International Journal of Pure and Applied Mathematics
Volume 5 No. 3 2003, 283-299

AN APPROXIMATE INERTIAL PROXIMAL METHOD USING THE ENLARGEMENT OF A MAXIMAL MONOTONE OPERATOR

A. Moudafi[1 §], E. Elisabeth

Scientific Department
University of the French West Indies and Guiana
B.P. 7209, 97275 Schoelcher Cedex, Martinique, FRANCE
and
Université des Antilles et de la Guyane
Département Scientifique Interfacultaires
B.P. 7209, 97275 Schoelcher Cedex, Martinique, F.W.I.
[1]e-mail: abdellatif.moudafi@martinique.univ-ag.fr

Abstract: An approximate procedure for solving the problem of finding a zero of a maximal monotone operator is proposed and its convergence is established under various conditions. More precisely, it is shown that this method weakly converges under natural assumptions and strongly converges provided that either the inverse of the involved operator is Lipschitz continuous around zero or the interior of the solution set is nonempty. A particular attention is given to the convex minimization case.

AMS Subject Classification: 90C25; 49M45, 65C25
Key Words: monotone operators, elargements, proximal point algorithm, local Lipschitz continuity, approximate subdifferential, convergence, convex minimization

Received: January 31, 2003

§Correspondence author

1. Introduction and Preliminaries

In this paper we will focus our attention on the classical problem of finding a zero of a maximal monotone operator T on a real Hilbert space $\mathcal{H}$:

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad Tx \ni 0. \tag{1.1}$$

This is a well-known problem which includes, as special cases, optimization and min-max problems, complementarity problems, and variational inequalities.

One of the fondamental approaches to solving (1.1) is the proximal method, which generates the next iterates x_{k+1} by solving the subproblem

$$0 \in \lambda_k T(x) + (x - x_k), \tag{1.2}$$

where x_k is the current iterate and λ_k is a regularization parameter. The literature on this subject is vast (see [8] for a survey).

Recently, an inertial proximal algorithm was proposed by Alvarez in [1] in the context of convex minimization. Afterwards, Attouch and Alvarez considered its extension to maximal monotone operators in [2]. It works as follows. Given $x_{k-1}, x_k \in \mathcal{H}$ and two parameters $\alpha_k \in [0, 1[$ and $\lambda_k > 0$, find $x_{k+1} \in \mathcal{H}$ such that

$$\lambda_k T(x_{k+1}) + x_{k+1} - x_k - \alpha_k(x_k - x_{k-1}) \ni 0. \tag{1.3}$$

It is well known that the proximal iteration may be interpreted as an implicit one-step discretisation method for the evolution differential inclusion

$$\frac{dx}{dt}(t) + T(x(t)) \ni 0 \quad \text{a.e.} \quad t \geq 0, \tag{1.4}$$

where the parameter λ_k is a (variable) stepsize. While the inspiration for (1.3) comes from the implicit discretization of the differential system of the second-order in time, namely

$$\frac{d^2x}{dt^2}(t) + \gamma\frac{dx}{dt}(t) + T(x(t)) \ni 0 \quad \text{a.e.} \quad t \geq 0, \tag{1.5}$$

where $\gamma > 0$ is a damping or a friction parameter.

Under appropriate conditions on α_k and λ_k Attouch and Alvarez proved that if the solution set $S = T^{-1}(0)$ is nonempty, then for every sequence $\{x_k\}$ generated by (1.3), there exists an $\bar{x} \in S$ such that $\{x_k\}$ converges to $\bar{x}$ weakly in $\mathcal{H}$ as $k \to \infty$.

For developing implementable computational techniques, it is of particular importance to treat the case when (1.3) is solved approximately. Before introducing our approximate method, let us recall the following concepts which are of common use in the context of convex and nonlinear analysis. Troughout, $\mathcal{H}$ is a real Hilbert space, $\langle \cdot, \cdot \rangle$ denotes the associated scalar product and $|\cdot|$ stands for the corresponding norm. An operator is said to be monotone if

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

It is said to be maximal monotone if, in addition, the graph, $\{(x, y) \in \mathcal{H} \times \mathcal{H} : y \in T(x)\}$, is not properly contained in the graph of any other monotone operator. It is well-known that for each $x \in \mathcal{H}$ and $\lambda > 0$ there is a unique $z \in \mathcal{H}$ 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 satisfies: $z = J_\lambda^T z$, if and only if, $0 \in Tz$. Let us also recall a notion which is clearly inspired by the approximate subdifferential. In [16], Iusem, Burachik and Svaiter defined $T^\varepsilon(x)$, an ε-enlargement of a monotone operator T, as

$$T^\varepsilon(x) := \{v \in \mathcal{H}; \langle u - v, y - x \rangle \geq -\varepsilon \quad \forall y, u \in T(y)\}, \tag{1.6}$$

where $\varepsilon \geq 0$. Since T is assumed to be maximal monotone, $T^0(x) = T(x)$, for any x. Furthermore, directly from the definition it follows that

$$0 \leq \varepsilon_1 \leq \varepsilon_2 \Rightarrow T^{\varepsilon_1}(x) \subset T^{\varepsilon_2}(x).$$

Thus T^ε is an enlargement of T. The use of elements in T^ε instead of T allows an extra degree of freedom, which is very useful in various applications. On the other hand, setting $\varepsilon = 0$ one retrieves the original operator T, so that the classical method can be also treated. For all these reasons, we consider the following scheme: find $x_{k+1} \in \mathcal{H}$ such that

$$\lambda_k T^{\varepsilon_k}(x_{k+1}) + x_{k+1} - y_k \ni 0, \tag{1.7}$$

where $y_k := x_k + \alpha_k(x_k - x_{k-1}), \lambda_k, \alpha_k, \varepsilon_k$ are nonnegative real numbers.

We will impose the following tolerance criteria on the term ε_k which are standard in the literature:

$$\sum_{k=1}^{+\infty} \lambda_k \varepsilon_k < +\infty, \tag{1.8}$$

and

$$\sqrt{\lambda_k \varepsilon_k} \leq \delta_k \|x_{k+1} - y_k\| \quad \text{with} \quad \sum_{k=1}^{+\infty} \delta_k < +\infty. \tag{1.9}$$

The first condition is typically needed to establish global convergence, while the second is required for local linear rate of convergence result under additional natural assumptions.

The remainder of the paper is organized as follows: In section 2, we present a weak convergence result for the sequence generated by (1.7) under criterion (1.8). We also consider various conditions for which the convergence is strong. In section 3, we present an application to convex minimization and study the convergence of a perturbed version of (1.7) as well as conditions ensuring the convergence for both the values and the iterates.

2. The Main Results

2.1. A Weak Convergence Result

Theorem 2.1. *Let $\{x_k\} \subset \mathcal{H}$ be a sequence such that*

$$0 \in x_{k+1} - x_k - \alpha_k(x_k - x_{k-1}) + \lambda_k T^{\varepsilon_k}(x_{k+1}), k = 1, 2, \cdots,$$

where $T : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ is a maximal monotone operator with $S := T^{-1}(0) \neq \emptyset$, and the parameters α_k, λ_k and ε_k satisfy:

1. $\exists \lambda > 0$ *such that* $\forall k \in I\!N^*, \lambda_k \geq \lambda$.
2. $\exists \alpha \in [0, 1[$ *such that* $\forall k \in I\!N^*, 0 \leq \alpha_k \leq \alpha$.
3. $\sum_{k=1}^{+\infty} \lambda_k \varepsilon_k < +\infty$.

If the following condition holds

$$\sum_{k=1}^{+\infty} \alpha_k |x_k - x_{k-1}|^2 < +\infty, \tag{2.10}$$

then, there exists $\bar{x} \in S$ such that $\{x_k\}$ weakly converges to $\bar{x}$ as $k \to +\infty$.

Proof. Fix $x \in S = T^{-1}(0)$, since $0 \in x_{k+1} - x_k - \alpha_k(x_k - x_{k-1}) + \lambda_k T^{\varepsilon_k}(x_{k+1})$, from definition (1.6) it follows that

$$\langle x_{k+1} - x_k - \alpha_k(x_k - x_{k-1}), x_{k+1} - x \rangle \leq \lambda_k \varepsilon_k.$$

Define the auxiliary real sequence $\varphi_k := \frac{1}{2}|x_k - x|^2$. It is direct to check that

$$\begin{aligned}\langle x_{k+1} - x_k - \alpha_k(x_k - x_{k-1}), x_{k+1} - x \rangle \;=\; & \varphi_{k+1} - \varphi_k + \frac{1}{2}|x_{k+1} - x_k|^2 \\ & -\alpha_k \langle x_k - x_{k-1}, x_{k+1} - x \rangle,\end{aligned}$$

and since

$$
\begin{aligned}
&\langle x_k - x_{k-1}, x_{k+1} - x\rangle \\
&\qquad = \langle x_k - x_{k-1}, x_k - x\rangle + \langle x_k - x_{k-1}, x_{k+1} - x_k\rangle \\
&\qquad\qquad = \varphi_k - \varphi_{k-1} + \frac{1}{2}|x_k - x_{k-1}|^2 + \langle x_k - x_{k-1}, x_{k+1} - x_k\rangle,
\end{aligned}
$$

it follows that

$$
\begin{aligned}
&\varphi_{k+1} - \varphi_k - \alpha_k(\varphi_k - \varphi_{k-1}) \\
\leq\ & -\frac{1}{2}|x_{k+1} - x_k|^2 + \alpha_k\langle x_k - x_{k-1}, x_{k+1} - x_k\rangle \\
& + \frac{\alpha_k}{2}|x_k - x_{k-1}|^2 + \lambda_k\varepsilon_k \\
=\ & -\frac{1}{2}|x_{k+1} - x_k - \alpha_k(x_k - x_{k-1})|^2 \\
& + \frac{\alpha_k + \alpha_k^2}{2}|x_k - x_{k-1}|^2 + \lambda_k\varepsilon_k.
\end{aligned}
$$

Hence

$$
\varphi_{k+1} - \varphi_k - \alpha_k(\varphi_k - \varphi_{k-1}) \leq -\frac{1}{2}|v_{k+1}|^2 + \alpha_k|x_k - x_{k-1}|^2 + \lambda_k\varepsilon_k. \tag{2.11}
$$

Setting $\theta_k := \varphi_k - \varphi_{k-1}$ and $\delta_k := \alpha_k|x_k - x_{k-1}|^2 + \lambda_k\varepsilon_k$, we obtain

$$
\theta_{k+1} \leq \alpha_k\theta_k + \delta_k \leq \alpha_k[\theta_k]_+ + \delta_k,
$$

where $[t]_+ := \max(t, 0)$, and consequently

$$
[\theta_{k+1}]_+ \leq \alpha[\theta_k]_+ + \delta_k,
$$

with $\alpha \in [0, 1[$ given by (2).

The rest of the proof follows that given in [2] and is presented here for completeness and to convey the idea in [2]. The latter inequality yields

$$
[\theta_{k+1}]_+ \leq \alpha^k[\theta_1]_+ + \sum_{i=0}^{k-1}\alpha^i\delta_{k-i},
$$

and therefore

$$
\sum_{k=1}^{\infty}[\theta_{k+1}]_{\leq}\frac{1}{1-\alpha}([\theta_1]_+ + \sum_{k=1}^{\infty}\delta_k),
$$

which is finite thanks to (3) and (2.10). Consider the sequence defined by $t_k := \varphi_k - \sum_{i=1}^{k}[\theta_i]_+$. Since $\varphi_k \geq 0$ and $\sum_{i=1}^{k}[\theta_i]_+ < +\infty$, it follows that t_k is bounded from below. But

$$t_{k+1} = \varphi_{k+1} - [\theta_{k+1}]_+ - \sum_{i=1}^{k}[\theta_i]_+ \leq \varphi_{k+1} - \varphi_{k+1} + \varphi_k - \sum_{i=1}^{k}[\theta_i]_+ = t_k,$$

so that $\{t_k\}$ is nonincreasing. We thus deduce that $\{t_k\}$ is convergent and so is $\{\varphi_k\}$. On the other hand, from (2.11) we obtain the estimate

$$\frac{1}{2}|v_{k+1}|^2 \leq \varphi_k - \varphi_{k+1} + \alpha[\theta_k]_+ + \delta_k.$$

Passing to the limit in the latter inequality and taking into account that $\{\varphi_k\}$ converges, $[\theta_k]_+$ and δ_k go to zero as k tends to $+\infty$, we obtain

$$\lim_{k \to +\infty} v_{k+1} = 0.$$

Now let $\bar{x}$ be a weak cluster point of $\{x_k\}$. There exists a subsequence $\{x_\nu\}$, which converges weakly to $\bar{x}$ and satisfies $v_{\nu+1} + \lambda_\nu T^{\varepsilon_\nu}(x_{\nu+1}) \ni 0$. By definition (1.6), we have

$$\langle -\frac{v_{\nu+1}}{\lambda_\nu} - z, x_{\nu+1} - y \rangle \geq -\varepsilon_\nu \quad \forall z \in T(y).$$

Passing to the limit, as $\nu \to +\infty$, we obtain

$$\langle -z, \bar{x} - y \rangle \geq 0,$$

this being true for any $z \in T(y)$, from maximal monotonicity of T, it follows that $0 \in T(\bar{x})$, that is $\bar{x} \in S$. We end the proof by applying the well-known Opial Lemma [12]. □

Remark 2.1. 1. Although condition (2.10) involves the iterates that are a priori unknown. In practice, as it was stressed by Alvarez and Attouch, it is easy to enforce it by applying an appropriate on-line rule. Furthermore, condition (2.10) is automatically satisfied in some special cases (see Proposition 2.1, [2]).

2. Under assumptions of theorem 2.1 and in view of its proof, it is clear that $\{x_k\}$ is *bounded* if and only if, there exists at least one solution to (1.1).

2.2. Strong Convergence Results

First, we give a result showing that the criterion (1.9) ensures the strong convergence of the sequence generated by:

$$x_{k+1} - y_k + \lambda_k T^{\varepsilon_k}(x_{k+1}) \ni 0, \tag{2.12}$$

where $y_k = x_k + \alpha_k(x_k - x_{k-1})$, ε_k and α_k are positive reals.

Theorem 2.2. *Let $\{x_k\}$ be any sequence generated by (1.7) using criterion (1.9) with $\{\lambda_k\}$ nondecreasing ($\lambda_k \uparrow \lambda_\infty \le +\infty$). Assume that $\{x_k\}$ is bounded, $\exists \alpha \in [0,1[$ such that $\forall k \in I\!N^*, 0 \le \alpha_k \le \alpha$, and T^{-1} is Lipschitz continuous around zero, i.e., problem (1.1) has the unique solution, say $\bar{x}$, and there exist some constants $a > 0$ and $\tau > 0$ such that*

$$|v| \le \tau, v \in T(y) \Rightarrow |y - \bar{x}| \le a|v|.$$

Then, there is a real $\eta \in]0,1[$ and a range $K \in I\!N^*$ such that

$$|x_{k+1} - \bar{x}| \le \eta |x_k - \bar{x}| + (2\delta_k + \frac{a}{\lambda_k})\alpha_k |x_k - x_{k-1}| \quad \forall k \ge K.$$

Furthermore, if we assume that $\lim_{k\to+\infty} \frac{\alpha_k}{\lambda_k}|x_k - x_{k-1}| = 0$, then, the sequence $\{x_k\}$ strongly converges to $\bar{x}$.

Proof. The first part of the proof follows that given in [15] and is presented here for completeness. The sequence $\{x_k\}$ being bounded, also satisfies condition 3 of Theorem 2.1 for $\varepsilon_k = \delta_k |x_{k+1} - y_k|$, so the conclusions of Theorem 2.1 are in force. Now, let $\tilde{x}_{k+1}$ be the exact solution of the k-th inertial proximal method, that is

$$\lambda_k T\tilde{x}_{k+1} + \tilde{x}_{k+1} - y_k \ni 0. \tag{2.13}$$

Definition of T^{ε_k} combined with relations (2.12) and (2.13) leads to

$$\langle y_k - x_{k+1} - (y_k - \tilde{x}_{k+1}), x_{k+1} - \tilde{x}_{k+1}\rangle \ge -\lambda_k \varepsilon_k,$$

which implies that $|\tilde{x}_{k+1} - x_{k+1}|^2 \le \lambda_k \varepsilon_k$. The latter together with (1.9) yield

$$|\tilde{x}_{k+1} - x_{k+1}| \le \delta_k |x_{k+1} - y_k|. \tag{2.14}$$

Therefore

$$|\tilde{x}_{k+1} - y_k| \le |\tilde{x}_{k+1} - x_{k+1}| + |x_{k+1} - y_k| \le (1 + \delta_k)|x_{k+1} - y_k|,$$

which, using the convergence of $v_{k+1} := x_{k+1} - y_k \to 0$ (Theorem 2.1), implies that $|\tilde{x}_{k+1} - y_k| \to 0$. Because $\{\lambda_k\}$ is bounded away from zero, we further conclude that $w_k = \dfrac{1}{\lambda_k}(y_k - \tilde{x}_{k+1}) \to 0$. Using the Lipschitz continuity of T^{-1} around 0, we have, for indices k sufficiently large, that:

$$|\tilde{x}_{k+1} - \bar{x}| \le a|w_k| = \frac{a}{\lambda_k}|\tilde{x}_{k+1} - y_k|. \tag{2.15}$$

We further obtain

$$\begin{aligned} |y_k - \bar{x}|^2 &= |y_k - \tilde{x}_{k+1}|^2 + |\tilde{x}_{k+1} - \bar{x}|^2 \\ &\quad +2\langle y_k - \tilde{x}_{k+1}, \tilde{x}_{k+1} - \bar{x}\rangle \\ &= |y_k - \tilde{x}_{k+1}|^2 + |\tilde{x}_{k+1} - \bar{x}|^2 + 2\lambda_k\langle w_k, \tilde{x}_{k+1} - \bar{x}\rangle \\ &\ge |y_k - \tilde{x}_{k+1}|^2 + |\tilde{x}_{k+1} - \bar{x}|^2 \\ &\ge (1 + (\frac{\lambda_k}{a})^2)|\tilde{x}_{k+1} - \bar{x}|^2. \end{aligned} \tag{2.16}$$

Hence, by setting $\mu_k := \frac{a}{\sqrt{a^2+\lambda_k^2}}$, we obtain $|\tilde{x}_{k+1} - \bar{x}| \le \mu_k|y_k - \bar{x}|$.

Using the latter relation and (2.14), we further obtain

$$\begin{aligned} |x_{k+1} - \bar{x}| &\le |x_{k+1} - \tilde{x}_{k+1}| + |\tilde{x}_{k+1} - \bar{x}| \\ &\le \delta_k|x_{k+1} - y_k| + \mu_k|y_k - \bar{x}|. \end{aligned} \tag{2.17}$$

Similarly,

$$\begin{aligned} |x_{k+1} - y_k| &\le |x_{k+1} - \tilde{x}_{k+1}| + |\tilde{x}_{k+1} - y_k| \\ &\le \delta_k|x_{k+1} - y_k| + |y_k - \bar{x}|, \end{aligned}$$

where also (2.16) was used in the last inequality.

Therefore

$$|x_{k+1} - y_k| \le \frac{1}{1-\delta_k}|y_k - \bar{x}|.$$

Now, combining the latter relation with (2.17) and using the triangular inequality, we obtain

$$|x_{k+1} - \bar{x}| \le \theta_k|y_k - \bar{x}| \le \theta_k|x_k - \bar{x}| + \theta_k\alpha_k|x_k - x_{k-1}|,$$

where $\theta_k := \dfrac{\delta_k}{1-\delta_k} + \mu_k$, from which we deduce easily, by taking into account conditions on δ_k and λ_k, the existence of a range K such that

$$|x_{k+1} - \bar{x}| \leq \eta |x_k - \bar{x}| + \beta_k,$$

where $\eta \in]0,1[$ and $\beta_k := (2\delta_k + \dfrac{a}{\lambda_k})\alpha_k |x_k - x_{k-1}|$.

Hence,

$$|x_k - \bar{x}| \leq \eta^k |x_0 - \bar{x}| + \sum_{j=1}^{k} \eta^j \beta_{k-j}.$$

The result follows from Ortega and Rheinboldt [14, p. 338], since by hypothesis $\lim_{k\to+\infty} \beta_k = 0$. □

Remark 2.2. 1. When $\alpha_k = 0$, we obtain the linear convergence estimate obtained by Solodov and Svaiter [15] which is strictly better than the one for the classical proximal algorithm, namely, $\tilde{\theta}_k = \dfrac{\mu_k + \delta_k}{1-\delta_k}$ ([14], Theorem 2).

2. The condition of Lipschitz continuity above holds true if T^{-1} is globally Lipschitz continuous which is satisfied, for instance, when T is strongly monotone.

3. The condition of local Lipschitz continuity above is also satisfied if T^{-1} is differentiable at 0, that is, $T^{-1}(0) = \{\bar{x}\}$ and $\exists A : \mathcal{H} \to \mathcal{H}$ a continuous linear transformation such that, for $\delta > 0$

$$T^{-1}(0) - \bar{x} + Aw \subset o(|w|)\mathcal{B} \quad \text{if} \quad |w| \leq \delta,$$

where $\mathcal{B}$ stands for the closed unit ball.

We close this section with a special result showing that the Inertial Proximal Method can converge in finitely many iterations:

Theorem 2.3. *Let $\{x_k\}$ be any sequence generated by (1.8) under the criterion (1.9) with λ_k bounded away from zero. Suppose that $\{x_k\}$ is bounded and that*

$$\exists \bar{x} \in \mathcal{H} \quad \textit{such that} \quad 0 \in \operatorname{int} T\bar{x}. \tag{2.18}$$

Then, for all k sufficiently large, it holds that

$$J^T_{\lambda_k}(x_k + \alpha_k(x_k - x_{k-1})) = \bar{x}.$$

Moreover, the sequence $\{x_k\}$ strongly converges to $\bar{x}$.

Proof. Let $\tilde{x}_{k+1}$ be the exact solution of the k-th inertial proximal method, that is

$$\lambda_k T\tilde{x}_{k+1} + \tilde{x}_{k+1} - y_k \ni 0.$$

Hypothesis (2.18) ([14], Theorem 3 imply the existence of a positive real ε such that

$$|x| \le \varepsilon \Rightarrow T^{-1}x = \{\bar{x}\}.$$

But the hypothesis of the present theorem recovers those of Theorem 2.1. therefore, we know that $\frac{1}{\lambda_k}(y_k - \tilde{x}_{k+1}) \to 0$. Since $\tilde{x}_{k+1} \in T^{-1}(\frac{1}{\lambda_k}(y_k - \tilde{x}_{k+1}))$, assumption (2.18) implies that, for k sufficiently large, $\tilde{x}_{k+1} = \bar{x}$. Thus, the inertial proximal method in its exact form converges to $\bar{x}$ in a finite number of iterations from any starting points x_0 and x_1, from which we deduce:

$$|x_{k+1} - \bar{x}|^2 = |x_{k+1} - \tilde{x}_{k+1}|^2 \le \lambda_k \varepsilon_k\,,$$

for all k sufficientely large. The strong convergence of the sequence $\{x_k\}$ to $\bar{x}$ follows by passing to the limit in the last inequality, since condition (1.8) implies that $\lim_{k\to+\infty} \lambda_k \varepsilon_k = 0$. □

3. Convex Minimization

3.1. Approximate Methods

An interesting case is obtained by taking $T = \partial f$, ∂f stands for the subdifferential of a proper convex lower-semicontinuous function $f : X \to I\!R \cup \{+\infty\}$. Indeed, ∂f is well-known to be a maximal monotone operator and problem (1.1) reduces to the one of finding a minimizer of the function f.

In [1], Alvarez proposed the following approximate inertial proximal method:

$$\lambda_k \partial_{\varepsilon_k} f(x_{k+1}) + x_{k+1} - x_k - \alpha_k(x_k - x_{k-1}) \ni 0, \tag{3.19}$$

where $\partial_{\varepsilon_k} f$ is the approximate subdifferential of f. Since in the case $T = \partial f$ the enlargement given in (1.6) is larger than the the approxiamte subdifferential, i.e. $\partial_\varepsilon f \subset (\partial f)^\varepsilon$, we can write $\partial_{\varepsilon_k} f(x_{k+1}) \subset (\partial f)^{\varepsilon_k}(x_{k+1})$, which leads to

$$\lambda_k (\partial f)^{\varepsilon_k}(x_{k+1}) + x_{k+1} - x_k - \alpha_k(x_k - x_{k-1}) \ni 0, \tag{3.20}$$

which is a particular case of the method proposed in this paper with $T = \partial f$. As a consequence of Theorem 2.1 and Theorem 2.2, we obtain the following convergence result, which recover and completes a result by Alvarez [1].

Corollary 3.1. *Let $\{x_k\} \subset \mathcal{H}$ be a sequence such that*

$$0 \in x_{k+1} - x_k - \alpha_k(x_k - x_{k-1}) + \lambda_k \partial_{\varepsilon_k} f(x_{k+1}), k = 1, 2, \cdots, \tag{3.21}$$

where f is a proper closed convex function with $\operatorname{Arg\,min} f \neq \emptyset$, and the parameters α_k, λ_k and ε_k satisfy:

1. $\exists \lambda > 0$ *such that* $\forall k \in \mathbb{N}, \lambda_k \geq \lambda$.
2. $\exists \alpha \in [0, 1[$ *such that* $\forall k \in \mathbb{N}, 0 \leq \alpha_k \leq \alpha$.
3. $\sum_{k=1}^{+\infty} \lambda_k \varepsilon_k < +\infty$.

If the following condition holds

$$\sum_{k=1}^{+\infty} \alpha_k |x_k - x_{k-1}|^2 < +\infty, \tag{3.22}$$

then $\{x_k\}$ weakly converges to a minimizer of f and $\lim_{k \to +\infty} f(x_k) = \inf_{x \in \mathcal{H}} f(x)$. Moreover, if we replace condition 3 of Theorem 2.1 by criterion (1.9) and we assume in addition that ∂f^* is Lipschitz continuous around zero, the convergene is strong. The function f^* stands for the Fenchel conjuguate of f, namely, $f(x^*) = \sup_{x \in \mathcal{H}}(\langle x^*, x\rangle - f(x))$.

Remark 3.1. The formula $\lim_{k \to +\infty} f(x_k) = \inf_{x \in \mathcal{H}} f(x)$ is obtained from relation (3.19), definition of the approximate subdifferential, lower semicontinuity of the function f and the fact that $\lim_{k \to +\infty} v_{k+1} = 0$ (see the proof of Theorem 2.1).

3.2. A Perturbed Inertial Proximal Method

When f is nonsmooth, subproblems (3.19) may be very hard to solve and several authors proposed to approximate f by a sequence f_k of more tractable convex functions (see for example [6], [7] and [11]).

In this section we consider $f_k, k = 2, 3 \cdots$, a monotone decreasing sequence of proper closed convex function on $\mathcal{H}$, f be such that $f = cl(\inf_k f_k)$. Now, x_0, x_1 be given in $\mathcal{H}$, let us consider a sequence $\{x_k\}$ generated by the following perturbed method:

$$0 \in x_k - x_{k-1} - \alpha_k(x_{k-1} - x_{k-2}) + \lambda_k \partial_{\varepsilon_k} f_k(x_k), k = 2, 3, \cdots, \tag{3.23}$$

where the parameters λ_k, ε_k and α_k are nonnegative real numbers.

Theorem 3.1. *Suppose that the sequence $\{x_k\}$ is bounded and that the following condition holds true*

$$\lim_{k\to+\infty} \varepsilon_k = 0 \quad \textit{and} \quad \lim_{k\to+\infty} \frac{\alpha_k}{\sqrt{\lambda_k}}|x_{k-1} - x_{k-2}| = 0, \tag{3.24}$$

then $\lim_{k\to+\infty} f_k(x_k) = \inf f$ *and every weak cluster point of* $\{x_k\}$ *is a minimizer of* f.

Proof. For all $x \in \mathcal{H}$, by definition of the approximate subdifferential, we have

$$f_k(x) \geq f_k(x_k) - \frac{1}{\lambda_k}\langle x_k - x_{k-1} - \alpha_k(x_{k-1} - x_{k-2}), x - x_k\rangle - \varepsilon_k. \tag{3.25}$$

Setting $x = x_{k-1}$ and taking into account the following inequalities

$$\begin{aligned} -\langle x_k - x_{k-1} - \alpha_k(x_{k-1} - x_{k-2}), x_{k-1} - x_k\rangle & \\ \geq\ & \frac{1}{2}|x_k - x_{k-1}|^2 \\ -\ & \alpha_k\langle x_{k-1} - x_{k-2}, x_k - x_{k-1}\rangle \\ =\ & -\frac{\alpha_k^2}{2}|x_{k-1} - x_{k-2}|^2 \\ +\ & \frac{1}{2}|x_k - x_{k-1} - \alpha_k(x_{k-1} - x_{k-2})|^2, \end{aligned}$$

we infer that

$$f_k(x_k) + \frac{1}{2\lambda_k}|v_k|^2 \leq f_k(x_{k-1}) + \frac{\alpha_k^2}{2\lambda_k}|x_{k-1} - x_{k-2}|^2 + \epsilon_k,$$

where $v_k := \frac{1}{2}|x_k - x_{k-1} - \alpha_k(x_{k-1} - x_{k-2})|$.

This, combined with the fact that f_k decreases yields

$$f_k(x_k) + \frac{1}{2\lambda_k}|v_k|^2 \leq f_{k-1}(x_{k-1}) + \frac{\alpha_k^2}{2\lambda_k}|x_{k-1} - x_{k-2}|^2 + \epsilon_k. \tag{3.26}$$

Now, let $\mathcal{I} = \{k \in I\!N, f_k(x_k) > f_{k-1}(x_{k-1})\}$.

- If $\mathcal{I}$ is not finite

  Passing to the limit in the inequality (2.26) and in the light of (3.24), we get $\lim_{k\in\mathcal{I}} \dfrac{1}{\lambda_k}|v_k| = 0$. On the other hand, since $f_k \downarrow f = cl(\inf_k f_k)$, we have that $\{f_k\}$ converges to f in the Mosco epi-convergence sense (see [3], Theorem 3.20). Namely, for any $z \in \mathcal{H}$, the following statements hold true

  1. $\exists z_k$ satisfying $z = s - \lim_{k\to+\infty} z_k$ and $\limsup_{k\to+\infty} f_k(z_k) \leq f(z)$.
  2. $\forall z_k$ with $z = w - \lim_{k\to+\infty} z_k$ one has $\liminf_{k\to+\infty} f_k(z_k) \geq f(z)$.

  By setting $x = z_k$ with $\lim_{k\to+\infty} z_k = z$ and $\limsup_{k\to+\infty} f_k(z_k) \leq f(z)$, and by passing to the limit in (3.25), we obtain

  $$\limsup_{k\in\mathcal{I}} f_k(x_k) \leq \inf f. \tag{3.27}$$

  - If $\mathcal{I}$ has an empty or finite complement, we get trivialy

    $$\limsup_{k\to+\infty} f_k(x_k) \leq \inf f. \tag{3.28}$$

  - Otherwise, for every $k \in I\!N$, we define $i(k) \in \mathcal{I}$ by

    $$i(k) = \begin{cases} k & \text{if} \quad k \in \mathcal{I}\,, \\ \max\{l; l < k, l \in S\} & \text{if} \quad k \notin \mathcal{I}\,. \end{cases}$$

    We have $\lim_{k\to+\infty} i(k) = +\infty$ and $f_k(x_k) \leq f_{i(k)}(x_{i(k)})$, which combined with (2.27) gives relation (2.28).

- If $\mathcal{I}$ is empty or finite.

  For every k large enough, we have $f_k(x_k) \leq f_{k-1}(x_{k-1})$. So the sequence $\{f_k(x_k)\}$ converges in $\overline{I\!R}$. If $\lim_{k\to+\infty} f_k(x_k) = -\infty$, then, thanks to the fact that $\{f_k\}$ decreases, we have $\lim_{k\to+\infty} f_k(x_k) = \inf f = -\infty$. Otherwise, from (3.26), we infer that $\lim_{k\to+\infty} \dfrac{1}{\lambda_k}|v_k| = 0$, and as in the previous case, we get directly (3.28).

  In all cases we have proved relation (2.28). From which we deduce

  $$\lim_{k\to+\infty} f_k(x_k) = \inf f,$$

  because the sequence $\{f_k\}$ decreases.

To conclude, we use an Mosco epi-convergence argument. Let $\bar{x}$ be any weak cluster point of $\{x_k\}$ and $\{x_\nu\}$ a subsequence, which weakly converges to $\bar{x}$. According to definition of the Mosco epi-convergence we can write

$$f(\bar{x}) \leq \liminf_{\nu \to +\infty} f_\nu(x_\nu) = \inf f.$$

That is $\bar{x}$ is a minimizer of f. □

Proposition 3.1. *If, in addition to the hypothesis of Corollary 3.1, we assume for every solution $x \in S$, that*

$$\sum_{k=2}^{+\infty} \lambda_k \max\{(f_k(x) - f_k(x_k)), 0\} < +\infty. \tag{3.29}$$

Then the whole sequence $\{x_k\}$ generated by (3.23) converges weakly to a minimizer of f.

Proof. From the proof of Theorem 2.1, relation (2.25) can be rewritten as

$$\lambda_k(f_k(x) - f_k(x_k)) \geq \varphi_k - \varphi_{k-1} - \alpha_k(\varphi_{k-1} - \varphi_{k-2}) + \frac{1}{2}|v_k|^2 - \alpha_k|x_{k-1} - x_{k-2}|^2 - \lambda_k\varepsilon_k.$$

Taking any $x \in S$ and according to relation (3.29), we obtain

$$\varphi_k - \varphi_{k-1} - \alpha_k(\varphi_{k-1} - \varphi_{k-2}) \leq -\frac{1}{2}|v_k|^2 + \delta_k,$$

with $\delta_k = \lambda_k \max\{(f_k(x) - f_k(x_k)), 0\} + \alpha_k|x_{k-1} - x_{k-2}|^2 + \lambda_k\varepsilon_k$ which is nothing but relation (2.11). So, from the proof of Theorem 2.1, we deduce that $\varphi_k(x) = \lim_{k \to +\infty} \frac{1}{2}|x_k - x|^2 = 0$, for all $x \in S$. This together with Theorem 3.1 gives the convergence of the whole sequence $\{x_k\}$ by applying Opial Lemma. □

Remark 3.2. It is worth mentioning that (3.29) is satisfied when $f_k = f$, for all k. Indeed, in this case $f(x) - f(x_k) \leq 0$. Furthermore, for the sequence of barrier functions, see the example below (resp. Tikhonov regularization, namely $f_k(\cdot) = f(\cdot) + \nu_k|\cdot|^2$), this assumption amounts to imposing a rate of convergence on the barrier (resp. Thikonov) parameters. More precisely, it can be proved easily that, if $\sum_{k=2}^{+\infty} \lambda_k \nu_k < +\infty$, then the functions f_k satisfy (3.29). This is, for instance, the case when $\{\lambda_k\}$ is bounded and $\nu_k = \frac{1}{k^\beta}, \forall k$ with $\beta > 1$.

To conclude this section, let us now give some examples of such perturbation.

Example 1. *Constrained convex minimization.* In the classical constrained minimization problem, we are given a convex function h from $\mathcal{H}$ to $I\!R$, and a nonempty closed convex set C, and we wish to find a solution of

$$\min_{x\in\mathcal{H}}(h(x)+\delta_C(x)). \tag{3.30}$$

An example of such an approximation of $f := h + \delta_C$ is obtained by taking an interior approximation of the feasible set C, namely $f_k := h + \delta_{C_k}$, where $\{C_k\}$ is a sequence of nonempty closed convex sets chosen in such a way that

$$C = \overline{\cup C_k} \quad \text{and} \quad C_k \subset C_{k+1} \subset C \text{ for } \ k = 1, 2, \cdots, \tag{3.31}$$

where C is described by finitely many convex inequalities: $g_i(x) \leq 0, i = 1, \cdots$. An other example of such an approximation uses the sequence of the inverse barrier functions associated with the closed convex set C, namely

$$f_k := h + \phi_k \quad \text{with} \quad \phi_k(x) = -\nu_k \sum_{i=1}^{m} \frac{1}{\ln(-g_i(x))}, \quad x \in \text{int } \ C, \tag{3.32}$$

and $+\infty$ otherwise, where the sequence $\{\nu_k\}$ of positive barrier parameters is strictly decreasing to 0.

Indeed, it is easy to check, in the both cases, that $f_k \downarrow f$.

Acknowledgement

The authors are grateful to F. Alvarez for his valuable comments and remarks.

References

[1] F. Alvarez, On the minimizing property of a second ordre disipative system in Hilbert space, *SIAM J. of Control and Optimization*, **38**, No. 4 (2000), 1102-1119.

[2] F. Alvarez, H. Attouch, An inertial proximal method for maximal monotone operators via discretization of a non linear oscillator with damping, *Set Valued Analysis*, **9** (2001), 3-11.

[3] H. Attouch, *Variational Convergence for Functions and Operators*, Appl. Math., Pitman, London (1984).

[4] H. Attouch, F. Alvarez, The heavy ball with friction method dynamical system for convex constrained minimization problems, In: *Proceedings of the 9-th Belgian-French-German Conference on Optimization*, Namur (1999); Lecture Notes In Econom. and Math. Systems, Springer, Berlin, **481** (2000), 25-35.

[5] H. Attouch, X. Cabot, Redont, The dynamic of elastic shocks via epigraphical regularization of a differential inclusion. Barrier and penalty approximations, *Preprint*, Université Montpellier II-ACSIOM (2001).

[6] B. Lemaire, Coupling optimization methods and variational convergence, In: *Trends in Math. Optim. International Series of Numerical Math.*, **84** (C), Birkhäuser Verlag (1988), 163-179.

[7] B. Lemaire, *About the Convergence of the Proximal Method, Advances in Optimization*, Lectures Notes Economics and Math. Systems, **382**, Springer-Verlag (1992).

[8] B. Lemaire, The proximal method, In: *New Methods in Optimization and their Industrial Uses*, Int. Ser. Numer. Math., **87**, Birkhäuser Verlag (1989), 73-89.

[9] B. Martinet, Régularisation d'inéquations variationnelles par approximations successives, *Rev. Francaise Inf. Rech. Oper.*, (1970), 154-159.

[10] A. Moudafi, Coupling proximal methods and variational convergence, *ZOR-Methods and Models of Operations Research* (1993), 269-280.

[11] H. Nguyen, J.J. Strodiot, G. Salmon, Coupling the auxiliary problem principle and epiconvergence theory to solve general variational inequalities, *Journal Optim. Theor. Appli.*, **104**, No. 3 (2000), 629-657.

[12] Z. Opial, Weak convergence of the sequence of successive approximations for nonexpansive mappings, *Bulletin of the American mathematical Society*, **73** (1967), 591-597.

[13] J.M. Ortega, W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Acad. Press (1970).

[14] R.T. Rockafellar, Monotone operator and the proximal point algorithm, *SIAM J. Control. Opt.*, **14**, No. 5 (1976), 877-898.

[15] M.V. Sodolov, B.F. Svaiter, A comparaison of rates of convergence of two inexact proximal point algorithms, *Nonlinear Optimization and related Topics, Appl. Optim.*, Kluwer, **36** (2000).

[16] B.F. Svaiter, R.S. Burachik, A.N. Iusem, Enlargement of maximal monotone operators with application to variational inequalities, *Set-Valued Analysis*, **5** (1997), 159-180.