Mathematical Programming manuscript No.
(will be inserted by the editor)

**Nguyen Thi Thu Van · J.J. Strodiot ·
V.H. Nguyen**

A Bundle Method for Solving Equilibrium Problems

Received: August 15, 2005 / Accepted: May 16, 2006

Abstract We present a bundle method for solving nonsmooth convex equilibrium problems based on the auxiliary problem principle. First, we consider a general algorithm that we prove to be convergent. Then we explain how to make this algorithm implementable. The strategy is to approximate the nonsmooth convex functions by piecewise linear convex functions in such a way that the subproblems are easy to solve and the convergence is preserved. In particular, we introduce a stopping criterion which is satisfied after finitely many iterations and which gives rise to $\Delta-$stationary points. Finally, we apply our implementable algorithm for solving the particular case of single-valued and multivalued variational inequalities and we find again the results obtained recently by Salmon et al. [18].

Nguyen Thi Thu Van
Department of Mathematics, University of Namur (FUNDP), Namur, Belgium, and Faculty of Mathematics and Informatics, University of Natural Sciences, Vietnam National University, Ho Chi Minh City, Vietnam. Research supported by a FUNDP Doctoral Institutional Scholarship CA-04-314, CA-05-303 and CA-06-689, and a CIUF-CUD Scholarship CUI-UNV02 and CUI-MT03.
E-mail: thu.van.nguyen@fundp.ac.be

J.J. Strodiot
Department of Mathematics, University of Namur (FUNDP), Namur, Belgium.
E-mail: jjstrodiot@fundp.ac.be

V.H. Nguyen
Department of Mathematics, University of Namur (FUNDP), Namur, Belgium.
E-mail: van.hien.nguyen@fundp.ac.be

Keywords Equilibrium problems · Variational inequalities · Bundle methods

Mathematics Subject Classification (2000) 90C25 · 49J40

1 Introduction

Let C be a nonempty closed convex subset of $I\!R^n$ and let $f : C \times C \to I\!R$ be a continuous function satisfying $f(x, x) = 0$ for all $x \in C$ and $f(x, \cdot)$ convex on C for all $x \in C$. The equilibrium problem (for short, EP) consists of

$$(EP) \qquad \text{finding} \quad x^* \in C \text{ such that } f(x^*, y) \geq 0 \text{ for all } y \in C.$$

This problem is very general in the sense that it includes, as particular cases, the optimization problem, the variational inequality problem, the Nash equilibrium problem in noncooperative games, the fixed point problem, the nonlinear complementarity problem and the vector optimization problem (see, for instance, [3], [10] and the references quoted therein). The interest of this problem is that it unifies all these particular problems in a convenient way. Moreover, many methods devoted to solving one of these problems can be extended, with suitable modifications, to solving the general equilibrium problem.

In this paper we suppose that there exists at least one solution to problem (EP). In particular, it is true when C is compact. Other existence results for this problem can be found, for instance, in [3], [9].

Most of the methods for solving equilibrium problems are derived from fixed-point formulations of (EP). Since $f(x, x) = 0$ for all $x \in C$, obviously $x^* \in C$ is a solution of problem (EP) if and only if x^* is a solution of problem $\min_{x \in C} f(x^*, x)$. Then, given $x^0 \in C$, the corresponding algorithm generates a sequence $\{x^k\}$ defined, for all $k \in I\!N$, by

$$x^{k+1} = \arg\min_{x \in C} f(x^k, x). \tag{1}$$

However it is more convenient to use, as for variational inequality problems (see Cohen [4]), the auxiliary problem principle which is based on the following fixed point property: $x^* \in C$ is a solution of problem (EP) if and only if it is a solution of

$$\min_{y \in C} \{\epsilon\, f(x^*, y) + h(y) - h(x^*) - \langle \nabla h(x^*), y \rangle\}, \tag{2}$$

where $\epsilon > 0$ and $h : C \to I\!R$ is a strongly convex differentiable function. A typical example of function h is the squared norm. Then the corresponding fixed point iteration is: *Given* $x^k \in C$, *find* $x^{k+1} \in C$ *the solution of*

$$(\mathrm{P}_k) \quad \min_{y\in C}\{\epsilon\, f(x^k, y) + h(y) - h(x^k) - \langle \nabla h(x^k), y\rangle\}.$$

Observe that this problem has a unique solution since h is strongly convex. This algorithm has been introduced by Mastroeni who proved its convergence in [16], Theorem 3.1 under the assumptions that f is strongly monotone on $C \times C$, in the sense that there exists $\gamma > 0$ such that

$$f(x, y) + f(y, x) \leq -\gamma\|y - x\|^2,\ \forall x, y \in C, \tag{3}$$

and that f satisfies the property: there exist $c, d > 0$ such that

$$\forall x, y, z \in C \quad f(x, y) + f(y, z) \geq f(x, z) - c\,\|y - x\|^2 - d\,\|z - y\|^2. \tag{4}$$

When

$$f(x, y) = \langle F(x), y - x\rangle + \varphi(y) - \varphi(x), \quad \forall x, y \in C \tag{5}$$

with $F : C \to I\!R^n$ a continuous mapping and $\varphi : C \to I\!R$ a continuous convex function, problem (EP) is reduced to the generalized variational inequality problem (GVIP):

Find $x^* \in C$ such that, for all $y \in C$, $\quad \langle F(x^*), y - x^*\rangle + \varphi(y) - \varphi(x^*) \geq 0.$

In that case, the auxiliary equilibrium problem principle algorithm becomes: *Given* $x^k \in C$, *find* $x^{k+1} \in C$ *the solution of the problem*

$$\min_{y\in C} \epsilon\,[\varphi(y) + \langle F(x^k), y - x^k\rangle] + h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k\rangle. \tag{6}$$

It is easy to see that (3) and (4) are satisfied when F is strongly monotone and Lipschitz continuous, respectively.

However these assumptions are too strong in that case. Indeed, Zhu and Marcotte ([19], Theorem 3.2) proved that the sequence $\{x^k\}$ generated by the auxiliary problem principle converges to a solution of problem (GVIP) when F is co-coercive on C in the sense that

$$\exists\, \gamma > 0 \quad \forall x, y \in C \quad \langle F(y) - F(x), y - x\rangle \geq \gamma\,\|F(y) - F(x)\|^2. \tag{7}$$

It is obvious that F co-coercive on C does not imply, in general, that the corresponding function f defined by (5) is strongly monotone on $C \times C$ (for instance, take $F = 0$ and observe that $f(x, y) + f(y, x) = 0$). So one of the aims of this paper is to obtain the convergence of Mastroeni's algorithm under assumptions weaker than (3) and (4) in such a way that Zhu and Marcotte's result can be derived as a particular case.

Concerning the implementation of the previous algorithm, the subproblems (P_k) can be difficult to solve when the convex function $f(x^k, \cdot)$ is nonlinear. It is the case when f is given by (5) with φ a nonsmooth convex function. In that case, our strategy is to approximate the function $f(x^k, \cdot)$ by another convex function so that the subproblems (P_k) become easy to solve and the convergence is preserved under the same assumptions as in the exact case. The approximation will be done by using an extension of the bundle method developed in [18] for problem (GVIP).

Let us mention that this strategy has been used by Konnov [13] at the lower level of a combined relaxation method for finding equilibrium points. More precisely, given $x^k \in C$, Konnov considers successive linearizations of the function $f(x^k, \cdot)$ in order to construct a convex piecewise linear approximation $\bar{f}_k$ of $f(x^k, \cdot)$ such that the solution y^k of subproblem (P_k) with $f(x^k, \cdot)$ replaced by $\bar{f}_k$ satisfies the property:

$$f(x^k, y^k) \leq \mu \bar{f}_k(y^k) \quad (0 < \mu < 1). \tag{8}$$

Then this solution y^k is used to compute a direction g^k in the subdifferential of the function $-f(\cdot, y^k)$ at x^k, and a steplength $\sigma_k = f(x^k, y^k)/\|g^k\|^2$ (if $g^k \neq 0$). Finally the next iterate x^{k+1} is defined as the projection over C of the vector $x^k - \gamma_k \sigma_k g^k$ where $0 < \gamma_k < 2$. Observe that this step is well defined when $f(\cdot, y)$ is concave on C for all $y \in C$. In this paper we do not assume this property, so we do not consider Konnov's projection step and instead of this step, we set $x^{k+1} = y^k$. In other terms, our method is simply an implementable version of Mastroeni's auxiliary problem principle.

To summarize our approach, first we study the convergence of the algorithm when $f(x^k, \cdot)$ is approximated from below by any function $\bar{f}_k$ which satisfies the inequality (8) and then we present an implementable method to construct a broad class of convex piecewise linear functions $\bar{f}_k$ approximating $f(x^k, \cdot)$. An advantage of our approach is that it allows to limit the size of the bundle used to obtain $\bar{f}_k$.

Another way for solving problem (EP) is to transform it into a variational inequality problem (see [12], Thm 2.1.2) and to use a bundle type method for solving this equivalent problem. This method is interesting when C is compact because in that case there exist efficient variants of the bundle method allowing to obtain a complexity analysis. In these methods the level sets of the piecewise linear models are used to construct the successive iterates (see [14] and [11] for more details). This approach has been used by Gol'shtein [7] for solving problem (EP) when C is compact and $f(x, \cdot)$ satisfies a Lipschitz condition with a constant L independent on x. These conditions can be taken into account by our convergence theory.

Finally to show the interest of our general algorithm, first we apply it to problem $(GVIP)$ with the purpose to find again the convergence theorem obtained in [18]. Then we consider the following multivalued variational inequality problem (MVIP):

$$\text{Find } x^* \in C \text{ and } r^* \in F(x^*) \text{ such that, for all } y \in C, \quad \langle r^*, y - x^* \rangle \geq 0,$$

where C is a nonempty closed convex subset of $I\!R^n$ and $F : C \to 2^{I\!R^n}$ is a co-coercive continuous multivalued mapping with compact values. This problem is a particular instance of problem (EP) when the function f is defined, for all $x, y \in C$, by

$$f(x, y) = \sup_{\xi \in F(x)} \langle \xi, y - x \rangle.$$

For this problem, we use a very simple approximating function and we derive a convergence result from our general theory.

The paper is organized as follows. In section 2, we consider a general algorithm for solving problem (EP) where the convex function $f(x^k, \cdot)$ is approximated, and we prove that it is convergent to a solution of problem (EP). In section 3, we present an implementable version of this general algorithm by using a bundle strategy. In particular, we introduce a stopping criterion and we study the convergence properties of the resulting algorithm. Finally, in section 4, first we find again the convergence results obtained in [18] for problem $(GVIP)$ and then we present a realization of the general algorithm for solving problem (MVIP).

2 A General Algorithm

From now on, we impose that the gradient ∇h is Lipschitz continuous on C with constant $\Lambda > 0$. We also denote by $\beta > 0$ the modulus of the strongly convex function h. In this section, we consider the general equilibrium problem (EP) and the algorithm introduced by Mastroeni for solving it where the parameter $\epsilon = \epsilon_k > 0$ is allowed to vary at each iteration. This algorithm can be expressed as follows: *Given* $x^k \in C$, *find* $x^{k+1} \in C$ *the solution of problem* (P_k).

As explained before in Section 1, the function $f(x^k, \cdot)$, denoted f_k in the sequel, is replaced in problem (P_k) by another convex function $\bar{f}_k$ in such a way that the new problem

$$(\bar{P}_k) \quad \min_{y \in C} \{ \epsilon_k \bar{f}_k(y) + h(y) - h(x^k) - \langle \nabla h(x^k), y \rangle \}$$

is easier to solve and that the corresponding algorithm:

Given $x^k \in C$, find $x^{k+1} \in C$ solution of problem $(\bar{P}_k)$

generates a sequence $\{x^k\}$ converging to some solution of problem (EP).

To obtain the convergence of this algorithm, we introduce some conditions on the approximating function $\bar{f}_k$.

Definition 1 Let $\mu \in (0,1]$ and $x^k \in C$. A convex function $\bar{f}_k : C \to I\!R$ is a $\mu-$approximation of f_k at x^k if $\bar{f}_k \leq f_k$ on C and if

$$f_k(y^k) \leq \mu \bar{f}_k(y^k), \tag{9}$$

where y^k is the unique solution of problem $(\bar{P}_k)$.

Since $f_k(x^k) = 0$, and $\bar{f}_k(x^k) \leq f_k(x^k)$, inequality (9) implies that $f_k(x^k) - f_k(y^k) \geq \mu[\bar{f}_k(x^k) - \bar{f}_k(y^k)]$, i.e., that the reduction on f_k is greater than a fraction of the reduction obtained by using the approximating function $\bar{f}_k$. This is motivated by the fact that, at iteration k, the objective is to minimize the function f_k (see (1)). Moreover, we observe that $\bar{f}_k = f_k$ is a 1-approximation of f_k at x^k.

Using this definition, the approximate auxiliary equilibrium principle algorithm can be expressed as follows:

A General Algorithm.
Step 1. *Let $x^0 \in C$ and $\mu \in (0,1]$. Set $k = 0$.*
Step 2. *Find $\bar{f}_k$ a $\mu-$approximation of f_k at x^k and denote by x^{k+1} the unique solution of problem $(\bar{P}_k)$.*
Step 3. *Increase k by 1 and go to Step 2.*

The convergence of this general algorithm is established in two steps. First we examine the convergence of the algorithm when the sequence $\{x^k\}$ is bounded and $\|x^{k+1} - x^k\| \to 0$. Then in a second theorem, we give conditions to obtain that these two properties are satisfied.

Theorem 1 *Assume that $\epsilon_k \geq \underline{\epsilon} > 0$ for all $k \in I\!N$. If the sequence $\{x^k\}$ generated by the General Algorithm is bounded and is such that $\|x^{k+1} - x^k\| \to 0$, $k \in I\!N$, then every limit point of $\{x^k\}_{k \in I\!N}$ is a solution of problem (EP).*

Proof. Let x^* be a limit point of $\{x^k\}_{k \in I\!N}$ and let $\{x^k\}_{k \in I\!K \subset I\!N}$ be some subsequence converging to x^*. Since $\|x^{k+1} - x^k\| \to 0$, we also have $\{x^{k+1}\}_{k \in I\!K} \to x^*$. Hence, as $\bar{f}_k \leq f_k$ and $f_k(x^{k+1}) \leq \mu \bar{f}_k(x^{k+1})$, we obtain

$$\frac{1}{\mu} f_k(x^{k+1}) \leq \bar{f}_k(x^{k+1}) \leq f_k(x^{k+1}).$$

Now $f_k(x^{k+1}) = f(x^k, x^{k+1}) \to f(x^*, x^*) = 0$ for $k \to +\infty$ because $x^k \to x^*$, $x^{k+1} \to x^*$ for $k \to +\infty$, $k \in K$, and f is continuous. Hence $\bar{f}_k(x^{k+1}) \to 0$

for $k \to +\infty$. On the other hand, since x^{k+1} solves the convex optimization problem $(\bar{P}_k)$, we have

$$0 \in \partial\{\epsilon_k(\bar{f}_k + \psi_C)\}(x^{k+1}) - \nabla h(x^k) + \nabla h(x^{k+1}),$$

i.e.,

$$\nabla h(x^k) - \nabla h(x^{k+1}) \in \partial\{\epsilon_k(\bar{f}_k + \psi_C)\}(x^{k+1}),$$

where ψ_C denotes the indicator function associated with C ($\psi_C(x) = 0$ if $x \in C$ and $+\infty$ otherwise). Using the definition of the subdifferential, we obtain

$$\forall y \in C \quad \bar{f}_k(y) - \bar{f}_k(x^{k+1}) \geq \frac{1}{\epsilon_k}\langle \nabla h(x^k) - \nabla h(x^{k+1}), y - x^{k+1}\rangle. \qquad (10)$$

Applying the Cauchy-Schwarz inequality and the properties $\bar{f}_k \leq f_k$ and ∇h is Lipschitz continuous on C with constant $\Lambda > 0$, we obtain successively for all $y \in C$,

$$\begin{aligned} f_k(y) - \bar{f}_k(x^{k+1}) &\geq -\frac{1}{\epsilon_k}\|\nabla h(x^k) - \nabla h(x^{k+1})\| \, \|y - x^{k+1}\| \\ &\geq -\frac{\Lambda}{\epsilon_k}\|x^k - x^{k+1}\| \, \|y - x^{k+1}\|. \end{aligned}$$

Taking the limit on $k \in I\!\!K$, we deduce

$$\forall y \in C \quad f(x^*, y) \geq 0,$$

because f is continuous, $\bar{f}_k(x^{k+1}) \to 0$, $\|x^k - x^{k+1}\| \to 0$, $\|y - x^{k+1}\| \to \|y - x^*\|$ and $\epsilon_k \geq \underline{\epsilon} > 0$. But this means that x^* is a solution of problem (EP). □

In the next theorem, we give conditions to obtain that the sequence $\{x^k\}$ is bounded and that $\|x^{k+1} - x^k\| \to 0$.

Theorem 2 *Assume that there exist γ, c, $d > 0$ and a nonnegative function $g : C \times C \to I\!\!R$ such that for all $x, y, z \in C$,*

$$(i) \quad f(x, y) \geq 0 \Rightarrow f(y, x) \leq -\gamma\, g(x, y);$$

$$(ii) \quad f(x, z) - f(y, z) - f(x, y) \leq c\, g(x, y) + d\, \|z - y\|^2.$$

If the sequence $\{\epsilon_k\}_{k \in I\!\!N}$ is nonincreasing and $\epsilon_k < \dfrac{\beta\mu}{2d}$ for all k and if $\dfrac{c}{\gamma} \leq \mu \leq 1$, then the sequence $\{x^k\}_{k \in I\!\!N}$ generated by the General Algorithm is bounded and $\lim_{k \to +\infty} \|x^{k+1} - x^k\| = 0$.

Proof. Let x^* be a solution of problem (EP) and consider for each $k \in I\!N$ the Lyapounov function $\Gamma^k : C \times C \to I\!R$ defined for all $y, z \in C$, by

$$\Gamma^k(y,z) = h(z) - h(y) - \langle \nabla h(y), z - y\rangle + \frac{\epsilon_k}{\mu} f(z,y). \tag{11}$$

Since h is strongly convex with modulus $\beta > 0$, we have immediately that, for all $x \in C$,

$$\Gamma^k(x^k, x^*) \geq \frac{\beta}{2}\|x^k - x^*\|^2. \tag{12}$$

Noticing that $\epsilon_{k+1} \leq \epsilon_k$ for all $k \in I\!N$, the difference $\Gamma^{k+1}(x^k, x^*) - \Gamma^k(x^k, x^*)$ can then be evaluated as follows:

$$\begin{aligned}\Gamma^{k+1}(x^{k+1}, x^*) - \Gamma^k(x^k, x^*) &\leq h(x^k) - h(x^{k+1}) + \langle \nabla h(x^k), x^{k+1} - x^k\rangle \\ &\quad + \langle \nabla h(x^k) - \nabla h(x^{k+1}), x^* - x^{k+1}\rangle \\ &\quad + \frac{\epsilon_k}{\mu}\{f(x^*, x^{k+1}) - f(x^*, x^k)\} \\ &= s_1 + s_2 + s_3,\end{aligned}$$

with

$$\begin{aligned} s_1 &= h(x^k) - h(x^{k+1}) + \langle \nabla h(x^k), x^{k+1} - x^k\rangle, \\ s_2 &= \langle \nabla h(x^k) - \nabla h(x^{k+1}), x^* - x^{k+1}\rangle, \\ s_3 &= \frac{\epsilon_k}{\mu}\{f(x^*, x^{k+1}) - f(x^*, x^k)\}.\end{aligned}$$

For s_1, we easily derive from the strong convexity of h that

$$s_1 \leq -\frac{\beta}{2}\|x^{k+1} - x^k\|^2. \tag{13}$$

For s_2, we obtain, taking $y = x^*$ in (10)

$$\begin{aligned} s_2 = \langle \nabla h(x^k) - \nabla h(x^{k+1}), x^* - x^{k+1}\rangle &\leq \epsilon_k\{\bar{f}_k(x^*) - \bar{f}_k(x^{k+1})\} \\ &\leq \epsilon_k\{f(x^k, x^*) - \tfrac{1}{\mu} f(x^k, x^{k+1})\},\end{aligned}$$

because $\bar{f}_k \leq f(x^k, \cdot)$ and (9) hold. Then, using assumption (ii), we deduce that

$$\begin{aligned} s_2 + s_3 &\leq \epsilon_k\{f(x^k, x^*) - \frac{1}{\mu} f(x^k, x^{k+1})\} + \frac{\epsilon_k}{\mu}\{f(x^*, x^{k+1}) - f(x^*, x^k)\} \\ &= \frac{\epsilon_k}{\mu}\{f(x^*, x^{k+1}) - f(x^*, x^k) - f(x^k, x^{k+1})\} + \epsilon_k f(x^k, x^*) \\ &\leq \frac{\epsilon_k}{\mu}\{c\, g(x^*, x^k) + d\, \|x^{k+1} - x^k\|^2\} + \epsilon_k f(x^k, x^*).\end{aligned}$$

Consequently, we have

$$\Gamma^{k+1}(x^{k+1},x^*) - \Gamma^k(x^k,x^*) \le \epsilon_k f(x^k,x^*) - \frac{1}{2}(\beta - 2\frac{\epsilon_k d}{\mu})\|x^{k+1}-x^k\|^2 + \frac{\epsilon_k c}{\mu} g(x^*,x^k).$$

Applying assumption (i) with $x = x^*$ and $y = x^k$, since $f(x^*,x^k) \ge 0$, we obtain

$$f(x^k,x^*) \le -\gamma\, g(x^*,x^k).$$

Finally, we have that

$$\Gamma^{k+1}(x^{k+1},x^*) - \Gamma^k(x^k,x^*) \le -\frac{1}{2}(\beta - 2\frac{\epsilon_k d}{\mu})\|x^{k+1}-x^k\|^2 - \epsilon_k(\gamma - \frac{c}{\mu})g(x^*,x^k). \tag{14}$$

Since $\epsilon_k < \dfrac{\beta\mu}{2d}$ for all k and $\mu \ge \dfrac{c}{\gamma}$, from (12) and (14), it follows that $\{\Gamma^k(x^k,x^*)\}_{k\in\mathbb{N}}$ is a nonincreasing sequence bounded below by 0. Hence, it is convergent in $\mathbb{R}$. Using again (12), we deduce that the sequence $\{x^k\}_{k\in\mathbb{N}}$ is bounded and, passing to the limit in (14), that the sequence $\{\|x^{k+1} - x^k\|\}_{k\in\mathbb{N}}$ converges to zero. □

Combining Theorems 1 and 2, we deduce the following theorem.

Theorem 3 *Assume that $\epsilon_k \ge \underline{\epsilon} > 0$ for all $k \in \mathbb{N}$ and that all assumptions of Theorem 2 are fulfilled, then the sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by the General Algorithm converges to a solution of problem (EP).*

Remark 21 *The same result as Theorem 2 can also be obtained when condition (ii) is replaced by the following condition:*

(iii) $\qquad f(x,z) - f(y,z) - f(x,y) \le c\, g(x,y) + d\,\|z-y\|,$

and when the series $\sum_{k=0}^{+\infty}\epsilon_k^2$ is convergent. If, in addition, $g(x,y) = 0$ and $\sum_{k=0}^{+\infty}\epsilon_k = +\infty$, then the convergence of the sequence $\{x^k\}$ to a solution of (EP) can be proved as in [18] by using the gap function $l(x) = -f(x,x^*)$ where x^* is a solution of (EP).

So in order to obtain the convergence of the General Algorithm, we need conditions (i) and (ii) or conditions (i) and (iii). Condition (i) is a monotonicity condition. Indeed, when $g = 0$, this condition means that f is pseudomonotone and when $g(x,y) = \|x-y\|^2$ that f is strongly pseudomonotone with modulus γ. Conditions (ii) and (iii) are Lipschitz-type conditions. The link between conditions (i) and (ii) or (iii) is made by the function g whose choice depends on the structure of the problem. So, for example, when $f(x,y) = \varphi(x) - \varphi(y)$ with $\varphi : C \to \mathbb{R}$ a continuous convex function, i.e., when problem (EP) is a constrained convex optimization problem, it suffices to choose $g(x,y) = 0$ for all $x, y \in C$ to obtain that (i), (ii) and (iii) are satisfied.

Other sufficient conditions to get conditions $(i), (ii)$ and (iii) are given in the next two propositions.

Proposition 1 *If f is pseudomonotone and $f(x,\cdot)$ is Lipschitz continuous on C uniformly in x, then conditions (i) and (iii) are satisfied with $g(x,y) = 0$.*

Proof. Let $x, y, z \in C$. Since $f(y,y) = 0$, we have

$$\begin{aligned} f(x,z) - f(y,z) - f(x,y) &= f(x,z) - f(x,y) + f(y,y) - f(y,z) \\ &\leq 2L\|z - y\|, \end{aligned}$$

where L denotes the Lipschitz constant of $f(x,\cdot)$. $\square$

Proposition 2 *If f is strongly monotone and if (4) holds, then conditions (i) and (ii) are satisfied with $g(x,y) = \|x - y\|^2$.*

Proof. If $f(x,y) \geq 0$, then by the strong monotonicity of f, we have

$$f(y,x) \leq -f(x,y) - \gamma\|x - y\|^2 \leq -\gamma\|x - y\|^2 = -\gamma\, g(x,y).$$

Condition (ii) is immediate from (4). $\square$

As a consequence of this proposition, Theorem 3 is also valid under assumptions (3) and (4). In particular, when $\mu = 1$, the conditions imposed on the parameters are $\varepsilon_k < \beta/(2d)$ for all k and $c/\gamma \leq 1$, and Theorem 3.1 of Mastroeni [16] is recovered. So, when $\mu = 1$, Theorem 3 can be considered as a generalization of this theorem.

Finally, we consider the case where f is given by (5) and we introduce the following definition: F is $\varphi-$co-coercive on C if there exists $\gamma > 0$ such that for all $x, y \in C$, if $\langle F(x), y - x\rangle + \varphi(y) - \varphi(x) \geq 0$ holds, then

$$\langle F(y), y - x\rangle + \varphi(y) - \varphi(x) \geq \gamma\,\|F(y) - F(x)\|^2. \tag{15}$$

It is easy to prove that if F is co-coercive on C, then F is φ-co-coercive on C. Indeed, if F is co-coercive on C, then there exists $\gamma > 0$ such that

$$\forall x, y \in C \quad \langle F(x) - F(y), x - y\rangle \geq \gamma\,\|F(x) - F(y)\|^2.$$

But then, if $\langle F(x), y - x\rangle + \varphi(y) - \varphi(x) \geq 0$, we have

$$\begin{aligned} \langle F(y), y - x\rangle + \varphi(y) - \varphi(x) &= \langle F(y) - F(x), y - x\rangle + \langle F(x), y - x\rangle \\ &\quad + \varphi(y) - \varphi(x) \geq \gamma\,\|F(y) - F(x)\|^2, \end{aligned}$$

i.e., inequality (15). Now in order to find again Zhu and Marcotte's convergence result ([19] Theorem 3.2) from our Theorem 3, we need the following proposition where another choice of g is necessary to obtain (i) and (ii).

Proposition 3 *Let $f(x,y) = \langle F(x), y-x\rangle + \varphi(y) - \varphi(x)$ where $F : C \to I\!R^n$ is continuous and $\varphi : C \to I\!R$ is convex. If F is $\varphi-$co-coercive on C, then there exist a nonnegative function $g : C \times C \to I\!R$ and $\gamma > 0$ such that for all $x, y, z \in C$ and for all $\nu > 0$,*

$$f(x,y) \geq 0 \Rightarrow f(y,x) \leq -\gamma\, g(x,y),$$
$$f(x,z) - f(y,z) - f(x,y) \leq \frac{1}{2\nu}\, g(x,y) + \frac{\nu}{2}\, \|z-y\|^2.$$

Proof. Using the definition of f and the $\varphi-$co-coercivity of F on C, there exists $\gamma > 0$ such that for all $x \in C$

$$f(x,y) \geq 0 \Rightarrow f(y,x) \leq -\gamma \|F(y) - F(x)\|^2.$$

On the other hand, we have for any $\nu > 0$,

$$\begin{aligned} f(x,z) - f(y,z) - f(x,y) &= \langle F(x) - F(y), z - y\rangle \\ &\leq \frac{1}{2\nu}\|F(x) - F(y)\|^2 + \frac{\nu}{2}\|z-y\|^2. \end{aligned}$$

So, with $g(x,y) = \|F(y) - F(x)\|^2$, we obtain the two inequalities. □

Using this proposition, Theorem 3.2 of [19] can be derived from our Theorem 3 with $\mu = 1$. Indeed, by choosing $\nu = 1/(2\gamma)$, we obtain $c = 1/(2\nu) = \gamma$ and $d = \nu/2 = 1/(4\gamma)$. Then conditions $c/\gamma \leq 1$ and $\varepsilon_k < \beta/(2d)$ of Theorem 3 reduce to $\varepsilon_k < 2\beta\gamma$, which is exactly the condition imposed by Zhu and Marcotte in their convergence theorem.

3 A Bundle Algorithm

In order to obtain an implementable algorithm, we have now to say how to construct a $\mu-$approximation $\bar{f}_k$ of f_k at x^k such that problem $(\bar{P}_k)$ is easier to solve than problem (P_k). Here we assume that $\mu \in (0,1)$. In that purpose, we observe that if $\bar{f}_k$ is a piecewise linear convex function of the form

$$\bar{f}_k(y) = \max_{1 \leq j \leq p} \{a_j^T y + b_j\},$$

where $a_j \in I\!R^n,\ b_j \in I\!R$ for $j = 1, \dots, p$, the problem $(\bar{P}_k)$ is equivalent to the problem

$$(\mathrm{QP}_k) \quad \begin{cases} \min \{\varepsilon_k v + h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k \rangle\} \\ \text{s.t.} \;\; v \geq a_j^T y + b_j,\ j = 1, \dots, p \\ \qquad y \in C. \end{cases}$$

When h is the squared norm and C is a closed convex polyhedron, this problem becomes quadratic.

There exist many efficient numerical methods for solving such a problem. When $\bar{f}_k$ is a piecewise linear convex function, it is judicious to construct $\bar{f}_k$, piece by piece, by generating successive models

$$\bar{f}_k^i,\ i = 1, 2, \dots$$

until (if possible) $\bar{f}_k^{i_k}$ is a $\mu-$approximation of f_k at x^k for some $i_k \geq 1$. For $i = 1, 2, \dots,$ we denote by y_k^i the unique solution of the problem

$$(P_k^i) \quad \min_{y \in C} \{\epsilon_k \bar{f}_k^i(y) + h(y) - h(x^k) - \langle \nabla h(x^k), y \rangle\},$$

and we set $\bar{f}_k = \bar{f}_k^{i_k}$ and $x^{k+1} = y_k^{i_k}$.
In order to obtain a $\mu-$approximation $\bar{f}_k^{i_k}$ of f_k at x^k, we have to impose some conditions on the successive models $\bar{f}_k^i$, $i = 1, 2, \dots$. However, before presenting them, we need to define the affine functions $l_k^i,\ i = 1, 2, \dots$ by

$$l_k^i(y) = \bar{f}_k^i(y_k^i) + \langle \gamma_k^i, y - y_k^i \rangle \quad \forall y \in C,$$

where $\gamma_k^i = \dfrac{1}{\epsilon_k}[\nabla h(x^k) - \nabla h(y_k^i)]$. By optimality of y_k^i, we have

$$\gamma_k^i \in \partial(\bar{f}_k^i + \psi_C)(y_k^i). \tag{16}$$

It is then easy to observe that

$$l_k^i(y_k^i) = \bar{f}_k^i(y_k^i) \quad \text{and} \quad l_k^i(y) \leq \bar{f}_k^i(y) \text{ for all } y \in C. \tag{17}$$

Now we assume that the following conditions inspired for [5] are satisfied by the convex models $\bar{f}_k^i$,

$(C1)$ $\bar{f}_k^i \leq f_k$ on C for $i = 1, 2, \dots$
$(C2)$ $\bar{f}_k^{i+1} \geq f_k(y_k^i) + \langle s(y_k^i), \cdot - y_k^i \rangle$ on C for $i = 1, 2, \dots$
$(C3)$ $\bar{f}_k^{i+1} \geq l_k^i$ on C for $i = 1, 2, \dots,$

where $s(y_k^i)$ denotes the subgradient of f_k available at y_k^i.

Several models fulfill these conditions. For example, for the first model $\bar{f}_k^1$, we can take the linear function

$$\bar{f}_k^1(y) = f_k(x^k) + \langle s(x^k), y - x^k \rangle \text{ for all } y \in C.$$

Since $s(x^k) \in \partial f_k(x^k)$, condition $(C1)$ is satisfied for $i=1$. For the next models $\bar{f}_k^i, i=2,\ldots,$ there exist several possibilities. A first example is to take for $i=1,2,\ldots$

$$\bar{f}_k^{i+1}(y) = \max\{l_k^i(y), f_k(y_k^i) + \langle s(y_k^i), y - y_k^i\rangle\}. \tag{18}$$

Conditions $(C2),(C3)$ are obviously satisfied and condition $(C1)$ is also satisfied for $i=2,3,\ldots$, because each linear piece of these functions are below f_k. Another example is to take for $i=1,2,\ldots$

$$\bar{f}_k^{i+1}(y) = \max_{0\le j\le i}\{f_k(y_k^j) + \langle s(y_k^j), y - y_k^j\rangle\} \tag{19}$$

where $y_k^0 = x^k$. Since $s(y_k^j) \in \partial f_k(y_k^j)$ for $j=0,\ldots,i$ and since $\bar{f}_k^{i+1} \ge \bar{f}_k^i \ge l_k^i$, it is easy to see that conditions $(C1)-(C3)$ are satisfied.
Comparing (18) and (19), we can say that l_k^i plays the same role as the i linear functions $f_k(y_k^j) + \langle s(y_k^j), y - y_k^j\rangle,\ j=0,\ldots,i-1$. It is the reason why this function l_k^i is called the aggregate affine function (see, e.g., [5]). The first example (18) is interesting from the numerical point of view, because its use allows to limit the number of linear constraints in subproblems (QP_k).

Now the algorithm allowing to pass from x^k to x^{k+1}, i.e., to make what is called a serious step, can be expressed as follows.

Serious Step Algorithm.
Let $x^k \in C$ and $\mu \in (0,1)$. Set $i=1$.
Step 1. *Choose* $\bar{f}_k^i$ *a convex function that satisfies* $(C1)-(C3)$ *and solve problem* (P_k^i) *to get* y_k^i.
Step 2. *If* $f_k(y_k^i) \le \mu \bar{f}_k^i(y_k^i)$*, then set* $x^{k+1} = y_k^i$, $i_k = i$ *and STOP;* x^{k+1} *is a serious step.*
Step 3. *Increase* i *by 1 and go to Step 1.*

Our aim is now to prove that if x^k is not a solution of problem (EP) and if the models $\bar{f}_k^i, i=1,\ldots$ satisfy $(C1)-(C3)$, then there exists $i_k \ge 1$ such that $\bar{f}_k^{i_k}$ is a $\mu-$approximation of f_k at x^k, i.e., that the STOP occurs at Step 2 after finitely many iterations.

To prove that, we need a lemma whose proof uses the following functions:

$$\tilde{l}_k^i(y) = l_k^i(y) + \frac{1}{\epsilon_k}\{h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k\rangle\},$$
$$\tilde{f}_k^i(y) = \bar{f}_k^i(y) + \frac{1}{\epsilon_k}\{h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k\rangle\}.$$

Using (16) and (17), we obtain:

$$\begin{aligned}
\tilde{l}_k^i(y) - \tilde{l}_k^i(y_k^i) &= l_k^i(y) + \frac{1}{\epsilon_k}\{h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k\rangle\} \\
&\quad - l_k^i(y_k^i) - \frac{1}{\epsilon_k}\{h(y_k^i) - h(x^k) - \langle \nabla h(x^k), y_k^i - x^k\rangle\} \\
&= \bar{f}_k^i(y_k^i) + \langle \gamma_k^i, y - y_k^i\rangle + \frac{1}{\epsilon_k}\{h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k\rangle\} \\
&\quad - \bar{f}_k^i(y_k^i) - \frac{1}{\epsilon_k}\{h(y_k^i) - h(x^k) - \langle \nabla h(x^k), y_k^i - x^k\rangle\} \\
&= \frac{1}{\epsilon_k}\{h(y) - h(y_k^i) - \langle \nabla h(y_k^i), y - y_k^i\rangle\}.
\end{aligned}$$

Consequently, we obtain

$$\tilde{l}_k^i(y) = \tilde{l}_k^i(y_k^i) + \frac{1}{\epsilon_k}\{h(y) - h(y_k^i) - \langle \nabla h(y_k^i), y - y_k^i\rangle\}. \tag{20}$$

Moreover from (17) and $(C3)$, we have

$$\tilde{f}_k^i(x^k) = \bar{f}_k^i(x^k) \tag{21}$$

$$\tilde{l}_k^i(y_k^i) = \tilde{f}_k^i(y_k^i) \tag{22}$$

$$\tilde{l}_k^i \leq \tilde{f}_k^{i+1} \quad \text{on } C. \tag{23}$$

Lemma 1 *Assume that the models $\bar{f}_k^i$, $i \in I\!N_0$ satisfy conditions $(C1)-(C3)$ and let, for each i, y_k^i be the unique solution of problem (P_k^i). Then*

(i) $f_k(y_k^i) - \bar{f}_k^i(y_k^i) \to 0$,

(ii) $y_k^i \to \bar{y}_k \equiv \arg\min_{y \in C}\{\epsilon_k f_k(y) + h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k\rangle\}$,

where $i \to +\infty$.

Proof. (i) To obtain the first statement, we use the following three steps.

(1) The sequence $\{\tilde{l}_k^i(y_k^i)\}_{i \in I\!N_0}$ is convergent and $y_k^{i+1} - y_k^i \to 0$ when $i \to +\infty$.

For all i, we have

$$\begin{aligned}
0 = f_k(x^k) &\geq \bar{f}_k^{i+1}(x^k) && \text{(by } C1) \\
&= \tilde{f}_k^{i+1}(x^k) && \text{(by (21))} \\
&\geq \tilde{f}_k^{i+1}(y_k^{i+1}) && \text{(by definition of } y_k^{i+1}) \\
&= \tilde{l}_k^{i+1}(y_k^{i+1}) && \text{(by (22))} \\
&\geq \tilde{l}_k^i(y_k^{i+1}) && \text{(by (23))} \\
&= \tilde{l}_k^i(y_k^i) + \frac{1}{\epsilon_k} D_h(y_k^{i+1}, y_k^i) && \text{(by (20))} \\
&\geq \tilde{l}_k^i(y_k^i) + \frac{\beta}{2\epsilon_k}\|y_k^{i+1} - y_k^i\|^2 && \text{(by strong convexity of } h \text{ on } C) \\
&\geq \tilde{l}_k^i(y_k^i)
\end{aligned}$$

where $D_h(y,z) = h(y) - h(z) - \langle \nabla h(z), y - z\rangle$. From these relations, we have for all i, that

$$\tilde{l}_k^{i+1}(y_k^{i+1}) \geq \tilde{l}_k^i(y_k^i).$$

So, the sequence $\{\tilde{l}_k^i(y_k^i)\}_{i\in I\!N_0}$ is nonincreasing and bounded above by 0. Consequently $\{\tilde{l}_k^i(y_k^i)\}_{i\in I\!N_0}$ is convergent and $y_k^{i+1} - y_k^i \to 0$ when $i \to +\infty$.

(2) The sequence $\{y_k^i\}_{i\in I\!N_0}$ is bounded.
We have (for y fixed)

$$\begin{aligned}
&f_k(y) + \frac{1}{\epsilon_k}\{h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k\rangle\} \\
&\quad\geq \bar{f}_k^{i+1}(y) + \frac{1}{\epsilon_k}\{h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k\rangle\} \quad \text{(by } C1) \\
&\quad= \tilde{f}_k^{i+1}(y) \\
&\quad\geq \tilde{l}_k^i(y) \quad \text{(by (23))} \\
&\quad= \tilde{l}_k^i(y_k^i) + \frac{1}{\epsilon_k}\{h(y) - h(y_k^i) - \langle \nabla h(y_k^i), y - y_k^i\rangle\} \quad \text{(by (20))} \\
&\quad\geq \tilde{l}_k^i(y_k^i) + \frac{\beta}{2\epsilon_k}\|y - y_k^i\|^2 \quad \text{(by } h \text{ is strongly convex on } C).
\end{aligned}$$

Since the sequence $\{\tilde{l}_k^i(y_k^i)\}_{i\in I\!N_0}$ is convergent, the sequence $\{y - y_k^i\}_{i\in I\!N_0}$ is bounded and thus the sequence $\{y_k^i\}_{i\in I\!N_0}$ is also bounded.

(3) $f_k(y_k^{i+1}) - \bar{f}_k^{i+1}(y_k^{i+1}) \to 0$.
We have successively

$$\begin{aligned}
\langle s(y_k^i), y_k^{i+1} - y_k^i\rangle &\leq \bar{f}_k^{i+1}(y_k^{i+1}) - f_k(y_k^i) \quad \text{(by } C2) \\
&\leq f_k(y_k^{i+1}) - f_k(y_k^i) \quad \text{(by } C1) \\
&\leq \langle s(y_k^{i+1}), y_k^{i+1} - y_k^i\rangle \quad \text{(by definition of } s(y_k^{i+1})).
\end{aligned}$$

Since $\{y_k^i\}_{i\in I\!N_0}$ is bounded, then, by Theorem 24.7 in [17], the set $\cup_i \partial f_k(y_k^i)$ is bounded and thus the sequence $\{s(y_k^i)\}_{i\in I\!N_0}$ is bounded. So, we obtain

$$\bar{f}_k^{i+1}(y_k^{i+1}) - f_k(y_k^i) \to 0 \text{ and } f_k(y_k^{i+1}) - f_k(y_k^i) \ \to \ 0,$$

and consequently,

$$f_k(y_k^{i+1}) - \bar{f}_k^{i+1}(y_k^{i+1}) = f_k(y_k^{i+1}) - f_k(y_k^i) + f_k(y_k^i) - \bar{f}_k^{i+1}(y_k^{i+1}) \ \to \ 0.$$

(ii) $y_k^i \to \bar{y}_k \equiv \arg\min_{y\in C}\{\epsilon_k f_k(y) + h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k\rangle\}$.
Since the sequence $\{y_k^i\}_{i\in I\!N_0}$ is bounded, it remains to prove that every limit point y_k^* of this sequence is equal to $\bar{y}_k$, i.e., that

$$\frac{1}{\epsilon_k}\{\nabla h(x^k) - \nabla h(y_k^*)\} \in \partial(f_k + \psi_C)(y_k^*)$$

or, by definition of the subdifferential, we obtain from (16) and $(C1)$ that

$$\forall y \in C \quad f_k(y) \geq \bar{f}_k^i(y) \geq \bar{f}_k^i(y_k^i) + \frac{1}{\epsilon_k}\langle \nabla h(x^k) - \nabla h(y_k^i), y - y_k^i \rangle,$$

i.e.,

$$\begin{aligned} \forall y \in C \quad f_k(y) \geq [\bar{f}_k^i(y_k^i) - f_k(y_k^i)] + [f_k(y_k^i) - f_k(y_k^*)] \\ + f_k(y_k^*) + \frac{1}{\epsilon_k}\langle \nabla h(x^k) - \nabla h(y_k^i), y - y_k^i \rangle. \end{aligned} \tag{24}$$

Since y_k^* is a limit point of $\{y_k^i\}_{i \in I\!N_0}$, there exists $I\!K \subseteq I\!N_0$ such that

$$y_k^i \to y_k^* \text{ for } i \in I\!K, i \to +\infty.$$

Taking the limit (for $i \in I\!K$) of both sides of (24), we obtain, for all $y \in C$, that

$$\begin{aligned} f_k(y) \geq \lim_i [\bar{f}_k^i(y_k^i) - f_k(y_k^i)] + \lim_i [f_k(y_k^i) - f_k(y_k^*)] + f_k(y_k^*) \\ + \frac{1}{\epsilon_k} \lim_i \langle \nabla h(x^k) - \nabla h(y_k^i), y - y_k^i \rangle. \end{aligned}$$

Since $\lim_i [\bar{f}_k^i(y_k^i) - f_k(y_k^i)] = 0$ by (i), $\lim_i [f_k(y_k^i) - f_k(y_k^*)] = 0$ because f_k is continuous, and ∇h is continuous at y_k^*, we deduce that

$$f_k(y) \geq f_k(y_k^*) + \frac{1}{\epsilon_k}\langle \nabla h(x^k) - \nabla h(y_k^*), y - y_k^* \rangle \quad \forall y \in C.$$

This completes the proof. □

Theorem 4 *Assume x^k is not a solution of problem (EP). Then the serious step algorithm stops after finitely many iterations i_k with $\bar{f}_k^{i_k}$ a $\mu-$ approximation of f_k at x_k and with $x^{k+1} = y_k^{i_k}$.*

Proof. Suppose, to get a contradiction, that the STOP never occurs. Then

$$f_k(y_k^i) > \mu \bar{f}_k^i(y_k^i) \geq \mu \bar{f}_k^i(y_k^i) \quad \text{for all } i \in I\!N_0. \tag{25}$$

Moreover, by Lemma 1, $y_k^i \to \bar{y}_k$. Then taking the limit of both members of (25), we obtain

$$f_k(\bar{y}_k) \geq \mu f_k(\bar{y}_k)$$

because f_k is continuous over C and $f_k(y_k^i) - \bar{f}_k^i(y_k^i) \to 0$. Hence, since $\mu < 1$, we deduce that $f_k(\bar{y}_k) \geq 0$.

On the other hand, by definition of $\bar{y}_k$ (see Lemma 1), we have, for all $y \in C$, that

$$\begin{aligned} \epsilon_k f_k(\bar{y}_k) + h(\bar{y}_k) - h(x^k) - \langle \nabla h(x^k), \bar{y}_k - x^k \rangle \leq \epsilon_k\, f_k(y) + h(y) - h(x^k) \\ - \langle \nabla h(x^k), y - x^k \rangle. \end{aligned}$$

If we choose $y = x^k$ and observe that $f_k(x^k) = 0$, then this inequality becomes

$$\epsilon_k f_k(\bar{y}_k) \leq -h(\bar{y}_k) + h(x^k) + \langle \nabla h(x^k), \bar{y}_k - x^k \rangle.$$

Finally, using the strong convexity of h yields

$$0 \leq \epsilon_k f_k(\bar{y}_k) \leq -\frac{\beta}{2}\|\bar{y}_k - x^k\|.$$

Consequently $\|\bar{y}_k - x^k\| = 0$ and thus $x^k = \bar{y}_k$. But this means that x^k is a solution of problem (EP), which contradicts the assumption of the theorem. So the Serious Step Algorithm stops after finitely many iterations. □

Incorporating the Serious Step Algorithm in Step 1 of the General Algorithm, we obtain the following algorithm.

Bundle Algorithm for solving problem (EP).
Let an initial point $x^0 \in C$, together with a tolerance $\mu \in (0,1)$ and a positive sequence $\{\epsilon_k\}_{k \in I\!N}$. Set $y_0^0 = x^0$ and $k = 0, i = 1$.
Step 1. *Choose a piecewise linear convex function* $\bar{f}_k^i$ *satisfying* $(C1) - (C3)$ *and solve*

(P_k^i) $\quad \min_{y \in C} \{ \epsilon_k \bar{f}_k^i(y) + h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k \rangle \}$

to obtain the unique optimal solution $y_k^i \in C$.
Step 2. *If*

$$f_k(y_k^i) \leq \mu \bar{f}_k^i(y_k^i), \tag{26}$$

then set $x^{k+1} = y_k^i$, $y_{k+1}^0 = x^{k+1}$, *increase* k *by 1 and set* $i = 0$.
Step 3. *Increase* i *by 1 and go to Step 1.*

From Theorems 3 and 4, we obtain the following convergence results.

Theorem 5 *If after some* k *has been reached, the criterion (26) is never satisfied, then* x^k *is a solution of problem (EP).*

Theorem 6 *Assume that* $\epsilon_k \geq \underline{\epsilon} > 0$ *for all* $k \in I\!N$ *and that all assumptions of Theorem 2 are fulfilled, and that the sequence* $\{x^k\}$ *generated by the Bundle Algorithm is infinite. Then* $\{x^k\}$ *converges to some solution of problem (EP).*

For practical implementation, it is necessary to give a stopping criterion. In order to present it, we introduce the definition of a stationary point.

Definition 2 Let $\Delta \geq 0$. A point $x^* \in I\!R^n$ is called a $\Delta-$stationary point of problem (EP) if $x^* \in C$ and if

$$\exists\, \gamma \in \partial_{\Delta}(f_{x^*} + \psi_C)(x^*) \text{ such that } \|\gamma\| \leq \Delta.$$

Using the definition of the Δ−subdifferential of the convex function $f_{x^*}+\psi_C$, we obtain that if x^* is a Δ−stationary point of problem (EP), then

$$\forall y \in C \quad f_{x^*}(y) \geq f_{x^*}(x^*) + \langle \gamma, y - x^* \rangle - \epsilon \geq -\Delta \|y - x^*\| - \Delta,$$

where we have used $f_{x^*}(x^*) = 0$, the Cauchy-Schwarz inequality and $\|\gamma\| \leq \epsilon$. Observe that if $\Delta = 0$, then a Δ−stationary point of problem (EP) is a solution of problem (EP). Now to prove that the iterate x^k generated by the bundle algorithm is a Δ-stationary point of problem (EP) for k large enough, we need the following results.

Proposition 4 *Let y_k^i be the solution of problem (P_k^i) and let*

$$\gamma_k^i = \frac{1}{\epsilon_k}[\nabla h(x^k) - \nabla h(y_k^i)] \text{ and } \delta_k^i := \langle \gamma_k^i, y_k^i - x^k \rangle - \bar{f}_k^i(y_k^i). \tag{27}$$

Then

$$\delta_k^i \geq 0 \text{ and } \gamma_k^i \in \partial_{\delta_k^i}(f_k + \psi_C)(x^k).$$

Proof. By optimality of y_k^i, we obtain that

$$0 \in \epsilon_k \partial(\bar{f}_k^i + \psi_C)(y_k^i) + \nabla h(y_k^i) - \nabla h(x^k),$$

i.e.,

$$\gamma_k^i \in \partial(\bar{f}_k^i + \psi_C)(y_k^i).$$

Hence by definition of the subdifferential and since $\bar{f}_k^i \leq f_k$, we have, for all $x \in C$

$$f_k(x) \geq \bar{f}_k^i(x) \geq \bar{f}_k^i(y_k^i) + \langle \gamma_k^i, x - y_k^i \rangle. \tag{28}$$

In particular for $x = x^k$, and noting that $f_k(x^k) = 0$, we deduce that

$$0 \geq \bar{f}_k^i(y_k^i) + \langle \gamma_k^i, x^k - y_k^i \rangle,$$

i.e., that $\delta_k^i \geq 0$.
On the other hand, from (28) and the definition of δ_k^i, we can write for all $x \in C$,

$$f_k(x) \geq \bar{f}_k^i(y_k^i) + \langle \gamma_k^i, x - y_k^i \rangle = f_k(x^k) + \langle \gamma_k^i, x - x^k \rangle - \delta_k^i,$$

i.e., that $\gamma_k^i \in \partial_{\delta_k^i}(f_k + \psi_C)(x^k)$. □

Theorem 7 *Assume that $\epsilon_k \geq \underline{\epsilon} > 0$ for all $k \in I\!N$ and that all assumptions of Theorem 2 hold. Let $\{x^k\}$ be the sequence generated by the Bundle Algorithm.*
(i) If $\{x^k\}$ is infinite, then the sequences $\{\gamma_k^{i_k}\}_k$ and $\{\delta_k^{i_k}\}_k$ converge to zero.
(ii) If $\{x^k\}$ is finite with k the latest index, then the sequences $\{\gamma_k^i\}_i$ and $\{\delta_k^i\}_i$ converge to zero.

Proof. (i) Since $\{x^k\}_k$ is infinite, it follows from Theorem 6 that $\{x^k\}$ converges to some solution x^* of problem (EP).
On the other hand, we have, for all k

$$0 \le \|\gamma_k^{i_k}\| = \|\frac{\nabla h(x^k) - \nabla h(y_k^{i_k})}{\epsilon_k}\| \le \frac{\Lambda}{\epsilon}\|x^k - y_k^{i_k}\| = \frac{\Lambda}{\epsilon}\|x^k - x^{k+1}\|,$$

because ∇h is Lipschitz-continuous with constant $\Lambda > 0, \epsilon_k \ge \epsilon > 0$ and $y_k^{i_k} = x^{k+1}$. Since $\|x^{k+1} - x^k\| \to 0$, we obtain that the sequence $\{\gamma_k^{i_k}\}_k$ converges to zero. Moreover, since

$$|\langle \gamma_k^{i_k}, y_k^{i_k} - x^k\rangle| \le \|\gamma_k^{i_k}\| \, \|y_k^{i_k} - x^k\| = \|\gamma_k^{i_k}\| \, \|x^{k+1} - x^k\|,$$

we also obtain that $\langle \gamma_k^{i_k}, y_k^{i_k} - x^k\rangle \to 0$ when $k \to +\infty$. Finally, by definition of x^{k+1} and (26), we have

$$\frac{1}{\mu} f_k(x^{k+1}) \le \bar{f}_k^{i_k}(x^{k+1}) \le f_k(x^{k+1}). \tag{29}$$

But $f_k(x^{k+1}) = f(x^k, x^{k+1}) \to f(x^*, x^*) = 0$ by continuity of f_k, so that (29) implies that $\bar{f}_k^{i_k}(x^{k+1}) \to 0$. Consequently, we obtain that $\delta_k^i \to 0$ when $k \to +\infty$.

(ii) Let k be the latest index of the sequence $\{x^k\}$. Then x^k is a solution of problem (EP) by Theorem 5 and $\{y_k^i\}_i$ converges to $\bar{y}_k$ when $i \to +\infty$ by Lemma 1. Hence $x^k = \bar{y}_k$ and $\|x^k - y_k^i\| \to 0$ when $i \to +\infty$. But this means that $\{\gamma_k^i\}_i$ converges to zero. Moreover, by Lemma 1, for $i \to +\infty$, we have $f_k(y_k^i) - \bar{f}_k^i(y_k^i) \to 0$ and thus $\bar{f}_k^i(y_k^i) = \bar{f}_k^i(y_k^i) - f_k(y_k^i) + f_k(y_k^i) \to 0$ because f_k is continuous and $f_k(y_k^i) = f(x^k, y_k^i) \to f(x^k, x^k) = 0$. Consequently $\delta_k^i \to 0$ when $i \to +\infty$. □

Thanks to Proposition 4 and Theorem 7, we can easily introduce a stopping criterion in the Bundle Algorithm just after Step 1 as follows.
Compute γ_k^i and δ_k^i by using (27). If $\|\gamma_k^i\| \le \Delta$ and $\delta_k^i \le \Delta$, then STOP; x^k is a $\Delta-$stationary point of problem (EP). Otherwise, go to Step 2 of the Bundle Algorithm.

Let us mention that this criterion is a generalization of the classical stopping test for bundle methods in optimization (see, e.g., [15]).

4 Application to variational inequality problems

First we apply the Bundle Algorithm for solving problem (GVIP) under the assumption that $F : I\!R^n \to I\!R^n$ is a continuous mapping and $\varphi : I\!R^n \to I\!R$ a convex function. As we know it, this problem is a particular case of problem

(EP) corresponding to the function f defined, for all $x, y \in I\!R^n$, by $f(x, y) = \langle F(x), y - x\rangle + \varphi(y) - \varphi(x)$. Since the function φ may be nondifferentiable, we choose as model $\bar{f}_k^i$, the function

$$\bar{f}_k^i(y) = \theta_k^i(y) - \varphi(x^k) + \langle F(x^k), y - x^k\rangle,$$

where θ_k^i is a piecewise linear convex function which approximates φ at x^k. Moreover, we assume that this function θ_k^i satisfies the three following conditions:

$(C'1)$ $\theta_k^i \leq \varphi$ on C for $i = 1, 2, \ldots$

$(C'2)$ $\theta_k^{i+1} \geq \varphi(y_k^i) + \langle s'(y_k^i), \cdot - y_k^i\rangle$ on C for $i = 1, 2, \ldots$

$(C'3)$ $\theta_k^{i+1} \geq {l'}_k^i$ on C for $i = 1, 2, \ldots$

where ${l'}_k^i(y) = \theta_k^i(y_k^i) + \langle \gamma_k^i - F(x^k), y - y_k^i\rangle$ and $s(y_k^i)$ denotes the subgradient of φ available at y_k^i.

With these choices, problem (P_k^i) is equivalent to the problem

$$\min_{y\in C} \{\, \epsilon_k \theta_k^i(y) + \epsilon_k\langle F(x^k), y - x^k\rangle + h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k\rangle\},$$

and (26) becomes

$$\varphi(x^k) - \varphi(y_k^i) \geq \mu\,[\varphi(x^k) - \theta_k^i(y_k^i)] + (1-\mu)\langle F(x^k), y_k^i - x^k\rangle.$$

Finally, the Bundle Algorithm can be particularized as follows:

Bundle Algorithm for solving problem (GVIP).

Let an initial point $x^0 \in C$, together with a tolerance $\mu \in (0, 1)$ and a positive sequence $\{\epsilon_k\}_{k\in I\!N}$. Set $y_0^0 = x^0$ and $k = 0$, $i = 1$.

Step 1. *Choose a piecewise linear convex function θ_k^i satisfying $(C'1) - (C'3)$ and solve*

$$\min_{y\in C} \{\, \epsilon_k \theta_k^i(y) + \epsilon_k\langle F(x^k), y - x^k\rangle + h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k\rangle\} \quad (30)$$

to obtain the unique optimal solution $y_k^i \in C$.

Step 2. *If*

$$\varphi(x^k) - \varphi(y_k^i) \geq \mu\,[\varphi(x^k) - \theta_k^i(y_k^i)] + (1-\mu)\,\langle F(x^k), y_k^i - x^k\rangle, \quad (31)$$

then set $x^{k+1} = y_k^i, y_{k+1}^0 = x^{k+1}$, increase k by 1 and set $i = 0$.

Step 3. *Increase i by 1 and go to Step 1.*

This algorithm was presented by Salmon et al. in [18] and proven to be convergent under the assumption that F is $\varphi-$co-coercive on C. Thanks to Proposition 3, we can deduce from Theorem 6 the convergence theorems obtained in [18] for the bundle method applied for solving problem (GVIP) (see Theorems 4.2 and 4.3 in [18]).

Theorem 8 *Assume that the sequence $\{\epsilon_k\}$ is nonincreasing and satisfies $0 < \epsilon \leq \epsilon_k$ for all k. If F is $\varphi-$co-coercive on C with $\gamma > \dfrac{\epsilon_0}{2\,\beta\,\mu^2}$, and if the sequence $\{x^k\}$ generated by the Bundle Algorithm for solving (GVIP) is infinite, then the sequence $\{x^k\}$ converges to some solution of problem (GVIP).*

Proof. From Theorem 6 and Proposition 3, we only have to prove that if $\gamma > \dfrac{\epsilon_0}{2\beta\mu^2}$ then there exists $\tau > 0$ such that $\epsilon_0 < \dfrac{\beta\mu}{\tau}$ and $\mu \geq \dfrac{1}{2\tau\gamma}$. Since $\epsilon_0 < 2\beta\mu^2\gamma$, it is sufficient to set $\tau = \dfrac{1}{2\mu\,\gamma} > 0$ to obtain the two inequalities. □

As a second application, we apply the general algorithm to the multivalued variational inequality problem (MVIP). This problem corresponds to problem (EP) with the function f defined, for all $x, y \in C$, by $f(x,y) = \sup_{\xi \in F(x)} \langle \xi, y - x\rangle$ where $F : C \to 2^{I\!R^n}$ is a continuous multivalued mapping with compact values. Thanks to Proposition 23 in [2], it is easy to see that f is continuous on $C \times C$. At iteration k, we consider the approximating function $\bar{f}_k(y) = \langle \xi^k, y - x^k\rangle$ with $\xi^k \in F(x^k)$. Here, we assume that at least one element of $F(x)$ is available for each $x \in C$. When h is the squared norm, the subproblem $(\bar{P}_k)$ becomes

$$\min_{y \in C}\{\epsilon_k \langle \xi^k, y - x^k\rangle + \frac{1}{2}\|y - x^k\|^2\}. \tag{32}$$

We observe that the optimality conditions associated with (32) are

$$\langle \epsilon_k \xi^k + y^k - x^k, y - y^k\rangle \geq 0 \quad \forall y \in C, \tag{33}$$

where y^k is a solution of $(\bar{P}_k)$. In other words, y^k is the orthogonal projection of the vector $x^k - \epsilon_k \xi^k$ over C. This problem is a particular convex quadratic programming problem whose solution can be found explicitly when C has a special structure as a box, a ball, ... Without loss of generality, we can assume that $y^k \neq x^k$. Indeed, if $y^k = x^k$, then it is easy to see that x^k is a solution of (MVIP).

Our aim is first to find conditions to ensure that the function $\bar{f}_k$ defined above is a μ-approximation of f_k at x^k and then to apply Theorem 3 to get the convergence of the sequence $\{x^k\}$. In that purpose, we introduce the following definitions.

Definition 3 Let C be a nonempty closed convex subset of $I\!R^n$ and let $F : C \to 2^{I\!R^n}$.

(i) F is strongly monotone on C if $\exists\,\alpha > 0$ such that $\forall x, y \in C$, $\forall \xi_1 \in F(x)$, $\forall \xi_2 \in F(y)$, one has

$$\langle \xi_1 - \xi_2, x - y\rangle \geq \alpha \|x - y\|^2.$$

(ii) F is Lipschitz continuous on C if $\exists\, L > 0$ such that $\forall x, y \in C$, one has

$$g(x, y) \leq L\|x - y\|,$$

where

$$g(x, y) := \sup_{\xi_1 \in F(x)} \inf_{\xi_2 \in F(y)} \|\xi_1 - \xi_2\|^2. \tag{34}$$

(iii) F is co-coercive on C if $\exists\,\gamma > 0$ such that $\forall x, y \in C$, $\forall \xi_1 \in F(x)$, $\forall \xi_2 \in F(y)$, one has

$$\langle \xi_1 - \xi_2, x - y\rangle \geq \gamma g(x, y)\,.$$

In the next proposition, we present the main property of the function $\bar{f}_k$.

Proposition 5 *Assume F is co-coercive on C with constant $\gamma > 0$. Let $\mu \in (0, 1)$ and $x^k \in C$. If $\varepsilon_k \leq 4\gamma(1-\mu)$, then the function $\bar{f}_k(y) = \langle \xi_k, y - x^k\rangle$ with $\xi_k \in F(x^k)$ is a μ-approximation of f_k at x^k, i.e., $\bar{f}_k \leq f_k$ and $f_k(y^k) \leq \mu \bar{f}_k(y^k)$ where y^k is a solution of problem $(\bar{P}_k)$.*

Proof. Let $\mu \in (0, 1)$ and ξ_k, $\xi \in F(x^k)$. From (33) with $y = x^k$, we deduce that

$$\epsilon_k \langle \xi_k, y^k - x^k\rangle \leq -\|x^k - y^k\|^2 < 0. \tag{35}$$

Using successively the co-coercivity of F, the definition of g in (34) and the Cauchy Schwarz inequality. We have, for every $\eta \in F(y^k)$ and for any $\nu > 0$, that

$$\begin{aligned}\langle \xi - \xi_k, y^k - x^k\rangle &= \langle \xi - \eta, y^k - x^k\rangle + \langle \eta - \xi_k, y^k - x^k\rangle \\ &\leq -\gamma g(x^k, y^k) + \|\eta - \xi_k\|\,\|y^k - x^k\| \\ &\leq -\gamma g(x^k, y^k) + (1/2\nu)\|\eta - \xi_k\|^2 + \nu/2\|y^k - x^k\|^2.\end{aligned}$$

Taking the infimum on $\eta \in F(y^k)$ and using (35), we obtain

$$\begin{aligned}\langle \xi - \xi_k, y^k - x^k\rangle &\leq -\gamma g(x^k, y^k) + \tfrac{1}{2\nu} \inf_{\eta \in F(y^k)} \|\eta - \xi_k\|^2 + \tfrac{\nu}{2}\|y^k - x^k\|^2 \\ &\leq (\tfrac{1}{2\nu} - \gamma)\, g(x^k, y^k) - \frac{\nu\varepsilon_k}{2}\, \langle \xi_k, y^k - x^k\rangle\end{aligned}$$

for all $\nu > 0$. Choosing $\nu = 1/(2\gamma)$, we can write

$$\langle \xi, y^k - x^k\rangle \leq (1 - \frac{\varepsilon_k}{4\gamma})\, \langle \xi_k, y^k - x^k\rangle.$$

Finally, taking the supremum on $\xi \in F(x^k)$, and using the condition $\varepsilon_k < 4\gamma(1-\mu)$, we deduce the thesis. $\square$

Since $\bar{f}_k$ is a μ-approximation of f_k at x^k for a suitable value of ε_k, using this approximating function, the general algorithm becomes:
Given $x^k \in C$ and $\varepsilon_k > 0$, choose $\xi_k \in F(x^k)$ and solve the problem

$$\min_{y \in C} \{\epsilon_k \langle \xi_k, y - x^k \rangle + \frac{1}{2}\|y - x^k\|^2\}$$

to get x^{k+1}.

In particular case, when F is co-coercive on C, the assumptions (i) and (ii) of Theorem 2 are satisfied.

Proposition 6 *Let $f(x,y) = \sup_{\xi \in F(x)} \langle \xi, y - x \rangle$ and g defined by (34). Then*
(i) for every $x, y, z \in C$ and for any $\nu > 0$,

$$f(x,z) - f(y,z) - f(x,y) \leq \frac{1}{2\nu}\, g(x,y) + \frac{\nu}{2}\, \|z - y\|^2,$$

(ii) if F is co-coercive on C with constant γ, then for every $x, y \in C$,

$$f(x, y \geq 0 \Rightarrow f(y,x) \leq -\gamma g(x,y).$$

Finally, for the sequence $\{x^k\}$ generated by this algorithm, we obtain the following convergence theorem.

Theorem 9 *Assume F is co-coercive on C with constant $\gamma > 0$. Let $\{\varepsilon_k\}$ be a nonincreasing sequence bounded away from 0. If $\varepsilon_k < 4(2-\sqrt{3})\,\gamma$ for all k, then the sequence $\{x^k\}$ converges to some solution x^* of problem (MVIP).*

Proof. Since $\beta = 1$, from Propositions 5 and 6, and from Theorem 3, we only have to prove that there exist $\mu \in (0,1)$, and $\nu > 0$ such that

$$\varepsilon_k \leq 4\gamma(1-\mu), \quad \varepsilon_k < \frac{\mu}{\nu}, \quad \frac{1}{2\nu\gamma} \leq \mu.$$

Choosing the smallest possible ν, we obtain $\nu = 1/(2\mu\gamma)$. Then the previous conditions become:

$$\varepsilon_k \leq 4\gamma(1-\mu) \quad \text{and} \quad \varepsilon_k < 2\mu^2\gamma. \tag{36}$$

It is easy to see that the maximum of the function $r(\mu) = \min\{4\gamma(1-\mu), 2\mu^2\gamma\}$ occurs at $\mu = \sqrt{3}-1$ and has $4(2-\sqrt{3})\gamma$ for optimal value. So the conditions (36) are satisfied with this value of μ if $\varepsilon_k < 4(2-\sqrt{3})\,\gamma$. $\square$

When F is singlevalued, the approximating function $\bar{f}_k$ coincides with f_k. In that case, $\mu = 1$ and Proposition 5 must not be considered. This means that only the second inequality in (36) must be retained, i.e.,

$$\varepsilon_k < 2\gamma.$$

An interesting particular case is when F is strongly monotone (with constant $\alpha > 0$) and Lipschitz continuous (with constant $L > 0$) on C. Then, for all $x, y \in C$,

$$g(x, y) \leq L^2 \|x - y\|^2.$$

Hence, F being strongly monotone, we have, for all $x, y \in C$ and $\xi \in F(x), \eta \in F(y)$, that

$$\langle \xi - \eta, x - y \rangle \geq \alpha \|x - y\|^2 \geq \frac{\alpha}{L^2} g(x, y).$$

But this means that F is co-coercive on C with constant $\gamma = \alpha / L^2$. Then Theorem 9 becomes:

Theorem 10 *Assume F is strongly monotone (with constant $\alpha > 0$) and Lipschitz continuous (with constant $L > 0$) on C. Let $\{\varepsilon_k\}$ be a nonincreasing sequence bounded away from 0. If $\varepsilon_k < 4(2 - \sqrt{3}) \frac{\alpha}{L^2}$ for all k, then the sequence $\{x^k\}$ converges to the unique solution x^* of problem (MVIP). When F is singlevalued, the same property holds but with $\varepsilon_k < \frac{2\alpha}{L^2}$ for all k.*

Let us mention that when F is singlevalued, we retrieve a classical result for variational inequalities (see, for instance, [1]). In the multivalued case, our algorithm has been studied by El Farouq in [6] but under the assumption that the series $\sum \varepsilon_k^2$ is convergent, and thus that the sequence $\{\varepsilon_k\}$ converges to 0.

Acknowledgements We are very grateful to Prof. Roberto Cominetti and the referees for their really helpful and constructive comments; in particular to one of the referees for bringing to our attention references [7] and [13].

References

1. Anh, P.N., Muu, L.D.: Coupling the Banach Contraction Mapping Principle and the Proximal Point Algorithm for Solving Monotone Variational Inequalities. Acta Math. Vietnam. **29**, 119–133 (2004)
2. Aubin, J.P., Ekeland, I.: Applied Nonlinear Analysis. John Wiley and Sons Inc., New York (1984)
3. Blum, E., Oettli, W.: From Optimization and Variational Inequalities to Equilibrium Problems. The Math. Stud. **63**, 123–145 (1994)
4. Cohen, G.: Auxiliary Problem Principle Extended to Variational Inequalities. J. Optimization Theory Appl. **59**, 325–333 (1988)

5. Correa, R., Lemaréchal, C.: Convergence of Some Algorithms for Convex Minimization. Math. Program. **62**, 261–275 (1993)
6. El Farouq, N.: Pseudomonotone Variational Inequalities: Convergence of the Auxiliary Problem Method. J. Optimization Theory Appl. **111**, 305–326 (2001)
7. Gol'shtein, E.G.: A Method for Solving Variational Inequalities Defined by Monotone Mappings. Comput. Math. Math. Phys. **42** (7), 921–930 (2002)
8. Hiriart-Urruty, J.B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms. Springer-Verlag, Berlin (1993)
9. Iusem, A., Sosa, W.: New Existence Results for Equilibrium Problem. Nonlinear Anal., Theory Methods Appl. **52**, 621–635 (2003)
10. Iusem, A., Sosa, W.: Iterative Algorithms for Equilibrium Problems. Optimization **52**, 301–316 (2003)
11. Kiwiel, K.C.: Proximal Level Bundle Methods for Convex Nondifferentiable Optimization, Saddle Point Problems and Variational Inequalities. Math. Program. **69** (1), 89–109 (1995)
12. Konnov, I.V.: Combined Relaxation Methods for Variational Inequalities. Springer-Verlag, Berlin (2001)
13. Konnov, I.V.: The Application of a Linearization Method to Solving Nonsmooth Equilibrium Problems. Russ. Math. **40** (12), 54–62 (1996)
14. Lemaréchal, C., Nemirovskii, A., Nesterov, Y.: New Variants of Bundle Methods. Math. Program. **69** (1), 111–147 (1995)
15. Lemaréchal, C., Strodiot, J.J., Bihain, A.: On a Bundle Method for Nonsmooth Optimization. In: O.L Mangasarian, R.R. Meyer, and S.M Robinson, (eds.), Nonlinear Programming 4. Academic Press, 1981, pp. 245–282
16. Mastroeni, G.: On Auxiliary Principle for Equilibrium Problems. In: P. Daniele, F. Giannessi and A. Maugeri, (eds.), Equilibrium Problems and Variational Models. Kluwer Academic Publishers, Dordrecht, 2003, pp. 289–298
17. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton, New Jersey (1970).
18. Salmon, G., Strodiot, J.J., Nguyen, V.H.: A Bundle Method for Solving Variational Inequalities. SIAM J. Optim. **14** (3), 869–893 (2004)
19. Zhu, D., Marcotte, P.: Co-coercivity and Its Role in the Convergence of Iterative Schemes for Solving Variational Inequalities. SIAM J. Optim. **6**, 714–726 (1996)