Gap functions and penalization for solving equilibrium problems with nonlinear constraints*

Giancarlo Bigi† Mauro Passacantando‡

Abstract. The paper deals with equilibrium problems (EPs) with nonlinear convex constraints. First, EP is reformulated as a global optimization problem introducing a class of gap functions, in which the feasible set of EP is replaced by a polyhedral approximation. Then, an algorithm is given for solving EP through a descent type procedure, which exploits also exact penalty functions, and its global convergence is proved. Finally, the algorithm is tested on a network oligopoly problem with nonlinear congestion constraints.

Keywords. Equilibrium problems gap function exact penalization descent direction.

1 Introduction

In this paper we consider the following *equilibrium problem*

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

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. It is well-known (see e.g. [2]) that *(EP)* provides a general setting which includes several problems such as scalar and vector optimization, variational inequality, fixed point, complementarity, and Nash equilibrium problems in noncooperative games.

Several methods to solve equilibrium problems have been proposed, often extending those originally conceived for optimization problems or variational inequalities (see, for instance, [9, 14]) to the framework of more general equilibrium problems. Well-known solution methods are the so-called descent methods, which are based on the reformulation of the equilibrium problem as a global optimization problem through appropriate gap functions [1, 4, 15, 16, 18, 23, 24]. Most approaches need to minimize a convex function over C in order to evaluate the gap function, and the evaluation could be computationally expensive when the feasible region C is described by nonlinear convex inequalities. Therefore, we introduce a family of gap functions which rely

*This paper has been published in Computational Optimization and Applications, vol. 53 (2012), pp. 323–346, see: `https://rd.springer.com/article/10.1007/s10589-012-9481-z`.

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

‡Dipartimento di Matematica Applicata, Università di Pisa, via Buonarroti 1/c, 56127 Pisa, Italia.

on a polyhedral approximation of C rather than on the feasible region itself, and we develop a method based on the minimization of convex functions over polyhedra. In Section 2 these gap functions are introduced, considering f along with an additional regularizing bifunction, and some properties about their continuity and generalized directional differentiability are given. Moreover, we prove that monotonicity type assumptions on f guarantee that each stationary point of a gap function is actually a solution of the equilibrium problem. This result extends to equilibrium problems a similar one developed in [21] for variational inequalities. Section 3 provides a solution method which does not require the above "stationarity property", relying on a concavity type assumption on f. Moreover, unlike most of the available algorithms, we consider a search direction which could be unfeasible, so that the introduction of an exact penalty function is required. The direction is indeed a descent one if either the regularization or the penalization parameter is small enough. Therefore, the algorithm exploits fixed values for the two parameters as long as they provide a descent direction and it decreases both of them otherwise. Section 4 provides the results of some numerical tests, which have been performed applying the algorithm to a problem of production competition over a network under the Nash-Cournot equilibrium framework.

Throughout all the paper the following basic assumptions are made:

- *The set C is given by the intersection of a bounded polyhedron D and a convex set given through convex inequalities, namely $C = D \cap \widetilde{C}$ with*

$$D = \{y \in \mathbb{R}^n : \langle a_j, y\rangle \leq b_j \quad j = 1, \dots, r_1, \quad \langle a_j, y\rangle = b_j \quad j = r_1 + 1, \dots, r\}$$

  for some $a_j \in \mathbb{R}^n$ and $b_j \in \mathbb{R}$, and

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

  where $c_i : \mathbb{R}^n \to \mathbb{R}$ are twice continuously differentiable (nonlinear) convex functions.

- *The vectors a_j with $j = r_1 + 1, \dots, r$ are linearly independent and there exists $\hat{y} \in D$ such that $\langle a_j, \hat{y}\rangle < b_j$ for all $j = 1, \dots, r_1$ and $c_i(\hat{y}) < 0$ for all $i = 1, \dots, m$.*

- *The bifunction $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, $f(x, \cdot)$ is convex and $f(x, x) = 0$ for all $x \in D$.*

It is well-known (see e.g. [10]) that the above assumptions guarantee the existence of at least one solution of *(EP)*.

2 Gap functions

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 are tools 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 regularity properties, auxiliary bifunctions are generally exploited together with f. While the most used regularizing bifunction is $h(x,y) = ||y-x||_2^2/2$, in this paper we consider any continuously differentiable bifunction $h : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ such that

- $h(x,y) \geq 0$ 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 e.g. [11, 18]) is

$$\phi_\alpha(x) = -\min_{y \in C} \{f(x,y) + \alpha\, h(x,y)\}. \tag{1}$$

Computing $\phi_\alpha(x)$ involves the solution of a convex optimization problem with nonlinear constraints. Thus, we consider a modification of the above gap function, which is obtained replacing the feasible region C by its polyhedral approximation at each considered point, namely

$$\varphi_\alpha(x) = -\min_{y \in P(x)} \{f(x,y) + \alpha\, h(x,y)\}, \tag{2}$$

where

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

Since the constraining functions c_i are convex, then $C \subseteq P(x) \subseteq D$ holds for all $x \in \mathbb{R}^n$, that is $P(x)$ is a bounded polyhedral outer approximation of the feasible region C at the point x. Moreover, $x \in C$ if and only if $x \in P(x)$.

Since the objective function $f(x,\cdot) + \alpha\, h(x,\cdot)$ is strictly convex and $P(x)$ is compact, there exists a unique optimal solution $y_\alpha(x)$ of the optimization problem which defines the gap function (2). Therefore, it can be written as

$$\varphi_\alpha(x) = -f(x, y_\alpha(x)) - \alpha h(x, y_\alpha(x)), \tag{3}$$

and $y_\alpha(x)$ satisfies the optimality condition

$$\langle \nabla_y f(x, y_\alpha(x)) + \alpha \nabla_y h(x, y_\alpha(x)), y - y_\alpha(x) \rangle \geq 0, \qquad \forall\, y \in P(x). \tag{4}$$

Let $\Lambda_\alpha(x)$ denote the set of Lagrange multipliers associated to $y_\alpha(x)$, i.e. the set of the vectors $(\lambda, \mu) \in \mathbb{R}^m_+ \times \mathbb{R}^r$ such that $\mu_1, \dots, \mu_{r_1} \geq 0$ and

$$\begin{cases} \nabla_y f(x, y_\alpha(x)) + \alpha\, \nabla_y h(x, y_\alpha(x)) + \displaystyle\sum_{i=1}^m \lambda_i \nabla c_i(x) + \sum_{j=1}^r \mu_j\, a_j = 0, \\ \lambda_i\, [c_i(x) + \langle \nabla c_i(x), y_\alpha(x) - x \rangle] = 0, \quad i = 1, \dots, m, \\ \mu_j\, [\langle a_j, y_\alpha(x) \rangle - b_j] = 0, \quad j = 1, \dots, r_1. \end{cases}$$

A fixed point reformulation of *(EP)* holds relying on the optimal map y_α, which is single-valued under our assumptions.

Lemma 2.1. *Given any $\alpha > 0$, x^* solves (EP) if and only if $y_\alpha(x^*) = x^*$.*

Proof. If x^* solves *(EP)*, then it minimizes $f(x^*, \cdot)$ over C since $f(x^*, x^*) = 0$. Thus, there exist Lagrange multiplier vectors $\lambda^* \in \mathbb{R}^m_+$ and $\mu^* \in \mathbb{R}^r$ such that $\mu_1, \ldots, \mu_{r_1} \geq 0$ and

$$\begin{cases} \nabla_y f(x^*, x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) + \sum_{j=1}^{r} \mu_j^* a_j = 0, \\ \lambda_i^* \, c_i(x^*) = 0, \quad i = 1, \ldots, m, \\ \mu_j \left[\langle a_j, x^* \rangle - b_j \right] = 0 \quad j = 1, \ldots, r_1, \\ c_i(x^*) \leq 0, \quad i = 1, \ldots, m, \\ \langle a_j, x^* \rangle \leq b_j, \quad j = 1, \ldots, r_1, \\ \langle a_j, x^* \rangle = b_j, \quad j = r_1 + 1, \ldots, r. \end{cases}$$

Setting $g_i(y) = c_i(x^*) + \langle \nabla c_i(x^*), y - x^* \rangle$, then we have $g_i(x^*) = c_i(x^*)$ and $\nabla g_i(y) = \nabla c_i(x^*)$ for all $y \in \mathbb{R}^n$ and $i = 1, \ldots, m$. Hence, the above system can be equivalently stated as

$$\begin{cases} \nabla_y f(x^*, x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{r} \mu_j^* a_j = 0, \\ \lambda_i^* \, g_i(x^*) = 0, \quad i = 1, \ldots, m, \\ \mu_j \left[\langle a_j, x^* \rangle - b_j \right] = 0 \quad j = 1, \ldots, r_1, \\ g_i(x^*) \leq 0, \quad i = 1, \ldots, m, \\ \langle a_j, x^* \rangle \leq b_j, \quad j = 1, \ldots, r_1, \\ \langle a_j, x^* \rangle = b_j, \quad j = r_1 + 1, \ldots, r, \end{cases}$$

which are the Karush-Kuhn-Tucker conditions for the problem of minimizing $f(x^*, \cdot)$ over $P(x^*)$. Since this is a convex problem, x^* solves it and therefore

$$f(x^*, y_\alpha(x^*)) \geq f(x^*, x^*) = 0.$$

Moreover, condition (4) for $y = x = x^*$ reads

$$\langle \nabla_y f(x^*, y_\alpha(x^*)) + \alpha \nabla_y h(x^*, y_\alpha(x^*)), x^* - y_\alpha(x^*) \rangle \geq 0.$$

Since $f(x^*, \cdot) + \alpha h(x^*, \cdot)$ is convex and $f(x^*, x^*) = h(x^*, x^*) = 0$, we have

$$\begin{aligned} 0 &\geq f(x^*, y_\alpha(x^*)) + \alpha\, h(x^*, y_\alpha(x^*)) \\ &+ \langle \nabla_y f(x^*, y_\alpha(x^*)) + \alpha\, \nabla_y h(x^*, y_\alpha(x^*)), x^* - y_\alpha(x^*) \rangle. \end{aligned}$$

The above inequalities imply $h(x^*, y_\alpha(x^*)) = 0$ since h is non-negative on $D \times D$. Moreover, the assumptions on h imply that x^* is the unique minimizer of $h(x^*, \cdot)$ over D and hence $y_\alpha(x^*) = x^*$.

Now, suppose $y_\alpha(x^*) = x^*$. Since $x^* \in P(x^*)$, then $x^* \in C$. Moreover, condition (4) for $x = x^*$ reads

$$\langle \nabla_y f(x^*, x^*), y - x^* \rangle \geq 0, \qquad \forall\, y \in P(x^*).$$

Since $C \subseteq P(x^*)$ and $f(x^*, \cdot)$ is convex, we have

$$f(x^*, y) \geq f(x^*, x^*) + \langle \nabla_y f(x^*, x^*), y - x^* \rangle \geq 0, \qquad \forall\, y \in C,$$

i.e. x^* solves *(EP)*. □

Since the solutions of *(EP)* coincide with the fixed points of the optimal map y_α, they actually minimize φ_α over C.

Theorem 2.1. *Given any $\alpha > 0$, φ_α is a gap function for (EP), i.e.*

a) $\varphi_\alpha(x) \geq 0$ for all $x \in C$;

b) x^* solves (EP) if and only if $x^* \in C$ and $\varphi_\alpha(x^*) = 0$.

Proof. a) If $x \in C$, then $x \in P(x)$. Thus, $\varphi_\alpha(x) \geq -f(x, x) - \alpha h(x, x) = 0$.

b) If x^* solves *(EP)*, then $x^* \in C$ and Lemma 2.1 implies $y_\alpha(x^*) = x^*$. Hence,

$$\varphi_\alpha(x^*) = -f(x^*, x^*) - \alpha\, h(x^*, x^*) = 0.$$

Now, suppose $x^* \in C$ and $\varphi_\alpha(x^*) = 0$. Thus, we have

$$f(x^*, y) + \alpha\, h(x^*, y) \geq -\varphi_\alpha(x^*) = 0, \qquad \forall\, y \in P(x^*).$$

Since $C \subseteq P(x^*)$, x^* minimizes $f(x^*, \cdot) + \alpha h(x^*, \cdot)$ over C and therefore the first order optimality condition reads

$$\langle \nabla_y f(x^*, x^*) + \alpha\, \nabla_y h(x^*, x^*), y - x^* \rangle \geq 0, \qquad \forall\, y \in C.$$

Since $f(x^*, \cdot)$ is convex and $\nabla_y h(x^*, x^*) = 0$, we have

$$f(x^*, y) \geq f(x^*, x^*) + \langle \nabla_y f(x^*, x^*), y - x^* \rangle \geq 0, \qquad \forall\, y \in C,$$

i.e. x^* solves (EP). □

In order to achieve continuity and generalized differentiability properties of φ_α, the map y_α has to be continuous in light of equality (3).

Lemma 2.2. *Given any $\alpha > 0$, the map y_α is continuous on $\mathbb{R}^n$.*

Proof. The set-valued map $x \longmapsto P(x)$ is continuous on $\mathbb{R}^n$ (see [21]). Moreover, f is continuous and the map y_α is single-valued and it is also bounded since $y_\alpha(x) \in P(x) \subseteq D$ for all $x \in \mathbb{R}^n$. Hence, [13, Corollary 8.1] guarantees that y_α is continuous on $\mathbb{R}^n$. □

The gap function φ_α is locally Lipschitz continuous near any $x \in \mathbb{R}^n$, and therefore its generalized directional derivative

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

at x in any direction $d \in \mathbb{R}^n$ is finite. Furthermore, an upper estimate of the directional derivative at x in the particular direction $y_\alpha(x) - x$ is available.

Theorem 2.2. *Let $\alpha > 0$. Then,*

a) φ_α is locally Lipschitz continuous on $\mathbb{R}^n$;

b) the inequality

$$\varphi_\alpha^\circ(x; y_\alpha(x) - x) \leq -\langle \nabla_x f(x, y_\alpha(x)) + \alpha \nabla_x h(x, y_\alpha(x)), y_\alpha(x) - x \rangle \tag{5}$$

holds for any $x \in D$.

Proof. *a)* Introducing $g_i(x, y) := c_i(x) + \langle \nabla c_i(x), y - x \rangle$ for $i = 1, \dots, m$ and

$$g(x, y) = (g_1(x, y), \dots, g_m(x, y)),$$

the optimization problem in (2) can be written as

$$\min \{ \; f(x, y) + \alpha h(x, y) \;\; : \;\; g(x, y) \leq 0, \;\; y \in D\},$$

and its dual problem is

$$\sup \{ \; \inf \; \{f(x, y) + \alpha h(x, y) + \langle u, g(x, y) \rangle \;\; : \;\; y \in D\} \;\; : \;\; u \in \mathbb{R}^m_+\}.$$

By the assumptions $g_i(x, \hat{y}) \leq c_i(\hat{y}) < 0$ for all $x \in D$ and all $i = 1, \dots, m$. Thus, the set $U_\alpha(x)$ of the optimal solutions of the dual problem is nonempty for all $x \in D$ and $(y_\alpha(x), u_\alpha(x))$ is a saddle point of the Lagrangian function

$$L(x, y, u) = f(x, y) + \alpha h(x, y) + \langle u, g(x, y) \rangle,$$

for any $u_\alpha(x) \in U_\alpha(x)$, i.e.

$$L(x, y_\alpha(x), u) \leq L(x, y_\alpha(x), u_\alpha(x)) \leq L(x, y, u_\alpha(x)), \qquad \forall \; y \in D, \; \forall \; u \in \mathbb{R}^m_+.$$

Since $L(x, y_\alpha(x), u_\alpha(x)) = f(x, y_\alpha(x)) + \alpha h(x, y_\alpha(x)) = -\varphi_\alpha(x)$, we get

$$-L(x, y, u_\alpha(x)) \leq \varphi_\alpha(x) \leq -L(x, y_\alpha(x), u), \qquad \forall \; y \in D, \; \forall \; u \in \mathbb{R}^m_+.$$

Similarly, given any $z \in D$ and any $u_\alpha(z) \in U_\alpha(z)$, we have

$$-L(z, y, u_\alpha(z)) \leq \varphi_\alpha(z) \leq -L(z, y_\alpha(z), u), \qquad \forall\, y \in D,\ \forall\, u \in \mathbb{R}^m_+. \tag{6}$$

Therefore, choosing $y = y_\alpha(x)$ and $u = u_\alpha(z)$, we get

$$\begin{aligned}
\varphi_\alpha(x) - \varphi_\alpha(z) &\leq L(z, y_\alpha(x), u_\alpha(z)) - L(x, y_\alpha(x), u_\alpha(z)) \\
&= f(z, y_\alpha(x)) + \alpha h(z, y_\alpha(x)) - [f(x, y_\alpha(x)) + \alpha h(x, y_\alpha(x))] \\
&\quad + \langle u_\alpha(z), g(z, y_\alpha(x)) - g(x, y_\alpha(x))\rangle \\
&\leq f(z, y_\alpha(x)) + \alpha h(z, y_\alpha(x)) - [f(x, y_\alpha(x)) + \alpha h(x, y_\alpha(x))] \\
&\quad + \|u_\alpha(z)\|_2\, \|g(z, y_\alpha(x)) - g(x, y_\alpha(x))\|_2
\end{aligned}$$

Let $\bar{x} \in \mathbb{R}^n$ be fixed. The mean value theorem guarantees that

$$\begin{aligned}
& f(z, y_\alpha(x)) + \alpha h(z, y_\alpha(x)) - [f(x, y_\alpha(x)) + \alpha h(x, y_\alpha(x))] \\
&= \langle \nabla_x f(z', y_\alpha(x)) + \alpha \nabla_x h(z', y_\alpha(x)), z - x\rangle
\end{aligned}$$

holds for some z' in the line segment between z and x. Since y_α, $\nabla_x f$, and $\nabla_x h$ are continuous, there exist $L_1 > 0$ and $\delta_1 > 0$ such that

$$f(z, y_\alpha(x)) + \alpha h(z, y_\alpha(x)) - [f(x, y_\alpha(x)) + \alpha h(x, y_\alpha(x))] \leq L_1\, \|z - x\|_2$$

holds for all $x, z \in B(\bar{x}, \delta_1)$. On the other hand, the functions g_i are continuously differentiable with respect to the first variable, hence there exist $L_2 > 0$ and $\delta_2 > 0$ such that

$$\|g(z, y_\alpha(x)) - g(x, y_\alpha(x))\|_2 \leq L_2\, \|z - x\|_2$$

holds for all $x, z \in B(\bar{x}, \delta_2)$. Moreover, [12, Lemma 2] guarantees that there exist $L_3 > 0$ and $\delta_3 > 0$ such that $\|u_\alpha(z)\|_2 \leq L_3$ holds for all $z \in B(\bar{x}, \delta_3)$ and all $u_\alpha(z) \in U_\alpha(z)$. Therefore, the last three inequalities imply that

$$\varphi_\alpha(x) - \varphi_\alpha(z) \leq (L_1 + L_2\, L_3)\, \|z - x\|_2$$

holds for all $x, z \in B(\bar{x}, \delta)$, where $\delta = \min\{\delta_1, \delta_2, \delta_3\}$.

b) Set $d := y_\alpha(x) - x$ and $z_t := z + t\, d$ for any $z \in \mathbb{R}^n$ and $t > 0$, and consider any $u_\alpha(z_t) \in U_\alpha(z_t)$. Arguing as in a), we get

$$\begin{aligned}
\varphi_\alpha(z_t) - \varphi_\alpha(z) &\leq f(z, y_\alpha(z_t)) + \alpha h(z, y_\alpha(z_t)) - \big[f(z_t, y_\alpha(z_t)) \\
&\quad + \alpha h(z_t, y_\alpha(z_t))\big] + \langle u_\alpha(z), g(z, y_\alpha(z_t)) - g(z_t, y_\alpha(z_t))\rangle.
\end{aligned}$$

The mean value theorem guarantees that

$$\begin{aligned} & f(z,y_\alpha(z_t))+\alpha h(z,y_\alpha(z_t))-[f(z_t,y_\alpha(z_t))+\alpha h(z_t,y_\alpha(z_t))] \\ & = \langle \nabla_x f(\tilde{z}(z,t),y_\alpha(z_t))+\alpha\nabla_x h(\tilde{z}(z,t),y_\alpha(z_t)),z-z_t\rangle \\ & = t\,\langle -\nabla_x f(\tilde{z}(z,t),y_\alpha(z_t))-\alpha\nabla_x h(\tilde{z}(z,t),y_\alpha(z_t)),d\rangle \end{aligned}$$

holds for some $\tilde{z}(z,t)$ in the line segment between z and z_t. Similarly, applying the mean value theorem to g_i, we get

$$\begin{aligned} g_i(z,y_\alpha(z_t))-g_i(z_t,y_\alpha(z_t)) & = \langle \nabla_x g_i(\tilde{z}_i'(z,t),y_\alpha(z_t)),z-z_t\rangle \\ & = -t\,\langle \nabla_x g_i(\tilde{z}_i'(z,t),y_\alpha(z_t)),d\rangle \end{aligned}$$

for some $\tilde{z}_i'(z,t)$ in to the line segment between z and z_t.

By the definition of the generalized directional derivative there exist two sequences $z^k\to x$, $t_k\downarrow 0$ such that $\varphi_\alpha^\circ(x;d)=\lim_{k\to\infty}t_k^{-1}\left[\varphi_\alpha(z^k_{t_k})-\varphi_\alpha(z^k)\right]$. Exploiting the last three formulas above with $z=z^k$ and $t=t_k$ (and therefore $z_t=z^k_{t_k}$), we get

$$\begin{aligned} \frac{\varphi_\alpha(z^k_{t_k})-\varphi_\alpha(z^k)}{t_k} & \le \langle -\nabla_x f(\tilde{z}(z^k,t_k),y_\alpha(z^k_{t_k}))-\alpha\nabla_x h(\tilde{z}(z^k,t_k),y_\alpha(z^k_{t_k})),d\rangle \\ & \quad -\langle u_\alpha(z^k),w_\alpha(x,z^k_{t_k},t_k)\rangle. \end{aligned}$$

where $w_\alpha(x,z^k_{t_k},t_k)=(\langle \nabla_x g_i(\tilde{z}_i'(z^k,t_k),y_\alpha(z_{t_k})),d\rangle)_{i=1,\dots,m}$. Since $z^k\to x$ and $t_k\downarrow 0$, then $z^k_{t_k}\to x$, $\tilde{z}(z^k,t_k)\to x$, and $y_\alpha(z^k_{t_k})\to y_\alpha(x)$ by Lemma 2.2. Hence, we get

$$\begin{aligned} & \lim_{k\to\infty}\langle -\nabla_x f(\tilde{z}(z^k,t_k),y_\alpha(z^k_{t_k}))-\alpha\nabla_x h(\tilde{z}(z^k,t_k),y_\alpha(z^k_{t_k})),y_\alpha(x)-x\rangle \\ & = -\langle \nabla_x f(x,y_\alpha(x))+\alpha\nabla_x h(x,y_\alpha(x)),d\rangle \end{aligned}$$

since $\nabla_x f$ is continuous. [12, Lemma 2] guarantees that $U_\alpha(z)$ is uniformly bounded on a neighborhood of x and closed at x. Hence, taking a subsequence if necessary, there exists $\hat{u}\in U_\alpha(x)$ such that $u_\alpha(z^k)\to\hat{u}$. Moreover, we get

$$\begin{aligned} \lim_{k\to\infty} -\langle u_\alpha(z^k),w_\alpha(x,z^k_{t_k},t_k)\rangle & = -\langle \hat{u},\nabla_x g(x,y_\alpha(x))\,d\rangle \\ & = -\sum_{i=1}^{m}\hat{u}_i\langle d,\nabla^2 c_i(x)\,d\rangle \le 0, \end{aligned}$$

since $\tilde{z}_i'(z^k,t_k)\to x_i$, $\nabla_x g$ is continuous, and all the c_i's are convex functions. Therefore, $\varphi_\alpha^\circ(x;d)\le -\langle \nabla_x f(x,y_\alpha(x))+\alpha\,\nabla_x h(x,y_\alpha(x)),d\rangle$. □

Theorem 2.1 allows to formulate *(EP)* as the global optimization problem

$$\min\{\ \varphi_\alpha(x)\ :\ x \in C\ \}. \tag{7}$$

However, most optimization algorithms lead only to a stationary point. Actually, any stationary point of φ_α solves (7) and therefore *(EP)* under suitable assumptions on f, which anyway do not guarantee the convexity of φ_α.

Theorem 2.3. *Suppose*

$$\langle \nabla_x f(x,y) + \nabla_y f(x,y), y - x\rangle > 0, \qquad \forall\ x, y \in D \text{ with } x \neq y. \tag{8}$$

a) If $x \in C$ is not a solution of (EP), then $y_\alpha(x) - x$ is a descent direction for φ_α at x, i.e.

$$\varphi^\circ_\alpha(x; y_\alpha(x) - x) < 0.$$

b) If $x^* \in C$ is a stationary point of φ_α over C, i.e.

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

then x^* solves (EP).

Proof. a) By Lemma 2.1 $y_\alpha(x) - x \neq 0$. Therefore, considering (5), (8) and any $(\lambda, \mu) \in \Lambda_\alpha(x)$ we have

$$\begin{aligned}
\varphi^\circ_\alpha(x; y_\alpha(x) - x) &\leq \langle -\nabla_x f(x, y_\alpha(x)) - \alpha \nabla_x h(x, y_\alpha(x)), y_\alpha(x) - x\rangle \\
&< \langle \nabla_y f(x, y_\alpha(x)) + \alpha \nabla_y h(x, y_\alpha(x)), y_\alpha(x) - x\rangle \\
&= \langle -\sum_{i=1}^{m} \lambda_i \nabla c_i(x) - \sum_{j=1}^{r} \mu_j\, a_j, y_\alpha(x) - x\rangle \\
&= -\sum_{i=1}^{m} \lambda_i \langle \nabla c_i(x), y_\alpha(x) - x\rangle + \sum_{j=1}^{r} \mu_j\, [\langle a_j, x\rangle - b_j] \\
&= \sum_{i=1}^{m} \lambda_i\, c_i(x) + \sum_{j=1}^{r} \mu_j\, [\langle a_j, x\rangle - b_j] \leq 0.
\end{aligned}$$

b) Since there exists $\hat{y} \in \mathbb{R}^n$ such that $c_i(\hat{y}) < 0$ for any $i = 1, ..., m$, the Bouligand tangent cone of $\widetilde{C}$ at x^* is the set

$$T(\widetilde{C}, x^*) = \{y \in \mathbb{R}^n\ :\ \langle \nabla c_i(x^*), y\rangle \leq 0, \quad i \ \text{ s.t. } \ c_i(x^*) = 0\}.$$

Since $\hat{y} \in D$, the Bouligand tangent cone of C at x^* is the set

$$T(C, x^*) = T(\widetilde{C}, x^*) \cap \mathit{cone}\,(D - x^*),$$

where *cone* denotes the cone generated by a set. Theferore, we have

$$P(x^*) \subseteq x^* + T(C, x^*).$$

Since $\widetilde{C}$ is convex, $T(\widetilde{C}, x^*) = cl\ cone\,(\widetilde{C} - x^*)$ and therefore $C = D \cap \widetilde{C}$ guarantees

$$P(x^*) \subseteq x^* + cl\,\{\tau\,(y - x^*), \quad y \in C, \quad \tau > 0\},$$

where *cl* denotes the closure of a set. Moreover, the stationarity of x^* for φ_α over C and the positive homogeneity of $\varphi^\circ_\alpha(x^*; \cdot)$ imply

$$\varphi^\circ_\alpha(x^*, y - x^*) \geq 0, \qquad \forall\, y \in P(x^*).$$

If x^* were not a solution of *(EP)*, then $\varphi^\circ_\alpha(x^*; y_\alpha(x^*) - x^*) < 0$ would hold by *a)* in contradiction with the above inequality for $y = y_\alpha(x^*)$. □

Condition (8) was introduced in [18], and named strict ∇-monotonicity later [1], in order to obtain the same properties of Theorem 2.3 for the gap function (1). When *(EP)* is actually a variational inequality, i.e. $f(x, y) = \langle F(x), y - x\rangle$ for some $F : \mathbb{R}^n \to \mathbb{R}^n$, condition (8) is equivalent to require that ∇F is positive definite. Therefore, Theorem 2.3 generalizes Theorem 2.11 of [21], which stated the analogous result for variational inequalities only.

Replacing strict ∇-monotonicity with ∇-monotonicity, i.e. the condition

$$\langle \nabla_x f(x, y) + \nabla_y f(x, y), y - x\rangle \geq 0, \qquad \forall x, y \in D, \tag{9}$$

which means weakening the assumption to positive semidefiniteness in the case of variational inequalities, Theorem 2.3 is no longer true as the following example shows.

Example 2.1. Consider *(EP)* with $n = 2$, $m = 1$, $f(x, y) = x_1 - y_1 + x_2 - y_2$, $c_1(x) = x_1^2 + x_2^2 - 1$ and $D = [-1, 1] \times [-1, 1]$. Therefore, the feasible region C is the unit ball, which is a subset of the given box D, and $x^* = (\sqrt{2}/2, \sqrt{2}/2)$ is the unique solution of *(EP)*. Notice that f satisfies (9) but not (8) since

$$\nabla_x f(x, y) + \nabla_y f(x, y) = (1, 1) + (-1, -1) = (0, 0) \qquad \forall\, x, y \in \mathbb{R}^2.$$

Furthermore, we have

$$P(x) = \{y \in [-1, 1]^2 \ : \ 2x_1y_1 + 2x_2y_2 \leq 1 + x_1^2 + x_2^2\}.$$

Considering $h(x, y) = [(x_1 - y_1)^2 + (x_2 - y_2)^2]/2$, we have

$$\varphi_\alpha(x) = \max\{y_1 + y_2 - \alpha[(y_1 - x_1)^2 + (y_2 - x_2)^2)]/2 \ : \ y \in P(x)\} - x_1 - x_2.$$

Since $\hat{y}_\alpha(x) = (x_1 + 1/\alpha, x_2 + 1/\alpha)$ maximizes the objective function over the whole $\mathbb{R}^2$, it is easy to check that $y_\alpha(x) = \hat{y}_\alpha(x)$ and $\varphi_\alpha(x) = 1/\alpha$ if $\alpha \in [1/2, \sqrt{2}]$ and $x \in [-1 - 1/\alpha, 1 - 1/\alpha]^2$. Therefore, considering any fixed $\alpha \in (2 - \sqrt{2}, \sqrt{2}]$, there exists no descent direction for φ_α at any point $x \in (-1, 1 - 1/\alpha)^2 \cap C$, as x is stationary for φ_α though it does not solve *(EP)*.

Different descent methods for solving *(EP)* have been proposed, relying on the minimization of the gap function (1) for some fixed parameter α [4, 15, 18] or the minimization of the corresponding D-gap function for some fixed pair of parameters [16, 23, 24]. Combinations of ∇-monotonicity and strict ∇-monotonicity assumptions on f and the auxiliary bifunctions have been considered in order to guarantee the so-called "stationarity property" of Theorem *2.3 b)*, i.e., that all the stationary points of the considered gap function are actually solutions of *(EP)*. Indeed, under such assumptions any suitable (local) minimization algorithm could be applied for solving *(EP)*. Furthermore, the classical monotonicity condition

$$f(x, y) + f(y, x) \leq 0 \tag{10}$$

or suitable reinforcements have also been exploited in some papers [4, 15, 18].

In the next section we propose a solution method which does not need this stationarity property, though our assumptions are not stronger than those in the above papers. Our key assumption is a concavity type condition[1] [1], namely

$$f(x, y) + \langle \nabla_x f(x, y), y - x \rangle \geq 0, \qquad \forall\ x, y \in D, \tag{11}$$

which, for instance, is satisfied in Example 2.1. Actually, no relationship holds between the strict ∇-monotonicity condition (8) and our assumption (see also Examples 3.2 in [1]), while it implies the ∇-monotonicity condition (9) [1, Theorem 3.1]. Furthermore, no relationships holds also with the monotonicity condition (10) as the bifunctions of Example 2.2 in [1] show.

3 Solution method

We aim at developing a solution method based on a descent type procedure related to the family of gap functions φ_α, following the approach developed in [1, 25] without considering any approximation of the feasible region C. The basic idea is to use the vector $y_\alpha(x) - x$ as a search direction at the current point x. However, $y_\alpha(x)$ belongs to the approximating polyhedral set $P(x)$ while it does not necessarily lie in C and thus the new point could be unfeasible. Following the penalization approach proposed in [21] for variational inequalities, an exact penalty function can be exploited instead of the function φ_α itself, namely

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

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 $\bar{\alpha} > 0$, the exactness of the penalization is achieved (and therefore the penalty function is a gap function) when the parameter ε is sufficiently small [7]. Actually, the penalty function keeps being a gap function for the same range of ε also decreasing the parameter α.

Lemma 3.1. *Given any $\bar{\alpha} > 0$ and any $p \in [1, \infty]$, there exists $\bar{\varepsilon} > 0$ such that*

a) $\psi_{\alpha,\varepsilon,p}(x) \geq 0$ *for all* $x \in D$*,*

[1] if $f(\cdot, y)$ is concave for all $y \in D$, then f satisfies (11).

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

for all $\alpha \in [0, \bar{\alpha}]$ and $\varepsilon \in (0, \bar{\varepsilon})$.

Proof. Consider any compact set D' such that it contains D in its interior, namely $D \subset \operatorname{int} D'$, and the penalty function

$$\widetilde{\psi}_{\alpha,\varepsilon,p}(x) = \varphi_\alpha(x) + \frac{1}{\varepsilon} \|(c^+(x), d^+(x), e(x))\|_p,$$

where $d^+(x) = (d_1^+(x), \ldots, d_{r_1}^+(x))$ with $d_j^+(x) = \max\{0, \langle a_j, x\rangle - b_j\}$ and $e(x) = (e_{r_1+1}, \ldots, e_r(x))$ with $e_j(x) = \langle a_j, x\rangle - b_j$. By [7, Proposition 8 and Theorems 11 and 12] there exists $\bar{\varepsilon} > 0$ such that

$$\operatorname{argmin}\{\ \varphi_{\bar{\alpha}}(x)\ :\ x \in C\ \} = \operatorname{argmin}\{\ \widetilde{\psi}_{\bar{\alpha},\varepsilon,p}(x)\ :\ x \in \operatorname{int} D'\ \}$$

holds for any $\varepsilon \in (0, \bar{\varepsilon})$. Take any global minimizer $\hat{x}$ of $\varphi_{\bar{\alpha}}$ or equivalently of $\widetilde{\psi}_{\bar{\alpha},\varepsilon,p}$. Since $\hat{x} \in C$, then $\varphi_{\bar{\alpha}}(\hat{x}) = 0$ guarantees also $\widetilde{\psi}_{\bar{\alpha},\varepsilon,p}(\hat{x}) = 0$. Therefore, $\widetilde{\psi}_{\bar{\alpha},\varepsilon,p}(x) \geq 0$ for all $x \in \operatorname{int} D'$ and Theorem 2.1 implies that $x^* \in \operatorname{int} D'$ and $\widetilde{\psi}_{\bar{\alpha},\varepsilon,p}(x^*) = 0$ if and only if x^* solves *(EP)*. Taken any $\alpha \in [0, \bar{\alpha}]$, then $\varphi_\alpha(x) \geq \varphi_{\bar{\alpha}}(x)$ and thus $\widetilde{\psi}_{\alpha,\varepsilon,p}(x) \geq \widetilde{\psi}_{\bar{\alpha},\varepsilon,p}(x)$ for any $x \in \mathbb{R}^n$. Note that $\widetilde{\psi}_{\alpha,\varepsilon,p}$ coincides with $\psi_{\alpha,\varepsilon,p}$ on D, and thus *a*) and *b*) follow immediately. □

Lemma 3.1 provides a whole family of gap functions to exploit within a descent framework. While $y_\alpha(x)$ is computed through φ_α (see (2) and (3)), the descent of the direction is tested on the penalized gap function $\psi_{\alpha,\varepsilon,p}$ checking whether or not

$$\psi^\circ_{\alpha,\varepsilon,p}(x; y_\alpha(x) - x) < 0 \tag{12}$$

holds. Computing the value of the generalized directional derivative may be not easy. Anyway, condition (11) allows to achieve an upper estimate for the generalized directional derivative.

Lemma 3.2. *If f satisfies (11), then*

$$\psi^\circ_{\alpha,\varepsilon,p}(x; y_\alpha(x) - x) \leq -\psi_{\alpha,\varepsilon,p}(x) - \alpha\, [h(x, y_\alpha(x)) + \langle \nabla_x h(x, y_\alpha(x)), y_\alpha(x) - x\rangle]$$

holds for any $x \in D$, $\alpha > 0$, $\varepsilon > 0$ and $p \in [1, \infty]$.

Proof. Since it is convex, the function $v(x) = \|c^+(x)\|_p$ is regular in the Clarke sense, i.e. $v^\circ(x; \cdot) = v'(x; \cdot)$ where $v'(x; \cdot)$ denotes the standard directional derivative. Moreover, it holds (see [6]):

$$v'(x; y_\alpha(x) - x) = \begin{cases} \displaystyle\sum_{i=1}^m \xi_i(x) & \text{if } p = 1, \\ \left[\displaystyle\sum_{i=1}^m (c_i^+(x))^{p-1}\xi_i(x)\right] \cdot \|c^+(x)\|_p^{1-p} & \text{if } p \in (1,\infty), x \notin C, \\ \left[\displaystyle\sum_{i=1}^m (\xi_i(x))^p\right]^{1/p} & \text{if } p \in (1,\infty), x \in C, \\ \displaystyle\max_{i \in I_\infty(x)} \xi_i(x) & \text{if } p = \infty, \end{cases}$$

where

$$\xi_i(x) = \begin{cases} 0 & \text{if } i \in I_-(x) := \{i : c_i(x) < 0\}, \\ \max\{0, \langle \nabla c_i(x), y_\alpha(x) - x\rangle\} & \text{if } i \in I_0(x) := \{i : c_i(x) = 0\}, \\ \langle \nabla c_i(x), y_\alpha(x) - x\rangle & \text{if } i \in I_+(x) := \{i : c_i(x) > 0\}, \end{cases}$$

and $I_\infty(x) = \{i :\ c_i^+(x) = \|c^+(x)\|_\infty\}$. Moreover, $y_\alpha(x) \in P(x)$ implies that $\langle \nabla c_i(x), y_\alpha(x) - x\rangle \leq -c_i(x)$, and hence $\xi_i(x) = 0$ if $i \in I_0(x)$ and $\xi_i(x) \leq -c_i(x)$ if $i \in I_+(x)$. If $p = 1$, then

$$v'(x; y_\alpha(x) - x) = \sum_{i=1}^m \xi_i(x) \leq -\sum_{i \in I_+(x)} c_i(x) = -\sum_{i=1}^m c_i^+(x) = -v(x).$$

If $p \in (1,\infty)$ and $x \notin C$, then

$$\begin{aligned} v'(x; y_\alpha(x) - x) &= \left[\sum_{i=1}^m (c_i^+(x))^{p-1}\xi_i(x)\right] / \|c^+(x)\|_p^{p-1} \\ &\leq -\left[\sum_{i \in I_+(x)} (c_i^+(x))^{p-1}c_i(x)\right] / \|c^+(x)\|_p^{p-1} \\ &= -\|c^+(x)\|_p = -v(x). \end{aligned}$$

If $p \in (1,\infty)$ and $x \in C$, then

$$v'(x; y_\alpha(x) - x) = \left[\sum_{i=1}^m (\xi_i(x))^p\right]^{1/p} = 0 = -v(x).$$

Finally, if $p = \infty$ then

$$v'(x; y_\alpha(x) - x) = \max_{i \in I_\infty(x)} \xi_i(x) \leq \max_{i \in I_\infty(x)} -c_i^+(x) = -\|c^+\|_\infty = -v(x).$$

Hence, we have $v^\circ(x; y_\alpha(x) - x) = v'(x; y_\alpha(x) - x) \le -v(x)$. Moreover, we obtain

$$\begin{aligned} \varphi_\alpha^\circ(x; y_\alpha(x) - x) &\le -\langle \nabla_x f(x, y_\alpha(x)) + \alpha \nabla_x h(x, y_\alpha(x)), y_\alpha(x) - x \rangle \\ &\le f(x, y_\alpha(x)) - \alpha \langle \nabla_x h(x, y_\alpha(x)), y_\alpha(x) - x \rangle \\ &= -\varphi_\alpha(x) - \alpha \left[h(x, y_\alpha(x)) + \langle \nabla_x h(x, y_\alpha(x)), y_\alpha(x) - x \rangle \right], \end{aligned}$$

where the first inequality follows from Theorem 2.2*b)* and the second one from condition (11). Therefore, we have

$$\begin{aligned} &\psi_{\alpha,\varepsilon,p}^\circ(x; y_\alpha(x) - x) \le \varphi_\alpha^\circ(x; y_\alpha(x) - x) + \tfrac{1}{\varepsilon}\, v^\circ(x; y_\alpha(x) - x) \\ &\le -\varphi_\alpha(x) - \alpha \left[h(x, y_\alpha(x)) + \langle \nabla_x h(x, y_\alpha(x)), y_\alpha(x) - x \rangle \right] - \tfrac{1}{\varepsilon}\, v(x) \\ &= -\psi_{\alpha,\varepsilon,p}(x) - \alpha \left[h(x, y_\alpha(x)) + \langle \nabla_x h(x, y_\alpha(x)), y_\alpha(x) - x \rangle \right]. \end{aligned}$$

□

One way to force the decrease of the gap function along a descent direction is to compare the above upper estimate with the value of the gap function itself, i.e.

$$-\psi_{\alpha,\varepsilon,p}(x) - \alpha \left[h(x, y_\alpha(x)) + \langle \nabla_x h(x, y_\alpha(x)), y_\alpha(x) - x \rangle \right] \le -\eta\, \psi_{\alpha,\varepsilon,p}(x) \tag{13}$$

where $\eta \in (0, 1)$ is a fixed parameter. If x is feasible and does not solve *(EP)*, then (13) guarantees that $y_\alpha(x) - x$ is a descent direction for $\psi_{\alpha,\varepsilon,p}$ at x. Indeed, inequality (13) holds at a feasible point x whenever the regularization parameter α is small enough. On the contrary, if x is not feasible, it may happen $\psi_{\alpha,\varepsilon,p}(x) < 0$ when the penalization parameter ε is not below the threshold of exactness, and therefore (13) may be useless. Anyway, $y_\alpha(x) - x$ is a descent direction also in this case regardless of (13) if ε is small enough. Any $(\lambda, \mu) \in \Lambda_\alpha(x)$ provides an upper bound for the appropriate ε, relying on the vector $\lambda^+ \in \mathbb{R}^m$, whose components are given by

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

Theorem 3.1. *Suppose that f satisfies (11).*

a) If $x \in C$ does not solve (EP) and $\eta \in (0, 1)$, then (13) holds for any $\varepsilon > 0$, $p \in [1, \infty]$, and any sufficiently small α.

b) If $x \in D \setminus C$ and $(\lambda, \mu) \in \Lambda_\alpha(x)$, then (12) holds for any $\alpha > 0$, $p \in [1, \infty]$ and ε such that $1/\varepsilon > \|\lambda^+\|_q$, where $\|\cdot\|_q$ is the dual norm of $\|\cdot\|_p$.

Proof. a) Since $x \in C$, then $\psi_{\alpha,\varepsilon,p}(x) = \varphi_\alpha(x) > 0$ for any $\alpha > 0$, $\varepsilon > 0$ and $p \in [1, \infty]$. By contradiction, suppose that there exists a sequence $\alpha_k \downarrow 0$ such that (13) does not hold for

$\alpha = \alpha_k$. Thus,

$$\begin{aligned}\psi_{\alpha_1,\varepsilon,p}(x) &\leq \psi_{\alpha_k,\varepsilon,p}(x)\\ &\leq -\frac{\alpha_k}{1-\eta}\left[h(x, y_{\alpha_k}(x)) + \langle \nabla_x h(x, y_{\alpha_k}(x)), y_{\alpha_k}(x) - x\rangle\right].\end{aligned}$$

Since D is bounded and $y_{\alpha_k}(x) \in D$ for all k , then we obtain the contradiction $\psi_{\alpha_1,\varepsilon,p}(x) \leq 0$ simply taking the limit in the above inequalities.

b) Since $f(x, \cdot)$ is convex, then

$$0 = f(x, x) \geq f(x, y) + \langle \nabla_y f(x, y), x - y\rangle$$

holds for all $y \in D$, and hence (11) implies that also

$$\langle \nabla_x f(x, y) + \nabla_y f(x, y), y - x\rangle \geq 0$$

holds for all $y \in D$. Exploiting the upper estimate provided by Theorem 2.2 *b)*, we get

$$\begin{aligned}\varphi^\circ_\alpha(x; y_\alpha(x) - x) &\leq -\langle \nabla_x f(x, y_\alpha(x)) + \alpha \nabla_x h(x, y_\alpha(x)), y_\alpha(x) - x\rangle\\ &\leq \langle \nabla_y f(x, y_\alpha(x)) + \alpha \nabla_y h(x, y_\alpha(x)), y_\alpha(x) - x\rangle\\ &= -\sum_{i=1}^{m} \lambda_i \langle \nabla c_i(x), y_\alpha(x) - x\rangle - \sum_{j=1}^{r} \mu_j \langle a_j, y_\alpha(x) - x\rangle\\ &= \sum_{i=1}^{m} \lambda_i\, c_i(x) + \sum_{j=1}^{r} \mu_j \left[\langle a_j, x\rangle - b_j\right]\\ &\leq \sum_{i=1}^{m} \lambda_i^+\, c_i^+(x) = \langle \lambda^+, c^+(x)\rangle.\end{aligned}$$

Considering the function $v(x) = \|c^+(x)\|_p$, the above inequalities together with $v^\circ(x; y_\alpha(x)-x) \leq -v(x)$ (see the proof of Lemma 3.2) allow to get

$$\begin{aligned}\psi^\circ_{\alpha,\varepsilon,p}(x; y_\alpha(x) - x) &\leq \varphi^\circ_\alpha(x; y_\alpha(x) - x) + \tfrac{1}{\varepsilon}\, v^\circ(x; y_\alpha(x) - x)\\ &\leq \langle \lambda^+, c^+(x)\rangle - \tfrac{1}{\varepsilon}\, \|c^+(x)\|_p\\ &\leq \|\lambda^+\|_q\, \|c^+(x)\|_p - \tfrac{1}{\varepsilon}\, \|c^+(x)\|_p\\ &= \left(\|\lambda^+\|_q - \tfrac{1}{\varepsilon}\right) \|c^+(x)\|_p < 0\end{aligned}$$

where the last inequality follows from the assumption $1/\varepsilon > \|\lambda^+\|_q$. □

The above results provide the key tools to devise a solution method. Given values for α and ε, the corresponding penalty function $\psi_{\alpha,\varepsilon,p}$ is exploited as long as three conditions hold: its value at the current point is positive; the penalization parameter ε is small enough with respect to the magnitude of a vector of multipliers corresponding to the linearized constraints at the current point; the search direction is indeed a descent direction and the decrease of the value of the penalty function is large enough according to (13). When any of the three conditions fails to hold, a null step is performed simply decreasing both parameters simultaneously.

Algorithm

(0) Choose $p \in [1, \infty]$, $\beta, \gamma, \delta, \eta \in (0, 1)$, sequences $\alpha_k, \varepsilon_k \downarrow 0$, $x^0 \in D$ and set $k = 1$.

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

(2) Compute $y^j = \arg\min\{f(z^j, y) + \alpha_k h(z^j, y) \ : \ y \in P(z^j)\}$ and λ^j any Lagrange multiplier vector corresponding to the linearized constraints.

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

(4) If the following relations hold

a) $\psi_{\alpha_k,\varepsilon_k,p}(z^j) > 0$,

b) $1/\varepsilon_k \geq \|(\lambda^j)^+\|_q + \delta$,

c) $-\psi_{\alpha_k,\varepsilon_k,p}(z^j) - \alpha_k \left[h(z^j, y^j) + \langle \nabla_x h(z^j, y^j), y^j - z^j \rangle\right] \leq -\eta \, \psi_{\alpha_k,\varepsilon_k,p}(z^j)$

then compute the smallest non-negative integer s such that

$$\psi_{\alpha_k,\varepsilon_k,p}(z^j + \gamma^s \, d^j) - \psi_{\alpha_k,\varepsilon_k,p}(z^j) \leq -\beta \, \gamma^{2s} \, \|d^j\|_2,$$

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.

Convergence to a solution of *(EP)* is achieved considering separetely the case in which the parameters actually go to zero from the case in which they are updated a finite number of times.

Theorem 3.2. *If f satisfies (11), then either the algorithm 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. First, we prove that the line search procedure in step 4 is always finite. By contradiction, assume that there exist k and j such that

$$\psi_{\alpha_k,\varepsilon_k,p}(z^j + \gamma^s \, d^j) - \psi_{\alpha_k,\varepsilon_k,p}(z^j) > -\beta \, \gamma^{2s} \, \|d^j\|_2$$

holds for all $s \in \mathbb{N}$. Therefore, we have

$$\psi^\circ_{\alpha_k,\varepsilon_k,p}(z^j; d^j) \geq \limsup_{s \to \infty} \ \gamma^{-s}(\psi_{\alpha_k,\varepsilon_k,p}(z^j + \gamma^s \, d^j) - \psi_{\alpha_k,\varepsilon_k,p}(z^j)) \geq 0,$$

which is impossible since Theorem 3.1 guarantees $\psi^\circ_{\alpha_k,\varepsilon_k,p}(z^j; d^j) < 0$.

If the algorithm stops at z^j after a finite number of iterations, then the stopping criterion and Lemma 2.1 guarantee 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 $\{x^{k_\ell}\}$, we have $x^{k_\ell} \to x^*$. Since $\alpha_k, \varepsilon_k \downarrow 0$, Lemma 3.1 guarantees that there exists k' such that $\psi_{\alpha_k,\varepsilon_k,p}$ is a gap function for all $k \geq k'$ and in particular there exists ℓ' such that $\psi_{\alpha_{k_\ell},\varepsilon_{k_\ell},p}(x^{k_\ell}) > 0$ for all $\ell \geq \ell'$. Lemma 2 in [12] guarantees that $\{\lambda^{k_\ell}\}$ is bounded for ℓ sufficiently large, thus there exists ℓ'' such that $1/\varepsilon_{k_\ell} \geq \|(\lambda^{k_\ell})^+\|_q + \delta$ for all $\ell \geq \ell''$. Choosing $\bar{\ell} := \max\{\ell', \ell''\}$, then we have both $1/\varepsilon_{k_\ell} \geq \|(\lambda^{k_\ell})^+\|_q + \delta$ and $\psi_{\alpha_{k_\ell},\varepsilon_{k_\ell},p}(x^{k_\ell}) > 0$ for all $\ell \geq \bar{\ell}$. By the rule in step 4 condition c) fails at $z^j = x^{k_\ell}$ for $\ell \geq \bar{\ell}$ and hence

$$\begin{aligned} 0 < \psi_{\alpha_{k_{\bar{\ell}}},\varepsilon_{k_{\bar{\ell}}},p}(x^{k_\ell}) &\leq \psi_{\alpha_{k_\ell},\varepsilon_{k_\ell},p}(x^{k_\ell}) \\ &< -\frac{\alpha_{k_\ell}}{(1-\eta)} \left[h(x^{k_\ell}, y^{k_\ell}) + \langle \nabla_x h(x^{k_\ell}, y^{k_\ell}), y^{k_\ell} - x^{k_\ell} \rangle \right]. \end{aligned}$$

Since x^{k_ℓ} and y^{k_ℓ} belong to the bounded set D, the continuity of h and $\nabla_x h$ guarantee that the sequence $\{h(x^{k_\ell}, y^{k_\ell}) + \langle \nabla_x h(x^{k_\ell}, y^{k_\ell}), y^{k_\ell} - x^{k_\ell} \rangle\}$ is bounded from above. Thus, we get $\psi_{\alpha_{k_{\bar{\ell}}},\varepsilon_{k_{\bar{\ell}}},p}(x^*) = 0$ taking the limit as $\ell \to +\infty$, and therefore x^* solves *(EP)*.

Now, suppose that the algorithm generates an infinite sequence $\{z^j\}$ for some fixed k. Therefore, we can set $\alpha = \alpha_k$ and $\varepsilon = \varepsilon_k$ as these values do not change anymore, and let z^* be a cluster point of $\{z^j\}$: taking the appropriate subsequence $\{z^{j_\ell}\}$, we have $z^{j_\ell} \to z^*$. Expoiting Lemma 2.2, $z^{j_\ell} \to z^*$ implies also $d^{j_\ell} \to d^* = y_\alpha(z^*) - z^*$.

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

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

Taking the limit as $\ell \to +\infty$, we get $t_{j_\ell} \to 0$ since $d^* \neq 0$. Moreover, the inequality

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

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

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

where ξ^{j_ℓ} is a generalized gradient of $\psi_{\alpha,\varepsilon,p}$ at $z^{j_\ell} + \theta_{j_\ell}\, t_{j_\ell}\, \gamma^{-1}\, d^{j_\ell}$. Hence, we get

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

On the other hand, we also have

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

Thus, we get

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

Moreover, $z^{j_\ell} + \theta_{j_\ell}\, t_{j_\ell}\, \gamma^{-1}\, d^{j_\ell} \to z^*$ as $\ell \to +\infty$. Since $\psi^\circ_{\alpha,\varepsilon,p}$ is upper semicontinuous as function of $(z;d)$ (see e.g. [5]), we get

$$\psi^\circ_{\alpha,\varepsilon,p}(z^*;d^*) \geq \limsup_{\ell\to+\infty} \psi^\circ_{\alpha,\varepsilon,p}\left(z^{j_\ell} + \theta_{j_\ell}\, t_{j_\ell}\, \gamma^{-1}\, d^{j_\ell}; d^{j_\ell}\right) \geq 0. \tag{14}$$

On the other hand, if $z^* \in C$ then $\psi_{\alpha,\varepsilon,p}(z^*) = \varphi_\alpha(z^*) > 0$ since z^* does not solve *(EP)*. Moreover, the three conditions at step 4 are satisfied for all ℓ, hence we have

$$-\psi_{\alpha,\varepsilon,p}(z^{j_\ell}) - \alpha\, \left[h(z^{j_\ell}, y^{j_\ell}) + \langle \nabla_x h(z^{j_\ell}, y^{j_\ell}), y^{j_\ell} - z^{j_\ell}\rangle\right] \leq -\eta\, \psi_{\alpha,\varepsilon,p}(z^{j_\ell}).$$

Thus, taking the limit the upper estimate provided in Lemma 3.2 gives

$$\begin{aligned} \psi^\circ_{\alpha,\varepsilon,p}(z^*;d^*) &\leq -\psi_{\alpha,\varepsilon,p}(z^*) - \alpha\, \left[h(z^*, y_\alpha(z^*)) + \langle \nabla_x h(z^*, y_\alpha(z^*)), d^*\rangle\right] \\ &\leq -\eta\, \psi_{\alpha,\varepsilon,p}(z^*) < 0, \end{aligned}$$

which contradicts (14). Therefore, $z^* \notin C$. Since $1/\varepsilon \geq \|(\lambda^{j_\ell})^+\|_q + \delta$, then taking the limit as $\ell \to +\infty$ (eventually considering a subsequence) provides $1/\varepsilon \geq \|(\lambda^*)^+\|_q + \delta$ for some $(\lambda^*, \mu^*) \in \Lambda_\alpha(z^*)$. Thus, Theorem 3.1*b)* guarantees $\psi^\circ_{\alpha_k,\varepsilon_k,p}(z^*;d^*) < 0$, contradicting again (14). Therefore, z^* solve *(EP)*. □

When *(EP)* is the variational inequality associated to the operator F, condition (11) is equivalent to require that ∇F is positive semidefinite, while the algorithm presented in [21] for variational inequalities requires positive definiteness. Updating the parameters α and ε, which on the contrary are kept fixed in [21], is the key feature to devise a solution method that converges under weaker assumptions. Furthermore, the above algorithm involves an inexact line search while the algorithm in [21] needs the rather theoretical exact line search.

4 Numerical tests

We applied the algorithm to solve a problem of production competition over a network under the Nash-Cournot equilibrium framework. We considered a modification of the oligopolistic model originally proposed in [17]. The same commodity is produced by n firms, which compete over quantity in a noncooperative fashion. Given a transportation network (N, A), the firms and the markets are located at some sets of nodes $I \subset N$ and $J \subset N$, respectively. Each firm $i \in I$ chooses the quantity x_{ij} to supply to each market $j \in J$ and how to ship it, by choosing the quantities v^i_a to be sent on each arc $a \in A$; the goal of the firm i is to maximize its profit given by

$$\sum_{j\in J} x_{ij} p_j \left(\sum_{\ell\in I} x_{\ell j}\right) - \sum_{a\in A} s_a\, v^i_a - \pi_i \left(\sum_{j\in J} x_{ij}\right),$$

where $p_j : \mathbb{R}_+ \to \mathbb{R}_+$ is the inverse demand function for market j, that is $p_j(z)$ denotes the unitary price at which the market j requires a total quantity z, s_a is the unitary transportation cost on arc a, and $\pi_i : \mathbb{R}_+ \to \mathbb{R}_+$ is the production cost function of firm i. On the other hand, each firm is subject to flow-conservation constraints

$$(Ev^i)_k = \begin{cases} -\sum_{j\in J} x_{kj} & \text{if } k = i \\ 0 & \text{if } k \notin J \\ x_{ik} & \text{if } k \in J \end{cases} \qquad k \in N, \tag{15}$$

where E is the node-arc incidence matrix of the network and $v^i = (v^i_a)_{a\in A}$. Moreover, it has a bounded production capacity, i.e.

$$\sum_{j\in J} x_{ij} \le q_i, \tag{16}$$

where q_i denotes the maximum quantity that firm i may produce. Finally, a public authority selects a set $\mathcal{R}$ of paths and imposes upper bounds on the congestion of these selected paths, namely

$$\sum_{a\in r} t_a(v) \le T_r \qquad \forall\, r \in \mathcal{R}, \tag{17}$$

where t_a denotes the (convex) travel time function on arc a and T_r is the maximum travel time on path r.

An equilibrium state is reached when the production levels and the flows are such that no firm would increase its profit by changing its own production and shipping choices while the other firms keep the same ones. Finding such an equilibrium can be formulated as a Generalized Nash Equilibrium Problem (GNEP), i.e. a noncooperative game in which the strategy set of each player (firm), as well as his payoff function, depends on the strategies of all players (see [8] and references therein). More precisely, the congestion constraints (17), which are shared by all the players, make the problem a jointly convex GNEP. It is well known that normalized equilibria of a jointly convex GNEP, as introduced in [20], are the solutions of a suitable equilibrium problem (see e.g. [22]). In our case, setting $x = (x_{ij})_{i\in I, j\in J}$, $v = (v^i)_{i\in I}$ and analogously y and w, normalized equilibria coincide with the solutions of *(EP)* where the feasible set C is defined by constraints (15)–(17) and the bifunction f is given by:

$$\begin{aligned} f((x,v),(y,w)) = \; & \sum_{i\in I} \Bigg[\sum_{j\in J} x_{ij} p_j \left(\sum_{\ell\in I} x_{\ell j} \right) - \sum_{j\in J} y_{ij} p_j \left(y_{ij} + \sum_{\ell\in I, \ell\neq i} x_{\ell j} \right) \\ & + \sum_{a\in A} s_a \left(w^i_a - v^i_a \right) + \pi_i \left(\sum_{j\in J} y_{ij} \right) - \pi_i \left(\sum_{j\in J} x_{ij} \right) \Bigg] \end{aligned}$$

We applied our algorithm to a problem with 3 firms, 2 markets, and the transportation network of Figure 1 with $I = \{1, 2, 3\}$ and $J = \{13, 14\}$.

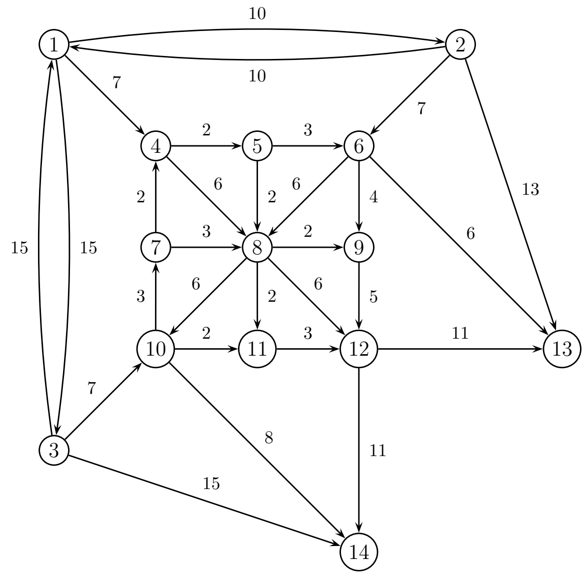
We assumed that both markets have the same inverse demand function

$$p_j(z) = p(z) = \rho^{1/\tau}(z + \sigma)^{-1/\tau},$$

and that the cost functions have the form

$$\pi_i(x_i) = \pi_i\, x_i + (1 + \delta_i)^{-1} K_i^{-\delta_i} x_i^{1+\delta_i},$$

Figure 1: Transportation network.



where $\rho = 5000$ and $\tau = 1.1$ have been selected as in [19] as well as the values for the parameters of the cost functions, which are shown in Table 1. We chose $\sigma = 0.01$ rather than $\sigma = 0$: while the effect on the equilibrium values is negligible, the problem is well defined also for a zero total production, which could be selected as the starting point of the algorithm.

Table 1: Parameters of cost functions.

i	π_i	K_i	δ_i
1	10	5	5/6
2	6	5	1
3	2	5	5/4

Since the functions π_i and p have been chosen convex and differentiable, the function $z \mapsto z\, p(z)$ is therefore concave. Thus, the bifunction $f(\cdot, (y, w))$ turns out to be concave for any (y, w) and hence assumption (11) of the algorithm is satisfied.

Regarding the constraints, we set production bounds q_i of the firms all equal to 40. Moreover, we considered the travel time functions introduced by the U.S. Bureau of Public Roads [3]

$$t_a(v) = f_a \left[1 + 0.15 \left(\frac{\sum_{i \in I} v_a^i}{C_a} \right)^4 \right],$$

where f_a denotes the free flow time and C_a the capacity of arc a. The values of f_a are displayed in Figure 1 and the capacity C_a has been set to 5 for all the arcs. Congestion constraints have been set on the 8 paths listed in Table 2, and the values T_r have been chosen equal to 50 for all the paths.

Table 2: Paths with congestion constraint.

	1	2	3	4	5	6	7	8
path	4-5-6	7-8-9	10-11-12	10-7-4	5-8-11	6-9-12	4-8-12	6-8-10

The algorithm has been implemented in MATLAB 7.10.0; the built-in function `fmincon` from the Optimization Toolbox was exploited to evaluate the gap function φ_α and to compute the direction $y_\alpha(x) - x$. The value 10^{-2} was used as the threshold for the stopping criterion of the algorithm at step 3. We chose the regularizing bifunction $h(x, y) = \|y - x\|_2^2/2$.

After some preliminary tests we set parameters of the algorithm as follows: $\beta = 0.1$, $\gamma = 0.7$, $\eta = 0.9$, $\alpha_k = 1/3^k$, $\varepsilon_k = 1/k^2$, and $p = \infty$. Running the algorithm with these values of parameters and a zero total production and flow as the starting point, it performed 11 iterations, 3 just updated the parameters α and ε (null steps), and the gap function had to be evaluated 14 times. The solution found is given in Tables 3 and 4. As shown in the Tables some bounds on production and travel times are tight.

Table 3: Equilibrium solution found (supplied quantities).

	markets		total
firms	13	14	production
1	20.2087	15.6148	35.8235
2	27.9749	12.0251	40.0000
3	13.5517	26.4483	40.0000

Table 4: Equilibrium solution found (path flows and travel times).

path	4-5-6	7-8-9	10-11-12	10-7-4	5-8-11	6-9-12	4-8-12	6-8-10
flow	27.89	0.00	27.70	0.00	0.90	23.48	12.75	13.03
travel time	50.00	5.00	50.00	5.00	4.00	50.00	50.00	50.00

Subsequently, we randomly selected 50 starting points and we ran the algorithm for different choices of the parameters β, γ, η, α_k, ε_k, and p. Results with respect to different values of β and γ are given in Table 5: in each row the minimum, average and maximum number of iterations, null steps and solved optimization problems are given. The results suggest to choose $\beta < \gamma$. A good choice seems to be a small value of β, close to zero, and a value of γ close to one.

The impact of different values of η is shown in Table 6, in which the minimum, average and maximum number of iterations, null steps and solved optimization problems are given. Results show that the closer η is to one, the better is the behaviour of the algorithm.

Table 7 reports the results with respect to different values of parameters α_k and ε_k. According to the results, it is better to choose an exponentially decreasing sequence for α_k, while ε_k should be a polynomially decreasing one.

Table 8 shows the impact of different choice of the value of p, which seems not to have a significant impact on the behaviour of the algorithm.

Table 5: Results with different values of parameters β and γ.

		iterations			null steps			opt. problems		
β	γ	min	avg	max	min	avg	max	min	avg	max
0.1	0.1	9	19.36	44	2	2.94	4	10	30.46	79
0.1	0.3	9	12.74	19	2	2.84	4	10	17.5	29
0.1	0.5	9	10.66	14	2	2.68	4	10	13.24	20
0.1	0.7	9	9.94	14	2	2.62	4	10	13.22	19
0.1	0.9	9	10.22	13	2	2.62	4	10	16.16	23
0.3	0.1	10	21.94	52	2	2.94	4	11	36.48	96
0.3	0.3	9	13	22	2	2.84	4	11	18.82	36
0.3	0.5	9	10.5	15	2	2.68	4	11	13.86	23
0.3	0.7	9	9.94	13	2	2.62	4	11	14.74	25
0.3	0.9	9	10.08	13	2	2.62	4	11	19.28	46
0.5	0.1	10	23.46	52	2	2.94	4	13	40.06	96
0.5	0.3	9	13.38	22	2	2.84	4	11	20.14	36
0.5	0.5	9	10.72	16	2	2.68	4	11	14.84	25
0.5	0.7	9	10.22	14	2	2.62	4	11	16.72	40
0.5	0.9	9	10.16	13	2	2.62	4	12	23.2	77
0.7	0.1	12	25.48	52	2	2.94	4	15	45.18	122
0.7	0.3	9	13.62	22	2	2.84	4	11	20.96	36
0.7	0.5	9	11.1	17	2	2.68	4	12	16.64	46
0.7	0.7	9	10.56	15	2	2.62	4	11	19.32	59
0.7	0.9	9	10.32	14	2	2.62	4	12	27.8	123

Finally, we selected production capacities for each firm in the range $[20, 50]$ and congestion bounds for any path in the range $[30, 100]$ and we ran the algorithm with the original values

Table 6: Results with different values of parameter η.

	iterations			null steps			opt. problems		
η	min	avg	max	min	avg	max	min	avg	max
0.1	10	22.66	33	1	1.40	3	13	83.68	157
0.2	11	21.48	33	1	1.44	3	13	77.38	157
0.3	10	20.42	33	1	1.54	3	12	71.44	157
0.4	10	22.28	32	1	1.48	3	12	82.46	157
0.5	10	20.20	32	1	1.58	4	12	68.76	157
0.6	9	13.30	27	1	1.92	3	10	25.30	111
0.7	9	11.82	24	1	2.12	3	12	16.96	93
0.8	9	11.22	14	1	2.46	3	11	14.66	33
0.9	9	9.94	14	2	2.62	4	10	13.22	19

Table 7: Results with different values of parameters α_k and ε_k.

		iterations			null steps			opt. problems		
α_k	ε_k	min	avg	max	min	avg	max	min	avg	max
$1/k$	$1/k$	21	31.12	71	7	20.52	61	24	39.32	101
$1/k$	$1/k^2$	24	28.30	40	4	4.36	7	33	57.70	152
$1/k$	$1/3^k$	34	51.54	79	2	2.12	3	77	176.18	333
$1/k^2$	$1/k$	11	16.76	23	4	9.26	16	12	20.12	27
$1/k^2$	$1/k^2$	11	13.70	20	2	3.38	7	13	19.92	51
$1/k^2$	$1/3^k$	14	15.80	28	2	2.14	4	17	25.86	140
$1/3^k$	$1/k$	9	17.24	23	2	9.24	16	10	20.82	27
$1/3^k$	$1/k^2$	9	9.94	14	2	2.62	4	10	13.22	19
$1/3^k$	$1/3^k$	9	12.08	19	1	1.88	3	12	18.60	68

Table 8: Results with different values of parameter p.

	iterations			null steps			opt. problems		
p	min	avg	max	min	avg	max	min	avg	max
1	9	10.2	14	2	2.66	4	11	13.92	21
2	9	10.08	14	2	2.66	4	10	13.58	21
3	9	10.02	14	2	2.64	4	10	13.42	19
4	9	9.92	14	2	2.62	4	10	13.22	19
5	9	9.94	14	2	2.62	4	10	13.22	19
10	9	9.94	14	2	2.62	4	10	13.22	19
∞	9	9.94	14	2	2.62	4	10	13.22	19

of parameters: $\beta = 0.1$, $\gamma = 0.7$, $\eta = 0.9$, $\alpha_k = 1/3^k$, $\varepsilon_k = 1/k^2$, $p = \infty$, and the zero vector as starting point. The results are shown in Table 9, which shows the effect of capacity and congestion bounds on the production values, path flows, and path travel times at the equilibrium solution. In particular, for each given value of the capacities and of the congestion bounds, the quantities supplied by the firms to each markets are reported on the first line, the total quantities supplied on the second line, and the path flows on the third line, where a bold font denotes that the congestion constraint is active on the correponding path.

Acknowledgements. The authors wish to thank the two anonymous referees for the careful reviews and valuable comments, which helped them to improve the paper.

References

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

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

[3] Bureau of Public Roads: Traffic Assignment Manual. U.S. Dept. of Commerce, Urban Planning Division, Washington D.C. (1964)

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

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

[6] Di Pillo, G., Grippo, L.: Exact penalty functions in constrained optimization. SIAM J. Control Optim. **27**, 1333–1360 (1989)

[7] Di Pillo, G., Facchinei, F.: Exact penalty functions for nondifferentiable programming problems. In Nonsmooth Optimization and Related Topics, pp. 89–107. Plenum Press, New York (1989)

[8] Facchinei, F., Kanzow, C.: Generalized Nash equilibrium problems. 4OR **5**, 173–210 (2007)

[9] Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer, Berlin (2003)

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

[11] Fukushima, M.: Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. Math. Program. **53**, 99–110 (1992)

Table 9: The effect of capacity and congestion bounds on the production values, path flows, and path travel times at the equilibrium solution.

bounds				quantities supplied to each market total quantities supplied path flows							
q_1	q_2	q_3	T_r								
40	40	40	50	20.21	15.61		27.97	12.03		13.55	26.45
					35.82			40.00			40.00
				27.89	0.00	**27.70**	0.00	0.90	**23.48**	**12.75**	**13.03**
40	40	40	30	20.52	15.25		28.33	11.67		12.01	27.99
					35.78			40.00			40.00
				24.03	0.00	**24.03**	0.00	0.00	**19.86**	11.88	**12.31**
40	40	40	100	20.25	16.09		26.65	13.35		14.49	25.51
					36.34			40.00			40.00
				33.62	0.00	29.35	0.00	0.74	26.71	**15.72**	**15.72**
50	50	50	50	19.06	14.42		32.46	13.97		13.92	28.77
					33.48			46.43			42.69
				27.83	0.00	**27.83**	0.00	0.00	**23.48**	13.99	**14.98**
50	50	50	100	18.58	15.28		32.24	14.77		15.53	28.22
					33.86			47.00			43.74
				33.55	0.00	31.06	0.00	0.00	**28.65**	15.28	**16.16**
30	30	30	50	16.40	13.60		19.04	10.96		11.56	18.44
					30.00			30.00			30.00
				27.96	0.00	23.97	0.00	1.71	21.92	**12.75**	**12.75**
30	30	30	30	16.51	13.49		19.79	10.21		11.10	18.90
					30.00			30.00			30.00
				24.18	0.00	**23.62**	0.00	2.56	**19.86**	**10.57**	**10.86**
20	20	50	50	10.83	9.17		13.89	6.11		18.61	26.59
					20.00			20.00			45.20
				27.83	0.00	**27.83**	9.38	0.00	6.11	9.16	15.27
20	50	20	50	11.17	8.83		30.33	19.05		5.83	14.17
					20.00			49.38			20.00
				27.46	0.00	15.43	0.00	8.87	**23.48**	3.71	**19.70**
20	50	20	100	11.15	8.85		29.89	20.11		6.08	13.92
					20.00			50.00			20.00
				31.83	0.00	11.48	0.68	8.85	27.05	0.00	22.02

[12] Hogan, W.: Directional derivatives for extremal-value functions with applications to the completely convex case. Oper. Res. **21**, 188–209 (1973)

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

[14] Konnov, I.V.: Combined Relaxation Methods for Variational Inequalities. Springer, Berlin (2001)

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

[16] 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)

[17] Marcotte, P.: Algorithms for the network oligopoly problem. J Oper. Res. Soc. **38**, 1051-1065 (1987)

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

[19] Murphy, F.H., Sherali, H.D., Soyster, A.L.: A mathematical programming approach for determining oligopolistic market equilibrium. Math. Program. **24**, 92–106 (1982)

[20] Rosen, J.B.: Existence and uniqueness of equilibrium points for concave $n-$games. Econometrica **33**, 520–534 (1965)

[21] 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)

[22] von Heusinger, A., Kanzow, C.: Optimization reformulations of the generalized Nash equilibrium problem using Nikaido-Isoda-type functions. Comput. Optim. Appl. **43**, 353–377 (2009)

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

[24] 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)

[25] Zhu, D.L., Marcotte, P.: Modified descent methods for solving the monotone variational inequality problem. Oper. Res. Lett. **14**, 111–120 (1993)