

# Convergence Properties of Feasible Descent Methods for Solving Variational Inequalities in Banach Spaces

DAO LI ZHU

*CRT, Université de Montréal, Montréal, Canada H3C 3J7*

PATRICE MARCOTTE

marcotte@iro.umontreal.ca

*DIRO, Université de Montréal, Montréal, Canada H3C 3J7*

*Received April 22, 1996; Revised January 13, 1997; Accepted February 4, 1997*

**Abstract.** This work is concerned with the analysis of convergence properties of feasible descent methods for solving monotone variational inequalities in Banach spaces.

**Keywords:** Banach spaces, variational inequalities, Frank-Wolfe algorithm, global convergence, projection

## 1. Introduction

Let  $X$  be a reflexive Banach space,  $X^*$  the topological dual space of continuous linear functionals defined on  $X$ ,  $C$  a closed, convex and nonempty subset of  $X$  and  $F$  a continuously differentiable (in the Gâteaux sense) operator from  $C$  into  $X^*$ . In this paper, we analyze the convergence behaviour of descent algorithms for finding points of  $C$  that satisfy the variational inequality:

$$\langle F(x^*), x - x^* \rangle \geq 0 \quad \forall x \in C. \quad (1)$$

We assume that the solution set of (1) is nonempty. To find a solution, we introduce a nonnegative and continuously differentiable functional  $\phi : C \rightarrow \mathbb{R}^+$  such that  $\phi(x) = 0$  if and only if  $x$  is a solution of the variational inequality. Such merit functions and the associated descent algorithms have received considerable attention in the recent literature on finite dimensional variational inequalities. A solution can then be obtained by globally minimizing  $\phi$  over  $C$ . Merit functions such as  $\phi$ , differentiable or not, are sometimes referred to as gap functions, and have been given attention in the recent literature (see the surveys by Fukushima [9] and Larsson and Patriksson [10]).

The sequence of iterates converging to a solution will be recursively defined by the relation:

$$x^{k+1} = x^k + t_k(Y(x^k) - x^k) \quad (2)$$

where  $Y(x^k)$  is a point in  $C$ ,  $t_k$  is a suitably chosen stepsize and the feasible direction  $d^k = Y(x^k) - x^k$  is a descent direction for  $\phi$  at  $x^k$ , i.e.:

$$\phi'(x^k; d^k) = \langle \phi'(x^k), d^k \rangle < 0. \quad (3)$$

We also assume that the mapping  $Y$  possesses the fixed point property:

$$Y(x^*) = x^*. \quad (4)$$

Our contribution is twofold. First we prove that, if  $\phi$  satisfies some general descent condition, the sequence of iterates defined by (2) converges to a solution of the variational inequality. Next we show that several algorithms satisfy the given descent condition. These are: a variant of the Frank-Wolfe (conditional gradient) method, Fukushima's modified projection method [8], the general descent method of Zhu and Marcotte [18], and the auxiliary principle descent method [2] [17]. The first three of these methods have so far been only proposed for finite-dimensional spaces.

The remainder of the paper will address the choice of the mapping  $Y$  and the stepsize  $t_k$ . But first, let us recall some concepts that will be used throughout the paper.

*Definition 1.* A functional  $f : C \subseteq X \rightarrow R$  is Lipschitz continuous on  $C$  if there exists a positive constant  $L_f$  such that, for all  $x, y$  in  $C$ :

$$|f(x) - f(y)| \leq L_f \|x - y\|. \quad (5)$$

Similarly, an operator  $F : C \rightarrow X^*$  is Lipschitz continuous on  $C$  if there exists a positive constant  $L_F$  such that, for all  $x, y$  in  $C$ :

$$\|F(x) - F(y)\| \leq L_F \|x - y\| \quad (6)$$

where  $\|\cdot\|$  denotes both the norm in  $X$  and its associated norm in  $X^*$ .

*Definition 2.* A functional  $f : C \subseteq X \rightarrow R$  is convex on  $C$  if there exists a nonnegative constant  $b$  such that, for all  $x, y$  in  $C$  and all  $\alpha$  in  $[0, 1]$ :

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) - \frac{1}{2}b\alpha(1 - \alpha)\|x - y\|^2. \quad (7)$$

If  $b$  is positive, we say that  $f$  is strongly convex with modulus  $b$ .

*Definition 3.* An operator  $F$  from  $C \subseteq X$  into  $X^*$  is monotone on  $C$  if there exists a nonnegative constant  $\mu$  such that, for all  $x, y$  in  $C$ :

$$\langle F(x) - F(y), x - y \rangle \geq \mu \|x - y\|^2. \quad (8)$$

If  $\mu$  is positive, we say that  $F$  is strongly monotone with modulus  $\mu$ .

If  $F$  is strongly monotone on  $C$  with modulus  $\mu$  and continuous differentiable on an open set containing  $C$ , then

$$\langle F'(x)h, h \rangle \geq \mu \|h\|^2 \quad \forall x \in C, \quad \forall h \in X. \quad (9)$$

*Definition 4.* An operator  $F$  from  $C \subseteq X$  into  $X^*$  is pseudomonotone on  $C$  if there exists a nonnegative constant  $\beta$  such that, for all  $x, y$  in  $C$ :

$$\langle F(x), y - x \rangle \geq 0 \quad \Rightarrow \quad \langle F(y), y - x \rangle \geq \beta \|y - x\|^2. \quad (10)$$

If  $\beta$  is positive, we say that  $f$  is strongly pseudomonotone with modulus  $\beta$ .

*Definition 5.* A functional  $f$  is subweakly continuous on  $C$  if for any sequence  $\{x^k\}_{k \in \mathcal{K}}$  weakly converging to  $\bar{x} \in C$ , there exists a subsequence  $\{x^{k'}\}_{k' \in \mathcal{K}}$  such that  $f(x^{k'}) \rightarrow f(\bar{x})$  for  $k' \in \mathcal{K}$ .

If  $A \subset X$ , the polar of  $A$  is defined to be the set

$$A^\circ = \{x^* \in X^* \mid \langle x^*, x \rangle \leq 0, \forall x \in A\}.$$

The normal cone to a convex set  $A$  at  $x \in A$  is defined as:

$$N_A(x) = \{x^* \in X^* \mid \langle x^*, y - x \rangle \leq 0, \forall y \in A\}$$

Finally, the tangent cone to  $A$  at  $x$  is defined by the relation

$$T_A(x) = N_A(x)^\circ.$$

We close this section with two classical lemmas.

**LEMMA 1** *Let  $f$  be continuously differentiable on an open set containing  $C$  and its derivative  $f'$  be Lipschitz continuous with Lipschitz constant  $L_{f'}$ . Then*

$$f(x) - f(y) \leq \langle f'(y), x - y \rangle + \frac{L_{f'}}{2} \|x - y\|^2. \quad (11)$$

**LEMMA 2** *Let  $f$  be continuously differentiable on an open set containing  $C$  and strongly convex with modulus  $b$  on  $C$ . Then  $f'$  is strongly monotone on  $C$  with modulus  $b$  and*

$$f(x) - f(y) \geq \langle f'(y), x - y \rangle + \frac{b}{2} \|x - y\|^2. \quad (12)$$

## 2. Convergence analysis of the generic algorithm

In this section we analyze the convergence properties of the algorithm based on the iteration (2) for minimizing the functional  $\phi$  on  $C$ , which is equivalent to finding a point  $x$  in  $C$  satisfying  $\phi(x) = 0$ . In the sequel we require that the following descent condition hold:

$$\phi'(x^k; d^k) = \phi'(x^k; Y(x^k) - x^k) \leq -\phi(x^k). \quad (13)$$

Three stepsize rules will be considered. They are:

**Optimal stepsize rule:**  $t_k \in \arg \min_{t \in [0,1]} \phi(x^k + td^k)$ .

**Constant stepsize rule:**  $t_k = t_0$  for all  $k$ .

**Open loop stepsize rule:**  $t_k \in (0, 1]$  and:

$$\sum_{k=0}^{\infty} t_k = \infty \quad \sum_{k=0}^{\infty} t_k^2 < \infty. \quad (14)$$

As mentioned in Dunn [4] a sequence satisfying the assumptions of the open loop stepsize rule can be obtained through the recursion:

$$\begin{aligned} t_0 &= 1 \\ t_{k+1} &= t_k - t_k^2/2. \end{aligned} \quad (15)$$

The following lemmas on real sequences will be used in the convergence proofs.

**LEMMA 3 (Polyak [15])** *Let  $\{\epsilon_k\}$  be a sequence of nonnegative numbers and  $\{u_k\}$  a sequence of positive numbers such that:*

$$u_{k+1} \leq u_k - \epsilon_k u_k^{1+p} \quad (16)$$

*for some positive constant  $p$ . Then we have:*

$$u_k \leq u_0 (1 + p u_0^p \sum_{i=0}^{k-1} \epsilon_i)^{-1/p}. \quad (17)$$

*In particular, if  $\epsilon_k = \epsilon$  and  $p = 1$ , we obtain:*

$$u_k \leq u_0 / (1 + \epsilon k u_0). \quad (18)$$

**LEMMA 4** *Let  $\{t_k\}$  satisfy the open loop stepsize rule, and  $\{u_k\}$  be a sequence of non-negative numbers such that*

$$u_{k+1} \leq u_k - t_k u_k + b t_k^2, \quad (19)$$

*where  $b$  is positive number. Then the entire sequence  $\{u_k\}$  converges to zero.*

**Proof:** We first show, by contradiction, that there exists a subsequence  $\{u_k\}$  converging to zero. If no such subsequence exists, then there is a positive number  $\eta$  such that

$$\sum_{k=0}^{\infty} t_k u_k \geq \sum_{k=0}^{\infty} t_k \eta = \infty. \quad (20)$$

However, for any positive integer  $N$ , one has

$$\sum_{k=0}^N t_k u_k \leq \sum_{k=0}^N (u_k - u_{k+1}) + \sum_{k=0}^N b t_k^2 = u_0 - u_{N+1} + b \sum_{k=0}^N t_k^2.$$

Since the sequence  $\{u_k\}$  is nonnegative and the series  $\sum_{k=0}^{\infty} t_k^2$  is convergent, we obtain a contradiction with (20).

Now, let condition (19) hold and assume that the entire sequence does not converge to zero, i.e., there exists a positive number  $\eta'$  and a subsequence  $\{u_k\}_{k \in \mathcal{K}}$  such that  $u_k \geq \eta'$  for all  $k$  in  $\mathcal{K}$ . Let  $\{u_k\}_{k \in \mathcal{K}'}$  be a subsequence converging to zero. Select an index  $K'$  in  $\mathcal{K}'$  such that for all  $k$  larger or equal to  $K'$  we have

$$u_k \geq \eta' \quad (k \in \mathcal{K}) \quad u_k < \frac{\eta'}{4} \quad (k \in \mathcal{K}') \quad \sum_{k=K'}^{\infty} bt_k^2 < \frac{\eta'}{2}.$$

Next we select an index  $K''$  ( $K'' > K'$ ) in  $\mathcal{K}$ . From inequality (19), we obtain

$$u_{K''} - u_{K'} \leq \sum_{k=K'}^{\infty} bt_k^2 < \frac{\eta'}{2}.$$

But  $u_{K''} - u_{K'} > \eta' - \frac{\eta'}{4} > \frac{3\eta'}{4}$  which implies the desired contradiction.  $\blacksquare$

We are now in position to prove the main convergence theorems. Let us first study the convergence of the algorithm under the optimal stepsize rule.

**THEOREM 1** *Let  $\phi$  be continuously differentiable on the bounded set  $C$  and satisfy the descent condition (13). Let  $\{x^k\}$  be the sequence generated by the iterative scheme (2) under the optimal stepsize rule.*

- (i) *If  $\phi$  is subweakly continuous on  $C$ , then the sequence  $\{\phi(x^k)\}$  converges to zero and every weak cluster point of the sequence  $\{x^k\}$  solves the variational inequality (1). In addition, if  $\phi'$  is Lipschitz continuous on  $C$ , with Lipschitz constant  $L$ , then  $\phi(x^k) = O(1/k)$ .*
- (ii) *If there exist a solution  $x^*$  and positive numbers  $p$  and  $\nu_p$  such that:*

$$\phi(x) \geq \nu_p \|x - x^*\|^p \quad \forall x \in C, \quad (21)$$

*then the entire sequence  $\{x^k\}$  converges towards the solution  $x^*$ .*

- (iii) *Let  $\phi$  satisfy the condition:*

$$\phi(x) \geq \nu \|x - x^*\|^2 \quad (22)$$

*for some positive constant  $\nu$ . If the mapping  $Y$  is Lipschitz continuous, with Lipschitz constant  $L_Y$ , then the solution  $x^*$  of the variational inequality is unique and there exists a number  $\gamma \in (0, 1)$  such that, for all  $k \geq 0$ :  $\phi(x^{k+1}) \leq \gamma \phi(x^k)$  and  $\|x^k - x^*\| = O(\gamma^{k/2})$ .*

**Proof:** (i) Under the optimal stepsize rule, one has that  $\phi(x^{k+1})$  is a nonnegative decreasing sequence of real numbers. Let  $\phi^*$  be its limiting value. We have, for all  $t \in [0, 1]$ :

$$\begin{aligned} \phi(x^{k+1}) - \phi(x^k) &\leq \phi(x^k + td^k) - \phi(x^k) \\ &= \langle \phi'(\xi), td^k \rangle \\ &= t \langle \phi'(x^k), d^k \rangle + t \langle \phi'(\xi) - \phi'(x^k), d^k \rangle \\ &\leq -t\phi(x^k) + tD \|\phi'(\xi) - \phi'(x^k)\|, \end{aligned}$$

where  $\xi = x^k + \theta d^k$  for some  $\theta$  in  $[0, t]$  and  $D$  denotes the diameter of the bounded set  $C$ . This yields, for positive values of  $t$ :

$$0 < \phi(x^k) \leq \frac{\phi(x^k) - \phi(x^{k+1})}{t} + D\|\phi'(\xi) - \phi'(x^k)\|. \quad (23)$$

Let  $\epsilon$  be a predetermined positive number. Since  $\phi'$  is uniformly continuous on the bounded set  $C$ , there exists a positive number  $\delta$  such that, whenever  $t \leq \delta$ :

$$0 < \phi(x^k) \leq \frac{\phi(x^k) - \phi(x^{k+1})}{t} + \epsilon.$$

Since  $\phi(x^k) - \phi(x^{k+1})$  converges to zero, it follows that:

$$0 \leq \liminf_{k \rightarrow \infty} \phi(x^k) \leq \limsup_{k \rightarrow \infty} \phi(x^k) \leq \epsilon$$

and:  $\lim_{k \rightarrow \infty} \phi(x^k) = 0$ . Now the sequence  $\{x^k\}$  is bounded and must therefore contain at least one cluster point in the weak topology. Let  $\bar{x}$  be such a point, and denote by  $\{x^k\}_{k \in \mathcal{K}}$  a subsequence that converges weakly to  $\bar{x}$ . Since  $C$  is convex and weakly closed, the point  $\bar{x}$  is in  $C$ . From the subweak continuity of  $\phi$ , we conclude that  $\phi(\bar{x}) = 0$ , as claimed.

Now we assume that  $\phi'$  is Lipschitz continuous with constant  $L$  on  $C$ . Let us introduce the quadratic polynomial  $\Lambda(t)$  defined as:

$$\Lambda(t) = \phi(x^k) + t\langle \phi'(x^k), d^k \rangle + \frac{t^2 L \|d^k\|^2}{2}. \quad (24)$$

We have:

$$\phi(x^{k+1}) = \min_{t \in [0, 1]} \phi(x^k + td^k) \leq \min_{t \in [0, 1]} \Lambda(t). \quad (25)$$

The minimum of the above polynomial is achieved at

$$t_k^* = \min\{1, -\langle \phi'(x^k), d^k \rangle / L \|d^k\|^2\},$$

which is positive, in view of (13). If the term  $-\langle \phi'(x^k), d^k \rangle / L \|d^k\|^2$  is less than one, we obtain:

$$\phi(x^{k+1}) \leq \Lambda(t_k^*) = \phi(x^k) - \frac{\langle \phi'(x^k), d^k \rangle^2}{2L \|d^k\|^2}. \quad (26)$$

If, on the other hand,  $-\langle \phi'(x^k), d^k \rangle / L \|d^k\|^2$  is larger or equal to one, we set  $t_k^*$  to one and obtain:

$$\phi(x^{k+1}) \leq \Lambda(1) \leq \phi(x^k) + \langle \phi'(x^k), d^k \rangle + \frac{L}{2} \|d^k\|^2 \quad (27)$$

$$\leq \phi(x^k) + \frac{\langle \phi'(x^k), d^k \rangle}{2}. \quad (28)$$

In both cases,  $\{\phi(x^k)\}$  is monotone decreasing. This implies that  $\langle \phi'(x^k), d^k \rangle$  converges to zero and, by (13), that  $\phi(x^k)$  converges to zero as well. The inequalities (26) and (28) then imply that, for  $k$  sufficiently large, there holds:

$$-\langle \phi'(x^k), d^k \rangle \leq \max\{\sqrt{2LD}(\phi(x^k) - \phi(x^{k+1}))^{\frac{1}{2}}, 2(\phi(x^k) - \phi(x^{k+1}))\} \quad (29)$$

$$\leq \sqrt{2LD}(\phi(x^k) - \phi(x^{k+1}))^{\frac{1}{2}}, \quad (30)$$

where  $D$ , as before, denotes the diameter of the set  $C$ .

Combining this inequality with (13), we obtain:

$$\phi(x^{k+1}) \leq \phi(x^k) - \frac{(\phi(x^k))^2}{2LD^2}. \quad (31)$$

Now the sequence  $\{\phi(x^k)\}$  satisfies the assumptions of Lemma 3, and we can conclude that  $\phi(x^k) = O(1/k)$ .

(ii) The proof is straightforward.

(iii) We first show that there exists an iteration index  $K$  and a positive constant  $c$  such that, for all  $k$  greater or equal to  $K$ :  $c\phi(x^k) \geq \|d^k\|^2$ . Indeed, under the stated assumptions, one has:

$$\|Y(x^k) - x^*\|^2 = \|Y(x^k) - Y(x^*)\|^2 \leq L_Y \|x^k - x^*\|^2 \leq \frac{L_Y}{\nu} \phi(x^k) \quad (32)$$

and:

$$\|d^k\|^2 = \|Y(x^k) - x^k\|^2 = \|(Y(x^k) - x^*) + (x^* - x^k)\|^2 \quad (33)$$

$$\leq 2[\|Y(x^k) - x^*\|^2 + \|x^* - x^k\|^2] \quad (34)$$

$$\leq c\phi(x^k) \quad (35)$$

with  $c = 2(L_Y + 1)/\nu$ . Next, since  $\phi'$  is Lipschitz continuous on  $C$  there must exist, in view of Lemma 1, an index  $K$  such that, for all  $k$  greater or equal to  $K$ :

$$\phi(x^{k+1}) \leq \phi(x^k) + t\langle \phi'(x^k), d^k \rangle + \frac{t^2 L \|d^k\|^2}{2} \quad (36)$$

$$\leq \phi(x^k) - t\phi(x^k) + \frac{t^2 c L \phi(x^k)}{2} \quad (37)$$

$$\leq \phi(x^k) - t\phi(x^k) + c't^2 \phi(x^k) \quad (38)$$

where  $c'$  is any positive number larger or equal to  $\max\{cL/2, 1/2\}$ . Therefore:

$$\phi(x^{k+1}) \leq \left[ \left(1 - \frac{1}{4c'}\right) + c' \left(\frac{1}{2c'} - t\right)^2 \right] \phi(x^k). \quad (39)$$

Setting  $t = 1/(2c') \leq 1$  in the above equation there comes:

$$\phi(x^{k+1}) \leq \left(1 - \frac{1}{4c'}\right) \phi(x^k) = \gamma \phi(x^k) \quad (40)$$

with  $\gamma < 1$ . Finally, under assumption (22), there clearly holds:

$$\|x^k - x^*\| \leq \frac{\gamma^{k/2}}{\nu} \phi(x^0), \quad (41)$$

i.e.,  $\|x^k - x^*\| = O(\gamma^{k/2})$ . ■

**THEOREM 2** *Let  $\phi$  satisfy the descent condition (13) and  $\phi'$  be Lipschitz continuous with Lipschitz constant  $L$  on  $C$ . Let  $\{x^k\}$  be the sequence generated by the iterative scheme (2) under the constant stepsize rule.*

(i) *If  $\phi$  is subweakly continuous on  $C$  and there exists a positive number  $\rho$  such that:*

$$\langle \phi'(x^k), d^k \rangle \leq -\rho \|d^k\|^2, \quad (42)$$

*then the sequence  $\phi(x^k) \rightarrow 0$  and every weak cluster point (at least one exists) of the sequence  $\{x^k\}$  is a solution of the variational inequality, whenever*

$$t_k = t_0 < \min\{1, 2\rho/L\}.$$

(ii) *If conditions (21) and (42) are satisfied, then the entire sequence  $\{x^k\}$  converges to the solution of the variational inequality.*

(iii) *If condition (22) is satisfied and the mapping  $Y$  is Lipschitz continuous, then there exist numbers  $\gamma$  and  $t_0$  in  $(0, 1)$  such that, for all indices  $k$ :  $\phi(x^{k+1}) \leq \gamma\phi(x^k)$  and  $\|x^k - x^*\| = O(\gamma^{k/2})$  whenever  $t_k \leq t_0$ .*

(iv) *If the mapping  $Y$  is pseudocontractive at  $x^*$ , i.e., for  $x$  sufficiently close to  $x^*$ , there exists a number  $\sigma \in (0, 1)$  such that:*

$$\|Y(x) - Y(x^*)\| \leq \sigma \|x - x^*\|, \quad (43)$$

*then the sequence  $\{x^k\}$  converges geometrically to  $x^*$ .*

**Proof:** (i) From Lemma 1, we have:

$$\phi(x^{k+1}) \leq \phi(x^k) + t_k \langle \phi'(x^k), d^k \rangle + \frac{L}{2} t_k^2 \|d^k\|^2. \quad (44)$$

Since (42) holds, we have:

$$\phi(x^{k+1}) \leq \phi(x^k) - \rho t_k \|d^k\|^2 + \frac{L}{2} t_k^2 \|d^k\|^2 \quad (45)$$

$$\leq \phi(x^k) - t_k \left( \rho - \frac{L}{2} t_k \right) \|d^k\|^2. \quad (46)$$

If  $t_k = t_0$  as defined in the statement of the theorem, we have that  $\phi(x^{k+1}) < \phi(x^k)$  and  $d^k \rightarrow 0$ . We conclude from (13) and (42) that  $\phi(x^k) \rightarrow 0$ . The remainder of the proof is similar to that of Theorem 1.

(ii) Under the additional condition (21), the convergence of the whole sequence is obvious.

(iii) Under the stated assumptions, the inequality (39) still holds, and it suffices to take

$$t_0 < 1/\max\{2L(L_Y + 1)/\nu, 1\}.$$

(iv) Under the assumption that  $Y$  is pseudocontractive, we can write:

$$\|x^{k+1} - x^*\| = \|x^k + t_k d^k - x^*\| \quad (47)$$

$$= \|x^k + t_0(Y(x^k) - x^k) - x^*\| \quad (48)$$

$$\leq (1 - t_0(1 - \sigma))\|x^k - x^*\|, \quad (49)$$

where  $1 - t_0(1 - \sigma) < 1$ . ■

**THEOREM 3** *Let  $\phi$  satisfy the descent condition (13) and  $\phi'$  be Lipschitz continuous on the bounded set  $C$ , with Lipschitz constant  $L$ . Let  $\{x^k\}$  be the sequence generated by the iterative scheme (2) under the open loop stepsize rule. Then  $\phi(x^k)$  converges to zero. Moreover (i) if  $\phi$  is subweakly continuous on  $C$ , then every weak cluster point (at least one exists) of the sequence  $\{x^k\}$  is a solution of the variational inequality (ii) if (21) is satisfied,  $\|x^k - x^*\|$  converges to zero.*

**Proof:** From Lemma 1 and (13), one obtains:

$$\phi(x^{k+1}) \leq \phi(x^k) + t_k \langle \phi'(x^k), d^k \rangle + \frac{L}{2} t_k^2 \|d^k\|^2 \quad (50)$$

$$\phi(x^{k+1}) \leq \phi(x^k) - t_k \phi(x^k) + \frac{L}{2} t_k^2 \|d^k\|^2. \quad (51)$$

Since  $\{\phi(x^k)\}$  is nonnegative and  $C$  is bounded, Lemma 4 implies that  $\phi(x^k)$  converges to zero. The remainder of the proof is similar to that of Theorem 1. ■

We close this section by stating a finite convergence result, based on the notions of sharpness and pseudomonotonicity<sup>+</sup>, and established by Marcotte and Zhu [13]. We recall that  $x^*$  is a sharp solution of the variational inequality (1) if there exists a positive constant  $\tau$  such that:

$$\langle F(x^*), x - x^* \rangle \geq \tau \|x - x^*\| \quad \forall x \in C \quad (52)$$

and that an operator  $F$  is pseudomonotone<sup>+</sup> on  $C$  if it is pseudomonotone on  $C$  and there holds the additional condition:

$$\langle F(y), y - x \rangle = 0 \quad \Rightarrow \quad F(y) = F(x). \quad (53)$$

We denote by  $p(x) = \text{Proj}_{T_C(x)}(-F(x))$  the projection of  $-F(x)$  over the tangent cone of the set  $C$  at the point  $x$ .

**THEOREM 4** (Marcotte and Zhu [13]) *Let  $F$  be single-valued, pseudomonotone<sup>+</sup>, uniformly continuous on  $C$  and the solution  $x^*$  of the variational inequality (1) be sharp and unique. Then the algorithm generates a sequence  $\{x^k\}$  converging to  $x^*$ . Moreover the sequence  $p(x^k)$  converges (strongly) to zero and there exists an index  $K$  such that, for all  $k \geq K$ ,  $x^k = x^*$ .*

### 3. Instantiation of the general convergence scheme

In this final section of the paper we consider various realizations of the general convergence framework. To simplify the exposition, we restrict our analysis to the basic algorithm incorporating the optimal stepsize rule. Note however that the analog of Theorem 3 holds for the LP-based and Fukushima methods, and that the analog of Theorem 2 holds for Fukushima's method.

#### 3.1. An LP-based algorithm for variational inequalities

Let  $F$  be monotone, continuously differentiable and weakly continuous on  $C$ . We take for  $\phi$  the gap functional:

$$\phi(x) = \max_{y \in C} \langle F(x), x - y \rangle \quad (54)$$

and make the assumption that the solution  $Y(x)$  of the above maximization problem is unique. The uniqueness of  $Y(x)$  implies that  $\phi$  is continuously differentiable on  $C$ , with derivative:

$$\phi'(x) = F(x) + F'(x)(x - Y(x)). \quad (55)$$

In this context, the assumptions of Theorem 1 hold, as shown below.

**THEOREM 5** *Let  $\phi$  be given by (54) and assume that  $Y(x)$  is single-valued. Then:*

- (i) *The descent condition (13) holds and  $\phi$  is subweakly continuous.*
- (ii) *If  $F'$  and  $Y$  are both Lipschitz continuous on  $C$ , so is  $\phi'$ .*
- (iii) *If  $F$  is strongly monotone on  $C$  then (22) holds.*
- (iv) *If  $\|x^k - Y(x^k)\|$  converges to zero, then  $p(x^k)$  converges to zero as well.*

**Proof:** (i) Since  $F$  is monotone, we have:

$$\langle \phi'(x), Y(x) - x \rangle = \langle F'(x)(x - Y(x)), Y(x) - x \rangle + \langle F(x), Y(x) - x \rangle \quad (56)$$

$$\leq \langle F(x), Y(x) - x \rangle \quad (57)$$

$$\leq -\phi(x). \quad (58)$$

For any sequence  $\{x^k\}$  converging weakly to  $\bar{x} \in C$ , we have that  $F(x^k) \rightarrow F(\bar{x})$ . Let  $\{x^{k'}\}$  be a subsequence of  $\{x^k\}$  such that  $Y(x^{k'})$  converges weakly to  $\bar{y} \in C$ . We can write:

$$\begin{aligned} & |\langle F(x^{k'}), x^{k'} - y \rangle - \langle F(\bar{x}), \bar{x} - y \rangle| \\ &= |\langle F(x^{k'}), x^{k'} - y \rangle - \langle F(\bar{x}), x^{k'} - y \rangle + \langle F(\bar{x}), x^{k'} - y \rangle - \langle F(\bar{x}), \bar{x} - y \rangle| \\ &\leq |\langle F(x^{k'}), x^{k'} - y \rangle - \langle F(\bar{x}), x^{k'} - y \rangle| + |\langle F(\bar{x}), x^{k'} - y \rangle - \langle F(\bar{x}), \bar{x} - y \rangle| \\ &\leq |\langle F(x^{k'}) - F(\bar{x}), x^{k'} - y \rangle| + |\langle F(\bar{x}), x^{k'} - \bar{x} \rangle| \\ &\leq \|F(x^{k'}) - F(\bar{x})\| \|x^{k'} - y\| + |\langle F(\bar{x}), x^{k'} - \bar{x} \rangle| \\ &\rightarrow 0 \end{aligned}$$

since  $F$  is weakly continuous and the set  $C$  bounded, and the sequence  $\{x^{k'}\}$  converges weakly to  $\bar{x}$ . Similarly:

$$\phi(x^{k'}) = \langle F(x^{k'}), x^{k'} - Y(x^{k'}) \rangle \rightarrow \langle F(\bar{x}), \bar{x} - \bar{y} \rangle. \quad (59)$$

Next we need to show that  $\bar{y} = Y(\bar{x})$ . Indeed:

$$\langle F(x^{k'}), x^{k'} - Y(x^{k'}) \rangle \geq \langle F(x^{k'}), x^{k'} - y \rangle \quad \forall y \in C \quad (60)$$

implies, from the previous limiting results, that:

$$\langle F(\bar{x}), \bar{x} - \bar{y} \rangle \geq \langle F(\bar{x}), \bar{x} - y \rangle \quad \forall y \in C, \quad (61)$$

i.e.,  $\bar{y} = Y(\bar{x})$ . We conclude that  $\phi(\bar{x}) = \langle F(\bar{x}), \bar{x} - \bar{y} \rangle$  and  $\phi(x^{k'}) \rightarrow \phi(\bar{x})$ . Hence  $\phi$  is subweakly continuous on  $C$ .

(ii) The proof is straightforward.

(iii) Let  $\beta$  denote the modulus of strong monotonicity of  $F$  and  $x^*$  the unique solution of the variational inequality. We have:

$$\phi(x) = \langle F(x), x - Y(x) \rangle \quad (62)$$

$$\geq \langle F(x), x - x^* \rangle \quad (63)$$

$$= \langle F(x) - F(x^*), x - x^* \rangle + \langle F(x^*), x - x^* \rangle \quad (64)$$

$$\geq \beta \|x - x^*\|^2 \quad (65)$$

and (22) holds with  $\nu = \beta$ .

(iv) By construction,  $Y(x^k)$  satisfies:

$$\langle F(x^k), Y(x^k) - x \rangle \leq 0 \quad \forall x \in C \quad (66)$$

or, equivalently:  $-F(x^k) \in N_C(Y(x^k))$ , where  $N_C(Y(x^k))$  denotes the normal cone to  $C$  at  $Y(x^k)$ . The Moreau decomposition yields

$$\begin{aligned} \|p(Y(x^k))\| &= \min_{-z \in F(Y(x^k)) + N_C(Y(x^k))} \|z\| \\ &\leq \|F(Y(x^k)) - F(x^k)\|. \end{aligned}$$

Since  $\|x^k - Y(x^k)\| \rightarrow 0$  and  $F$  is uniformly continuous, we conclude that  $p(x^k)$  converges to zero. ■

The above theorem, together with Theorems 1 and 4, implies the following convergence results.

**COROLLARY 1** *Let  $F$  be monotone and weakly continuous on  $C$ . Then the sequence  $\phi(x^k)$  generated by the optimal stepsize rule converges to zero and every weak cluster point of the sequence  $\{x^k\}$  (there exists at least one) is a solution of the variational inequality. Moreover if  $Y$  and  $F'$  are Lipschitz continuous on  $C$ ,  $\phi(x^k) = O(1/k)$ . If  $F$  is strongly*

monotone on  $C$ , we have that  $\phi(x^k) \leq \gamma\phi(x^k)$  and  $\|x^k - x^*\| = O(\gamma^{k/2})$  for some positive constant  $\gamma$  less than one. If  $x^*$  is a sharp solution of the variational inequality, then the convergence is finite.

**Remark.** The condition that the mapping  $Y$  be single-valued is a strong one, and not likely to hold in finite-dimensional applications, unless the set  $C$  is strongly convex. However it is satisfied naturally in the context of the multiclass variational inequality studied by Marcotte and Zhu [12]. In this latter work, an Armijo stepsize rule is proposed and its convergence properties analyzed. Numerical results on large scale multiclass traffic assignment problems are reported in Marcotte et al. [11].

### 3.2. Extension of Fukushima's method

For finite-dimensional, strongly monotone and continuously differentiable mappings, Fukushima [8] proposed the function:

$$\phi(x) = \max_{y \in C} \langle F(x), x - y \rangle - \frac{1}{2\alpha} \|x - y\|^2. \quad (67)$$

The solution  $Y(x)$  of this maximization problem is unique and given by the projection of  $x - \alpha F(x)$  onto the set  $C$ :

$$Y(x) = \text{Proj}_C(x - \alpha F(x)). \quad (68)$$

The function  $\phi$  is continuously differentiable. This result is also valid in Banach spaces. In particular, the derivative of  $\phi$  is given by

$$\phi'(x) = F(x) + F'(x)(x - Y(x)) + \frac{1}{\alpha}(Y(x) - x). \quad (69)$$

As before, we show that the assumptions of Theorem 1 hold.

**THEOREM 6** *Let  $\phi$  be defined by (67). Then  $Y(x)$  is single-valued. Also:*

- (i) *If  $F$  and  $F'$  are continuous on  $C$ , then  $\phi'(x)$  is continuous on  $C$ . If  $F$  and  $F'$  are Lipschitz continuous on  $C$ , so is  $\phi'(x)$ .*
- (ii) *Let  $F$  be strongly monotone with modulus  $\beta > 1/2\alpha$  on  $C$ ; then (13), (22) and (42) hold and  $Y$  is Lipschitz continuous on  $C$ .*
- (iii) *If the sequence  $\|x^k - Y(x^k)\|$  converges to zero and  $F$  is uniformly continuous on  $C$ , then  $p(Y(x^k))$  converges to zero as well.*

**Proof:** (i) Since projection mappings are nonexpansive, we have:

$$\|Y(y) - Y(x)\| \leq \|(y - \alpha F(y)) - (x - \alpha F(x))\|. \quad (70)$$

Hence the mapping  $Y$  is continuous (respectively Lipschitz continuous) whenever  $F$  is continuous (respectively Lipschitz continuous).

From the expression (69) for  $\phi'$  it follows that  $\phi'$  is continuous (respectively Lipschitz continuous) if  $F$  and  $F'$  are continuous (respectively Lipschitz continuous).

(ii) If  $F$  is strongly monotone, we have:

$$\begin{aligned} \langle \phi'(x), Y(x) - x \rangle &= \langle F'(x)(x - Y(x)) + F(x), Y(x) - x \rangle + \|Y(x) - x\|^2/\alpha \\ &\leq -(\beta - 1/2\alpha)\|Y(x) - x\|^2 + \langle F(x), Y(x) - x \rangle \\ &\quad + \|Y(x) - x\|^2/2\alpha \\ &= -(\beta - 1/2\alpha)\|Y(x) - x\|^2 - \phi(x) \end{aligned}$$

i.e., inequality (13) and (42) hold (if  $\beta > 1/2\alpha$ ). By definition of  $\phi$ , we then have that

$$\phi(x) \geq \langle F(x), x - x^* \rangle - \frac{1}{2\alpha}\|x - x^*\|^2 \geq (\beta - \frac{1}{2\alpha})\|x - x^*\|^2,$$

and (22) holds with  $\nu = \beta - 1/2\alpha$ . The uniqueness of  $Y$  is obvious.

(iii) The projection  $Y(x^k)$  can alternatively be defined as the unique solution of the variational inequality

$$\langle F(x^k) + \frac{1}{\alpha}(Y(x^k) - x^k), x - Y(x^k) \rangle \geq 0 \quad \forall x \in C.$$

Thus  $-(F(x^k) + \frac{Y(x^k) - x^k}{\alpha}) \in N_C(Y(x^k))$  and we obtain, using Moreau decomposition:

$$\begin{aligned} \|p(Y(x^k))\| &= \|P_{T_C(Y(x^k))}(-F(Y(x^k)))\| \\ &= \min_{-z \in F(Y(x^k)) + N_C(Y(x^k))} \|z\| \\ &\leq \|F(Y(x^k)) - F(x^k) - \frac{Y(x^k) - x^k}{\alpha}\|. \end{aligned}$$

Since  $\|x^k - Y(x^k)\|$  converges to zero and  $F$  is uniformly continuous on  $C$ , we conclude that  $p(Y(x^k))$  converges to zero. ■

**COROLLARY 2** *Let  $F$  be differentiable and strongly monotone with modulus  $\beta$  ( $\beta > 1/2\alpha$ ) on  $C$ . Then the sequence  $\{x^k\}$  generated by the optimal stepsize rule converges to the solution  $x^*$  of the variational inequality. If  $F$  and  $F'$  are Lipschitz continuous on  $C$  we have  $\phi(x^{k+1}) \leq \phi(x^k)$  and  $\|x^k - x^*\| = O(\gamma^{k/2})$  for some positive constant  $\gamma$  less than 1. If  $x^*$  is a sharp solution of the variational inequality, the convergence is finite.*

**Remark.** Convergence results for descent algorithms have been obtained under the weaker assumption that the mapping be monotone (see Zhu and Marcotte [19] and Patriksson [14]). However, the constraint  $C$  was required to be compact. It would be interesting to see if an equivalent result holds in infinite dimension.

### 3.3. A remark on optimization problems

Let us consider the optimization problem

$$\min_{x \in C} f(x) \quad (71)$$

where  $f$  is pseudoconvex and differentiable on  $C$ . It is well known that the problem (71) is equivalent to the variational inequality problem

$$\langle f'(x^*), x - x^* \rangle \geq 0 \forall x \in C. \quad (72)$$

Descent methods for solving optimization problems in infinite dimensional spaces have been addressed by several researchers, including Auslender [1], Valadier [16] and Dunn [4] [5] (extension of Frank-Wolfe and projected gradient methods), Zhu and Marcotte [17] (coupling descent methods with the auxiliary problem principle). In particular, Dunn [5] has given the convergence properties of linearization and gradient projection algorithms. It is interesting to note that these results can be recovered using the techniques of the present paper, if one defines the functional  $\phi$  as

$$\phi(x) = f(x) - f(x^*).$$

If  $f$  is convex and continuous,  $f$  is also weakly lower semicontinuous, see [6]. Hence  $\phi$  is naturally subweakly continuous in this context.

### Acknowledgments

This work was partially supported by NSERC grant A 5789 and FCAR (Québec).

### References

1. Auslender, A., *Optimisation, Méthodes numériques*, Masson: Paris, 1976.
2. Cohen, G., "Auxiliary problem principle and decomposition of optimization Problems," *Journal of Optimization Theory and Applications*, vol. 32, pp. 277–305, 1980.
3. Demyanov, V. F. and Rubinov, A. M., *Approximate Methods in Optimization problems*, American Elsevier: New York, 1970.
4. Dunn, J. C., "Conditional gradient algorithms with open loop step size rules," *Journal of Mathematical Analysis and Applications*, vol. 62, pp. 432–444, 1978.
5. Dunn, J. C., "Global and asymptotic convergence rate estimates for a class of projected gradient processes," *SIAM Journal on Control and Optimization*, vol. 19, pp. 368–400, 1981.
6. Ekeland, I. and Temam, R., "Convex analysis and variational problems," North-Holland Publishing Company: Amsterdam, Holland, 1970.
7. Frank, M. and Wolfe, P., "An algorithm for quadratic programming," *Naval Research Logistics Quarterly*, vol. 3, pp. 95–110, 1956.
8. Fukushima, M., "Equivalent differentiable optimization problems and descent method for asymmetric variational inequality problem," *Mathematical Programming*, vol. 53, pp. 161–220, 1992.
9. Fukushima, M., "Merit functions for variational inequality and complementarity problems," forthcoming in *Nonlinear Optimization and Applications*.
10. Larsson, T. and Patriksson, M., "A class of gap functions for variational inequalities," *Mathematical Programming*, vol. 64, pp. 53–79, 1994.

11. Marcotte, P., Nguyen, S. and Tanguay, K., "Implementation of an efficient algorithm for the multiclass traffic assignment problem," in: J.-B. Lesort, ed.: *Transportation and Traffic Theory. Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, Pergamon, Oxford, 1996, pp. 217 – 236.
12. Marcotte, P. and Zhu, D. L., "Equilibria with infinitely many differentiated classes of customers," *Proceedings of the 13th International Conference on Complementarity Problems*, J-S Pang and M. Ferris, eds., SIAM, Philadelphia (1997) pp. 234-288.
13. Marcotte, P. and Zhu, D. L., "Weak sharp solutions and finite convergence of algorithms for solving variational inequalities," forthcoming in *SIAM Journal on Optimization*.
14. Patriksson, M., "On the convergence of descent methods for monotone variational inequalities," *Operations Research Letters*, vol. 16, pp. 265–269, 1994.
15. Polyak, B. T., "Introduction to Optimization," Optimization Software Inc.: New York, 1987.
16. Valadier, M., "Extension d'un algorithme de Frank-Wolfe," *Revue Française de Recherche Opérationnelle*, vol. 36, pp. 251-253, 1965.
17. Zhu, D. L. and Marcotte, P., "Coupling the auxiliary problem principle with descent methods of pseudoconvex programming," *European Journal of Operational Research*, vol. 83, pp. 670-685, 1995.
18. Zhu, D. L. and Marcotte, P., "An extended descent framework for monotone variational inequalities," *Journal of Optimization Theory and Applications*, vol. 80, pp. 349-366, 1994.
19. Zhu, D. L. and Marcotte, P., "Modified descent methods for solving the monotone variational inequality problem," *Operations Research Letters*, vol. 14, pp. 111–120, 1993.