

Descent and Penalization Techniques for Equilibrium Problems with Nonlinear Constraints*

Giancarlo Bigi[†] Mauro Passacantando[‡]

Abstract. This paper deals with equilibrium problems with nonlinear constraints. Exploiting a gap function which relies on a polyhedral approximation of the feasible region, we propose two descent methods. They are both based on the minimization of a suitable exact penalty function, but they use different rules for updating the penalization parameter and they rely on different types of line search. The convergence of both algorithms is proved under standard assumptions.

Keywords. Equilibria, constraint linearization, gap function, exact penalization.

AMS Subject Classification: 90C33, 90C30.

1 Introduction

The equilibrium problem provides a general setting to formulate a large number of problems such as scalar and vector optimization, variational inequality, fixed point, complementarity, saddle points and noncooperative games in a unique format (see [1, 2] and references therein). A huge number of applications have been developed in different areas such as economics, environment, transportation, information technology and telecommunications: some recent papers focused on oligopolistic and spatial price markets [3–5], auction and financial markets [6–8], risk management [9], climate policies [10, 11], traffic and pricing over telecommunication networks or over public roads [12–14], clouding computing [15, 16], power allocation in radio systems [17, 18], internet advertising [19].

Several kinds of methods to solve equilibrium problems have been proposed (see, for instance, the recent survey [1]). One popular approach relies on the reformulation of the equilibrium problem as an optimization problem through appropriate gap or D-gap functions: many ad hoc descent methods for minimizing the chosen gap function have been developed (see [13, 14, 20–28]). Most of them require the computation of the minimum of a convex

*This paper has been published in *Journal of Optimization Theory and Applications*, vol. 164 (3), pp. 804–818, DOI: 10.1007/s10957-013-0473-7, see: <https://link.springer.com/article/10.1007/s10957-013-0473-7>.

[†]Dipartimento di Informatica, Università di Pisa, Largo B.Pontecorvo, 3, 56127 Pisa, Italia. email: giancarlo.biggi@unipi.it

[‡]Dipartimento di Informatica, Università di Pisa, Largo B.Pontecorvo, 3, 56127 Pisa, Italia. email: mauro.passacantando@unipi.it

function over the (convex) feasible region of the equilibrium problem just to evaluate the gap function at a given point. Therefore, this evaluation could be computationally expensive when the feasible region is described by nonlinear (convex) constraints.

Recently, a gap function, which uses a polyhedral approximation of the feasible region, has been introduced in [14]. This paper introduces two descent methods for solving equilibrium problems with nonlinear constraints, exploiting this new gap function. They are both based on a search direction which could be unfeasible, unlike most of the known algorithms. Therefore, some penalization techniques are needed: an exact penalty term is introduced and a descent direction for the penalized gap function is available, provided that the penalization parameter is small enough. It is worthy to remark that the penalization parameter is updated throughout the iterations of the algorithms whenever it is needed.

These new methods have some better features than most of the available descent methods. At each iteration, a convex optimization problem with linear constraints is solved instead of a convex problem with nonlinear constraints, as in [13, 21, 23, 25, 26]. Moreover, the key assumption for convergence is weaker than those for the methods proposed in [21, 25, 26]. Finally, the parameters are bounded away from zero while they might go to zero, and thus give numerical instability, in the methods developed in [13, 14, 23].

The paper is organized as follows: Section 2 recalls some well-known definitions and results about gap functions for equilibrium problems, and a key lemma on the penalized gap function is proved. Section 3 provides the two new solution methods and their convergence is proved under standard assumptions. Finally, Section 4 contains some final remarks and comparisons between the new methods and those already available in the literature.

2 Preliminaries

We consider the following equilibrium problem:

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

where $C \subseteq \mathbb{R}^n$ is closed and convex and $f : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ is a bifunction. Throughout the paper, the following basic assumptions are made:

- The feasible set C is bounded and it is defined by convex inequalities, i.e.,

$$C := \{x \in \mathbb{R}^n : c_i(x) \leq 0, \quad i = 1, \dots, m\},$$

where $c_i : \mathbb{R}^n \rightarrow \mathbb{R}$ are convex functions.

- The c_i 's are twice continuously differentiable and Slater constraint qualification holds, i.e., there is $\hat{x} \in \mathbb{R}^n$ such that $c_i(\hat{x}) < 0$ for all $i = 1, \dots, m$.
- The bifunction f is continuously differentiable, $f(x, \cdot)$ is convex, $f(x, x) = 0$ for all $x \in D$, where D is a bounded polyhedron containing C .

The above assumptions guarantee that (EP) admits at least one solution (see, for instance, [29]).

A function $g : C \rightarrow \mathbb{R}$ is said to be a *gap function* for (EP) iff g is non-negative on C , and x^* solves (EP) if and only if $x^* \in C$ and $g(x^*) = 0$. Thus, gap functions allow to

reformulate an equilibrium problem as a global optimization problem, whose optimal value is known a priori. In order to build gap functions with good smoothness properties, it is helpful to consider a continuously differentiable auxiliary bifunction $h : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ which satisfies the following conditions:

- $h(x, y) \geq 0$ for all $x, y \in D$ and $h(z, z) = 0$ for all $z \in D$,
- $h(x, \cdot)$ is strictly convex for all $x \in D$,
- $\nabla_y h(z, z) = 0$ for all $z \in D$,
- $\langle \nabla_x h(x, y) + \nabla_y h(x, y), y - x \rangle \geq 0$ for all $x, y \in D$.

If all the conditions on f and h which involve D actually hold on the whole \mathbb{R}^n , then any bounded polyhedron D such that $C \subseteq D$ can be considered.

Given any $\alpha > 0$, a well-known gap function (see, for instance, [26]) is

$$\phi_\alpha(x) = - \min_{y \in C} \{f(x, y) + \alpha h(x, y)\}.$$

Several descent methods based on the minimization of the gap function ϕ_α have been developed [13, 14, 21, 25, 26]. However, computing $\phi_\alpha(x)$ requires the solution of a convex optimization problem with nonlinear constraints, which may be computationally expensive.

Recently, the gap function ϕ_α has been modified by replacing the feasible region C by its polyhedral approximation at each considered point [14], namely introducing the function

$$\varphi_\alpha(x) = - \min_{y \in P(x)} \{f(x, y) + \alpha h(x, y)\}, \quad (1)$$

where

$$P(x) = \{y \in D : c_i(x) + \langle \nabla c_i(x), y - x \rangle \leq 0, \quad i = 1, \dots, m\}.$$

The polyhedron D guarantees the boundedness of $P(x)$: in fact, the linearization of the constraints alone could be not enough for it, even though C itself is bounded.

Since the inner optimization problem in (1) has a strictly convex objective function and a bounded feasible region, it admits a unique solution $y_\alpha(x)$. We remark that this modification of the gap function ϕ_α extends to (EP) a similar idea developed in [30] for variational inequalities.

Lemma 2.1 [14] *The following statements hold:*

- a) φ_α is locally Lipschitz continuous on D ;
- b) x^* solves (EP) if and only if $y_\alpha(x^*) = x^*$;
- c) φ_α is a gap function for (EP);
- d) If $x \in C$ does not solve (EP) and f is strictly ∇ -monotone on D , i.e.

$$\langle \nabla_x f(x, y) + \nabla_y f(x, y), y - x \rangle > 0, \quad \forall x, y \in D \text{ with } x \neq y, \quad (2)$$

then $y_\alpha(x) - x$ is a descent direction for φ_α at x .

It is worth noting that it is not possible to replace $P(x)$ by D in (1): φ_α would no longer be a gap function, and also the fixed-point reformulation in the above lemma would not hold.

The above results suggest to exploit $y_\alpha(x) - x$ as a search direction at a given iterate x in the minimization process of the gap function φ_α . However, the direction $y_\alpha(x) - x$ may be unfeasible because $y_\alpha(x)$ belongs to the approximating polyhedron $P(x)$, but not necessarily to the feasible set C . For this reason, the following exact penalty function has been introduced in [14]:

$$\psi_{\alpha,\varepsilon,p}(x) := \varphi_\alpha(x) + \frac{1}{\varepsilon} \|c^+(x)\|_p,$$

where $c^+(x) = (c_1^+(x), \dots, c_m^+(x))$ with $c_i^+(x) = \max\{0, c_i(x)\}$, $\varepsilon > 0$ and $p \in [1, \infty]$. Given any $\alpha > 0$, also the penalty function turns out to be a gap function, provided that the penalization parameter ε is small enough. In fact, well-known results about penalization [31] allow to prove the following key properties.

Lemma 2.2 *Given any $\alpha > 0$ and any $p \in [1, \infty]$, there exists $\bar{\varepsilon} > 0$ such that*

- a) $\psi_{\alpha,\varepsilon,p}(x) \geq 0$ for all $x \in D$,
- b) x^* solves (EP) if and only if $x^* \in D$ and $\psi_{\alpha,\varepsilon,p}(x^*) = 0$,

for all $\varepsilon \in]0, \bar{\varepsilon}[$.

Proof. Consider any compact set D' containing D in its interior, namely $D \subset \text{int } D'$. By [31, Proposition 8 and Theorems 11 and 12], there exists $\bar{\varepsilon} > 0$ such that

$$\text{argmin}\{ \varphi_\alpha(x) : x \in C \} = \text{argmin}\{ \psi_{\alpha,\varepsilon,p}(x) : x \in \text{int } D' \} \quad (3)$$

holds for any $\varepsilon \in]0, \bar{\varepsilon}[$. Consider any global minimizer \hat{x} of φ_α over C : we have both $\hat{x} \in C$ and $\varphi_\alpha(\hat{x}) = 0$, so that $\psi_{\alpha,\varepsilon,p}(\hat{x}) = \varphi_\alpha(\hat{x}) = 0$. Therefore, a) follows immediately since (3) implies that \hat{x} is also a minimizer of $\psi_{\alpha,\varepsilon,p}$ on D . As a consequence, Lemma 2.1 c) and (3) imply that b) holds as well. \square

Next section introduces two new solution methods for (EP) based on the minimization of the penalized gap function $\psi_{\alpha,\varepsilon,p}$, which are both convergent under assumption (2).

Actually, a method based on the minimization of the $\psi_{\alpha,\varepsilon,p}$ has already been proposed in [14]. The basic idea of the algorithm is the following: given values for α and ε , the new iterate is obtained moving away from the current iterate x^k along the direction $y_\alpha(x^k) - x^k$ if the decrease condition

$$\psi_{\alpha,\varepsilon,p}(x^k) - \alpha [h(x^k, y_\alpha(x^k)) + \langle \nabla_x h(x^k, y_\alpha(x^k)), y_\alpha(x^k) - x^k \rangle] \leq -\eta \psi_{\alpha,\varepsilon,p}(x^k) \quad (4)$$

is satisfied for some given $\eta \in]0, 1[$ along with two further technical conditions (related to ε). If this is not the case, a null step is performed simply decreasing both α and ε simultaneously before trying again to move away from x^k with the same procedure. The decrease is run according to sequences of parameters going to zero. Thus, if the algorithm performs an infinite sequence of null steps, it generates a sequence of iterates which converges to a solution of (EP) while the parameters α and ε go to zero.

On the other hand, the assumptions required by this latter method for convergence are different from those used in this paper. In fact, the above method from [14] is based on the following concavity-type condition:

$$f(x, y) + \langle \nabla_x f(x, y), y - x \rangle \geq 0 \quad \forall x, y \in D, \quad (5)$$

which is neither weaker nor stronger than condition (2) (see Examples 2.5 and 3.2 in [13]). Furthermore, the following example satisfies both the conditions (2) and (5), and it provides a case in which the above method could be numerically unstable since the parameters α and ε actually go to zero. On the contrary, this may never happen in the algorithms of this paper since α is kept fixed and ε is always updated a finite number of times.

Example 2.1 Consider (EP) with $n = 2$, $m = 1$,

$$f(x, y) = (x_1 + x_2)y_1 + (x_2 - x_1)y_2 - x_1^2 - x_2^2$$

and $c_1(x) = x_1^2 + x_2^2 - 1$. Therefore, the feasible region C is the unit ball and $x^* = (0, 0)$ is the unique solution of (EP). Consider $h(x, y) = \|y - x\|_2^2/2$ and the box $D = [-1, 1] \times [-1, 1]$ containing the feasible region. Note that f satisfies (2) since

$$\langle \nabla_x f(x, y) + \nabla_y f(x, y), y - x \rangle = \|y - x\|_2^2.$$

Furthermore, $f(\cdot, y)$ is concave for all $y \in \mathbb{R}^n$, hence

$$0 = f(y, y) \leq f(x, y) + \langle \nabla_x f(x, y), y - x \rangle, \quad \forall x, y \in D,$$

holds for all $x, y \in D$, i.e., f satisfies (5) as well.

Three cases can occur running the algorithm devised in [14] (see [14, Theorem 5]):

1. it finds x^* after a finite number of iterations;
2. it generates a sequence which converges to x^* while the parameters α and ε are updated only a finite number of times;
3. it generates a sequence which converges to x^* while α and ε go to zero.

Unless the starting point of the algorithm is x^* itself, the first two cases cannot happen.

The first case requires that $x^* = (0, 0)$ belongs to the segment with x and $y_\alpha(x)$ as extreme points for some x , i.e., $y_\alpha(x) = \gamma x$ for some $\gamma < 0$. Furthermore, $y_\alpha(x)$ is the projection of $z_\alpha(x)$ onto $P(x)$ with

$$z_\alpha(x) = (x_1 - (x_1 + x_2)/\alpha, x_2 - (x_2 - x_1)/\alpha).$$

In fact, we have

$$\begin{aligned} f(x, y) + \alpha h(x, y) &= (x_1 + x_2)y_1 + (x_2 - x_1)y_2 - x_1^2 - x_2^2 + \alpha[(y_1 - x_1)^2 + (y_2 - x_2)^2]/2 \\ &= \alpha[(y_1 - x_1 + (x_1 + x_2)/\alpha)^2 + (y_2 - x_2 + (x_2 - x_1)/\alpha)^2]/2 \\ &= \alpha\|y - z_\alpha(x)\|^2/2, \end{aligned}$$

and therefore

$$y_\alpha(x) = \operatorname{argmin}\{f(x, y) + \alpha h(x, y) : y \in P(x)\} = \operatorname{argmin}\{\|y - z_\alpha(x)\|^2 : y \in P(x)\}.$$

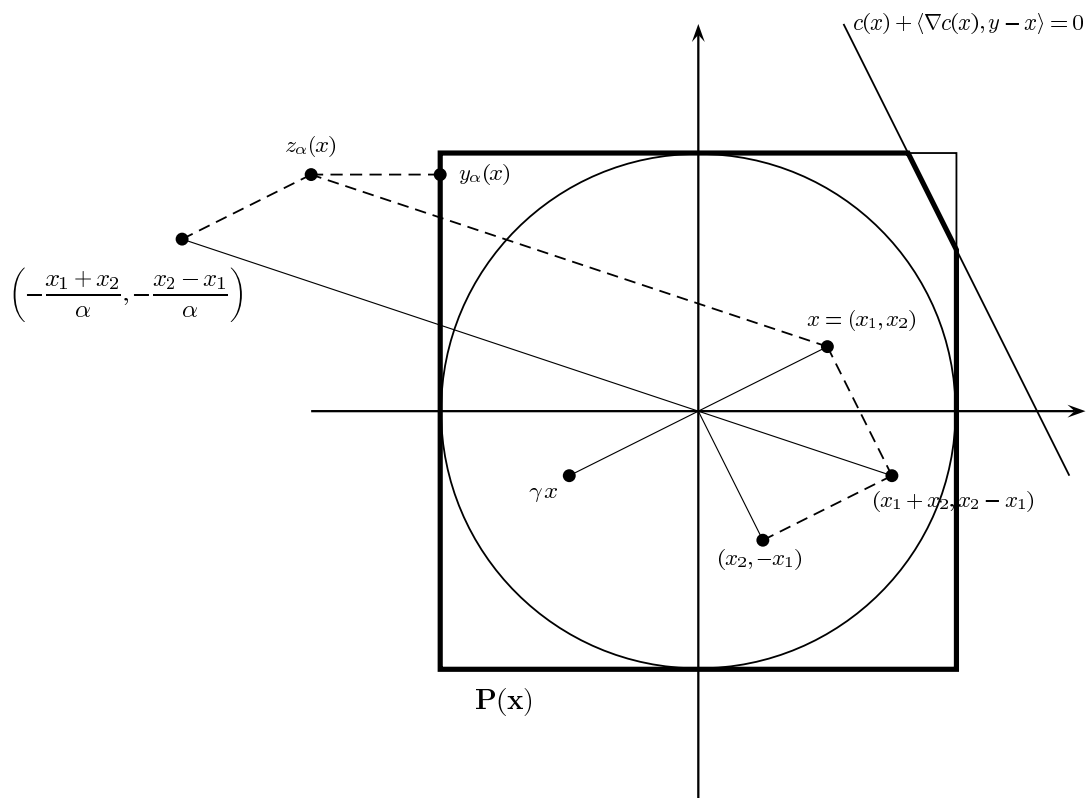


Figure 1: Case 1 is not possible (Example 2.1)

Figure 1 shows that no point γx can ever be the projection of $z_\alpha(x)$ onto $P(x)$. Therefore, $y_\alpha(x) \neq \gamma x$ for all $\gamma < 0$, which means that the first case can not occur.

In the second case the algorithm generates a sequence $x^k \rightarrow x^*$ while α is definitely fixed and the decrease condition (4) holds whenever k is sufficiently large. Clearly, $z_\alpha(x^k) \rightarrow x^*$ since $z_\alpha(x) \rightarrow 0$ as $x \rightarrow 0$. Hence, $z_\alpha(x^k) \in C \subset P(x^k)$ holds for any sufficiently large k , and it implies $y_\alpha(x^k) = z_\alpha(x^k)$ and therefore $\psi_{\alpha,\varepsilon,p}(x^k) = \varphi_\alpha(x^k) = \|x^k\|^2/\alpha$. However, the decrease condition (4) reads

$$0 \leq -\frac{\eta}{\alpha} \|x^k\|^2,$$

which is not possible. Therefore, also the second case can not occur.

As a result, only the third case can occur: the algorithm generates a sequence which converges to x^* while α and ε go to zero.

3 Descent Methods

While $y_\alpha(x) - x$ is a descent direction for φ_α at x , it is not necessarily so for any penalized gap function $\psi_{\alpha,\varepsilon,p}$. Indeed, the key result shows that it is a descent direction also for $\psi_{\alpha,\varepsilon,p}$ at x if the parameter ε is small enough. The generalized directional derivative of $\psi_{\alpha,\varepsilon,p}$ at x along the direction d , i.e.,

$$\psi_{\alpha,\varepsilon,p}^\circ(x; d) := \limsup_{\substack{z \rightarrow x \\ t \downarrow 0}} t^{-1} [\psi_{\alpha,\varepsilon,p}(z + td) - \psi_{\alpha,\varepsilon,p}(z)],$$

provides a convenient tool to check whether d a descent direction. In fact, if $\psi_{\alpha,\varepsilon,p}^\circ(x; d) < 0$, then $\psi_{\alpha,\varepsilon,p}(x + td) < \psi_{\alpha,\varepsilon,p}(x)$ holds whenever $t > 0$ is small enough.

Theorem 3.1 *Let $\alpha > 0$, $p \in [1, \infty]$ and $\Lambda_\alpha(x)$ be the set of all the Lagrange multipliers associated to $y_\alpha(x)$. If (2) holds and $x \in D$ does not solve (EP), then*

$$\psi_{\alpha,\varepsilon,p}^\circ(x; y_\alpha(x) - x) < 0,$$

and therefore $y_\alpha(x) - x$ is a descent direction for $\psi_{\alpha,\varepsilon,p}$ at x , provided that $1/\varepsilon \geq \|\lambda^+\|_q$, where

$$\lambda_i^+ = \begin{cases} \lambda_i, & \text{if } c_i(x) > 0, \\ 0, & \text{otherwise,} \end{cases}$$

for any given $\lambda \in \Lambda_\alpha(x)$, and $\|\cdot\|_q$ is the dual norm of $\|\cdot\|_p$.

Proof. Since x does not solve (EP), then $d := y_\alpha(x) - x \neq 0$. Considering the convex function $v(x) := \|c^+(x)\|_p$, then $v^\circ(x; d)$ coincides with its directional derivative $v'(x; d)$. Thus, the generalized directional derivative of $\psi_{\alpha,\varepsilon,p}$ satisfies the following inequality:

$$\psi_{\alpha,\varepsilon,p}^\circ(x; d) \leq \varphi_\alpha^\circ(x; d) + \frac{1}{\varepsilon} v'(x; d).$$

The following chain of inequalities and equalities holds:

$$\begin{aligned}
\varphi_\alpha^\circ(x; d) &\leq -\langle \nabla_x f(x, y_\alpha(x)) + \alpha \nabla_x h(x, y_\alpha(x)), d \rangle \\
&< \langle \nabla_y f(x, y_\alpha(x)) + \alpha \nabla_y h(x, y_\alpha(x)), d \rangle \\
&= -\sum_{i=1}^m \lambda_i \langle \nabla c_i(x), d \rangle \\
&= \sum_{i=1}^m \lambda_i c_i(x) \\
&\leq \sum_{i=1}^m \lambda_i^+ c_i^+(x) \\
&= \langle \lambda^+, c^+(x) \rangle.
\end{aligned}$$

The first inequality is actually Theorem 2 (b) in [14], while the second one is due to the strict ∇ -monotonicity of $f + \alpha h$. The subsequent equalities follow from the multipliers' rule and the complementarity slackness condition: in fact, we have

$$\nabla_y f(x, y_\alpha(x)) + \alpha \nabla_y h(x, y_\alpha(x)) + \sum_{i=1}^m \lambda_i \nabla c_i(x) = 0 \quad (6)$$

and

$$\lambda_i [c_i(x) + \langle \nabla c_i(x), y_\alpha(x) - x \rangle] = 0, \quad i = 1, \dots, m, \quad (7)$$

since λ is a Lagrange multiplier associated to $y_\alpha(x)$. Finally, the last inequality and equality follow immediately from the definitions. Moreover, the proof of Lemma 4 in [14] shows that $v'(x; d) \leq -v(x)$. Thus, we have:

$$\begin{aligned}
\psi_{\alpha, \varepsilon, p}^\circ(x; d) &< \langle \lambda^+, c^+(x) \rangle - \varepsilon^{-1} \|c^+(x)\|_p \\
&\leq \|\lambda^+\|_q \|c^+(x)\|_p - \varepsilon^{-1} \|c^+(x)\|_p \\
&= (\|\lambda^+\|_q - \varepsilon^{-1}) \|c^+(x)\|_p \\
&\leq 0,
\end{aligned}$$

since $1/\varepsilon \geq \|\lambda^+\|_q$. □

Note that $y_\alpha(x) - x$ is a descent direction for $\psi_{\alpha, \varepsilon, p}$ at a feasible point x for any $\varepsilon > 0$: in fact, $x \in C$ implies $\lambda^+ = 0$ and hence $1/\varepsilon \geq \|\lambda^+\|_q$ holds for all $\varepsilon > 0$. Under assumptions on f other than strict ∇ -monotonicity, this is no longer necessarily true (see [14, Theorem 4]).

The solutions of (EP) coincide with the global minima of $\psi_{\alpha, \varepsilon, p}$ on the set D , provided that ε is small enough (see Lemma 2.2). As a consequence of Theorem 3.1, a further result holds under the strict ∇ -monotonicity of f . In fact, provided that ε is small enough, if (2) holds, then the stationary points of $\psi_{\alpha, \varepsilon, p}$ on D , i.e., those $x^* \in D$ such that

$$\psi_{\alpha, \varepsilon, p}^\circ(x^*; y - x^*) \geq 0, \quad \forall y \in D,$$

solve (EP). Furthermore, an explicit bound on ε is also available.

Corollary 3.1 *Let $\alpha > 0$, $p \in [1, \infty]$. If (2) holds and*

$$1/\varepsilon \geq \sup \{ \|\lambda^+\|_q : \lambda \in \Lambda_\alpha(x), x \in D \},$$

then any stationary point of $\psi_{\alpha, \varepsilon, p}$ on D solves (EP).

Proof. The thesis follows immediately from Theorem 3.1, provided that the bound is finite. Consider the Lagrangian function

$$L(y) := f(x, y) + \alpha h(x, y) + \sum_{i=1}^m \lambda_i [c_i(x) + \langle \nabla c_i(x), y - x \rangle],$$

where $x \in D$ and $\lambda \in \Lambda_\alpha(x)$ are fixed. The complementarity slackness condition (7) implies $L(y_\alpha(x)) = -\varphi_\alpha(x)$, while the multipliers' rule (6) states that $y_\alpha(x)$ is a stationary point of L . Since L is a convex function, then $y_\alpha(x)$ minimizes L on \mathbb{R}^n . Considering any point $\hat{x} \in \mathbb{R}^n$ which satisfies the Slater constraint qualification (that is, $c_i(\hat{x}) < 0$ for all $i = 1, \dots, m$), the following chain of inequalities holds:

$$\begin{aligned} f(x, \hat{x}) + \alpha h(x, \hat{x}) + \varphi_\alpha(x) &\geq \sum_{i=1}^m \lambda_i [-c_i(x) - \langle \nabla c_i(x), \hat{x} - x \rangle] \\ &\geq -\sum_{i=1}^m \lambda_i c_i(\hat{x}) \\ &\geq \zeta \sum_{i=1}^m \lambda_i, \end{aligned}$$

where $\zeta := \min\{-c_1(\hat{x}), \dots, -c_m(\hat{x})\} > 0$. The first inequality simply states $L(y_\alpha(x)) \leq L(\hat{x})$, while the second one follows from the convexity of the functions c_i 's. Therefore, we have

$$\|\lambda\|_1 \leq [f(x, \hat{x}) + \alpha h(x, \hat{x}) + \varphi_\alpha(x)] / \zeta.$$

The maximum of the right-hand side over D is finite since f , h and φ_α are continuous and D is compact. Therefore, all the multipliers $\lambda \in \Lambda_\alpha(x)$ are uniformly bounded over $x \in D$. Since $\|\lambda^+\|_1 \leq \|\lambda\|_1$, the same property holds also for all the corresponding vectors λ^+ : thus, the bound in the statement is finite. \square

Therefore, any local minimization method could be directly applied for solving (EP) exploiting a unique penalized gap function, but the computation of the bound would be required. Actually, this is not necessary: a descent method can be devised moving away from the current iterate x^k along the direction $d^k = y_\alpha(x^k) - x^k$ after updating the penalization parameter ε just in case it is too big for d^k to be a descent direction. Clearly, $d^k = 0$, which means that x^k is a fixed point of the map y_α , guarantees that x^k solves (EP) .

Algorithm 1

- (0) Choose a sequence $\rho_j \downarrow 0$, $p \in [1, \infty]$, $\alpha > 0$, $\beta, \gamma \in]0, 1[$, $x^0 \in D$. Set $\varepsilon = \rho_0$ and $k = 0$.
- (1) Compute $y^k = \arg \min\{f(x^k, y) + \alpha h(x^k, y) : y \in P(x^k)\}$ and λ^k a corresponding Lagrange multiplier vector.
- (2) If $d^k := y^k - x^k = 0$, then STOP.
- (3) while $1/\varepsilon < \|(\lambda^k)^+\|_q$ do
 - set $\varepsilon = \rho_{j+1}$ and $j = j + 1$.
 end

(4) Compute the smallest non-negative integer s such that

$$\psi_{\alpha,\varepsilon,p}(x^k + \gamma^s d^k) - \psi_{\alpha,\varepsilon,p}(x^k) \leq -\beta \gamma^{2s} \|d^k\|_2,$$

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

Theorem 3.2 *If (2) holds, then either Algorithm 1 stops at a solution of (EP) after a finite number of iterations or ε is updated at most a finite number of times and Algorithm 1 produces a bounded sequence $\{x^k\}$, such that any of its cluster points solves (EP) and the sequence of the values $\{\psi_{\alpha,\varepsilon,p}(x^k)\}$ converges to 0.*

Proof. Since there exists $M > 0$ such that $\|\lambda\|_q \leq M$ holds for all $x \in D$ and all $\lambda \in \Lambda_\alpha(x)$ (see the proof of Corollary 3.1), the parameter ε is updated at most a finite number of times and step 3 may never loop indefinitely.

Next, we show that the line search procedure in step 4 is finite. By Theorem 3.1 and the choice of ε at step 3, $\psi_{\alpha,\varepsilon,p}^\circ(x^k; d^k) < 0$ holds for all k . By contradiction, suppose there exists an iteration k such that

$$\psi_{\alpha,\varepsilon,p}(x^k + \gamma^s d^k) - \psi_{\alpha,\varepsilon,p}(x^k) > -\beta \gamma^{2s} \|d^k\|_2$$

holds for all $s \in \mathbb{N}$. Then, taking the limit we get the contradiction

$$\psi_{\alpha,\varepsilon,p}^\circ(x^k; d^k) \geq \limsup_{s \rightarrow \infty} \gamma^{-s} [\psi_{\alpha,\varepsilon,p}(x^k + \gamma^s d^k) - \psi_{\alpha,\varepsilon,p}(x^k)] \geq 0.$$

If the algorithm stops at x^* after a finite number of iterations, then the stopping criterion guarantees that x^* solves (EP) thanks to Lemma 2.1.

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 D . Consider any cluster point x^* of the sequence. Taking the appropriate subsequence $\{x^\ell\}$, we have $x^\ell \rightarrow x^*$. Without any loss of generality, we can assume that ε is constant for all the iterations. Moreover, the continuity of the map y_α guarantees $d^\ell \rightarrow d^* = y_\alpha(x^*) - x^*$. We want to prove $d^* = 0$ and therefore that x^* solves (EP). By contradiction, suppose $d^* \neq 0$. Since the sequence $\{\psi_{\alpha,\varepsilon,p}(x^k)\}$ is monotone, decreasing and bounded below, it has a limit. Hence, we also have

$$\lim_{\ell \rightarrow \infty} [\psi_{\alpha,\varepsilon,p}(x^\ell) - \psi_{\alpha,\varepsilon,p}(x^{\ell+1})] = 0.$$

Moreover, the stepsize rule 4 guarantees

$$\psi_{\alpha,\varepsilon,p}(x^\ell) - \psi_{\alpha,\varepsilon,p}(x^{\ell+1}) \geq \beta t_\ell^2 \|d^\ell\|_2 > 0.$$

Therefore, $t_\ell \rightarrow 0$ as $\ell \rightarrow +\infty$ since $d^* \neq 0$. Moreover, the inequality

$$\psi_{\alpha,\varepsilon,p}(x^\ell + t_\ell \gamma^{-1} d^\ell) - \psi_{\alpha,\varepsilon,p}(x^\ell) > -\beta (t_\ell \gamma^{-1})^2 \|d^\ell\|_2 \quad (8)$$

holds for all $\ell \in \mathbb{N}$. Since $\psi_{\alpha,\varepsilon,p}$ is locally Lipschitz continuous, the mean value theorem guarantees that

$$\psi_{\alpha,\varepsilon,p}(x^\ell + t_\ell \gamma^{-1} d^\ell) - \psi_{\alpha,\varepsilon,p}(x^\ell) = \langle \xi^\ell, t_\ell \gamma^{-1} d^\ell \rangle, \quad (9)$$

where ξ^ℓ is a generalized gradient of $\psi_{\alpha,\varepsilon,p}$ at $x^\ell + \theta_\ell t_\ell \gamma^{-1} d^\ell$, holds for some $\theta_\ell \in]0, 1[$. Hence, (8) and (9) imply

$$\langle \xi^\ell, d^\ell \rangle > -\beta t_\ell \gamma^{-1} \|d^\ell\|_2.$$

On the other hand, we have

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

and thus

$$\psi_{\alpha,\varepsilon,p}^\circ(x^\ell + \theta_\ell t_\ell \gamma^{-1} d^\ell; d^\ell) > -\beta t_\ell \gamma^{-1} \|d^\ell\|_2.$$

Since $x^\ell \rightarrow x^*$, $d^\ell \rightarrow d^*$, and $t_\ell \rightarrow 0$, we get $x^\ell + \theta_\ell t_\ell \gamma^{-1} d^\ell \rightarrow x^*$. Since $\psi_{\alpha,\varepsilon,p}^\circ$ is upper semicontinuous as function of $(x; d)$ (see, for instance, [32]), taking the limit we get

$$\psi_{\alpha,\varepsilon,p}^\circ(x^*; d^*) \geq \limsup_{\ell \rightarrow \infty} \psi_{\alpha,\varepsilon,p}^\circ(x^\ell + \theta_\ell t_\ell \gamma^{-1} d^\ell; d^\ell) \geq 0. \quad (10)$$

Eventually taking a subsequence, $\lambda^\ell \rightarrow \lambda^*$ as $\ell \rightarrow +\infty$ for some $\lambda^* \in \mathbb{R}_+^m$ since the sequence $\{\lambda^\ell\}$ bounded. Moreover, we have $\lambda^* \in \Lambda_\alpha(x^*)$ since the set-valued map Λ_α is closed (see [33, Lemma 2]). Therefore, $1/\varepsilon \geq \|(\lambda^\ell)^+\|_q$ implies $1/\varepsilon \geq \|(\lambda^*)^+\|_q$, and hence Theorem 3.1 ensures $\psi_{\alpha,\varepsilon,p}^\circ(x^*; d^*) < 0$ in contradiction with (10).

Therefore, x^* solves (EP). By Lemma 2.2 we have $\psi_{\alpha,\varepsilon,p}(x^*) = 0$, and thus 0 is the limit of the whole sequence $\{\psi_{\alpha,\varepsilon,p}(x^k)\}$. \square

The algorithm does not require to check that the penalized gap function is positive at the current iterate x^k . Indeed, $\psi_{\alpha,\varepsilon,p}(x^k) < 0$ might happen at some iteration and a descent step could be taken as well. Anyhow, negative values may be met only as long as the penalization parameter ε keeps being updated, hence they are possible at most a finite number of times. In fact, after ε is fixed once and for all, the sequence $\{\psi_{\alpha,\varepsilon,p}(x^k)\}$ is monotone, decreasing and goes to 0, which could not happen with negative values.

The main difference between the above algorithm and the one in [14] is about the parameters' update. Parameters are updated to guarantee that d^k is a descent direction: in Algorithm 1 it is enough to check that ε is small enough, while in the algorithm of [14] also two further conditions must be met and α has to be updated too. This difference affects the behaviour of the parameters meaningfully: in Algorithm 1 α is fixed and ε is updated at most a finite number of times, while in the other algorithm α and ε change simultaneously and may actually go to zero (see Example 2.1).

In general, a different kind of line search can be considered too: the term $\|d\|_2$ may be replaced by the value of the considered gap function (see [13]). Applying this idea in the current framework requires some additional care as the solution strategy is based on the minimization of the penalized gap function $\psi_{\alpha,\varepsilon,p}$. While a descent direction can be obtained in the same way of Algorithm 1, lack of feasibility may create some further troubles. A line search based on the inequality

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

may not work. In fact, the right-hand size of the inequality has to be negative. Therefore, a further check has to be performed to guarantee that the penalized gap function is positive at the current iterate x^k : if it is not so, it is enough to decrease the penalization parameter ε enough to get a positive value (see Lemma 2.2). Anyway, feasibility may be not achieved all the same: the method could provide sequences for which the values of the gap function go

to zero, but whose cluster points are not feasible. Adding the penalty term $\|c^+(x)\|_p$ to the right-hand side of the line search inequality allows to get feasibility as well.

Algorithm 2

- (0) Choose a sequence $\rho_j \downarrow 0$, $p \in [1, \infty]$, $\alpha, \delta > 0$, $\beta, \gamma \in]0, 1[$, $x^0 \in D$. Set $\varepsilon = \rho_0$, $j = 0$ and $k = 0$.
- (1) Compute $y^k = \arg \min\{f(x^k, y) + \alpha h(x^k, y) : y \in P(x^k)\}$ and λ^k a corresponding Lagrange multiplier vector.
- (2) If $d^k := y^k - x^k = 0$, then STOP.
- (3) while $\psi_{\alpha, \varepsilon, p}(x^k) \leq 0$ or $1/\varepsilon < \|(\lambda^k)^+\|_q$ do
 set $\varepsilon = \rho_{j+1}$ and $j = j + 1$.
 end
- (4) Compute the smallest non-negative integer s such that

$$\psi_{\alpha, \varepsilon, p}(x^k + \gamma^s d^k) - \psi_{\alpha, \varepsilon, p}(x^k) \leq -\beta \gamma^{2s} [\psi_{\alpha, \varepsilon, p}(x^k) + \delta \|c^+(x^k)\|_p],$$

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

Theorem 3.3 *If (2) holds, then either Algorithm 2 stops at a solution of (EP) after a finite number of iterations or ε is updated at most a finite number of times and Algorithm 2 produces a bounded sequence $\{x^k\}$, such that any of its cluster points solves (EP) and the sequence of the values $\{\psi_{\alpha, \varepsilon, p}(x^k)\}$ converges to 0.*

Proof. The parameter ε is updated at most a finite number of times since all the multipliers $\lambda \in \Lambda_\alpha(x)$ are uniformly bounded over $x \in D$ (see the proof of Corollary 3.1). Furthermore, arguing as in the proof of Theorem 3.2, it is easy to show that the line search procedure in step 4 is finite.

If the algorithm stops at x^* after a finite number of iterations, then the stopping criterion guarantees that x^* solves (EP) thanks to Lemma 2.1.

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 D . Taking the appropriate subsequence $\{x^\ell\}$, we have $x^\ell \rightarrow x^*$. Without any loss of generality, we can assume that ε is constant for all the iterations. By contradiction, suppose that x^* does not solve (EP). If $x^* \in C$, then

$$\psi_{\alpha, \varepsilon, p}(x^*) + \delta \|c^+(x^*)\|_p = \psi_{\alpha, \varepsilon, p}(x^*) = \varphi_\alpha(x^*) > 0.$$

On the other hand, if $x^* \in D \setminus C$, then $\|c^+(x^*)\|_p > 0$ and $\psi_{\alpha, \varepsilon, p}(x^*) \geq 0$, since $\psi_{\alpha, \varepsilon, p}(x^\ell) > 0$. Therefore, we have

$$\psi_{\alpha, \varepsilon, p}(x^*) + \delta \|c^+(x^*)\|_p > 0$$

also in this case. Since the sequence $\{\psi_{\alpha, \varepsilon, p}(x^k)\}$ is monotone, decreasing and bounded below, it has a limit. Hence, we also have

$$\lim_{\ell \rightarrow \infty} [\psi_{\alpha, \varepsilon, p}(x^\ell) - \psi_{\alpha, \varepsilon, p}(x^{\ell+1})] = 0.$$

Moreover, the stepsize rule 4 guarantees

$$\psi_{\alpha,\varepsilon,p}(x^\ell) - \psi_{\alpha,\varepsilon,p}(x^{\ell+1}) \geq \beta t_\ell^2 [\psi_{\alpha,\varepsilon,p}(x^\ell) + \delta \|c^+(x^\ell)\|_p] > 0. \quad (11)$$

Since

$$\lim_{\ell \rightarrow \infty} \psi_{\alpha,\varepsilon,p}(x^\ell) + \delta \|c^+(x^\ell)\|_p = \psi_{\alpha,\varepsilon,p}(x^*) + \delta \|c^+(x^*)\|_p > 0,$$

then $t_\ell \rightarrow 0$ as $\ell \rightarrow +\infty$. Arguing as in the proof of Theorem 3.2, we get

$$\psi_{\alpha,\varepsilon,p}^\circ(x^*; y_\alpha(x^*) - x^*) \geq 0 \quad (12)$$

and $1/\varepsilon \geq \|(\lambda^*)^+\|_q$ for some $\lambda^* \in \Lambda_\alpha(x^*)$. Therefore, Theorem 3.1 guarantees

$$\psi_{\alpha,\varepsilon,p}^\circ(x^*; y_\alpha(x^*) - x^*) < 0,$$

which contradicts (12).

Therefore, x^* solves (EP). By Lemma 2.2 we have $\psi_{\alpha,\varepsilon,p}(x^*) = 0$, and thus 0 is the limit of the whole sequence $\{\psi_{\alpha,\varepsilon,p}(x^k)\}$. \square

Though Algorithm 2 updates the penalization parameter ε according to a different rule from Algorithm 1, it is still updated at most a finite number of times, thus preventing the possible numerical troubles due to arbitrarily small parameters.

4 Conclusions

In the paper, two globally convergent algorithms for solving equilibrium problems with nonlinear constraints have been developed. They are both based on the minimization of a suitable penalized gap function: at each iteration they solve a convex optimization problem with linear constraints, but the computation of the generalized derivative of the penalized gap function is not needed. The rule to update the penalization parameter and the line search are the core differences between the two algorithm.

This paper extends to equilibrium problems the ideas given in [30] for variational inequalities only. Moreover, explicit rules to update the penalization parameter ε throughout the iterations are given, while the algorithm in [30] requires the a priori knowledge of a suitable fixed ε . Furthermore, the new algorithms perform Armijo-type inexact line searches instead of the rather theoretical exact line search of [30].

The descent methods given in [13,21,23,25,26] do not perform any constraint linearization, and hence convex optimization problems with nonlinear constraints have to be solved at each iteration, while the algorithms of this paper provide this valuable feature. Moreover, the convergence of the methods proposed in [21,25,26] requires the strong ∇ -monotonicity of the equilibrium bifunction (see, for instance, [26, condition (15)]), which is a stronger assumption than the strict ∇ -monotonicity condition (2) used in this paper.

While the methods proposed in [13,14,23] converge under assumptions which are neither stronger nor weaker than condition (2), the behaviour of the regularization and penalization parameters can be compared. In the algorithms of this paper the regularization parameter α is fixed and the penalization parameter ε is updated at most a finite number of times, and thus they are both bounded away from zero. Conversely, α and ε change simultaneously and

may actually go to zero in [14], α is the unique parameter and may go to zero in [13], while α is fixed but ε (which actually plays the role of a further regularization parameter as no penalization is involved) always goes to zero in [23].

References

1. Bigi, G., Castellani, M., Pappalardo, M., Passacantando, M.: Existence and solution methods for equilibria. *Eur. J. Oper. Res.* **227**, 1–11 (2013)
2. Blum, E., Oettli, W.: From optimization and variational inequalities to equilibrium problems. *Math. Student* **63**, 123–145 (1994)
3. Konnov, I.V.: Spatial equilibrium problems for auction-type systems. *Russ. Math.* **52**, 30–44 (2008)
4. Mordukhovich, B.S., Outrata, J.V., Cervinka, M.: Equilibrium problems with complementarity constraints: case study with applications to oligopolistic markets. *Optimization* **56**, 479–494 (2007)
5. Nagurney, A.: Formulation and analysis of horizontal mergers among oligopolistic firms with insights into the merger paradox: a supply chain network perspective. *Comput. Manag. Sci.* **7**, 377–406 (2010)
6. Konnov, I.V.: On variational inequalities for auction market problems. *Optim. Lett.* **1**, 155–162 (2007)
7. Konnov, I.V.: Variational inequalities for modeling auction markets with price mappings. *The Open Operational Research Journal* **2**, 29–37 (2008)
8. Liu, Z., Nagurney, A.: Financial networks with intermediation and transportation network equilibria: a supernetwork equivalence and reinterpretation of the equilibrium conditions with computations. *Comput. Manag. Sci.* **4**, 243–281 (2007)
9. Miller, N., Ruszczyński, A.: Risk-adjusted probability measures in portfolio optimization with coherent measures of risk. *Eur. J. Oper. Res.* **191**, 193–206 (2008)
10. Drouet, L., Haurie, A., Moresino, F., Vial, J.-P., Vielle, M., Viguier, L.: An oracle based method to compute a coupled equilibrium in a model of international climate policy. *Comput. Manag. Sci.* **5**, 119–140 (2008)
11. Forgó, F., J. Fülöp, M. Prill, Game theoretic models for climate change negotiations. *Eur. J. Oper. Res.* **160**, 252–267 (2005)
12. Altman, E., Wynter, L.: Equilibrium, games, and pricing in transportation and telecommunication networks. *Networks and Spatial Economics* **4**, 7–21 (2004)
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. Ardagna, D., Panicucci, B., Passacantando, M.: A game theoretic formulation of the service provisioning problem in cloud systems. In: Proceedings of the 20th international conference on World Wide Web, Hyderabad, India, pp. 177–186 (2011)
16. Ardagna, D., Panicucci, B., Passacantando, M.: Generalized Nash Equilibria for the Service Provisioning Problem in Cloud Systems. *IEEE T. Serv. Comput.* **6**, 429–442 (2013)
17. Pang, J.-S., Scutari, G., Palomar, D.P., Facchinei, F.: Design of cognitive radio systems under temperature-interference constraints: a variational inequality approach. *IEEE T. Signal Proces.* **58**, 3251–3271 (2010)
18. Scutari, G., Palomar, D.P., Barbarossa, S.: Competitive optimization of cognitive radio MIMO systems via game theory. In: Convex optimization in signal processing and communications, Cambridge University Press, Cambridge, pp. 387–442 (2010)
19. Zhao, L., Nagurney, A.: A network equilibrium framework for internet advertising: models, qualitative analysis, and algorithms. *Eur. J. Oper. Res.* **187**, 456–472 (2008)
20. Cherugondi, C.: A note on D-gap functions for equilibrium problems. *Optimization* **62**, 211–226 (2013)
21. Chadli, O., Konnov, I.V., Yao, J.C.: Descent methods for equilibrium problems in a Banach space. *Comput. Math. Appl.* **48**, 609–616 (2004)
22. Di Lorenzo, D., Passacantando, M., Sciandrone, M., A convergent inexact solution method for equilibrium problems, *Optim. Methods Softw.*, DOI: 10.1080/10556788.2013.796376
23. Konnov, I.V., Ali, M.S.S.: Descent methods for monotone equilibrium problems in Banach spaces. *J. Comput. Appl. Math.* **188**, 165–179 (2006)
24. Konnov, I.V., Pinyagina, O.V.: D-gap functions for a class of equilibrium problems in Banach spaces. *Comput. Methods Appl. Math.* **3**, 274–286 (2003)
25. 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)
26. Mastroeni, G.: Gap functions for equilibrium problems. *J. Glob. Optim.* **27**, 411–426 (2003)
27. Zhang, L., Han, J.Y.: Unconstrained optimization reformulations of equilibrium problems. *Acta Math. Sin. (Engl. Ser.)* **25**, 343–354 (2009)
28. Zhang, L., Wu, S.-Y.: An algorithm based on the generalized D-gap function for equilibrium problems. *J. Comput. Appl. Math.* **231**, 403–411 (2009)
29. Fan, K.: A minimax inequality and applications. In: Inequalities, III, pp. 103–113. Academic Press, New York (1972)
30. Taji, K., Fukushima, M.: A new merit function and a successive quadratic programming algorithm for variational inequality problems. *SIAM J. Optim.* **6**, 704–713 (1996)

31. Di Pillo, G., Facchinei, F.: Exact penalty functions for nondifferentiable programming problems. In: *Nonsmooth Optimization and Related Topics*, pp. 89–107. Plenum, New York (1989)
32. Clarke, F.H.: *Optimization and Nonsmooth Analysis*. John Wiley and Sons, New York (1983)
33. Hogan, W.: Directional derivatives for extremal-value functions with applications to the completely convex case. *Oper. Res.* **21**, 188–209 (1973)