

a) If $x \in C$ does not solve (EP), then $\varphi_\alpha^\circ(x; y_\alpha(x) - x) < 0$;

b) If x^* is a stationary point of φ_α on C, i.e.,

$$\varphi_\alpha^\circ(x^*; y - x^*) \geq 0, \qquad \forall \ y \in C,$$

then x^* solves (EP).

Proof a) The assumption guarantees $y_\alpha(x) \neq x$. Then, $y_\alpha(x) - x$ satisfies

$$\begin{aligned}
\varphi_\alpha^\circ(x; y_\alpha(x) - x) &\leq f_x^\circ((x, y_\alpha(x)); x - y_\alpha(x)) \\
&\quad + \alpha \langle \nabla_x h(x, y_\alpha(x)), x - y_\alpha(x) \rangle \\
&< -f_y'((x, y_\alpha(x)); x - y_\alpha(x)) \\
&\quad + \alpha \langle \nabla_x h(x, y_\alpha(x)), x - y_\alpha(x) \rangle \qquad (10) \\
&\leq -f_y'((x, y_\alpha(x)); x - y_\alpha(x)) \\
&\quad - \alpha \langle \nabla_y h(x, y_\alpha(x)), x - y_\alpha(x) \rangle \\
&\leq 0,
\end{aligned}$$

where the first inequality is guaranteed by Theorem 2.2, the second by assumption (6), the third by the assumptions on h and the last by the first order optimality condition for $y_\alpha(x)$.

b) Suppose x^* does not solve (EP). Hence, a) implies $\varphi_\alpha^\circ(x^*; y_\alpha(x^*) - x^*) < 0$ in contradiction with the stationarity of x^*. $\qquad\square$

The above theorem provides a descent direction at non stationary points as well: a basic descent procedure can be therefore easily devised (see Algorithm 1 in the next section). When f takes the shape of a mixed variational inequality, the descent property given in Theorem 3.1 a) was already shown in [22, Corollary 4.1].

Weakening strict generalized ∇-monotonicity by replacing the strict inequality in (6) with an inequality, Theorem 3.1 is no longer true even if f is continuously differentiable (see the counterexample in [14]). Indeed, no strict inequality would appear in the chain of inequalities (10).

When f is continuously differentiable, a descent approach not relying on strict ∇-monotonicity has been proposed in [13]: the key assumption on f is the so-called c-monotonicity, i.e.,

$$f(x, y) \geq \langle \nabla_x f(x, y), x - y \rangle \qquad \forall\, x, y \in C.$$

When f is just locally Lipschitz, this kind of monotonicity condition can be addressed through its generalized directional derivatives in the following way.

Definition 3.2 f is called *generalized c-monotone* on C if

$$f(x, y) \geq f_x^\circ((x, y); x - y), \qquad \forall\, x, y \in C. \tag{11}$$

Proposition 3.2 *f is generalized c-monotone on C if any of the following conditions holds:*

a) f is continuously differentiable on $C \times C$ and c-monotone on C.

b) $f(x,y) = \langle F(x), y - x \rangle$, with F locally Lipschitz continuous on C, and any matrix $A \in \mathbb{R}^{n \times n}$ in the generalized Jacobian $\partial F(x)$ is positive semidefinite for any $x \in C$.

c) $f(\cdot, y)$ is concave on an open convex set D with $C \subseteq D$ for any $y \in C$.

Proof a) By definition, c-monotonicity and generalized c-monotonicity coincide for continuously differentiable bifunctions.

b) The positive semidefiniteness of A and (7) imply

$$f_x^\circ((x,y); x - y) = f(x,y) + \max_{A \in \partial F(x)} \langle A^T(y - x), x - y \rangle \leq f(x,y).$$

c) The same argument of Proposition 3.1 c) (see (8)) with an inequality rather than a strict inequality.   □

Notice that generalized c-monotonicity is neither stronger nor weaker than strict generalized ∇-monotonicity (see the counterexamples in [31] for the continuously differentiable case).

Under generalized c-monotonicity the equivalence between stationarity and global optimality is generally lost. Anyhow, descent techniques can be exploited all the same relying on the following inequalities and adjusting the parameter α accordingly.

Theorem 3.2 *Suppose (A1) and (A2) hold. If f is generalized c-monotone on C, then the following statements hold:*

a) *The inequality*

$$\varphi_\alpha^\circ(x; y_\alpha(x) - x) \le -\varphi_\alpha(x) - \alpha\, [h(x, y_\alpha(x)) +$$

$$+ \langle \nabla_x h(x, y_\alpha(x)), y_\alpha(x) - x \rangle] \tag{12}$$

holds for any $x \in C$;

b) *Let $\eta \in\,]0, 1[$. If C is bounded and $x \in C$ does not solve (EP), then there exists $\bar{\alpha} > 0$ such that*

$$-\varphi_\alpha(x) - \alpha\, [h(x, y_\alpha(x)) + \langle \nabla_x h(x, y_\alpha(x)), y_\alpha(x) - x \rangle] \le -\eta\, \varphi_\alpha(x) \tag{13}$$

holds for any $\alpha \in\,]0, \bar{\alpha}[$.

Proof a) The thesis follows from the chain of inequalities and equalities

$$\varphi_\alpha^\circ(x; y_\alpha(x) - x) \le f_x^\circ((x, y_\alpha(x)); x - y_\alpha(x)) + \alpha\, \langle \nabla_x h(x, y_\alpha(x)), x - y_\alpha(x) \rangle$$

$$\le f(x, y_\alpha(x)) + \alpha\, \langle \nabla_x h(x, y_\alpha(x)), x - y_\alpha(x) \rangle$$

$$= -\varphi_\alpha(x) - \alpha\, [h(x, y_\alpha(x)) + \langle \nabla_x h(x, y_\alpha(x)), y_\alpha(x) - x \rangle],$$

where the first inequality is guaranteed by Theorem 2.2, the second by assumption (11), and the equality by the definition of φ_α itself.

b) By contradiction, suppose there exists a sequence $\alpha_k \downarrow 0$ such that

$$-\varphi_{\alpha_k}(x) - \alpha_k\, [h(x, y_{\alpha_k}(x)) + \langle \nabla_x h(x, y_{\alpha_k}(x)), y_{\alpha_k}(x) - x \rangle] > -\eta\, \varphi_{\alpha_k}(x).$$

Then any large enough k satisfies $\alpha_k < 1$ and

$$0 < \varphi_1(x)$$

$$\le \varphi_{\alpha_k}(x)$$

$$< -\alpha_k\, [h(x, y_{\alpha_k}(x)) + \langle \nabla_x h(x, y_{\alpha_k}(x)), y_{\alpha_k}(x) - x \rangle] / (1 - \eta)$$

that is not possible because α_k goes to 0 and C is bounded. $\square$

As a consequence of (12) and (13), $y_\alpha(x) - x$ is a descent direction for φ_α at x whenever x does not solve (EP), and therefore $\varphi_\alpha(x) > 0$, provided that the parameter α is small enough. Therefore, a procedure based on descent directions can be devised updating α whenever necessary (see Algorithm 2 in the next section).

Remark 3.1 As already mentioned, the monotonicity conditions (6) and (11) are not related even when f is continuously differentiable. In our locally Lipschitz framework, (6) coincides with condition (2) of [25] and (11) with condition (8) of [25] if $f(\cdot, y)$ is concave on an open convex set D with $C \subseteq D$ for any $y \in C$.

Remark 3.2 When f takes the shape of a mixed variational inequality, inequality (12) extends the upper estimate of the directional derivative given by (4.7) in [22] under the strong monotonicity of the operator F, which implies that any matrix in the generalized Jacobian $\partial F(x)$ is positive definite for any $x \in C$. Indeed, (12) requires generalized c-monotonicity that amounts to the positive semidefiniteness of Jacobian matrices (see Proposition 3.2 b)) which in turn is equivalent to the monotonicity of F (see, for instance [22, Lemma 2.4]).

4 Algorithms

Both assumption (A1) and (A2) are fundamental in our nonsmooth setting. Therefore, throughout all the section we suppose that they both hold. Notice

that they provide properties that are always met by a continuously differentiable bifunction f (see Propositions 3.1 and 3.2), so that there is no need to underline them in smooth settings.

As briefly anticipated, Theorems 3.1 and 3.2 provide the tools to devise descent type methods under different assumptions, i.e., strict generalized ∇-monotonicity and generalized c-monotonicity.

Given any $\alpha > 0$, strict generalized ∇-monotonicity guarantees that $y_\alpha(x) - x$ is a descent direction unless x solves (EP), that is $y_\alpha(x) = x$. Therefore, a line search along $y_\alpha(x) - x$ can be performed to choose the next iterate until a solution of (EP) is found. This basic idea is detailed in Algorithm 1.

Algorithm 1

0. Choose $\alpha > 0$, $\beta, \gamma \in]0, 1[$, $x^0 \in C$ and set $k = 0$.

1. Compute $y^k = \arg\min\{f(x^k, y) + \alpha\, h(x^k, y) \ : \ y \in C\}$.

2. If $d^k := y^k - x^k = 0$, then STOP.

3. Compute the smallest non-negative integer s such that

$$\varphi_\alpha(x^k + \gamma^s\, d^k) - \varphi_\alpha(x^k) \leq -\beta\, \gamma^{2s}\, \varphi_\alpha(x^k),$$

set $t_k = \gamma^s$, $x^{k+1} = x^k + t_k\, d^k$, $k = k + 1$ and goto Step 1.

Theorem 4.1 *Suppose C is bounded and f is strictly generalized ∇-monotone on C. Then, either Algorithm 1 stops at a solution of (EP) after a finite number of iterations, or it produces a bounded sequence $\{x^k\}$ such that any of its cluster points solves (EP).*

Proof The line search procedure at Step 3 is always finite. By contradiction, suppose there exists an iteration k such that

$$\varphi_\alpha(x^k + \gamma^s\, d^k) - \varphi_\alpha(x^k) > -\beta\,\gamma^{2s}\,\varphi_\alpha(x^k)$$

holds for all $s \in \mathbb{N}$. Then, taking the maximum limit as $s \to +\infty$ yields

$$\varphi_\alpha^\circ(x^k; d^k) \geq \limsup_{s\to\infty}\ \left[\varphi_\alpha(x^k + \gamma^s\, d^k) - \varphi_\alpha(x^k)\right] / \gamma^s \geq 0$$

that contradicts Theorem 3.1 since it guarantees $\varphi_\alpha^\circ(x^k; d^k) < 0$.

If the algorithm stops at x^k after a finite number of iterations, then the stopping criterion guarantees that x^k solves (EP).

Now, suppose the algorithm generates an infinite sequence $\{x^k\}$: the sequence is bounded since x^k is a convex combination of x^{k-1} and $y_\alpha(x^k)$, which both belong to C. Consider any cluster point x^* of the sequence. Taking the appropriate subsequence $\{x^\ell\}$, $x^\ell \to x^*$ holds. Moreover, the continuity of the map y_α and the function φ_α guarantees $d^\ell \to d^* = y_\alpha(x^*) - x^*$ and $\varphi_\alpha(x^\ell) \to \varphi_\alpha(x^*)$. Proving that x^* solves (EP) is equivalent to proving $d^* = 0$ or $\varphi_\alpha(x^*) = 0$. By contradiction, suppose $d^* \neq 0$ and $\varphi_\alpha(x^*) > 0$. Since the sequence $\{\varphi_\alpha(x^k)\}$ is monotone, decreasing and bounded below, it has a limit and hence

$$\lim_{\ell\to\infty}\left[\varphi_\alpha(x^\ell) - \varphi_\alpha(x^{\ell+1})\right] = 0$$

holds as well. Moreover, the step size rule guarantees

$$\varphi_\alpha(x^\ell) - \varphi_\alpha(x^{\ell+1}) \geq \beta\, t_\ell^2\, \varphi_\alpha(x^\ell) > 0.$$

Therefore, $t_\ell \to 0$ as $\ell \to +\infty$ since $\varphi_\alpha(x^*) > 0$. Moreover, the inequality

$$\varphi_\alpha\left(x^\ell + t_\ell\,\gamma^{-1}\, d^\ell\right) - \varphi_\alpha(x^\ell) > -\beta\,(t_\ell\,\gamma^{-1})^2\,\varphi_\alpha(x^\ell) \tag{14}$$

holds for all $\ell \in \mathbb{N}$. Since φ_α is locally Lipschitz continuous by Theorem 2.1, the mean value theorem guarantees

$$\varphi_\alpha \left(x^\ell + t_\ell \, \gamma^{-1} \, d^\ell \right) - \varphi_\alpha(x^\ell) = \langle \xi^\ell, t_\ell \, \gamma^{-1} \, d^\ell \rangle, \tag{15}$$

where ξ^ℓ is a generalized gradient of φ_α at $x^\ell + \theta_\ell \, t_\ell \, \gamma^{-1} \, d^\ell$, holds for some $\theta_\ell \in (0, 1)$. Hence, (14) and (15) imply

$$\langle \xi^\ell, d^\ell \rangle > - \beta \, t_\ell \, \gamma^{-1} \, \varphi_\alpha(x^\ell).$$

On the other hand, by definition ξ^ℓ satisfies

$$\varphi_\alpha^\circ \left(x^\ell + \theta_\ell \, t_\ell \, \gamma^{-1} \, d^\ell; d^\ell \right) \geq \langle \xi^\ell, d^\ell \rangle,$$

and thus

$$\varphi_\alpha^\circ \left(x^\ell + \theta_\ell \, t_\ell \, \gamma^{-1} \, d^\ell; d^\ell \right) > - \beta \, t_\ell \, \gamma^{-1} \, \varphi_\alpha(x^\ell)$$

holds. Since $x^\ell \to x^*$, $d^\ell \to d^*$, and $t_\ell \to 0$, we get $x^\ell + \theta_\ell \, t_\ell \, \gamma^{-1} \, d^\ell \to x^*$. Since φ_α° is upper semicontinuous as function of $(x; d)$ [29, Proposition 2.1.1], taking the limit as $\ell \to +\infty$ yields

$$\varphi_\alpha^\circ(x^*; d^*) \geq \limsup_{\ell \to \infty} \varphi_\alpha^\circ \left(x^\ell + \theta_\ell \, t_\ell \, \gamma^{-1} \, d^\ell; d^\ell \right) \geq 0. \tag{16}$$

On the other hand, Theorem 3.1 ensures $\varphi_\alpha^\circ(x^*; d^*) < 0$ in contradiction with (16). Therefore, x^* solves (EP).   □

When f is generalized c-monotone, the choice of a unique parameter α may not be enough. In fact, $y_\alpha(x) - x$ is not necessarily a descent direction unless α is sufficiently small and the magnitude of smallness depends upon x. Therefore,

the basic idea of Algorithm 1 has to be enhanced: the direction $y_\alpha(x) - x$ is exploited to perform the line search as long as the sufficient decrease condition (13) is satisfied, otherwise α is reduced according to some given scheme. This idea is detailed in Algorithm 2.

Algorithm 2

0. Choose $\gamma, \eta \in]0, 1[$, $\beta \in]0, \eta[$, a sequence $\{\alpha_k\} \downarrow 0$, $x^0 \in C$ and set $k = 1$.

1. Set $z^0 = x^{k-1}$ and $j = 0$.

2. Compute $y^j = \arg\min\{f(z^j, y) + \alpha_k \, h(z^j, y) \; : \; y \in C\}$.

3. If $d^j := y^j - z^j = 0$, then STOP.

4. If $-\varphi_{\alpha_k}(z^j) - \alpha_k \left[h(z^j, y^j) - \langle \nabla_x h(z^j, y^j), z^j - y^j \rangle \right] < -\eta \, \varphi_{\alpha_k}(z^j)$,

   then compute the smallest non-negative integer s such that

   $$\varphi_{\alpha_k}(z^j + \gamma^s \, d^j) - \varphi_{\alpha_k}(z^j) \le -\beta \, \gamma^s \, \varphi_{\alpha_k}(z^j),$$

   set $t_j = \gamma^s$, $z^{j+1} = z^j + t_j \, d^j$, $j = j + 1$ and goto Step 2.

   else set $x^k = z^j$, $k = k + 1$ and goto Step 1.

Theorem 4.2 *Suppose C is bounded and f is generalized c-monotone on C. Then, either Algorithm 2 stops at a solution of (EP) after a finite number of iterations, or it produces either an infinite sequence $\{x^k\}$ or an infinite sequence $\{z^j\}$ such that any of its cluster points solves (EP).*

Proof The line search procedure at Step 4 is always finite. By contradiction, suppose there exist k and j such that

$$\varphi_{\alpha_k}(z^j + \gamma^s \, d^j) - \varphi_{\alpha_k}(z^j) > -\beta \, \gamma^s \, \varphi_{\alpha_k}(z^j)$$

holds for all $s \in \mathbb{N}$. Therefore, taking the maximum limit as $s \to +\infty$ yields

$$\varphi_{\alpha_k}^{\circ}(z^j; d^j) \geq \limsup_{s \to \infty} \gamma^{-s}(\varphi_{\alpha_k}(z^j + \gamma^s d^j) - \varphi_{\alpha_k}(z^j)) \geq -\beta \varphi_{\alpha_k}(z^j),$$

that is not possibile since Theorem 3.2 and $\eta > \beta$ guarantee the inequalities

$$\varphi_{\alpha_k}^{\circ}(z^j; d^j) \leq -\eta \varphi_{\alpha_k}(z^j) < -\beta \varphi_{\alpha_k}(z^j).$$

If the algorithm stops at z^j after a finite number of iterations, then the stopping criterion guarantees that z^j solves (EP).

Now, suppose that the algorithm generates an infinite sequence $\{x^k\}$. Let x^* be a cluster point of $\{x^k\}$: taking the appropriate subsequence $\{x^\ell\}$, then $x^\ell \to x^*$ holds. Since $\alpha_\ell \downarrow 0$, we can suppose $\alpha_\ell \leq 1$ for any ℓ without any loss of generality. Since the rule at Step 4 fails at $z^j = x^\ell$, the inequalities

$$0 < \varphi_1(x^\ell) \leq \varphi_{\alpha_\ell}(x^\ell) < -\alpha_\ell \left[h(x^\ell, y^\ell) + \langle \nabla_x h(x^\ell, y^\ell), y^\ell - x^\ell \rangle \right] / (1 - \eta)$$

hold. Since x^ℓ and y^ℓ belong to the bounded set C, the continuity of h and $\nabla_x h$ guarantee that the sequence $\{h(x^\ell, y^\ell) + \langle \nabla_x h(x^\ell, y^\ell), y^\ell - x^\ell \rangle\}$ is bounded from above. Thus, taking the limit as $\ell \to +\infty$, the continuity of φ_1 implies that $\varphi_1(x^*) = 0$, and therefore x^* solves (EP).

Now, suppose the algorithm generates an infinite sequence $\{z^j\}$ for some fixed k. Therefore, we can set $\alpha = \alpha_k$ as this value does not change anymore, and let z^* be a cluster point of $\{z^j\}$: taking the appropriate subsequence $\{z^\ell\}$, then $z^\ell \to z^*$ holds. The continuity of y_α implies $d^\ell \to d^* = y_\alpha(z^*) - z^*$.

By contradiction, suppose that z^* does not solve (EP), or equivalently $\varphi_\alpha(z^*) > 0$. The step size rule implies

$$\varphi_\alpha(z^\ell) - \varphi_\alpha(z^{\ell+1}) \geq \beta \, t_\ell \, \varphi_\alpha(z^\ell) > 0.$$

Taking the limit as $\ell \to +\infty$ yields $t_\ell \to 0$ since $\varphi_\alpha(z^*) > 0$. Moreover, the inequality

$$\varphi_\alpha\left(z^\ell + t_\ell\,\gamma^{-1}\,d^\ell\right) - \varphi_\alpha(z^\ell) > -\,\beta\,t_\ell\,\gamma^{-1}\,\varphi_\alpha(z^\ell)$$

holds for all $\ell \in \mathbb{N}$ large enough. Since φ_α is locally Lipschitz continuous, the mean value theorem guarantees that there exists $\theta_\ell \in (0,1)$ such that

$$\varphi_\alpha\left(z^\ell + t_\ell\,\gamma^{-1}\,d^\ell\right) - \varphi_\alpha(z^\ell) \leq (t_\ell\,\gamma^{-1})\,\varphi_\alpha^\circ\left(z^\ell + \theta_\ell\,t_\ell\,\gamma^{-1}\,d^\ell; d^\ell\right).$$

Since φ_α° is upper semicontinuous as function of $(z; d)$, the chain of inequalities

$$\varphi_\alpha^\circ(z^*; d^*) \geq \limsup_{\ell \to +\infty} \varphi_\alpha^\circ\left(z^\ell + \theta_\ell\,t_\ell\,\gamma^{-1}\,d^\ell; d^\ell\right) \geq -\beta\,\varphi_\alpha(z^*) \qquad (17)$$

follows. On the other hand, the condition at Step 4 is satisfied for all ℓ, hence

$$-\varphi_\alpha(z^\ell) - \alpha\,\left[h(z^\ell, y^\ell) + \langle \nabla_x h(z^\ell, y^\ell), y^\ell - z^\ell \rangle\right] \leq -\eta\,\varphi_\alpha(z^\ell).$$

Thus, taking the limit as $\ell \to +\infty$, the upper estimate of Theorem 3.2 and $\varphi_\alpha(z^*) > 0$ give

$$\varphi_\alpha^\circ(z^*; d^*) \leq -\varphi_\alpha(z^*) - \alpha\,\left[h(z^*, y_\alpha(z^*)) + \langle \nabla_x h(z^*, y_\alpha(z^*)), d^* \rangle\right]$$
$$\leq -\eta\,\varphi_\alpha(z^*)$$
$$< -\beta\,\varphi_\alpha(z^*),$$

which contradicts (17). Therefore, z^* solves (EP). $\qquad\qquad\square$

Notice that the two algorithms employ slightly different procedures for the line search. Indeed, Theorem 3.2 provides an estimate of the generalized directional derivative of φ_α that can be exploited to further control the decrease, while Theorem 3.1 guarantees just that $y_\alpha(x) - x$ is a descent direction.

5 Numerical Tests

To the best of our knowledge, benchmarks of test problems for (EP) are not yet available for the nonsmooth case. Therefore, we decided to test the two algorithms on a set of mathematical examples with box constraints and bifunctions given by the sum of multiplicatively separable nonsmooth terms with quadratic and bilinear ones. Precisely, we considered (EP) with the constraint $C = [0, b_1] \times \cdots \times [0, b_n]$ and the bifunction

$$f(x, y) = \langle P\,x + Q\,y + r, y - x \rangle + g_1(x)\,g_2(y) - g_1(y)\,g_2(x),$$

where $P, Q \in \mathbb{R}^{n \times n}$ are positive definite matrices, $r \in \mathbb{R}^n$ and

$$g_1(x) = \min\{\langle c_1, x \rangle,\ \langle d_1, x \rangle\}, \quad g_2(x) = \max\{\langle c_2, x \rangle,\ \langle d_2, x \rangle\}$$

for some $c_1, c_2, d_1, d_2 \in \mathbb{R}^n_+$. Notice that the key assumptions (6) and (11) of both algorithms are satisfied since $f(\cdot, y)$ is strongly concave for any $y \in C$ (indeed, the functions g_1 and g_2 are both non-negative on C, g_1 is concave while g_2 is convex, and the term $-\langle Px, x \rangle$ is strongly concave).

Instances have been produced relying on uniformly distributed pseudorandom numbers for the data of the bifunction f and the size of the box constraint C. Moreover, the formulas $P = a\,A\,A^T + b\,I$ and $Q = a'\,B\,B^T + b'\,I$ with pseudorandom entries for the matrices A and B allow producing also uniformly distributed pseudorandom minimum and maximum eigenvalues of P and Q by exploiting the coefficients a, a', b, b'. The choice of the ranges for the data and the eigenvalues are summarized in Table 1.

Table 1 Ranges for uniform distributions.

data	range
A_{ij}, B_{ij}	[0,50]
b_i	[10,15]
c_1, d_1	[0,1]
c_2, d_2	[0.5,2]
r	[-5,5]

eigenvalue	range
$\lambda_{min}(P)$	]0,10]
$\lambda_{max}(P)$	$[\lambda_{min}(P), 5 + \lambda_{min}(P)]$
$\lambda_{min}(Q)$	[0.5,1]
$\lambda_{max}(Q)$	$[\lambda_{min}(Q), 1 + \lambda_{min}(Q)]$

The algorithms have been implemented in MATLAB 7.10.0, choosing the auxiliary bifunction $h(x,y) = \|y - x\|_2^2/2$. Since the evaluation of the gap function φ_α and the computation of $y_\alpha(x)$ amount to solving a nonsmooth optimization problem, derivative-free methods are an appropriate tool. Specifically, direct search methods have been used relying on the built-in function `patternsearch` from the Optimization Toolbox together with its pattern `GSSPositiveBasis2N`. Finally, the value 10^{-3} was used as the threshold for the stopping criterion at step 2 of both algorithms: more precisely, the algorithms stopped whenever $\|d^k\|_\infty$ in Algorithm 1 or $\|d^j\|_\infty$ in Algorithm 2 was less or equal to 10^{-3}.

Computational tests have been carried out with $n = 10$ to analyse the sensitivity of the two algorithms with respect to their parameters and to compare their behaviour.

First, we ran Algorithm 1 for different choices of the parameters α, β and γ on a set of 100 random instances with random starting points. Results are given in Tables 2 and 3: each row reports the average and the minimum and maximum number of iterations, evaluations of the gap function (i.e., opti-

mization problem solved) and evaluations of the bifunction that have been performed. The results suggest that a value of α close to 1 and a large value of β are good choices. The choices $\gamma = 0.5$ and $\gamma = 0.7$ produce comparable performances while $\gamma = 0.9$ seems too large.

Table 2 Algorithm 1 with $\alpha = 1$: sensitivity with respect to β and γ.

β	γ	iterations			opt. problems			function evaluations		
		min	avg	max	min	avg	max	min	avg	max
0.1	0.5	3	18.07	78	3	18.37	78	2714	25008.43	103568
0.1	0.7	3	18.09	78	3	19.01	78	2714	25992.64	103568
0.1	0.9	3	21.07	78	3	24.40	96	2714	33309.90	143393
0.3	0.5	3	14.12	42	3	14.69	42	2714	20238.07	58614
0.3	0.7	3	13.95	42	3	15.41	42	2714	21239.42	58614
0.3	0.9	3	17.41	43	3	22.29	89	2714	30548.21	132419
0.5	0.5	3	11.13	25	3	12.08	25	2714	16967.44	33605
0.5	0.7	3	11.08	25	3	13.16	25	2714	18318.07	38236
0.5	0.9	3	14.25	27	3	20.78	68	2714	28534.42	103033
0.7	0.5	3	9.06	15	3	10.96	16	2714	15512.88	23588
0.7	0.7	3	9.06	15	3	11.85	24	2714	16533.97	38236
0.7	0.9	3	12.30	24	3	20.87	68	2714	28597.91	103033
0.9	0.5	3	8.08	10	3	12.64	19	2714	17797.70	27899
0.9	0.7	3	6.99	13	3	10.46	24	2714	14591.36	38236
0.9	0.9	3	10.35	19	3	21.38	57	2714	29015.21	87819

Afterwards, similar tests have been performed on Algorithm 2 for different choices of the parameters β, η and γ relying of the sequence $\alpha_k = 1/2^k$. Tables

Table 3 Algorithm 1 with $\beta = 0.9$ and $\gamma = 0.7$: sensitivity with respect to α.

	iterations			opt. problems			function evaluations		
α	min	avg	max	min	avg	max	min	avg	max
0.01	2	9.45	26	2	17.48	59	1785	24668.08	87660
0.1	2	9.20	27	2	16.74	52	1758	23556.82	73221
1	3	7.81	26	3	12.60	50	2575	17724.08	78488
10	3	12.13	37	3	21.13	97	3398	26980.33	122493

4 and 5 report the results of the tests and they suggest that close and large values for β and η and $\gamma = 0.5$ or $\gamma = 0.6$ are good choices.

Table 4 Algorithm 2 with $\gamma = 0.6$: sensitivity with respect to β and η.

		iterations			opt. problems			function evaluations		
β	η	min	avg	max	min	avg	max	min	avg	max
0.1	0.3	2	14.94	57	2	15.61	57	1927	21761.03	76493
0.1	0.5	2	14.94	57	2	15.61	57	1927	21761.03	76493
0.1	0.7	2	14.94	57	2	15.61	57	1927	21761.03	76493
0.1	0.9	3	19.58	100	3	20.86	100	2598	29323.04	146863
0.3	0.5	2	13.35	45	2	14.14	45	1927	19809.19	62235
0.3	0.7	2	13.35	45	2	14.14	45	1927	19809.19	62235
0.3	0.9	3	14.70	40	3	16.20	40	2598	23073.93	60288
0.5	0.7	2	10.94	26	2	12.06	26	1927	17128.34	36013
0.5	0.9	3	11.87	24	3	13.81	36	2598	20010.51	53968
0.7	0.9	3	9.83	20	3	12.76	36	2598	18590.27	53968

Finally, Table 6 reports the results of a comparison between Algorithms 1 and 2. According to the previous tests, we set $\alpha = 1$, $\beta = 0.9$ and $\gamma = 0.7$ for

Table 5 Algorithm 2 with $\beta = 0.5$ and $\eta = 0.7$: sensitivity with respect to γ.

γ	iterations			opt. problems			function evaluations		
	min	avg	max	min	avg	max	min	avg	max
0.5	2	9.31	23	2	10.49	23	1860	14290.57	30437
0.6	2	8.91	23	2	10.53	23	1860	14233.29	30437
0.7	2	10.06	23	2	12.90	38	1860	17431.60	51727
0.8	2	11.40	33	2	16.09	64	1860	21791.92	83276
0.9	2	13.29	28	2	23.37	96	1860	31672.35	136625

Algorithm 1 and $\alpha_k = 1/2^k$, $\beta = 0.5$, $\eta = 0.7$ and $\gamma = 0.6$ for Algorithm 2. We chose to focus the comparison on the minimum eigenvalue $\lambda_{min}(P)$ of P since it plays a relevant role in the convergence of the algorithms: indeed, Algorithm 2 converges if $\lambda_{min}(P) \geq 0$, while Algorithm 1 if $\lambda_{min}(P) > 0$. We ran each algorithm on a set of 100 random instances for given ranges to draw $\lambda_{min}(P)$ from. Each row corresponds to the choice of a range and it reports the average number of iterations, optimization problems and evaluations of the bifunctions required by a single instance. As expected, the results show that Algorithm 2 performs better and better than Algorithm 1 as $\lambda_{min}(P)$ gets closer and closer to 0, while their performances become at least comparable as it grows.

6 Conclusions

The paper studies gap functions and descent type methods for nonsmooth equilibrium problems. The gap function inherits local Lipschitz continuity from the equilibrium bifunction f; stationarity and descent properties are achieved

Table 6 Comparison between Algorithm 1 and Algorithm 2.

	iterations		opt. problems		function evaluations	
$\lambda_{min}(P)$	Alg. 1	Alg. 2	Alg. 1	Alg. 2	Alg. 1	Alg. 2
$]0, 0.1]$	18.40	12.98	39.20	15.97	47219.51	17034.09
$[0.1, 0.5]$	15.80	12.06	30.89	14.17	40503.23	15534.95
$[0.5, 1]$	12.85	12.24	23.36	16.09	34692.60	17769.72
$[1, 5]$	11.26	14.80	20.59	25.00	25375.46	25947.82

under new generalized monotonicity conditions on f, which involve generalized derivatives. This analysis leads to devise two globally convergent solution methods. Further work can be carried out trying to improve the results in a few directions.

The evaluation of the gap function is the most demanding task of the algorithms, since it amounts to solving a nonsmooth convex optimization problem. The numerical tests of Section 5 have been performed exploiting derivative-free direct search methods. Different nonsmooth optimization algorithms, such as bundle and dual methods, could be exploited as well, while ad hoc methods might be developed relying on the particular properties of the convex program at hand. Moreover, the algorithms rely on the exact evaluation of the gap function: inexact evaluations together with non-monotone line searches could be considered to make each iteration computationally less expensive (see [15]).

Linear approximations of nonlinear constraints could be exploited to introduce new classes of gap functions, which require to minimize a convex function just over a polyhedron. A combination of descent and penalization techniques could lead to alternative solution algorithms (see [14,16]).

The difference of a pair of gap functions, that is the so-called D-gap functions, could be analysed hopefully leading to solution methods based on unconstrained optimization techniques. Indeed, some interesting results have been developed in the smooth case (see, for instance, [4, Section 3.1.3]). Anyhow, the difference of functions is likely to bring in some additional difficulties in the development of nonsmooth stationarity and descent properties based on generalized monotonicity.

Finally, it would be very interesting to apply the approach of this paper to quasi-equilibria, that is equilibrium problems in which the feasible region changes together with the considered point. Indeed, they are much more challenging problems that have not received much attention up to now (see [32]).

References

1. Muu, L.D., Oettli, W.: Convergence of an adaptive penalty scheme for finding constrained equilibria. Nonlinear Anal. 1**8**, 1159-1166 (1992)

2. Blum, E., Oettli, W.: From optimization and variational inequalities to equilibrium problems. Math. Student **63**, 123-145 (1994)

3. Fan, K.: A minimax inequality and applications. In: Shisha, O. (ed.) Inequalities III, pp. 103-113. Academic Press, New York (1972)

4. Bigi, G., Castellani, M., Pappalardo, M., Passacantando, M.: Existence and solution methods for equilibria. Eur. J. Oper. Res. **227**, 1-11 (2013)

5. Pappalardo, M., Mastroeni, G., Passacantando, M.: Merit functions: a bridge between optimization and equilibria. 4OR **12**, 1-33 (2014)

6. Mastroeni, G.: Gap functions for equilibrium problems. J. Glob. Optim. **27**, 411-426 (2003)

7. Chadli, O., Konnov, I.V., Yao, J.C.: Descent methods for equilibrium problems in a Banach space. Comput. Math. Appl. **48**, 609-616 (2004)

8. Zhang, L., Wu, S.-Y.: An algorithm based on the generalized D-gap function for equilibrium problems. J. Comp. Appl. Math. **231**, 403-411 (2009)

9. Cherugondi, C.: A note on D-gap functions for equilibrium problems. Optimization **62**, 211-226 (2013)

10. Konnov. I.V., Pinyagina, O.V.: D-gap functions for a class of equilibrium problems in Banach spaces. Comput. Methods in Appl. Math. **3**, 274-286 (2003)

11. Konnov, I.V., Ali, M.S.S.: Descent methods for monotone equilibrium problems in Banach spaces. J. Comput. Appl. Math. **188**, 165-179 (2006)

12. Zhang, L., Wu, S.-Y., Fang, S.-C.: Convergence and error bound of a D-gap function based Newton-type algorithm for equilibrium problems. J. Ind. Manag. Opt. **6**, 333-346 (2010)

13. Bigi, G., Castellani, M., Pappalardo, M.: A new solution method for equilibrium problems. Optim. Methods Softw. **24**, 895-911 (2009)

14. Bigi, G., Passacantando, M.: Gap functions and penalization for solving equilibrium problems with nonlinear constraints. Comput. Optim. Appl. **53**, 323-346 (2012)

15. Di Lorenzo, D., Passacantando, M., Sciandrone, M.: A convergent inexact solution method for equilibrium problems. Optim. Methods Softw. **29**, 979-991 (2013)

16. Bigi G., Passacantando M.: Descent and Penalization Techniques for Equilibrium Problems with Nonlinear Constraints. J. Optim. Theory Appl. **164** 804-818 (2015)

17. Ng, K. F., Tan, L. L.: D-gap functions for nonsmooth variational inequality problems. J. Optim. Theory Appl. **133**, 77-97 (2007)

18. Ng, K. F., Tan, L. L.: Error bounds of regularized gap functions for nonsmooth variational inequality problems. Math. Program. **110**, 405-429 (2007)

19. Tan, L. L.: Regularized gap functions for nonsmooth variational inequality problems. J. Math. Anal. Appl. **334**, 1022-1038 (2007)

20. Xu, H.: Regularized gap functions and D-gap functions for nonsmooth variational inequalities. In: Optimization and related topics, pp. 153-176. Kluwer, Dordrecht (2001)

21. Konnov, I.V.: Descent method for nonsmooth variational inequalities. Comput. Math. Math. Phys. **46**, 1186-1192 (2006)

22. Konnov, I.V.: Descent Methods for Mixed Variational Inequalities with Non-Smooth Mappings. Contemp. Math. **568**, 121-138 (2012)

23. Konnov. I.V., Pinyagina, O.V.: Descent method with respect to the gap function for nonsmooth equilibrium problems. Russian Math. **47**, 67-73 (2003)

24. Konnov. I.V., Pinyagina, O.V.: A descent method with inexact linear search for nonsmooth equilibrium problems. Comput. Math. Math. Phys. **48**, 1777-1783 (2008)

25. Castellani M., Pappalardo, M.: Gap function for nonsmooth equilibrium problems, Taiwanese Journal of Mathematics **13**, 1837-1846 (2009)

26. Mastroeni, G.: On auxiliary principle for equilibrium problems. In: Daniele, P., Giannessi, F., Maugeri, A. (eds) Equilibrium problems and variational models, pp. 289-298. Kluwer Academic, Norwell (2003)

27. Hiriart-Urruty, J.-B., Lemarchal, C.: Convex analysis and minimization algorithms I. Springer, Berlin (1993)

28. Hogan, W.: Point-to-set maps in Mathematical Programming. SIAM Rev. **15**, 591-603 (1973)

29. Clarke, F.H.: Optimization and Nonsmooth Analysis. John Wiley and Sons, New York (1983)

30. Rockafellar, R.T.: Convex analysis, Princeton University Press, Princeton (1970)

31. Bigi G., Passacantando M.: Twelve monotonicity conditions arising from algorithms for equilibrium problems, Optim. Methods Softw. **30**, 323-337 (2015)

32. Strodiot, J.J., Nguyen, T.T.V., Nguyen, V.H.: A new class of hybrid extragradient algorithms for solving quasi-equilibrium problems. J. Glob. Optim. 56, 373-397 (2013)