Journal of Convex Analysis
Volume 14 (2007), No. 4, 869–878

A Proximal Method for Maximal Monotone Operators via Discretization of a First Order Dissipative Dynamical System

Paul-Emile Maingé
GRIMMAG, Université des Antilles-Guyane, Département Scientifique Interfacultaire, Campus de Schoelcher, 97230 Cedex, Martinique (F.W.I.)
paul-emile.mainge@martinique.univ-ag.fr

Abdellatif Moudafi
GRIMMAG, Université des Antilles-Guyane, Département Scientifique Interfacultaire, Campus de Schoelcher, 97230 Cedex, Martinique (F.W.I.)
abdellatif.moudafi@martinique.univ-ag.fr

Received: April 5, 2006
Revised manuscript received: October 12, 2006

We present an iterative method for finding zeroes of maximal monotone operators in a real Hilbert space. The underlying idea relies upon the discretization of a first order dissipative dynamical system which allows us to preserve the local feature, as well as to obtain convergence results. The main theorems do not only recover known convergence results of standard and inertial proximal methods, but also provide a theoretical basis for the application of new iterative methods.

Keywords: Monotone operators; standard and inertial proximal methods; minimization

2000 Mathematics Subject Classification: Primary 49J53, 65K10; Secondary 49M37, 90C25

1. Introduction and preliminaries

Throughout the paper $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is a real Hilbert space and $A : \mathcal{H} \to 2^{\mathcal{H}}$ is a maximal monotone operator, that is, A is monotone, namely $\forall x, y \in \mathcal{H}, \forall v \in A(x), \forall w \in A(y), \langle v - w, x - y \rangle \geq 0$, and the graph $GrA = \{(x, v) \in \mathcal{H} \times \mathcal{H}; v \in A(x)\}$ is not properly contained in the graph of any other monotone operator. The resolvent, $J_\lambda^A := (I + \lambda A)^{-1}$, of A with parameter $\lambda > 0$ is a nonexpansive mapping satisfying the following key property:

$$J_\lambda^A(x) = x \text{ if and only if } 0 \in A(x).$$

We are interested in solving the problem

$$(\mathcal{P}) \qquad \text{Find } x \in \mathcal{H} \text{ such that } 0 \in A(x).$$

To this end, we propose an algorithm based upon an implicit discretization of the following first order dissipative dynamical system introduced in [3]:

$$\left[\begin{array}{l} x^{(1)}(t) + \beta \nabla \phi(x(t)) + ax(t) + by(t) = 0, \\ y^{(1)}(t) + ax(t) + by(t) = 0, \\ x(0) = x_0, \quad y(0) = y_0, \end{array} \right. \tag{1}$$

ISSN 0944-6532 / \$ 2.50

where ϕ is assumed to be a convex differentiable function, the initial data x_0 and y_0 belong to $\mathcal{H}$ and where $\beta > 0$, $b > 0$ and $a + b > 0$.

It is worth mentioning that most of the existing numerical methods for minimizing a function are based upon a discretization of some continuous equations with appropriate convergence properties. The standard proximal point algorithm and the gradient method come from the first order steepest descent equation $x^{(1)}(t) + \nabla\phi(x(t)) = 0$. Conversely the inertial proximal method is inspired by the heavy ball with friction dynamical system $x^{(2)}(t)+\alpha x^{(1)}(t)+\nabla\phi(x(t)) = 0$. The latter system was introduced by Attouch et al. [6] for overcoming some of the drawbacks of the steepest descent method. It turns out that when ϕ is convex, the two trajectories of the last two equations weakly converge to minimizers of ϕ. Nevertheless, there is a drastic difference between them. By contrast with the steepest descent method, the heavy ball with friction dynamical system is no more a descent method. In the latter case, the function $\phi(x(t))$ does not decrease along the trajectories in general; it is the energy of the system $E(t) := \frac{1}{2}|x^{(1)}(t)|^2 + \phi(x(t))$ that is decreasing. This confers on this system interesting properties for the exploration of local minima of ϕ (see [6] for more details). It also appears that its trajectories may exhibit oscillations which are not desirable. In view of numerical optimization purposes, Alvarez and Pérez [4] have studied the "Continuous Newton" method $\nabla^2\phi(x(t))x^{(1)}(t) + \nabla\phi(x(t)) = 0$. If one combines this last system with the heavy ball system with friction, the system thus obtained,

$$x^{(2)}(t) + \alpha x^{(1)}(t) + \beta\nabla^2\phi(x(t))x^{(1)}(t) + \nabla\phi(x(t)) = 0, \tag{2}$$

inherits most of the advantages of the two preceding systems and corrects both of the above mentioned drawbacks: the term $\nabla^2\phi(x(t))x^{(1)}(t)$ is a clever geometric damping term, while the acceleration term $x^{(2)}(t)$ makes the Newton dynamical system well-posed, even if $\nabla^2\phi(x(t))$ is degenerate (see Attouch, Redont [7] for a first study of this question). It is worth mentionning that (2) is a second order system both in time and in space. Furthermore, it was proved that (2) is equivalent in some sense to the system (1) which is first-order in time and with no occurrence of the Hessian (see [3] for more details). This matter opens new interesting perspectives such as considering (2) for nonsmooth functions only lower semi-continuous or involving constraints, with clear applications to mechanics and PDE's (wave equations, shocks). It then seems natural to investigate implicit and/or non-implicit discretization of (1) for numerical and theoretical optimization purposes, because an accurate discrete version of an equation is supposed to preserve the essential properties of the continuous one. Motivated by the above arguments and inspired by the system (1), we introduce and analyze convergence properties of the following Dissipative Proximal Method, (DPM) for short, for the monotone inclusion $0 \in A(x)$:

Given parameters (λ_n), (θ_n) in $(0, +\infty)$, $\mu \in [0, 1]$, $\gamma \in (0, 2)$ and $\alpha \in (-\infty, +\infty)$, this iterative method generates two sequences (x_n) and (y_n) by:

$$\left[\begin{array}{l} \text{Initialization: } (x_0, y_0) \text{ in } \mathcal{H} \times \mathcal{H}. \\ \text{Step 1: } \; x_{n+1} := J^A_{\lambda_n}\left(x_n - \theta_n(\alpha x_n + \gamma y_n)\right). \\ \text{Step 2: } \; y_{n+1} = (1-\gamma)y_n - \alpha(\mu x_{n+1} + (1-\mu)x_n). \end{array}\right. \tag{3}$$

The goal of this paper is to provide a broad framework for the design and the analysis of algorithms based on a discretization of the system (1). This framework will not only lead to a unified convergence analysis of some existing algorithms in convex optimization and variational inequalities, but will also serve as a basis for the development of a new class of algorithms. More specifically, we would like to emphasize that the standard proximal point algorithm proposed in [16] is nothing but the special case of DPM when $\alpha = \gamma = 0$ and that the inertial proximal method initiated in [1] for convex minimization and further developed in [2] corresponds to the particular case of DPM where $\alpha = -1$, $\gamma = 1$ and $\mu = 0$.

Under suitable conditions on the parameters, we prove a discrete version of the main theorem in [3] in a more general context. More precisely, we prove that the proposed algorithm generates an asymptotically regular sequence (x_n) which weakly converges toward a solution of $0 \in A(x)$. An application to convex minimization and some concluding remarks are also provided.

2. Asymptotic Convergence

To begin with, let us state an elementary and key property.

Lemma 2.1. *For any $q \in A^{-1}(0)$, we have*

$$\begin{aligned} a_{n+1} - a_n \leq \left(\alpha + \gamma\left(\frac{1}{\theta_n} - \mu\alpha\right)\right) \langle q - x_{n+1}, x_{n+1} - x_n\rangle \\ + \langle x_{n+1} - x_n, y_{n+1} - y_n\rangle + \mu\alpha|x_{n+1} - x_n|^2, \end{aligned} \tag{4}$$

where $a_n := \langle q - x_n, \alpha x_n + \gamma y_n\rangle$.

Proof. First, we successively have

$$\begin{aligned} a_{n+1} - a_n &= \langle q - x_{n+1}, \alpha x_{n+1} + \gamma y_{n+1}\rangle - \langle q - x_n, \alpha x_n + \gamma y_n\rangle \\ &= \langle q - x_{n+1}, \alpha(x_{n+1} - x_n) + \gamma(y_{n+1} - y_n)\rangle + \langle x_n - x_{n+1}, \alpha x_n + \gamma y_n\rangle \\ &= \alpha\langle q - x_{n+1}, x_{n+1} - x_n\rangle + \gamma\langle q - x_{n+1}, y_{n+1} - y_n\rangle \\ &\quad + \langle x_n - x_{n+1}, \alpha x_n + \gamma y_n\rangle \end{aligned}$$

Moreover, by Step 2 in (3) we have

$$y_{n+1} = y_n - \gamma y_n - \alpha x_n - \alpha\mu(x_{n+1} - x_n),$$

hence

$$\alpha x_n + \gamma y_n = -(y_{n+1} - y_n) - \mu\alpha(x_{n+1} - x_n). \tag{5}$$

Consequently, we get

$$\begin{aligned} a_{n+1} - a_n &= \alpha\langle q - x_{n+1}, x_{n+1} - x_n\rangle + \gamma\langle q - x_{n+1}, y_{n+1} - y_n\rangle \\ &\quad + \langle x_n - x_{n+1}, -(y_{n+1} - y_n) - \mu\alpha(x_{n+1} - x_n)\rangle \\ &= \alpha\langle q - x_{n+1}, x_{n+1} - x_n\rangle + \gamma\langle q - x_{n+1}, y_{n+1} - y_n\rangle \\ &\quad + \langle x_{n+1} - x_n, y_{n+1} - y_n\rangle + \mu\alpha|x_{n+1} - x_n|^2. \end{aligned} \tag{6}$$

On the other hand, Step 1 in (3) amounts to

$$(I + \lambda_n A)x_{n+1} \ni x_n - \theta_n(\alpha x_n + \gamma y_n),$$

which by monotonicity of A yields

$$\langle x_{n+1} - x_n, x_{n+1} - q\rangle \leq -\theta_n \langle \alpha x_n + \gamma y_n, x_{n+1} - q\rangle,$$

hence by virtue of (5) we get

$$\begin{aligned}\langle x_{n+1} - x_n, x_{n+1} - q\rangle &\leq \theta_n \langle (y_{n+1} - y_n) + \mu\alpha(x_{n+1} - x_n), x_{n+1} - q\rangle \\ &= -\theta_n \langle y_{n+1} - y_n, q - x_{n+1}\rangle + \theta_n \mu\alpha \langle x_{n+1} - x_n, x_{n+1} - q\rangle,\end{aligned}$$

so that

$$\langle y_{n+1} - y_n, q - x_{n+1}\rangle \leq \left(\mu\alpha - \frac{1}{\theta_n}\right) \langle x_{n+1} - x_n, x_{n+1} - q\rangle. \tag{7}$$

Combining (6) and (7), we then infer the desired result □

The following elementary result will be useful to build the discrete energy.

Lemma 2.2. *For any $q \in A^{-1}(0)$, we have*

$$\begin{aligned}F_{n+1} - F_n \leq &\left(-\frac{1}{2} + \alpha^2 \left(\frac{1}{2} - \mu\right) + \mu\alpha\right) |x_{n+1} - x_n|^2 \\ &+ \gamma \left(\frac{\gamma}{2} - 1\right) |y_{n+1} - y_n|^2 \\ &+ (\alpha(\gamma(1-\mu) - 1) + 1) \langle x_{n+1} - x_n, y_{n+1} - y_n\rangle \\ &+ \left(1 - \alpha - \gamma\left(\frac{1}{\theta_n} - \mu\alpha\right)\right) \langle x_{n+1} - q, x_{n+1} - x_n\rangle.\end{aligned} \tag{8}$$

where $F_n = \frac{1}{2}|(q - x_n) + (\alpha x_n + \gamma y_n)|^2$.

Proof. To estimate the term $F_{n+1} - F_n$, let us set

$$\begin{aligned}b_n &= \frac{1}{2}|q - x_n|^2, \\ c_n &= \frac{1}{2}|\alpha x_n + \gamma y_n|^2.\end{aligned}$$

Then, we can write

$$\begin{aligned}b_{n+1} - b_n &= \frac{1}{2}\left(|q - x_{n+1}|^2 - |q - x_n|^2\right) \\ &= \langle x_n - x_{n+1}, q - \frac{x_{n+1} + x_n}{2}\rangle \\ &= \langle x_n - x_{n+1}, q - x_{n+1}\rangle + \langle x_n - x_{n+1}, \frac{x_{n+1} - x_n}{2}\rangle,\end{aligned} \tag{9}$$

namely

$$b_{n+1} - b_n = -\langle x_{n+1} - x_n, q - x_{n+1}\rangle - \frac{1}{2}|x_{n+1} - x_n|^2. \tag{10}$$

We also have

$$\begin{aligned}c_{n+1} - c_n &= \frac{1}{2}|\alpha x_{n+1} + \gamma y_{n+1}|^2 - \frac{1}{2}|\alpha x_n + \gamma y_n|^2 \\ &= \langle \alpha(x_{n+1} - x_n) + \gamma(y_{n+1} - y_n), \alpha\frac{x_{n+1} + x_n}{2} + \gamma\frac{y_{n+1} + y_n}{2}\rangle.\end{aligned} \tag{11}$$

Moreover, it is easily checked that

$$\alpha x_n + \gamma y_n = \alpha \frac{x_{n+1}+x_n}{2} + \gamma \frac{y_{n+1}+y_n}{2} - \frac{\alpha}{2}(x_{n+1}-x_n) - \frac{\gamma}{2}(y_{n+1}-y_n),$$

which in the light of (5) yields

$$\begin{aligned} &\alpha \frac{x_{n+1}+x_n}{2} + \gamma \frac{y_{n+1}+y_n}{2} \\ &= -(y_{n+1}-y_n) - \mu\alpha(x_{n+1}-x_n) + \frac{\alpha}{2}(x_{n+1}-x_n) + \frac{\gamma}{2}(y_{n+1}-y_n) \\ &= \left(\frac{\gamma}{2}-1\right)(y_{n+1}-y_n) + \alpha\left(\frac{1}{2}-\mu\right)(x_{n+1}-x_n). \end{aligned} \tag{12}$$

Consequently, we deduce

$$c_{n+1}-c_n = \left\langle \alpha(x_{n+1}-x_n)+\gamma(y_{n+1}-y_n), \left(\frac{\gamma}{2}-1\right)(y_{n+1}-y_n) + \alpha\left(\frac{1}{2}-\mu\right)(x_{n+1}-x_n)\right\rangle$$

that is

$$\begin{aligned} c_{n+1}-c_n = &\left(\alpha\left(\frac{\gamma}{2}-1\right) + \alpha\gamma\left(\frac{1}{2}-\mu\right)\right)\langle x_{n+1}-x_n, y_{n+1}-y_n\rangle \\ &+ \alpha^2\left(\frac{1}{2}-\mu\right)|x_{n+1}-x_n|^2 + \gamma\left(\frac{\gamma}{2}-1\right)|y_{n+1}-y_n|^2, \end{aligned} \tag{13}$$

or equivalently

$$\begin{aligned} c_{n+1}-c_n = &\,\alpha\left(\gamma(1-\mu)-1\right)\langle x_{n+1}-x_n, y_{n+1}-y_n\rangle \\ &+ \alpha^2\left(\frac{1}{2}-\mu\right)|x_{n+1}-x_n|^2 + \gamma\left(\frac{\gamma}{2}-1\right)|y_{n+1}-y_n|^2. \end{aligned} \tag{14}$$

Furthermore, it is easily observed that

$$F_{n+1}-F_n = (a_{n+1}-a_n) + (b_{n+1}-b_n) + (c_{n+1}-c_n),$$

where $a_n = \langle q - x_n, \alpha x_n + \gamma y_n\rangle$. Using Lemma 2.1 and thanks to (10) and (14), we finally obtain the desired result □

In order to prove our main convergence result, the idea is to apply the following well-known Opial lemma on weak convergence in Hilbert spaces.

Lemma 2.3. *Let $\mathcal{H}$ be a Hilbert space and (x_n) a sequence in $\mathcal{H}$ such that there exists a nonempty set $S \subset \mathcal{H}$ satisfying:*

(*i*) *For every $\tilde{x} \in S$, $lim_n|x_n - \tilde{x}|$ exists.*

(*ii*) *If (x_{n_ν}) weakly converges to $\tilde{x}$ for a subsequence $n_\nu \to +\infty$, then $\tilde{x} \in S$.*

Then, there exists $\bar{x} \in S$ such that (x_n) weakly converges to $\bar{x}$.

We are now in a position to prove the main convergence theorem.

Theorem 2.4. *Let $\gamma \in (0,2)$, $\mu \in [0,1]$, let α be any real number, suppose that $\lambda = \liminf_n \lambda_n > 0$ and assume the set of zeroes of the maximal monotone mapping A is nonempty. There exists a positive constant θ such that if (θ_n) is a nondecreasing sequence in $(0,\theta)$, then the sequences (x_n) and (y_n) generated by scheme (3) satisfy the following properties:*

(*i*) $\lim_{n\to+\infty} |x_{n+1} - x_n| = \lim_{n\to+\infty} |y_{n+1} - y_n| = 0.$
(*ii*) $\lim_{n\to+\infty} dist(0, A(x_n)) = 0.$
(*iii*) (x_n) *weakly converges to a zero of A.*

Proof. From Lemma 2.2, we have

$$\begin{aligned} F_{n+1} - F_n \le\ & d_1|x_{n+1} - x_n|^2 - d_2|y_{n+1} - y_n|^2 \\ & + d_3\langle x_{n+1} - x_n, y_{n+1} - y_n\rangle - \rho_n\langle x_{n+1} - q, x_{n+1} - x_n\rangle, \end{aligned} \tag{15}$$

where

$$\begin{aligned} d_1 &= \left(-\frac{1}{2} + \alpha^2\left(\frac{1}{2} - \mu\right) + \mu\alpha\right), \\ d_2 &= \gamma\left(1 - \frac{\gamma}{2}\right), \\ d_3 &= \left(\alpha(\gamma(1-\mu) - 1) + 1\right), \\ \rho_n &= -\left(1 - \alpha - \gamma\left(\frac{1}{\theta_n} - \mu\alpha\right)\right). \end{aligned}$$

Observing that

$$\langle x_{n+1} - q, x_{n+1} - x_n\rangle = -\frac{1}{2}|x_n - q|^2 + \frac{1}{2}|x_{n+1} - q|^2 + \frac{1}{2}|x_{n+1} - x_n|^2$$

and noting that (ρ_n) is a non-increasing sequence, since (θ_n) is non-decreasing, the previous inequality leads to

$$\begin{aligned} F_{n+1} - F_n \le\ & \left(d_1 - \frac{\rho_n}{2}\right)|x_{n+1} - x_n|^2 - d_2|y_{n+1} - y_n|^2 \\ & + d_3\langle x_{n+1} - x_n, y_{n+1} - y_n\rangle\frac{\rho_n}{2}|x_n - q|^2 - \frac{\rho_{n+1}}{2}|x_{n+1} - q|^2, \end{aligned} \tag{16}$$

or equivalently

$$\begin{aligned} & F_{n+1} + \frac{\rho_{n+1}}{2}|x_{n+1} - q|^2 - \left(F_n + \frac{\rho_n}{2}|x_n - q|^2\right) \\ \le\ & \left(d_1 - \frac{\rho_n}{2}\right)|x_{n+1} - x_n|^2 - d_2|y_{n+1} - y_n|^2 + d_3\langle x_{n+1} - x_n, y_{n+1} - y_n\rangle. \end{aligned} \tag{17}$$

Clearly, d_2 is a positive constant, provided that $\gamma \in (0,2)$, therefore the previous inequality may be equivalently rewritten as

$$\begin{aligned} & F_{n+1} + \frac{\rho_{n+1}}{2}|x_{n+1} - q|^2 - \left(F_n + \frac{\rho_n}{2}|x_n - q|^2\right) \\ \le\ & -d_2|(y_{n+1} - y_n) - \frac{d_3}{2d_2}(x_{n+1} - x_n)|^2 + \left(d_1 - \frac{\rho_n}{2} + \frac{d_3^2}{4d_2}\right)|x_{n+1} - x_n|^2. \end{aligned} \tag{18}$$

Now, let us define the discrete energy sequence by

$$E_n = F_n + \frac{\rho_n}{2}|x_n - q|^2 = \frac{1}{2}|(q - x_n) + (\alpha x_n + \gamma y_n)|^2 + \frac{\rho_n}{2}|x_n - q|^2 \tag{19}$$

and set

$$k_n = -\left(d_1 - \frac{\rho_n}{2} + \frac{d_3^2}{4d_2}\right). \tag{20}$$

It is then immediate that (18) is equivalent to

$$E_{n+1} - E_n + d_2|(y_{n+1} - y_n) - \frac{d_3}{2d_2}(x_{n+1} - x_n)|^2 + k_n|x_{n+1} - x_n|^2 \leq 0. \tag{21}$$

An elementary computation shows that

$$\begin{aligned} -k_n &= d_1 - \frac{\rho_n}{2} + \frac{d_3^2}{4d_2} \\ &= \alpha^2\left(\frac{1}{2} - \mu\right) + \mu\alpha - \frac{\alpha}{2} - \frac{\gamma}{2}\left(\frac{1}{\theta_n} - \mu\alpha\right) + \frac{(1 + \alpha(\gamma(1-\mu) - 1))^2}{2\gamma(2-\gamma)} \\ &= \frac{\gamma}{2}\left(\mu\alpha - \frac{2}{\gamma}\left(\frac{1}{2} - \mu\right)(\alpha - \alpha^2) + \frac{(1 + \alpha(\gamma(1-\mu) - 1))^2}{\gamma^2(2-\gamma)} - \frac{1}{\theta_n}\right) \end{aligned}$$

and

$$-\rho_n = \gamma\left(\frac{1-\alpha}{\gamma} + \mu\alpha - \frac{1}{\theta_n}\right).$$

It is obviously seen that there exists a positive constant θ such that

$$\frac{1}{\theta} > \mu\alpha + \frac{2}{\gamma}\left(\frac{1}{2} - \mu\right)(\alpha - \alpha^2) + \frac{(1 + \alpha(\gamma(1-\mu) - 1))^2}{\gamma^2(2-\gamma)}$$

and

$$\frac{1}{\theta} > \mu\alpha + \frac{1-\alpha}{\gamma}.$$

Clearly, for $\theta_n \in (0, \theta]$, the two sequences (ρ_n) and (k_n) are then positive. Consequently, (E_n) is a positive and non-increasing sequence, hence (E_n) is converging and it is thus bounded. Therefore, taking into account (21), we immediately obtain $\sum_{n\geq 0} k_n|x_{n+1} - x_n|^2 < \infty$ and $\sum_{n\geq 0} |(y_{n+1} - y_n) - \frac{d_3}{2d_2}(x_{n+1} - x_n)|^2 < \infty$. Noticing that (θ_n) is a converging sequence, because it is non-decreasing and $\theta_n \in (0, \theta]$, this shows that (k_n) converges to a positive real number. We then easily derive that $\sum_{n\geq 0} |x_{n+1} - x_n|^2 < \infty$ and $\sum_{n\geq 0} |y_{n+1} - y_n|^2 < \infty$, hence

$$\lim_{n\to+\infty} |x_{n+1} - x_n| = \lim_{n\to+\infty} |y_{n+1} - y_n| = 0.$$

As a consequence, by (5) we obtain $\lim_{n\to+\infty} |\alpha x_n + \gamma y_n| = 0$. Observing that (ρ_n) is also a converging sequence with $\rho_n > 0$ and bounded away from 0, by (19) we conclude that the sequence $(|x_n - q|)$ is convergent, hence condition (i) of Lemma 2.3 holds with

$S = A^{-1}(0)$. Next, since (x_n) is thus bounded, let $\bar{x}$ be any of its weak cluster points. By passing to the limit in

$$\frac{x_n - x_{n+1}}{\lambda_n} - \frac{\theta_n}{\lambda_n}(\alpha x_n + \gamma y_n) \in A(x_{n+1}),$$

and by taking into account the fact that $\lim_n |x_{n+1} - x_n| = \lim_n |\alpha x_n + \gamma y_n| = 0$ and that the graph of a maximal monotone operator is closed in $\mathcal{H} \times \mathcal{H}$ for the weak-strong topology, see for instance Brézis [9], we deduce that $0 \in A(\bar{x})$. Thus the condition (ii) of Lemma 2.3 is also satisfied, which proves the weak convergence of (x_n) □

It is worth noting that in the case $\gamma = 1$, $\alpha = -1$ and $\mu = 0$ our result improves Proposition 2.1 of [2].

Now we turn our attention to the convex minimization case. Given a convex lower semi-continuous proper function ϕ, our interest is in solving the convex minimization problem $\min_{x\in\mathcal{H}} \phi(x)$ via its optimality condition, namely

$$\text{find } \bar{x} \in \mathcal{H} \text{ such that } 0 \in \partial\phi(\bar{x}). \tag{22}$$

In this context, (DPM) is nothing but:

$$\left[\begin{array}{l} \text{Initialization: } (x_0, y_0) \text{ in } \mathcal{H} \times \mathcal{H}. \\ \text{Step 1: } x_{n+1} := Argmin\{\phi(x) + \dfrac{1}{2\lambda_n}|x - (x_n - \theta_n(\alpha x_n + \gamma y_n))\,|^2\}. \\ \text{Step 2: } y_{n+1} = (1-\gamma)y_n - \alpha(\mu x_{n+1} + (1-\mu)x_n). \end{array}\right. \tag{23}$$

By applying Theorem 2.4, we obtain:

Proposition 2.5. *Let $\gamma \in (0,2)$, $\mu \in [0,1]$, let α be any real number, suppose that $\lambda = \liminf_n \lambda_n > 0$ and that Argmin ϕ, the set of minimizers of ϕ on $\mathcal{H}$, is nonempty. There exists a positive constant θ such that if (θ_n) is a non-decreasing sequence in $(0,\theta)$, then the sequences (x_n) and (y_n) generated by scheme (3) satisfy the following properties:*

(i) $\lim_{n\to+\infty} |x_{n+1} - x_n| = \lim_{n\to+\infty} |y_{n+1} - y_n| = 0$.
(ii) $\lim_{n\to+\infty} \nabla\phi(x_n) = 0$.
(iii) *(x_n) weakly converges to a minimizer of the function ϕ.*

Finally, we would like to emphasize that algorithm (3) can also be used to compute the saddle points of a proper closed convex-concave function L, since finding saddle points of L, namely $(x^*, y^*) \in \mathcal{H}$ verifying $L(x^*, y) \leq L(x^*, y^*) \leq L(x, y^*)$ $\forall (x, y) \in H$, is equivalent to finding the zeroes of the corresponding maximal monotone operator $T_L := \partial_1 L \times \partial_2(-L)$, where $\partial_1 L$ (resp. $\partial_2 L$) stands for the subdifferential of the function L with respect to the first (resp. the second) variable.

Conclusion. The main purpose of this article is to establish the asymptotic convergence of some new implicit iterative methods for solving monotone inclusions. These algorithms, which generalize the classical proximal algorithm [16] and the inertial proximal method [2], are obtained by a discretization of a first order dissipative dynamical system. In the discrete setting and in the more general context of maximal monotone operators, we prove a similar result of that proposed in the continuous case by Alvarez et al. [3].

We did not consider errors, e_k, in the computation of the resolvent and/or replace A by its ε_k-enlargement, for a clearer presentation of the results. It is worth mentioning that the results are still valid for these type of approximate versions under conditions involving summability of the sequences e_k and $(\sqrt{\lambda_k \varepsilon_k})_{k\in\mathbb{N}}$. Indeed, for instance, the definition of the ε-enlargement, $A^\varepsilon(x)$, of the monotone operator A, introduced by [10], namely

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

where $\varepsilon \geq 0$, directly yields that $|J_{\lambda_k}^{A^{\varepsilon_k}}(z) - J_{\lambda_k}^{A}(z)|^2 \leq \lambda_k \varepsilon_k$.
To conclude, we think that the results obtained in this paper may inspire and pave the way for future research in this field, especially in developing new hybrid algorithms which admit less stringent requirements on solving the proximal subproblems in the spirit of Solodov and Svaiter [17] who showed that the tolerance requirements for solving the subproblems can be significantly relaxed if the solving of each subproblem is followed by a projection onto a certain hyperplane which separates the current iterate from the solution set of the problem.

Acknowledgements. The authors thank the anonymous referees for their careful reading of the paper.

References

[1] F. Alvarez: On the minimizing property of a second order dissipative dynamical system in Hilbert spaces, SIAM J. Control Optimization 38(4) (2000) 1102–1119.

[2] F. Alvarez, H. Attouch: An inertial proximal method for monotone operators via discretization of a nonlinear oscillator with damping, Set-Valued Anal. 9 (2001) 3–11.

[3] F. Alvarez, H. Attouch, J. Bolte, P. Redont: A second-order gradient-like dissipative dynamical system with Hessian driven damping. Application to optimization and mechanics, J. Math. Pures Appl., IX. Sér. 81(8) (2002) 747–779.

[4] F. Alvarez, J. M. Pérez: A dynamical system associated with Newton's method for parametric approximations of convex minimization problems, Appl. Math. Optimization 38 (1998) 193–217.

[5] Ph. N. Anh, L. D. Muu, V. H. Nguyen, J.-J. Strodiot: Contraction principle to implement the proximal point method for multivalued monotone variational inequalities, J. Optimization Theory Appl. 124(2) (2005) 285–306.

[6] H. Attouch, X. Goudon, P. Redont: The heavy ball system with friction method. I: The continuous dynamical system, Commun. Contemp. Math. 2(1) (2000) 1–34.

[7] H. Attouch, P. Redont: The second-order in time continuous Newton method, in: Approximation Optimization and Mathematical Economics, M. Lassonde (ed.), Physica, Heidelberg (2001) 25–36.

[8] A. Auslender: Optimisation. Méthodes Numériques, Applicable Mathematics Series, Pitman, London (1976).

[9] H. Brézis: Opérateurs Maximaux Monotones, Math. Stud. 5, North-Holland, Amsterdam (1973).

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

[11] F. Jules, P. E. Maingé: Numerical appraches to stationnary solution of a second order dissipative dynamical system, Optimization 51(2) (2002) 235–255.

[12] B. Lemaire: The proximal algorithm, in: New Methods in Optimization and their Industrial Uses, J.-P. Penot (ed.), Internat. Ser. Numer. Math. 87, Birkhäuser, Basel (1989) 73–87.

[13] A. Moudafi: A perturbed inertial proximal method for maximal monotone operators, Commun. Appl. Nonlinear Anal. 11(3) (2004) 101–108.

[14] A. Moudafi, E. Elisabeth: An approximate inertial proximal method using enlargement of a maximal monotone operator, Int. J. Pure Appl. Math. 5(3) (2003) 283–299.

[15] Z. Opial: Weak convergence of the sequence of successive approximations for nonexpansive mappings, Bull. Amer. Math. Soc. 73 (1967) 591–597.

[16] R. T. Rockafellar: Monotone operators and the proximal point algorithm, SIAM J. Control Optimization 14(5) (1976) 877–898.

[17] M. V. Solodov, B. F. Svaiter: A hybrid projection-proximal point algorithm, J. Convex Analysis 6(1) (1999) 59–70.