A penalization-gradient algorithm for variational inequalities

Abdellatif Moudafi, Eman Al-Shemas

Hindawi Publishing Corporation
International Journal of Mathematics and Mathematical Sciences
Volume 2011, Article ID 305856, 12 pages
doi:10.1155/2011/305856

Research Article

A Penalization-Gradient Algorithm for Variational Inequalities

Abdellatif Moudafi[1] **and Eman Al-Shemas**[2]

[1] *Département Scientifique Interfacultaires, Université des Antilles et de la Guyane, CEREGMIA, 97275 Schoelcher, Martinique, France*
[2] *Department of Mathematics, College of Basic Education, PAAET Main Campus-Shamiya, Kuwait*

Correspondence should be addressed to Abdellatif Moudafi, abdellatif.moudafi@martinique.univ-ag.fr

Received 11 February 2011; Accepted 5 April 2011

Academic Editor: Giuseppe Marino

This paper is concerned with the study of a penalization-gradient algorithm for solving variational inequalities, namely, find $\overline{x} \in C$ such that $\langle A\overline{x}, y - \overline{x}\rangle \geq 0$ for all $y \in C$, where $A : H \to H$ is a single-valued operator, C is a closed convex set of a real Hilbert space H. Given $\Psi : H \to \mathbb{R} \cup \{+\infty\}$ which acts as a penalization function with respect to the constraint $\overline{x} \in C$, and a penalization parameter β_k, we consider an algorithm which alternates a proximal step with respect to $\partial\Psi$ and a gradient step with respect to A and reads as $x_k = (I + \lambda_k\beta_k\partial\Psi)^{-1}(x_{k-1} - \lambda_k A x_{k-1})$. Under mild hypotheses, we obtain weak convergence for an inverse strongly monotone operator and strong convergence for a Lipschitz continuous and strongly monotone operator. Applications to hierarchical minimization and fixed-point problems are also given and the multivalued case is reached by replacing the multivalued operator by its Yosida approximate which is always Lipschitz continuous.

1. Introduction

Let H be a real Hilbert space, $A : H \to H$ a monotone operator, and let C be a closed convex set in H, we are interested in the study of a gradient-penalization algorithm for solving the problem of finding $\overline{x} \in C$ such that

$$\left\langle A\overline{x}, y - \overline{x}\right\rangle \geq 0 \quad \forall y \in C, \tag{1.1}$$

or equivalently

$$A\overline{x} + N_C(\overline{x}) \ni 0, \tag{1.2}$$

where N_C is the normal cone to a closed convex set C. The above problem is a variational inequality, initiated by Stampacchia [1], and this field is now a well-known branch of pure and applied mathematics, and many important problems can be cast in this framework.

In [2], Attouch et al., based on seminal work by Passty [3], solve this problem with a multivalued operator by using splitting proximal methods. A drawback is the fact that the convergence in general is only ergodic. Motivated by [2, 4] and by [5] where penalty methods for variational inequalities with single-valued monotone maps are given, we will prove that our proposed forward-backward penalization-gradient method (1.9) enjoys good asymptotic convergence properties. We will provide some applications to hierarchical fixed-point and optimization problems and also propose an idea to reach monotone variational inclusions.

To begin with, see, for instance [6], let us recall that an operator with domain $D(T)$ and range $R(T)$ is said to be monotone if

$$\langle u - v, x - y\rangle \geq 0 \quad \text{whenever } u \in T(x),\ v \in T(y). \tag{1.3}$$

It is said to be maximal monotone if, in addition, its graph, $\operatorname{gph} T := \{(x, y) \in H \times H : y \in T(x)\}$, is not properly contained in the graph of any other monotone operator. An operator sequence T_k is said to be graph convergent to T if $(\operatorname{gph}(T_k))$ converges to $\operatorname{gph}(T)$ in the Kuratowski-Painlevé's sense, that is, $\limsup_k \operatorname{gph}(T_k) \subset \operatorname{gph}(T) \subset \liminf_k \operatorname{gph}(T_k)$. It is well-known that for each $x \in H$ and $\lambda > 0$ there is a unique $z \in H$ such that $x \in (I + \lambda T)z$. The single-valued operator $J_\lambda^T := (I + \lambda T)^{-1}$ is called the resolvent of T of parameter λ. It is a nonexpansive mapping which is everywhere defined and is related to its Yosida approximate, namely $T_\lambda(x) := (x - J_\lambda^T(x))/\lambda$, by the relation $T_\lambda(x) \in T(J_\lambda^T(x))$. The latter is $1/\lambda$-Lipschitz continuous and satisfies $(T_\lambda)_\mu = T_{\lambda+\mu}$. Recall that the inverse T^{-1} of T is the operator defined by $x \in T^{-1}(y) \Leftrightarrow y \in T(x)$ and that, for all $x, y \in H$, we have the following key inequality

$$\left\| J_\lambda^T(x) - J_\lambda^T(y) \right\|^2 \leq \|x - y\|^2 + \left\| \left(I - J_\lambda^T\right)(x) - \left(I - J_\lambda^T\right)(y) \right\|^2. \tag{1.4}$$

Observe that the relation $(T_\lambda)_\mu(x) = T_{\lambda+\mu}(x)$ leads to

$$J_\mu^{T_\lambda}(x) = \frac{\lambda}{\lambda + \mu} x + \left(1 - \frac{\lambda}{\lambda + \mu}\right) J_{\lambda+\mu}^T(x). \tag{1.5}$$

Now, given a proper lower semicontinuous convex function $f : H \to \mathbb{R} \cup \{+\infty\}$, the subdifferential of f at x is the set

$$\partial f(x) = \{u \in H : f(y) \geq f(x) + \langle u, y - x\rangle\ \forall y \in H\}. \tag{1.6}$$

Its Moreau-Yosida approximate and proximal mapping f_λ and $\operatorname{prox}_{\lambda f}$ are given, respectively, by

$$f_\lambda(x) = \inf_{y \in H} \left\{ f(y) + \frac{1}{2\lambda} \|y - x\|^2 \right\}, \qquad \operatorname{prox}_{\lambda f}(x) = \operatorname*{argmin}_{y \in H} \left\{ f(y) + \frac{1}{2\lambda} \|y - x\|^2 \right\}. \tag{1.7}$$

We have the following interesting relation $(\partial f)_\lambda = \nabla f_\lambda$. Finally, given a nonempty closed convex set $C \subset H$, its indicator function is defined as $\delta_C(x) = 0$ if $x \in C$ and $+\infty$ otherwise. The projection onto C at a point u is $P_C(u) = \inf_{c \in C}\|u - c\|$. The normal cone to C at x is

$$N_C(x) = \{u \in H : \langle u, c - x\rangle \leq 0 \ \forall c \in C\} \tag{1.8}$$

if $x \in C$ and $\emptyset$ otherwise. Observe that $\partial \delta_C = N_C$, $\operatorname{prox}_{\lambda f} = J_\lambda^{\partial f}$, and $J_\lambda^{N_C} = P_C$.

Given some $x_{k-1} \in H$, the current approximation to a solution of (1.2), we study the penalization-gradient iteration which will generate, for parameters $\lambda_k > 0, \beta_k \to +\infty$, x_k as the solution of the regularized subproblem

$$\frac{1}{\lambda_k}(x_k - x_{k-1}) + Ax_{k-1} + \beta_k \partial \Psi(x_k) \ni 0, \tag{1.9}$$

which can be rewritten as

$$x_k = \left(I + \lambda_k \beta_k \partial \Psi\right)^{-1}(x_{k-1} - \lambda_k A x_{k-1}). \tag{1.10}$$

Having in view a large range of applications, we shall not assume any particular structure or regularity on the penalization function Ψ. Instead, we just suppose that Ψ is convex, lower semicontinuous and $C = \operatorname{argmin}\Psi \neq \emptyset$. We will denote by $\mathrm{VI}(A, C)$ the solution set of (1.2).

The following lemmas will be needed in our analysis, see for example [6, 7], respectively.

Lemma 1.1. *Let T be a maximal monotone operator, then $(\beta_k T)$ graph converges to $N_{T^{-1}(0)}$ as $\beta_k \to +\infty$ provided that $T^{-1}(0) \neq \emptyset$.*

Lemma 1.2. *Assume that α_k and δ_k are two sequences of nonnegative real numbers such that*

$$\alpha_{k+1} \leq \alpha_k + \delta_k. \tag{1.11}$$

If $\lim_{k\to+\infty}\delta_k = 0$, then there exists a subsequence of (α_k) which converges. Furthermore, if $\sum_{k=0}^{\infty}\delta_k < +\infty$, then $\lim_{k\to+\infty}\alpha_k$ exists.

2. Main Results

2.1. Weak Convergence

Theorem 2.1. *Assume that $\mathrm{VI}(A, C) \neq \emptyset$, A is inverse strongly monotone, namely*

$$\langle Ax - Ay, x - y\rangle \geq \frac{1}{L}\|Ax - Ay\|^2 \quad \forall x, y \in H, \ \textit{for some } L > 0. \tag{2.1}$$

If

$$\sum_{k=0}^{\infty}\left\|\overline{x}-J_{\lambda_k}^{\beta_k\partial\Psi}(\overline{x}-\lambda_k A\overline{x})\right\|<+\infty \quad \forall\overline{x}\in \mathrm{VI}(A,C), \tag{2.2}$$

and $\lambda_k \in]\varepsilon, 2/L-\varepsilon[$ *(where* $\varepsilon>0$ *is a small enough constant), then the sequence* $(x_k)_{k\in\mathbb{N}}$ *generated by algorithm* (1.9) *converges weakly to a solution of Problem* (1.2)*.*

Proof. Let $\overline{x}$ be a solution of (1.2), observe that $\overline{x}$ solves (1.2) if and only if $\overline{x}=(I+\lambda_k N_C)^{-1}(\overline{x}-\lambda_k A\overline{x})=P_C(\overline{x}-\lambda_k A\overline{x})$. Set $\overline{x}_k=(I+\lambda_k\beta_k\partial\Psi)^{-1}(\overline{x}-\lambda_k A\overline{x})$, by the triangular inequality, we can write

$$\|x_k-\overline{x}\|\le\|x_k-\overline{x}_k\|+\|\overline{x}_k-\overline{x}\|. \tag{2.3}$$

On the other hand, by virtue of (1.4) and (2.1), we successively have

$$\begin{aligned}\|x_k-\overline{x}_k\|^2 &\le \|x_{k-1}-\overline{x}-\lambda_k(Ax_{k-1}-A\overline{x})\|^2-\|x_{k-1}-x_k-\lambda_k(Ax_{k-1}-A\overline{x})+\overline{x}_k-\overline{x}\|^2\\ &\le \|x_{k-1}-\overline{x}\|^2-\lambda_k\left(\frac{2}{L}-\lambda_k\right)\|Ax_{k-1}-A\overline{x}\|^2\\ &\quad -\|x_{k-1}-x_k-\lambda_k(Ax_{k-1}-A\overline{x})+\overline{x}_k-\overline{x}\|^2.\end{aligned} \tag{2.4}$$

Hence

$$\begin{aligned}\|x_k-\overline{x}\| &< \sqrt{\|x_{k-1}-\overline{x}\|^2-\varepsilon^2\|Ax_{k-1}-A\overline{x}\|^2-\|x_{k-1}-x_k-\lambda_k(Ax_{k-1}-A\overline{x})+\overline{x}_k-\overline{x}\|^2}\\ &\quad+\|\overline{x}-\overline{x}_k\|.\end{aligned} \tag{2.5}$$

The later implies, by Lemma 1.2 and the fact that (2.2) insures $\lim_{k\to+\infty}\|\overline{x}-\overline{x}_k\|=0$, that the positive real sequence $(\|x_k-\overline{x}\|^2)_{k\in\mathbb{N}}$ converges to some limit $l(\overline{x})$, that is,

$$l(\overline{x})=\lim_{k\to+\infty}\|x_k-\overline{x}\|^2<+\infty, \tag{2.6}$$

and also assures that

$$\begin{gathered}\lim_{k\to+\infty}\|Ax_{k-1}-A\overline{x}\|^2=0,\\ \lim_{k\to+\infty}\|x_{k-1}-x_k-\lambda_k(Ax_{k-1}-A\overline{x})+\overline{x}_k-\overline{x}\|^2=0.\end{gathered} \tag{2.7}$$

Combining the two latter equalities, we infer that

$$\lim_{k\to+\infty}\|x_{k-1}-x_k\|^2=0. \tag{2.8}$$

Now, (1.9) can be written equivalently as

$$\frac{x_{k-1} - x_k}{\lambda_k} + Ax_k - Ax_{k-1} \in \left(A + \beta_k \partial\Psi\right)(x_k). \tag{2.9}$$

By virtue of Lemma 1.1, we have $(\beta_k \partial\Psi)$ graph converges to $N_{\operatorname{argmin}\Psi}$ because

$$(\partial\Psi)^{-1}(0) = \partial\Psi^*(0) = \operatorname{argmin}\Psi. \tag{2.10}$$

Furthermore, the Lipschitz continuity of A (see, e.g., [8]) clearly ensures that the sequence $(A + \beta_k \partial\Psi)$graph converges in turn to $A + N_{\operatorname{argmin}\Psi}$.

Now, let x^* be a cluster point of $\{x_k\}$. Passing to the limit in (2.9), on a subsequence still denoted by $\{x_k\}$, and taking into account the fact that the graph of a maximal monotone operator is weakly strongly closed in $H \times H$, we then conclude that

$$0 \in (A + N_C)x^*, \tag{2.11}$$

because A is Lipschitz continuous, (x_k) is asymptotically regular thanks to (2.8), and (λ_k) is bounded away from zero.

It remains to prove that there is no more than one cluster point, our argument is classical and is presented here for completeness.

Let $\tilde{x}$ be another cluster of $\{x_k\}$, we will show that $\tilde{x} = x^*$. This is a consequence of (2.6). Indeed,

$$l(x^*) = \lim_{k \to +\infty} \|x_k - x^*\|^2, \qquad l(\tilde{x}) = \lim_{k \to +\infty} \|x_k - \tilde{x}\|^2, \tag{2.12}$$

from

$$\|x_k - \tilde{x}\|^2 = \|x_k - x^*\|^2 + \|x^* - \tilde{x}\|^2 + 2\langle x_k - x^*, x^* - \tilde{x}\rangle, \tag{2.13}$$

we see that the limit of $\langle x_k - x^*, x^* - \tilde{x}\rangle$ as $k \to +\infty$ must exists. This limit has to be zero because x^* is a cluster point of $\{x_k\}$. Hence at the limit, we obtain

$$l(\tilde{x}) = l(x^*) + \|x^* - \tilde{x}\|^2. \tag{2.14}$$

Reversing the role of $\tilde{x}$ and x^*, we also have

$$l(x^*) = l(\tilde{x}) + \|x^* - \tilde{x}\|^2. \tag{2.15}$$

That is $\tilde{x} = x^*$, which completes the proof. □

Remark 2.2. (i) Note that, we can remove condition (2.2), but in this case we obtain that there exists a subsequence of (x_k) such that every weak cluster point is a solution of problem (1.2). This follows by Lemma 1.2 combined with the fact that $\overline{x} = J_{\lambda^*}^{\partial\delta_C}(\overline{x} - \lambda^* A\overline{x})$ and that

$(\beta_k\partial\Psi)$ graph converges to $\partial\delta_C$. The later is equivalent, see for example [6], to the pointwise convergence of $J_{\lambda_k}^{\beta_k\partial\Psi}$ to $J_{\lambda^*}^{\partial\delta_C}$ and therefore ensures that

$$\lim_{k\to+\infty}\left\|\overline{x}-J_{\lambda_k}^{\beta_k\partial\Psi}(\overline{x}-\lambda_k A\overline{x})\right\|=0. \tag{2.16}$$

(ii) In the special case $\Psi(x) = (1/2)\operatorname{dist}(x,C)^2$, (2.2) reduces to $\sum_{k=0}^{\infty}1/\beta_k < +\infty$, see Application (2) of Section 3.

Suppose now that $\Psi(x) = \operatorname{dist}(x,C)$, it well-known that $\operatorname{prox}_{\gamma\Psi}(x) = P_C(x)$ if $\operatorname{dist}(x,C)\le\gamma$. Consequently,

$$J_{\lambda_k}^{\beta_k\partial\Psi}(x)=P_C(x)\quad\text{if }\operatorname{dist}(x,C)\le\lambda_k\beta_k, \tag{2.17}$$

which is the case for all $k\ge\kappa$ for some $\kappa\in\mathbb{N}$ because (λ_k) is bounded and $\lim_{k\to+\infty}\beta_k=+\infty$. Hence $\lim_{k\to+\infty}\|\overline{x}-J_{\lambda_k}^{\beta_k\partial\Psi}(\overline{x}-\lambda_k A\overline{x})\|=0$, for all $k\ge\kappa$, and thus (2.2) is clearly satisfied.

The particular case $\Psi=0$ corresponds to the unconstrained case, namely, $C=H$. In this context the resolvent associated to $\beta_k\partial\Psi$ is the identity, and condition (2.2) is trivially satisfied.

2.2. Strong Convergence

Now, we would like to stress that we can guarantee strong convergence by reinforcing assumptions on A.

Proposition 2.3. *Assume that A is strong monotone with constant $\alpha>0$, that is,*

$$\langle Ax-Ay,x-y\rangle\ge\alpha\|x-y\|^2\quad\forall x,y\in H,\ \textit{for some }\alpha>0, \tag{2.18}$$

and Lipschitz continuous with constant $L>0$, that is,

$$\|Ax-Ay\|\le L\|x-y\|\quad\forall x,y\in H,\ \textit{for some }L>0. \tag{2.19}$$

If $\lambda_k\in\,]\varepsilon,2\alpha/L^2-\varepsilon[$ (where $\varepsilon>0$ is a small enough constant) and $\lim_{k\to+\infty}\lambda_k=\lambda^*>0$, then the sequence generated by (1.9) *strongly converges to the unique solution of* (1.2).

Proof. Indeed, by replacing inverse strong monotonicity of A by strong monotonicity and Lipschitz continuity, it is easy to see from the first part of the proof of Theorem 2.1 that the operator of $I-\lambda_k A$ satisfies

$$\|(I-\lambda_k A)(x)-(I-\lambda_k A)(y)\|^2\le\left(1-2\lambda_k\alpha+\lambda_k^2L^2\right)\|x-y\|^2. \tag{2.20}$$

Following the arguments in the proof of Theorem 2.1 to obtain

$$\|x_k - \overline{x}\| \leq \sqrt{1 - 2\lambda_k \alpha + \lambda_k^2 L^2}\|x_{k-1} - \overline{x}\| + \delta_k(\overline{x}) \quad \text{with } \delta_k(\overline{x}) := \left\|\overline{x} - J_{\lambda_k}^{\beta_k \partial\Psi}(\overline{x} - \lambda_k A\overline{x})\right\|. \tag{2.21}$$

Now, by setting $\Theta(\lambda) = \sqrt{1 - 2\lambda\alpha + \lambda^2 L^2}$, we can check that $0 < \Theta(\lambda) < 1$ if and only if $\lambda_k \in]0, 2\alpha/L^2[$, and a simple computation shows that $0 < \Theta(\lambda_k) \leq \Theta^* < 1$ with $\Theta^* = \max\{\Theta(\varepsilon), \Theta(2\alpha/L^2 - \varepsilon)\}$. Hence,

$$\|x_k - \overline{x}\| \leq (\Theta^*)^k \|x_0 - \overline{x}\| + \sum_{j=0}^{k-1} (\Theta^*)^j \delta_{k-j}(\overline{x}). \tag{2.22}$$

The result follows from Ortega and Rheinboldt [9, page 338] and the fact that $\lim_{k\to+\infty}\delta_k(\overline{x}) = 0$. The later follows thanks to the equivalence between graph convergence of the sequence of operators $(\beta_k \partial\Psi)$ to $\partial\delta_C$ and the pointwise convergence of their resolvent operators combined with the fact that $\lim_{k\to+\infty}\lambda_k = \lambda^*$. $\square$

3. Applications

(1) Hierarchical Convex Minimization Problems

Having in mind the connection between monotone operators and convex functions, we may consider the special case $A = \nabla\Phi$, Φ being a proper lower semicontinuous differentiable convex function. Differentiability of Φ ensures that $\nabla\Phi + N_{\operatorname{argmin}\Psi} = \partial(\Phi + \delta_{\operatorname{argmin}\Psi})$ and (1.2) reads as

$$\min_{x \in \operatorname{argmin}\Psi} \Phi(x). \tag{3.1}$$

Using definition of the Moreau-Yosida approximate, algorithm (1.9) reads as

$$x_k = \operatorname*{argmin}_{y \in H} \left\{ f(y) + \frac{1}{2\lambda_k} \left\| y - (I - \lambda_k A)x_{k-1} \right\|^2 \right\}. \tag{3.2}$$

In this case, it is well-known that the assumption (2.1) of inverse strong monotonicity of $\nabla\Phi$ is equivalent to its L-Lipschitz continuity. If further we assume $\sum_{k=1}^{\infty} \delta_k(\overline{x}) < +\infty$ for all $\overline{x} \in \mathrm{VI}(\nabla\Phi, C)$ and $\lambda_k \in]\varepsilon, 2/L - \varepsilon[$, then by Theorem 2.1 we obtain weak convergence of algorithm (3.2) to a solution of (3.1). The strong convergence is obtained, thanks to Proposition 2.3, if in addition Ψ is strongly convex (i.e., there is $\alpha > 0$;

$$(1-\mu)\Psi(x_1) + \mu\Psi(x_2) \geq \Psi((1-\mu)x_1 + \mu x_2) + \frac{\alpha}{2}\mu(1-\mu)\|x_1 - x_2\|^2 \tag{3.3}$$

for all $\mu \in [0,1]$, all $x_1, x_2 \in H$) and (λ_k) a convergent sequence with $\lambda_k \in]\varepsilon, 2\alpha/L^2 - \varepsilon[$. Note that strong convexity of Ψ is equivalent to α-strong monotonicity of its gradient. A concrete example in signal recovery is the Projected Land weber problem, namely,

$$\min_{x \in C} \Phi(x) := \frac{1}{2}\|Lx - z\|^2, \tag{3.4}$$

L being a linear-bounded operator. Set $A(x) := \nabla\Phi(x) = L^*(Lx - z)$. Consequently,

$$\forall x, y \in H \quad \|A(x) - A(y)\| = \|L^* L(x-y)\| \leq \|L\|^2 \|x - y\|, \tag{3.5}$$

and A is therefore Lipschitz continuous with constant $\|L\|^2$. Now, it is well-known that the problem possesses exactly one solution if L is bounded below, that is,

$$\exists \kappa > 0 \quad \forall x \in H \quad \|L(x)\| \geq \kappa\|x\|. \tag{3.6}$$

In this case, A is strongly monotone. Indeed, it is easily seen that f is strongly convex: consider $x, y \in H$ and $\mu \in]0,1[$, one has

$$\frac{\|\mu(Lx - z) + (1-\mu)(Ly - z)\|^2}{2} \leq \frac{\mu\|Lx - z\|^2}{2} + \frac{(1-\mu)\|Ly - z\|^2}{2} - \frac{\kappa^2\mu(1-\mu)\|x - y\|^2}{2}. \tag{3.7}$$

(2) Classical Penalization

In the special case where $\Psi(x) = (1/2)\operatorname{dist}(x, C)^2$, we have

$$\partial\Psi(x) = x - \operatorname{Proj}_C(x), \tag{3.8}$$

which is nothing but the classical penalization operator, see [10]. In this context, taking into account the fact that

$$((\partial f)_\lambda)_\mu = \nabla f_{\lambda+\mu}, \quad J_\lambda^{\partial f} = I - \lambda(\partial f)_\lambda = I - \lambda\nabla f_\lambda, \quad (\delta_C)_\lambda = \frac{1}{\lambda}\Psi, \tag{3.9}$$

and that $\overline{x}$ solves (1.2), and thus $\overline{x} = P_C(\overline{x} - \lambda_k A\overline{x})$, we successively have

$$\begin{aligned}
\|\overline{x}_k - \overline{x}\| &= \left\| J_{\lambda_k}^{\beta_k \partial \Psi}(\overline{x} - \lambda_k A\overline{x}) - J_{\lambda_k}^{N_C}(\overline{x} - \lambda_k A\overline{x}) \right\| \\
&= \lambda_k \left\| (\beta_k \partial \Psi)_{\lambda_k}(\overline{x} - \lambda_k A\overline{x}) - (N_C)_{\lambda_k}(\overline{x} - \lambda_k A\overline{x}) \right\| \\
&= \lambda_k \left\| \beta_k (\partial \Psi)_{\lambda_k \beta_k}(\overline{x} - \lambda_k A\overline{x}) - \nabla(\delta_C)_{\lambda_k}(\overline{x} - \lambda_k A\overline{x}) \right\| \\
&= \lambda_k \left\| \beta_k (\partial (\delta_C)_1)_{\lambda_k \beta_k}(\overline{x} - \lambda_k A\overline{x}) - \nabla(\delta_C)_{\lambda_k}(\overline{x} - \lambda_k A\overline{x}) \right\| \\
&= \lambda_k \left\| \beta_k \nabla(\delta_C)_{1+\lambda_k \beta_k}(\overline{x} - \lambda_k A\overline{x}) - \nabla(\delta_C)_{\lambda_k}(\overline{x} - \lambda_k A\overline{x}) \right\| \\
&= \lambda_k \left(\frac{1}{\lambda_k} - \frac{\beta_k}{1 + \lambda_k \beta_k} \right) \|(\overline{x} - \lambda_k A\overline{x}) - P_C(\overline{x} - \lambda_k A\overline{x})\| \\
&= \frac{1}{1 + \lambda_k \beta_k} \|\lambda_k A\overline{x}\| \le \frac{1}{\beta_k} \|A\overline{x}\|.
\end{aligned} \tag{3.10}$$

So condition on the parameters reduces to $\sum_{k=1}^{\infty} 1/\beta_k < +\infty$, and algorithm (1.9) is nothing but a relaxed projection-gradient method. Indeed, using (1.5) and the fact that $J_{\lambda}^{N_C} = P_C$, we obtain

$$x_k = \left(\frac{1}{1 + \lambda_k \beta_k} I + \frac{\lambda_k \beta_k}{1 + \lambda_k \beta_k} P_C \right)(I - \lambda_k A)x_{k-1}. \tag{3.11}$$

An inspection of the proof of Theorem 2.1 shows that the weak converges is assured with $\lambda_k \in]\varepsilon, 2/L - \varepsilon[$.

(3) A Hierarchical Fixed-Point Problem

Having in mind the connection between inverse strongly monotone operators and nonexpansive mappings, we may consider the following fixed-point problem:

$$(I - P)x + N_C(x) \ni 0, \tag{3.12}$$

with P a nonexpansive mapping, namely, $\|Px - Py\| \le \|x - y\|$.

It is well-known that $A = I - P$ is inverse strongly monotone with $L = 2$. Indeed, by definition of P, we have

$$\left\| (I - A)x - (I - A)y \right\| \le \left\| x - y \right\|. \tag{3.13}$$

On the other hand

$$\left\| (I - A)x - (I - A)y \right\|^2 = \left\| x - y \right\|^2 + \left\| Ax - Ay \right\|^2 - 2\langle x - y, Ax - Ay \rangle. \tag{3.14}$$

Combining the two last inequalities, we obtain

$$\langle x-y, Ax-Ay\rangle \geq \frac{1}{2}\|Ax-Ay\|^2. \tag{3.15}$$

Therefore, by Theorem 2.1 we get the weak convergence of the sequence (x_k) generated by the following algorithm:

$$x_k = \mathrm{prox}_{\beta_k\Psi}((I-\lambda_k)x_{k-1}+\lambda_k Px_{k-1}) \tag{3.16}$$

to a solution of (3.12) provided that $\sum_{k=1}^{\infty}\delta_k(\overline{x}) < +\infty$ for all $\overline{x}\in \mathrm{VI}(I-P,C)$ and $\lambda_k \in]\varepsilon, 1-\varepsilon[$. The strong convergence of (1.9) is obtained, by applying Proposition 2.3, for P a contraction mapping, namely, $\|Px-Py\| \leq \gamma\|x-y\|$ for $0<\gamma<1$ which is equivalent to the $(1-\gamma)$-strong monotonicity of $(I-P)$, and (λ_k) is a convergent sequence with $\lambda_k \in]\varepsilon, 2(1-\gamma)/(1+\gamma)^2-\varepsilon[$. It is easily seen that in this case $I-P$ is $(1+\gamma)$-Lipschitz continuous.

4. Towards the Multivalued Case

Now, we are interested in (1.2) when $A : H \to 2^H$ is a multi-valued maximal monotone operator. With the help of the Yosida approximate which is always inverse strongly monotone (and thus single-valued), we consider the following partial regularized version of (1.2):

$$A_\gamma x_\gamma^* + N_C\left(x_\gamma^*\right) \ni 0, \tag{4.1}$$

where A_γ stands for the Yosida approximate of A.

It is well-known that A_γ is inverse strongly monotone. More precisely, we have

$$\langle A_\gamma x - A_\gamma y, x-y\rangle \geq \gamma\|A_\gamma x - A_\gamma y\|^2. \tag{4.2}$$

Using definition of the Yosida approximate, algorithm (1.9) applied to (4.1) reads as

$$x_k^\gamma = (I+\lambda_k\beta_k\partial\Psi)^{-1}\left(\left(1-\frac{\lambda_k}{\gamma}\right)x_{k-1}^\gamma + \frac{\lambda_k}{\gamma}J_\gamma^A\left(x_{k-1}^\gamma\right)\right). \tag{4.3}$$

From Theorem 2.1, we infer that x_k^γ converges weakly to a solution $\overline{x}^\gamma$ provided that $\lambda_k \in]\varepsilon, 2\gamma-\varepsilon[$. Furthermore, it is worth mentioning that if A is strongly monotone, A_γ is also strongly monotone, and thus (4.1) has a unique solution $\overline{x}^\gamma$. By a result in [8, page 35], we have the following estimate:

$$\|\overline{x}-\overline{x}^\gamma\| \leq o\left(\sqrt{\gamma}\right). \tag{4.4}$$

Consequently, (4.3) provides approximate solutions to the variational inclusion (1.2) for small values of γ. Furthermore, when $A=\nabla\Phi$, we have

$$(\partial\Phi)_\gamma(\overline{x}) + N_C(\overline{x}) = \nabla\Phi_\gamma(\overline{x}) + N_C(\overline{x}) = \partial\left(\Phi_\gamma+\delta_C\right)(\overline{x}), \tag{4.5}$$

and thus (4.1) reduces to

$$\min_{x\in C}\Phi_\gamma(x). \tag{4.6}$$

If (3.1) and (4.1) are solvable, by ([11] Theorem 3.3), we have for all $\gamma > 0$

$$0 \le \min_{x\in C}\Phi(x) - \min_{x\in C}\Phi_\gamma(x) \le \gamma\|\overline{y}\|^2, \tag{4.7}$$

where $\overline{y} = \nabla\Phi(\overline{y})(\in -N_C(\overline{x}))$ with $\overline{x}$ a solution of (3.1). The value of (3.1) is thus close to those of (4.1) for small values of γ, and hence, this confirmed the pertinence of the proposed approximation idea to reach the multi-valued case. Observe that in this context, algorithm (4.3) reads as

$$x_k^\gamma = \operatorname{prox}_{\beta_k\Psi}\left(\left(1-\frac{\lambda_k}{\gamma}\right)x_{k-1}^\gamma + \frac{\lambda_k}{\gamma}\operatorname{prox}_{\gamma\Phi}\left(x_{k-1}^\gamma\right)\right). \tag{4.8}$$

5. Conclusion

The authors have introduced a forward-backward penalization-gradient algorithm for solving variational inequalities and studied their asymptotic convergence properties. We have provided some applications to hierarchical fixed-point and optimization problems and also proposed an idea to reach monotone variational inclusions.

Acknowledgment

We gratefully acknowledge the constructive comments of the anonymous referees which helped them to improve the first version of this paper.

References

[1] G. Stampacchia, "Formes bilinéaires coercitives sur les ensembles convexes," *Comptes Rendus de l'Académie des Sciences de Paris*, vol. 258, pp. 4413–4416, 1964.

[2] H. Attouch, M. O. Czarnecki, and J. Peypouquet, "Prox-penalization and splitting methods for constrained variational problems," *SIAM Journal on Optimization*, vol. 21, no. 1, pp. 149–173, 2011.

[3] G. B. Passty, "Ergodic convergence to a zero of the sum of monotone operators in Hilbert space," *Journal of Mathematical Analysis and Applications*, vol. 72, no. 2, pp. 383–390, 1979.

[4] B. Lemaire, "Coupling optimization methods and variational convergence," in *Trends in Mathematical Optimization*, vol. 84 of *International Series of Numerical Mathematics*, pp. 163–179, Birkhäuser, Basel, Switzerland, 1988.

[5] J. Gwinner, "On the penalty method for constrained variational inequalities," in *Optimization: Theory and Algorithms*, vol. 86 of *Lecture Notes in Pure and Applied Mathematics*, pp. 197–211, Dekker, New York, NY, USA, 1983.

[6] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, vol. 317, Springer, Berlin, Germany, 1998.

[7] P. L. Martinet, *Algorithmes pour la résolution de problèmes d'optimisation et de minimax*, thesis, Université de Grenoble, 1972.

[8] H. Brézis, *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*, North-Holland Mathematics Studies, no. 5, North-Holland, Amsterdam, The Netherlands, 1973.

[9] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, NY, USA, 1970.

[10] D. Pascali and S. Sburlan, *Nonlinear Mappings of Monotone Type*, Martinus Nijhoff, The Hague, The Netherlands, 1978.

[11] N. Lehdili, *Méthodes proximales de sélection et de décomposition pour les inclusions monotones*, thesis, Université de Montpellier, 1996.