Università di Pisa
Dipartimento di Informatica

Technical Report: TR-12-13

Descent and Penalization Techniques for Equilibrium Problems with Nonlinear Constraints

Giancarlo Bigi Mauro Passacantando

December 6, 2012

Address: Largo B.Pontecorvo, 3, 56127 Pisa, Italia.
Tel: +39 050 2212700 Fax: +39 050 2212726

Descent and Penalization Techniques for Equilibrium Problems with Nonlinear Constraints

Giancarlo Bigi Mauro Passacantando

Dipartimento di Informatica, Università di Pisa

Largo B.Pontecorvo, 3, 56127 Pisa, Italia.

{*giancarlo.bigi, mpassacantando*}*@di.unipi.it*

Abstract

This paper deals with equilibrium problems with nonlinear constraints. Exploiting the gap function recently introduced in [1], which rely 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*

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 [2, 3] 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 [4, 5, 6], auction and financial markets [7, 8, 9], risk management [10], climate policies [11, 12], traffic and pricing over telecommunication networks or over public roads [1, 13, 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 [2]). 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 [1, 14, 20, 21, 22, 23, 24, 25, 26, 27]). Most of them require the computation of the minimum of a convex 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 [1]. 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 [14, 21, 22, 24, 25]. Moreover, the key assumption for convergence is weaker than those for the methods proposed in [21, 24, 25]. Finally, the parameters are bounded away from zero while they might go to zero, and thus give numerical instability, in the methods developed in [1, 14, 22].

The paper is organized as follows: Section 2 recalls some well-known definitions and results about gap functions for equilibrium problems. 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:

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

where $C \subseteq \mathbb{R}^n$ is closed and convex and $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ 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.,

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

  where $c_i : \mathbb{R}^n \to \mathbb{R}$ 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 $i = 1, \ldots, m$.

- 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 [28]).

A function $g : C \to \mathbb{R}$ is said to be a *gap function* for (EP) if 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 $h : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ 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 $x, y \in D$.

Given any $\alpha > 0$, a well-known gap function (see, for instance, [25]) 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 [1, 14, 21, 24, 25]. However, computing $\phi_\alpha(x)$ require 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 [1], namely introducing the function

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

where

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

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 [29] for variational inequalities.

Lemma 2.1 *[1] The following statements hold:*

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

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

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

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 [1]:

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

where $c^+(x) = (c_1^+(x), \ldots, c_m^+(x))$ 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, the results about penalization in [30] (particularly, Proposition 8 and Theorems 11 and 12) allow to prove the following key properties just arguing as in the proof of Lemma 3 in [1].

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 $\varepsilon \in (0, \bar{\varepsilon})$.

The 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 [1]. However, the assumptions required by this latter method for convergence are different from those used in this paper. In fact, the method in [1] is based on the following concavity-type condition:

$$f(x, y) + \langle \nabla_x f(x, y), y - x \rangle \geq 0 \qquad \forall\, x, y \in D, \tag{3}$$

which is neither weaker nor stronger than condition (2). Moreover the following example shows that, if conditions (2) and (3) are both satisfied, then the method given in [1] could be numerically unstable because the parameter α might go to zero. This may never happen in the algorithms of this paper since the parameter α is kept fixed.

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)*. Notice that f satisfies both (2) and (3) since $f(\cdot, y)$ is concave for all y and

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

Consider $h(x, y) = [(x_1 - y_1)^2 + (x_2 - y_2)^2]/2$ and the box $D = [-1, 1] \times [-1, 1]$ containing the feasible region. Three cases are possible: either the algorithm of [1] finds x^* after a finite number of iterations, or it generates a sequence converging to x^* while the parameter α is updated only a finite number of times, or it generates a sequence converging to x^* while α goes to zero (see [1, Theorem 5]). We now show that the first two possibilities cannot happen.

Since $y_\alpha(x)$ is actually the projection of $z_\alpha(x) = (x_1 - (x_1 + x_2)/\alpha, x_2 - (x_2 - x_1)/\alpha)$ onto $P(x)$, it is not possible to have $y_\alpha(x) = \gamma x$ with $\gamma < 0$ for any $x \in D$, which would be necessary for the algorithm to identify x^* at a given iteration.

It is also impossible that the parameter α is updated only a finite number of times, being therefore definitely fixed. In fact, given any $\alpha > 0$, we have $z_\alpha(x) \to 0$ as $x \to 0$, and therefore $z_\alpha(x) \in P(x)$ whenever $\|x\|$ is small enough. In this situation, $y_\alpha(x) = z_\alpha(x)$ and the decrease condition (13) of [1] does not hold as it reads $\|x\| < 0$. Therefore, the algorithm would decrease α and keep the current iterate fixed until $z_\alpha(x) \notin P(x)$.

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 show 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.,

$$\psi^\circ_{\alpha,\varepsilon,p}(x;d) := \limsup_{\substack{z \to x \\ t \downarrow 0}} \; t^{-1} \left[\psi_{\alpha,\varepsilon,p}(z + t\,d) - \psi_{\alpha,\varepsilon,p}(z)\right],$$

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

Theorem 3.1 *Let $\alpha > 0$ and $p \in [1, \infty]$. If (2) holds and $x \in D$ does not solve (EP), then*

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

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

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

$\lambda \in \mathbb{R}^m_+$ is any Lagrange multiplier associated to $y_\alpha(x)$ 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:

$$\psi^\circ_{\alpha,\varepsilon,p}(x;d) \leq \varphi^\circ_\alpha(x;d) + \frac{1}{\varepsilon}\, v'(x;d).$$

The following chain of inequalities and equalities hold:

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

The first inequality is actually Theorem 2 (b) in [1] 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

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

and

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

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 [1] shows that $v'(x;d) \leq -v(x)$. Thus, we have:

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

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

Notice 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 $1/\varepsilon \geq \left\|\lambda^+\right\|_q$ holds for all $\varepsilon > 0$. Under assumptions on f other than strict ∇-monotonicity, this is no longer necessarily true (see [1, 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

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

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

Corollary 3.1 *Let $\alpha > 0$, $p \in [1, \infty]$ and $\Lambda_\alpha(x)$ be the set of all the Lagrange multipliers associated to $y_\alpha(x)$. If (2) holds and*

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

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

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 Algorithm 1 either stops at a solution of (EP) after a finite number of iterations or produces a sequence $\{x^k\}$ such that any of its cluster points solves (EP).*

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 [31, Lemma 3.3]), 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 $\psi^\circ_{\alpha,\varepsilon,p}(x^k; d^k) < 0$ 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

$$\psi^\circ_{\alpha,\varepsilon,p}(x^k; d^k) \geq \limsup_{s\to\infty}\ \gamma^{-s}\ \left[\psi_{\alpha,\varepsilon,p}(x^k + \gamma^s\, d^k) - \psi_{\alpha,\varepsilon,p}(x^k)\right] \geq 0.$$

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 $\{x^k\}$ and consider any cluster point x^* of the sequence. Taking the appropriate subsequence $\{x^\ell\}$, we have $x^\ell \to x^*$. Without loss of generality we can assume that ε is constant for all the iterations. Moreover, the continuity of the map y_α guarantees $d^\ell \to d^* = y_\alpha(x^*) - x^*$. 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 descreasing and bounded below, we have

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

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

$$\psi_{\alpha,\varepsilon,p}\left(x^\ell + t_\ell\, \gamma^{-1}\, d^\ell\right) - \psi_{\alpha,\varepsilon,p}(x^\ell) > -\,\beta\, (t_\ell\, \gamma^{-1})^2\, \|d^\ell\|_2 \tag{4}$$

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

$$\psi_{\alpha,\varepsilon,p}\left(x^\ell + t_\ell\, \gamma^{-1}\, d^\ell\right) - \psi_{\alpha,\varepsilon,p}(x^\ell) = \langle \xi^\ell, t_\ell\, \gamma^{-1}\, d^\ell \rangle, \tag{5}$$

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, (4) and (5) imply

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

On the other hand, we have

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

and thus

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

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 $\psi^\circ_{\alpha,\varepsilon,p}$ is upper semicontinuous as function of $(x; d)$ (see for instance [32]), taking the limit we get

$$\psi^\circ_{\alpha,\varepsilon,p}(x^*; d^*) \geq \limsup_{\ell\to\infty} \psi^\circ_{\alpha,\varepsilon,p}\left(x^\ell + \theta_\ell\, t_\ell\, \gamma^{-1}\, d^\ell; d^\ell\right) \geq 0. \tag{6}$$

Eventually taking a subsequence, $\lambda^\ell \to \lambda^*$ as $\ell \to +\infty$ 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 $\psi^\circ_{\alpha,\varepsilon,p}(x^*; d^*) < 0$ in contradiction with (6). □

The main difference between the above algorithm and the one in [1] 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 [1] 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 $\|d\|$ may be replaced by the value of the considered gap function (see [14]). 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

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

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 Algorithm 2 either stops at a solution of (EP) after a finite number of iterations or produces a sequence $\{x^k\}$ such that any of its cluster points solves (EP).*

Proof. Lemma 3.3 in [31] and Lemma 2.2 allow to check that ε is updated at most a finite number of times. 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 $\{x^k\}$ and consider any cluster point x^* of the sequence. Taking an appropriate subsequence $\{x^\ell\}$, we have $x^\ell \to x^*$. Without 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 descreasing and bounded below, we have

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

Moreover, the stepsize rule 4 guarantees

$$\psi_{\alpha,\varepsilon,p}(x^\ell) - \psi_{\alpha,\varepsilon,p}(x^{\ell+1}) \geq \beta\, t_\ell^2 \left[\psi_{\alpha,\varepsilon,p}(x^\ell) + \delta\|c^+(x^\ell)\|_p\right] > 0. \tag{7}$$

Since

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

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

$$\psi^\circ_{\alpha,\varepsilon,p}(x^*; y_\alpha(x^*) - x^*) \geq 0 \tag{8}$$

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

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

which contradicts (8). □

Though Agorithm 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 [29] for variational inequalities only. Moreover, explicit rules to update the penalization parameter ε throughout the iterations are given, while the algorithm in [29] require 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 [29].

The descent methods given in [14, 21, 22, 24, 25] 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, 24, 25] require the strong ∇-monotonicity of the equilibrium bifunction (see, for instance, [25, condition (15)]), which is a stronger assumption than the strict ∇-monotonicity condition (2) used in this paper.

While the methods proposed in [1, 14, 22] 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 [1], α is the unique parameter and may go to zero in [14], 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 [22].

References

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

[2] Bigi, G., Castellani, M., Pappalardo, M., Passacantando, M.: Existence and solution methods for equilibria. Eur. J. Oper. Res., to appear.

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

[4] Konnov, I.V.: Spatial equilibrium problems for auction-type systems. Russ. Math. **52**, 30–44 (2008)

[5] 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)

[6] 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)

[7] Konnov, I.V.: On variational inequalities for auction market problems. Optim. Lett. **1**, 155–162 (2007)

[8] Konnov, I.V.: Variational inequalities for modeling auction markets with price mappings. The Open Operational Research Journal **2**, 29–37 (2008)

[9] 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)

[10] Miller, N., Ruszczynski, A.: Risk-adjusted probability measures in portfolio optimization with coherent measures of risk. Eur. J. Oper. Res. **191**, 193–206 (2008)

[11] 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)

[12] Forgó, F., J. Fülöp, M. Prill, Game theoretic models for climate change negotiations. Eur. J. Oper. Res. **160**, 252–267 (2005)

[13] Altman, E., Wynter, L.: Equilibrium, games, and pricing in transportation and telecommunication networks. Networks and Spatial Economics **4** 7–21 (2004)

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

[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., DOI: 10.1109/TSC.2012.14

[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, DOI: 10.1080/02331934.2011.583987

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

[23] 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)

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

[26] Zhang, L., Han, J.Y.: Unconstrained optimization reformulations of equilibrium problems. Acta Math. Sin. (Engl. Ser.) **25**, 343–354 (2009)

[27] 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)

[28] Fan, K.: A minimax inequality and applications. In: Inequalities, III, pp. 103–113. Academic Press, New York (1972)

[29] 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)

[30] 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)

[31] Han, S.P.: A globally convergent method for nonlinear programming. J. Optim. Theor. Appl. **22**, 297–309 (1977)

[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)