SIAM J. OPTIM.
Vol. 14, No. 3, pp. 773–782

WEAK CONVERGENCE OF A RELAXED AND INERTIAL HYBRID PROJECTION-PROXIMAL POINT ALGORITHM FOR MAXIMAL MONOTONE OPERATORS IN HILBERT SPACE*

FELIPE ALVAREZ†

Abstract. This paper introduces a general implicit iterative method for finding zeros of a maximal monotone operator in a Hilbert space which unifies three previously studied strategies: relaxation, inertial type extrapolation and projection step. The first two strategies are intended to speed up the convergence of the standard proximal point algorithm, while the third permits one to perform inexact proximal iterations with fixed relative error tolerance. The paper establishes the global convergence of the method for the weak topology under appropriate assumptions on the algorithm parameters.

Key words. Hilbert space, maximal monotone operator, proximal point, inexact iteration, relative error, separating hyperplane, orthogonal projection, relaxation, weak convergence

AMS subject classifications. 90C25, 65K05, 47J25

DOI. 10.1137/S1052623403427859

1. Introduction.

From now on, $(H, \langle \cdot, \cdot \rangle)$ is a real Hilbert space and the set-valued mapping $A : H \rightrightarrows H$ is a *maximal monotone operator*, that is, A is monotone, i.e., $\forall x, y \in H$, $\forall v \in A(x)$, $\forall w \in A(y)$, $\langle v - w, x - y \rangle \geq 0$, and the graph $GrA = \{(x, v) \in H \times H \mid v \in A(x)\}$ is not properly contained in the graph of any other monotone operator. We are interested in the resolution of the inclusion problem

$$\text{Find } x \in H \text{ such that } 0 \in A(x), \tag{1.1}$$

which appears in a wide variety of equilibrium problems such as convex programming and monotone variational inequalities. This article establishes the asymptotic convergence, for the weak topology, of some implicit iterative methods for solving (1.1) under some implementable inexact conditions. These algorithms, which generalize the classical *Proximal Point Algorithm* (PPA), are of inertial type in the sense that they are obtained by discretization of a second-order-in-time dissipative dynamical system.

Recall that PPA, which was proposed in [15, 16] (inspired by [18]), generates a sequence $(x^k) \subset H$ by the successive approximation scheme

$$x^{k+1} = x^k - \lambda_k v^k, \quad v^k \in A(x^{k+1}), \quad k = 0, 1, \dots,$$

where $(\lambda_k) \subset \mathbb{R}_{++}$ is a sequence of positive regularization parameters. Equivalently,

$$x^{k+1} = J^A_{\lambda_k}(x^k), \tag{PPA}$$

where the single-valued (see [17]) function $J^A_\lambda := (I + \lambda A)^{-1} : H \to H$ is the *resolvent of A of parameter λ*. The resolvent is a nonexpansive mapping and, moreover,

$$J^A_\lambda(x) = x \text{ if and only if } 0 \in A(x). \tag{1.2}$$

*Received by the editors May 12, 2003; accepted for publication (in revised form) September 10, 2003; published electronically January 30, 2004. This work was partially supported by Fondecyt 1020610, ECOS-Conicyt C00E05, and Programa Iniciativa Científica Milenio P01-34.
http://www.siam.org/journals/siopt/14-3/42785.html

†Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático (CNRS UMR 2071), Universidad de Chile, Casilla 170/3, Correo 3, Santiago, Chile (falvarez@dim.uchile.cl).

See [7] for further details. PPA may be viewed as an implicit one-step discretization method for the first-order-in-time differential inclusion $\dot{x}(t) + A(x(t)) \ni 0$, a.e. $t \geq 0$, λ_k being interpreted as a step size parameter. Set $S := A^{-1}(\{0\})$. When $S \neq \emptyset$ and A is demipositive, it is proved in [8] that every solution of this differential inclusion converges weakly in H to a point in S. Concerning PPA, similar convergence results are established in [15, 16] for variational inequalities on bounded sets. The general case is treated in [22], where the equation $x^{k+1} = J^A_{\lambda_k}(x^k)$ is replaced by some inexact criteria, permitting approximate computations of resolvents. See [5] for a counterexample to strong convergence in the continuous case with A being the gradient of a convex function; the same counterexample works for PPA as shown in [13].

To motivate the so-called *Inertial Proximal Point Algorithm* (IPPA), consider the equation for an oscillator with damping and conservative restoring force: $\ddot{x}(t)+\gamma\dot{x}(t)+\nabla f(x(t)) = 0$, where $\gamma > 0$ and $f : H \to \mathbb{R}$ is differentiable. This dynamical system is called *Heavy Ball with Friction* (HBF), and it seems to have been considered for the first time in [21] in the context of optimization problems. The inertial nature of HBF can be exploited in numerical computations in order to accelerate the trajectories and speed up convergence; see [3, 25] for discussions in this direction. Concerning asymptotic convergence, it is proved in [1] that if f is convex (i.e., ∇f is monotone) and $(\nabla f)^{-1}(\{0\}) \neq \emptyset$, then each trajectory of HBF converges weakly in H to some $\hat{x} \in H$ with $\nabla f(\hat{x}) = 0$; see [4] for additional convergence results. Consider the implicit discretization of HBF: $(x^{k+1}-2x^k+x^{k-1})/h^2+\gamma(x^{k+1}-x^k)/h+\nabla f(x^{k+1}) = 0$, which can be rewritten as $x^{k+1} = x^k + \alpha(x^k - x^{k-1}) - \lambda\nabla f(x^{k+1})$, with $\lambda = h^2/(1 + \gamma h)$ and $\alpha = 1/(1 + \gamma h)$. In terms of resolvents, $x^{k+1} = J^{\nabla f}_{\lambda}(x^k + \alpha(x^k - x^{k-1}))$. Note that λ is no longer a step size but is indeed a regularization parameter that combines the damping factor γ and the actual step size $h > 0$.

Replacing ∇f with a maximal monotone operator A, and considering possibly variable parameters $\lambda_k > 0$ and $\alpha_k \in [0, 1)$, the previous discussion motivates the introduction of the inertial type iteration

$$\text{(IPPA)} \qquad x^{k+1} = J^A_{\lambda_k}(x^k + \alpha_k(x^k - x^{k-1})),$$

where the extrapolation term $\alpha_k(x^k - x^{k-1})$ is intended to speed up convergence. IPPA was first considered in [1] for a (nonsmooth) conservative operator $A = \partial f$, the subdifferential of a closed, proper, and convex function $f : H \to \mathbb{R} \cup \{\infty\}$; weak convergence toward a minimizer of f holds under suitable conditions (see [1, Thm. 3.1]). For the nonconservative case, a partial positive result for cocoercive operators is proved in [14], where comparisons with first-order-in-time methods are also given through some numerical tests, showing improvements in the speed of convergence.

The case of an arbitrary maximal monotone operator is treated in [2] under the conditions

$$\lambda := \inf_{k\geq 0} \lambda_k > 0, \tag{1.3}$$

$$\forall k \in \mathbb{N}, \alpha_k \in [0, 1) \quad \text{and} \quad \alpha := \sup_{k\geq 0} \alpha_k < 1, \tag{1.4}$$

$$\sum \alpha_k \|x^k - x^{k-1}\|^2 < \infty. \tag{1.5}$$

Since α_k may be chosen once x^{k-1} and x^k have been found, (1.5) is easy to implement in numerical computations. Furthermore, (1.5) holds automatically in some special situations that can be checked a priori; see, for instance, [1, Thm. 3.1], [2, Prop. 2.1], and Proposition 2.5 below.

From a different point of view, in order to accelerate the standard PPA, the following *Relaxed Proximal Point Algorithm* is proposed in [9] (partially based on [12]):

$$\text{(RPPA)} \qquad x^{k+1} = [(1-\rho_k)I + \rho_k J^A_{\lambda_k}](x^k),$$

where $\rho_k \in (0,2)$ is a *relaxation factor* which is supposed to satisfy

$$R_1 := \inf_{k\geq 0} \rho_k > 0 \text{ and } R_2 := \sup_{k\geq 0} \rho_k < 2. \tag{1.6}$$

The overrelaxation $\rho_k \in (1,2)$ may indeed speed up the convergence of the method; see, for instance, [6, pp. 129–131] and [10]. Weak convergence is proved in [9] for an inexact version of RPPA under a standard summable errors condition.

The first aim of this paper is to show that these two acceleration strategies may be coupled in an iteration of the type

$$\text{(RIPPA)} \qquad x^{k+1} = [(1-\rho_k)I + \rho_k J^A_{\lambda_k}](x^k + \alpha_k(x^k - x^{k-1})),$$

keeping the weak convergence property of the iterates.

On the other hand, from a practical point of view, it is interesting to consider inexact versions of IPPA and RIPPA. Concerning IPPA, a first positive answer is given in [1, Thm. 3.1] for minimization problems, where at each iteration ∂f is replaced with the approximate subdifferential $\partial_{\varepsilon_k} f$, under the hypothesis $\sum \varepsilon_k < \infty$. In this direction, a straightforward adaptation (see, for instance, [19]) of the proof of [2, Thm. 2.1] allows one to deal with the ε_k-enlargement A^{ε_k} of the original operator A. On the other hand, inexact iterations of RPPA are considered in [9], permitting additive residuals in the approximate computation of resolvents under a summability condition analogous to that considered in [22] for PPA. Nevertheless, such inexact criteria requiring summable errors are rather restrictive.

The second goal of this article is to extend the hybrid projection-proximal algorithm introduced in [23] to cover relaxed proximal iterations as RPPA and more generally RIPPA. This hybrid algorithm combines an inexact iteration of PPA with a projection step. In fact, the inexact PPA is used to construct a hyperplane that strictly separates the current iterate x^k from the solution set S; next, x^k is projected onto this separating hyperplane. This method has the remarkable property of permitting a fixed relative error tolerance in the inexact PPA iteration, a less stringent condition, without affecting the global convergence of the algorithm.

This paper is organized as follows. Section 2 introduces an inexact *Relaxed and Inertial Hybrid Projection-Proximal Point Algorithm*, for which weak convergence is proved under conditions (1.3)–(1.6), and then additional conditions on α_k are given in order to ensure (1.5) a priori. Next, a more standard inexact version of RIPPA is considered in section 3, for which weak convergence holds under appropriate summability conditions on the errors.

2. Relaxed and inertial projection-proximal iteration with constant relative error. In what follows, $\sigma \in [0,1)$ is a fixed relative error tolerance. Consider the following iterative scheme:

($\mathcal{A}_1^\rho$) Given $x^k, x^{k-1} \in H, \lambda_k > 0$, $\alpha_k \in [0,1)$, and $\rho_k \in (0,2)$, find $z^k \in H$ such that

$$(z^k - y^k)/\lambda_k + v^k = \eta^k, \text{ for some } v^k \in \rho_k A(z^k/\rho_k + (1-1/\rho_k)y^k), \tag{2.1}$$

where $y^k := x^k + \alpha_k(x^k - x^{k-1})$ and the residual $\eta^k \in H$ satisfies

$$\|\eta^k\| \leq \sigma \max\{\|z^k - y^k\|/\lambda_k, \|v^k\|\}. \tag{2.2}$$

($\mathcal{A}_2^\rho$) If $v^k = 0$ then set $x^n := y^k$ for all $n \geq k+1$ and stop.
Otherwise:

- Let $P_k : H \to H$ be the orthogonal projection operator onto the hyperplane

$$H_k = \{x \in H \mid \langle v^k, x - z^k\rangle = (1 - 1/\rho_k)\langle v^k, y^k - z^k\rangle\}. \tag{2.3}$$

- Set

$$x^{k+1} := y^k + \rho_k(P_k y^k - y^k) = y^k - \frac{\langle v^k, y^k - z^k\rangle}{\|v^k\|^2} v^k. \tag{2.4}$$

- Let $k \leftarrow k+1$ and return to ($\mathcal{A}_1^\rho$).

First, note that (2.1) amounts to $z^k = (1-\rho_k)y^k + \rho_k J^A_{\lambda_k}(y^k + (\lambda_k/\rho_k)\eta^k)$. Indeed, the latter is equivalent to $y^k + (\lambda_k/\rho_k)\eta^k \in (I + \lambda_k A)(z^k/\rho_k + (1 - 1/\rho_k)y^k)$, which can be written as $(y^k - z^k)/\lambda_k + \eta^k \in \rho_k A(z^k/\rho_k + (1 - 1/\rho_k)y^k)$, and this is exactly (2.1). In particular, the algorithm described above is well defined.

Notice that if $\eta^k = 0$, then $x^{k+1} = y^k - \lambda_k v^k = y^k - (y^k - z^k) = (1 - \rho_k)y^k + \rho_k J^A_{\lambda_k}(y^k)$. Therefore, ($\mathcal{A}_1^\rho$)–($\mathcal{A}_2^\rho$) with $\eta^k = 0$ becomes an exact iteration of RIPPA.

Taking $\sigma > 0$, $\alpha_k \equiv 0$, and $\rho_k \equiv 1$, one recovers the *Hybrid Projection-Proximal Point Algorithm* introduced in [23] (see also [24]), whose main feature is the fixed relative error tolerance given by (2.2). Concerning the projection step given by (2.4), this is necessary in general to ensure the boundedness of the iterates (see [23, p. 62]), even for minimization problems (see [11]).

Some elementary, and key, properties of the relative error criterion are summarized in the following lemma.

LEMMA 2.1. *Let $\sigma \in [0, 1)$. If $v = u + \eta$ with $\|\eta\| \leq \sigma \max\{\|u\|, \|v\|\}$, then*
(i) $\|v\| \leq \|u\|/(1 - \sigma)$,
(ii) $\langle v, u\rangle \geq (1 - \sigma)\|u\|\|v\|$.

Proof. Suppose $\|v\| > \|u\|$ so that $\|\eta\| \leq \sigma\|v\|$; then $\|v\| \leq \|u\| + \sigma\|v\|$, or equivalently $\|v\| \leq \|u\|/(1 - \sigma)$; otherwise, $\|v\| \leq \|u\|$. In any case, (i) holds. For (ii), it suffices to consider the case $\|v\| \leq \|u\|$, which implies $\langle v, u\rangle = \|u\|^2 + \langle \eta, u\rangle \geq (1 - \sigma)\|u\|^2 \geq (1 - \sigma)\|u\|\|v\|$. ☐

From (2.1), (2.2), and Lemma 2.1(i), it follows that $v^k = 0$ if and only if $z^k = y^k$. Then, if $v^{k_0} = 0$ for some k_0, then the algorithm ends with y^{k_0} satisfying $0 \in A(y^{k_0})$, a solution to (1.1).

THEOREM 2.2. *Let $(x^k) \subset H$ be a sequence generated by* (2.1)–(2.4)*, where $A : H \rightrightarrows H$ is a maximal monotone operator with $S := A^{-1}(\{0\}) \neq \emptyset$, $\sigma \in [0, 1)$, and the parameters α_k and ρ_k satisfy* (1.4) *and* (1.6)*, respectively. Under* (1.5)*, the following hold:*
(i) *For all $\bar{x} \in S$, $\|x^k - \bar{x}\|$ is convergent, and*

$$\lim_{k\to\infty} \|x^{k+1} - z^k/\rho_k - (1 - 1/\rho_k)y^k\| = 0. \tag{2.5}$$

(ii) *If in addition λ_k satisfies* (1.3)*, then* $\lim_{k\to\infty} \|v^k\| = 0$ *and there exists $x^* \in S$ such that $x^k \rightharpoonup x^*$ weakly in H as $k \to \infty$.*

Proof. From now on, assume that $v^k \neq 0$ for all $k \geq 1$; otherwise, the algorithm finishes in a finite number of iterations, providing a solution to (1.1).

Let $\bar{x} \in S = A^{-1}(\{0\})$ and define $\varphi_k := \frac{1}{2}\|x^k - \bar{x}\|^2$. It follows from (2.4) that

$$\begin{aligned}\varphi_{k+1} &= \tfrac{1}{2}\|y^k - \bar{x}\|^2 + \rho_k\langle P_k y^k - y^k, y^k - \bar{x}\rangle + \tfrac{\rho_k^2}{2}\|P_k y^k - y^k\|^2\\ &= \tfrac{1}{2}\|y^k - \bar{x}\|^2 - \rho_k\|P_k y^k - y^k\|^2 + \rho_k\langle P_k y^k - y^k, P_k y^k - \bar{x}\rangle + \tfrac{\rho_k^2}{2}\|P_k y^k - y^k\|^2\\ &= \tfrac{1}{2}\|y^k - \bar{x}\|^2 - \rho_k(1-\rho_k/2)\|P_k y^k - y^k\|^2 + \rho_k\langle P_k y^k - y^k, P_k y^k - \bar{x}\rangle.\end{aligned}$$

Next, notice that, by Lemma 2.1(i), $v^k \neq 0$ implies $(y^k - z^k)/\lambda_k \neq 0$ due to (2.1) and (2.2). Then, by virtue of Lemma 2.1(ii),

$$\ell_k(y^k) = \langle v^k, y^k - z^k\rangle \geq (1-\sigma)\|v^k\|\|y^k - z^k\| > 0, \tag{2.6}$$

where $\ell_k(x) = \langle v^k, x - z^k\rangle$. As $v^k \in \rho_k A(z^k/\rho_k + (1-1/\rho_k)y^k)$, the monotonicity of A gives $\langle v^k, \bar{x} - z^k/\rho_k - (1-1/\rho_k)y^k\rangle \leq 0$. Thus, $\bar{x}$ belongs to the half-space $H_k^{\leq} = \{x \in H \mid \ell_k(x) \leq (1-1/\rho_k)\ell_k(y^k)\}$. Therefore, since $\rho_k > 0$ and taking into account (2.6), the hyperplane H_k given by (2.3) strictly separates y^k from $\bar{x}$. Moreover, since the orthogonal projection of y^k onto H_k is also the orthogonal projection onto the half-space $H_k^{\leq}$, one gets $\langle P_k y^k - y^k, P_k y^k - \bar{x}\rangle \leq 0$. It follows that

$$\varphi_{k+1} \leq \tfrac{1}{2}\|y^k - \bar{x}\|^2 - \rho_k(1-\rho_k/2)\|P_k y^k - y^k\|^2. \tag{2.7}$$

But $\frac{1}{2}\|y^k - \bar{x}\|^2 = \varphi_k + \alpha_k\langle x^k - \bar{x}, x^k - x^{k-1}\rangle + \frac{\alpha_k^2}{2}\|x^k - x^{k-1}\|^2$. On the other hand, it is direct to verify that $\varphi_k = \varphi_{k-1} + \langle x^k - \bar{x}, x^k - x^{k-1}\rangle - \frac{1}{2}\|x^k - x^{k-1}\|^2$. Hence

$$\tfrac{1}{2}\|y^k - \bar{x}\|^2 = \varphi_k + \alpha_k(\varphi_k - \varphi_{k-1}) + \tfrac{\alpha_k + \alpha_k^2}{2}\|x^k - x^{k-1}\|^2. \tag{2.8}$$

Thus

$$\varphi_{k+1} \leq \varphi_k + \alpha_k(\varphi_k - \varphi_{k-1}) + \delta_k - \rho_k(1-\rho_k/2)\|P_k y^k - y^k\|^2, \tag{2.9}$$

where $\delta_k := \frac{\alpha_k+\alpha_k^2}{2}\|x^k - x^{k-1}\|^2$, which satisfies $\sum \delta_k < \infty$ thanks to (1.5) (recall that $\alpha_k \in [0,1)$). The following elementary result is a useful tool for proving convergence for this type of recursive finite difference inequality (see [1, 2]).

LEMMA 2.3. *Let $\varphi_k \geq 0$ and $\delta_k \geq 0$ be such that $\varphi_{k+1} \leq \varphi_k + \alpha_k(\varphi_k - \varphi_{k-1}) + \delta_k$ with $\sum \delta_k < \infty$, and $0 \leq \alpha_k \leq \alpha < 1$. Then the following hold:*

(i) *$\sum[\varphi_k - \varphi_{k-1}]_+ < \infty$, where $[t]_+ := \max\{t, 0\}$.*

(ii) *There exists $\varphi^* \geq 0$ such that $\lim_{k\to\infty}\varphi_k = \varphi^*$.*

Proof. Set $\theta_k = \varphi_k - \varphi_{k-1}$. Then $[\theta_{k+1}]_+ \leq \alpha[\theta_k]_+ + \delta_k$. This yields $[\theta_{k+1}]_+ \leq \alpha^k[\theta_1]_+ + \sum_{j=0}^{k-1}\alpha^j\delta_{k-j}$, so that $\sum[\theta_{k+1}]_+ \leq 1/(1-\alpha)\left([\theta_1]_+ + \sum\delta_k\right) < \infty$. Set $w_k := \varphi_k - \sum_{j=1}^{k}[\theta_j]_+$, which is bounded from below and nonincreasing. It follows that (w_k) is convergent; hence $\lim_{k\to\infty}\varphi_k = \sum_{j\geq 1}[\theta_j]_+ + \lim_{k\to\infty} w_k$. □

By virtue of Lemma 2.3 applied to (2.9), the sequence (φ_k) is convergent under (1.4) and (1.5). Since $\bar{x} \in S$ is arbitrary, the latter proves the first assertion in Theorem 2.2(i).

On the other hand, by Lemma 2.3(i), it follows from (2.9) that

$$\sum \rho_k(1-\rho_k/2)\|P_k y^k - y^k\|^2 \leq \varphi_1 + \alpha\sum[\varphi_k - \varphi_{k-1}]_+ + \sum\delta_k < \infty,$$

which amounts to

$$(1/R_2 - 1/2)\sum(\langle v^k, y^k - z^k\rangle/\|v^k\|)^2 < \infty, \tag{2.10}$$

with $R_2 = \sup_{k\geq 1}\rho_k < 2$ thanks to (1.6). By Lemma 2.1, it may be concluded from (2.10) that

$$\sum \lambda_k^2\|v^k\|^2 \leq \sum \|y^k - z^k\|^2/(1-\sigma)^2 < \infty. \tag{2.11}$$

It follows that

$$\lim_{k\to\infty}\langle v^k, y^k - z^k\rangle/\|v^k\| = \lim_{k\to\infty}\|y^k - z^k\| = \lim_{k\to\infty}\lambda_k\|v^k\| = 0. \tag{2.12}$$

By (2.4), the first limit in (2.12) ensures that $\lim_{k\to\infty}\|x^{k+1} - y^k\| = 0$. From this fact, together with the second limit in (2.12), it follows that (2.5) holds because $R_1 = \inf \rho_k > 0$ due to (1.6). This completes the proof of Theorem 2.2(i).

In order to prove Theorem 2.2(ii), the idea is to apply the following well-known result on weak convergence in Hilbert spaces, whose proof is given here for the convenience of the reader.

LEMMA 2.4 (Opial). *Let H be a Hilbert space and (x^k) a sequence such that there exists a nonempty set $S \subset H$ verifying the following:*

(a) *For every $\bar{x} \in S$, $\lim_{k\to\infty}\|x^k - \bar{x}\|$ exists.*

(b) *If $x^{k_j} \rightharpoonup \hat{x}$ weakly in H for a subsequence $k_j \to \infty$, then $\hat{x} \in S$.*

Then, there exists $x^* \in S$ such that $x^k \rightharpoonup x^*$ weakly in H as $k \to \infty$.

Proof. It suffices to prove the uniqueness of the weak cluster point. The original proof in [20] requires S to be closed and convex. The following argument (see [15, 22]) does not need that hypothesis. Let $\widehat{x}_1, \widehat{x}_2 \in S$ be two cluster points of (x^k) for the weak topology of H. Set $l_i := \lim_{k\to\infty}\|x^k - \widehat{x}_i\|^2$ for each $i = 1, 2$. Take a sequence $k_j \to \infty$ such that $x^{k_j} \rightharpoonup \widehat{x}_1$ weakly in H. But $\|x^k - \widehat{x}_1\|^2 - \|x^k - \widehat{x}_2\|^2 = \|\widehat{x}_1 - \widehat{x}_2\|^2 + 2\langle\widehat{x}_1 - \widehat{x}_2, \widehat{x}_2 - x^k\rangle$, so that $l_1 - l_2 = -\|\widehat{x}_1 - \widehat{x}_2\|^2$. Similarly, taking $k_m \to \infty$ such that $x^{k_m} \rightharpoonup \widehat{x}_2$, $l_1 - l_2 = \|\widehat{x}_1 - \widehat{x}_2\|^2$. Consequently, $\|\widehat{x}_1 - \widehat{x}_2\| = 0$. □

By Theorem 2.2(i), condition (a) of Lemma 2.4 holds with $S = A^{-1}(\{0\})$. Next, suppose (1.3) and let $\hat{x}$ be a weak cluster point of (x^k). By (2.5), $z^k/\rho_k + (1 - 1/\rho_k)y^k \rightharpoonup \hat{x}$. But

$$v^k/\rho_k \in A(z^k/\rho_k + (1 - 1/\rho_k)y^k), \tag{2.13}$$

with $v^k/\rho_k \to 0$ strongly in H thanks to the last limit in (2.12) together with (1.3) and (1.6). Since the graph of the maximal monotone operator A is closed in $H \times H$ for the weak-strong topology (see [7]), it is possible to pass to the limit in (2.13) to deduce that $0 \in A(\hat{x})$, i.e., $\hat{x} \in S$. Thus, condition (b) of Lemma 2.4 is also satisfied, which proves the weak convergence of (x^k). □

Remark 1. If (1.3) is replaced with

$$\sum \lambda_k^2 = \infty, \tag{2.14}$$

then it may be deduced from (2.11) that there exists a subsequence of (v^k) that converges strongly to 0. In the finite dimensional case, this is sufficient for the convergence of (x^k) (see [23, Rem. 2.3]). Indeed, take $v^{k_i} \to 0$ and assume that $\dim H < \infty$. By

Theorem 2.2(i), (x^k) is bounded so that one may assume that, up to a subsequence, $x^{k_i+1} \to \hat{x}$ for some $\hat{x} \in H$. By virtue of (2.5), one may let $k_i \to \infty$ in (2.13) to deduce that $0 \in A(\hat{x})$. Hence $\hat{x} \in S$ and, by Theorem 2.2(i), $\|x^k - \hat{x}\|$ is convergent. Therefore $\lim_{k\to\infty} \|x^k - \hat{x}\| = \lim_{i\to\infty} \|x^{k_i+1} - \hat{x}\| = 0$.

In practical computations, it is easy to enforce (1.5) by means of a dynamic rule to update the inertial parameter α_k, taking into account the current value of $\|x^k - x^{k-1}\|$. Furthermore, (1.5) holds a priori in some special cases as the next result shows, extending [2, Prop. 2.1].

PROPOSITION 2.5. *Under the assumptions of Theorem* 2.2 *with, in addition,* (α_k) *being nondecreasing (i.e.,* $\alpha_{k+1} \geq \alpha_k$*) and satisfying* $0 \leq \alpha_k \leq \alpha$ *for some* $\alpha \in [0, 1)$ *such that*

$$0 < p(\alpha) := 1/R_2 - 1/2 - (2/R_1 - 1/2)\alpha - (1 - 1/R_2)\alpha^2, \tag{2.15}$$

then $\sum \|x^k - x^{k-1}\|^2 < \infty$*. In particular,* (1.5) *holds and thus there exists* $\widehat{x} \in S$ *such that* $x^k \rightharpoonup \widehat{x}$ *weakly in* H *as* $k \to \infty$*.*

Proof. Noticing that $\rho_k^2\|P_k y^k - y^k\|^2 = \|x^{k+1} - y^k\|^2 = \|x^{k+1} - x^k\|^2 - 2\alpha_k\langle x^{k+1} - x^k, x^k - x^{k-1}\rangle + \alpha_k^2\|x^k - x^{k-1}\|^2 \geq (1 - \alpha_k)\|x^{k+1} - x^k\|^2 - \alpha_k(1 - \alpha_k)\|x^k - x^{k-1}\|^2$, it follows from (2.9) that

$$\begin{aligned} \varphi_{k+1} \leq \varphi_k + \alpha_k(\varphi_k - \varphi_{k-1}) + [(\alpha_k + \alpha_k^2)/2 + (1/\rho_k - 1/2)\alpha_k(1 - \alpha_k)]\|x^k - x^{k-1}\|^2 \\ - (1/\rho_k - 1/2)(1 - \alpha_k)\|x^{k+1} - x^k\|^2, \end{aligned}$$

where $\varphi_k := \frac{1}{2}\|x^k - \bar{x}\|^2$. This yields

$$\begin{aligned} \varphi_{k+1} - \alpha_k\varphi_k \leq \varphi_k - \alpha_k\varphi_{k-1} + \alpha_k[1/\rho_k + (1 - 1/\rho_k)\alpha_k]\|x^k - x^{k-1}\|^2 \\ - (1/\rho_k - 1/2)(1 - \alpha_k)\|x^{k+1} - x^k\|^2. \end{aligned}$$

Setting $\mu_k := \varphi_k - \alpha_k\varphi_{k-1} + \alpha_k[1/\rho_k + (1 - 1/\rho_k)\alpha_k]\|x^k - x^{k-1}\|^2$, and since $\alpha_{k+1} \geq \alpha_k$, we obtain

$$\mu_{k+1} \leq \mu_k + [\alpha_{k+1}/\rho_{k+1} + (1 - 1/\rho_{k+1})\alpha_{k+1}^2 + (1/\rho_k - 1/2)(\alpha_k - 1)]\|x^{k+1} - x^k\|^2.$$

But $\alpha_{k+1}/\rho_{k+1} + (1 - 1/\rho_{k+1})\alpha_{k+1}^2 \leq \alpha/R_1 + (1 - 1/R_2)\alpha^2$ and $(1/\rho_k - 1/2)(\alpha_k - 1) \leq (1/R_1 - 1/2)\alpha - 1/R_2 + 1/2$. Therefore $\mu_{k+1} \leq \mu_k - p(\alpha)\|x^{k+1} - x^k\|^2$, where $p(\alpha)$ is given by (2.15). Since $p(\alpha) > 0$, (μ_k) is nonincreasing, which implies $\varphi_k \leq \alpha\varphi_{k-1} + \mu_k \leq \alpha\varphi_{k-1} + \mu_1$. This gives $\varphi_k \leq \alpha^k\varphi_0 + \mu_1\sum_{j=0}^{k-1}\alpha^j \leq \alpha^k\varphi_0 + \mu_1/(1 - \alpha)$. Furthermore, it follows that $p(\alpha)\sum_{j=0}^{k}\|x^{j+1} - x^j\|^2 \leq \mu_1 - \mu_{k+1} \leq \mu_1 + \alpha\varphi_k \leq \alpha^{k+1}\varphi_0 + \mu_1/(1 - \alpha)$. This shows that $\sum \|x^k - x^{k-1}\|^2 \leq 2\mu_1/((1 - \alpha)p(\alpha))$. The conclusion follows by Theorem 2.2. ◻

Remark 2. Suppose $R_2 \geq 1$. Since $p(0) = 1/R_2 - 1/2 > 0$ thanks to (1.6), there exists a unique positive root $\alpha^* > 0$ of the quadratic polynomial $p(\alpha)$, and for all $\alpha \in [0, \alpha^*)$, $p(\alpha) > 0$. For instance, when $\rho_k \equiv 1$, one gets $p(\alpha) = 1/2 - (3/2)\alpha$ and so $\alpha^* = 1/3$.

3. An alternative inexact scheme with summable residuals. It follows from (2.2) and (2.11) that the sequence of residuals (η^k) associated with the sequence (x^k) generated by (2.1)–(2.4) satisfies $\sum \lambda_k^2\|\eta^k\|^2 < \infty$, and hence $\sum \|\eta^k\|^2 < \infty$ in view of (1.3). However, it may occur that $\sum \|\eta^k\| = \infty$; see [11] for an example with $\rho_k \equiv 1$ and $\alpha_k \equiv 0$, which is based on [13]. The constant relative error criterion (2.2) is thus less stringent than

$$\sum \lambda_k\|\eta^k\| < \infty. \tag{3.1}$$

On the other hand, the next result, which extends [9, Thm. 3], shows that under such a summability condition the projection step is not necessary for convergence.

THEOREM 3.1. *Let* $A : H \rightrightarrows H$ *be a maximal monotone operator with* $S := A^{-1}(\{0\}) \neq \emptyset$ *and* $(x^k) \subset H$ *a sequence satisfying*

$$(3.2)\quad (x^{k+1} - y^k)/\lambda_k + v^k = \eta^k \quad \text{for some } v^k \in \rho_k A(x^{k+1}/\rho_k + (1 - 1/\rho_k)y^k),$$

where $y^k = x^k + \alpha_k(x^k - x^{k-1})$, *and the parameters* λ_k, α_k, *and* ρ_k *satisfy* (1.3), (1.4), *and* (1.6), *respectively. Suppose* (1.5), (3.1), *and*

$$(3.3)\qquad \sum \lambda_k \|\eta^k\| \|y^k\| < \infty.$$

Then $v^k \to 0$ *strongly in* H *and there exists* $x^* \in S$ *such that* $x^k \rightharpoonup x^*$ *weakly in* H.

Proof. It is easy to see that (3.2) amounts to $x^{k+1} = (1 - \rho_k)y^k + \rho_k J^A_{\lambda_k}(y^k + (\lambda_k/\rho_k)\eta^k)$. Let (w^k) be the auxiliary sequence defined by

$$(3.4)\qquad w^k := (1 - \rho_k)y^k + \rho_k J^A_{\lambda_k}(y^k).$$

Since J^A_λ is nonexpansive,

$$(3.5)\qquad \|x^{k+1} - w^k\| \leq \lambda_k \|\eta^k\|.$$

On the other hand, (3.4) may be written as $w^k = y^k - \lambda_k \rho_k A_{\lambda_k}(y^k)$, where $A_\lambda : H \to H$ is given by $A_\lambda = \frac{1}{\lambda}(I - J^A_\lambda)$. Thanks to (1.2),

$$(3.6)\qquad 0 \in A(x) \text{ if and only if } A_\lambda(x) = 0.$$

Moreover, as J^A_λ is nonexpansive, A_λ is a cocoercive maximal monotone operator of parameter λ; that is,

$$(3.7)\qquad \forall x_1, x_2 \in H, \langle A_\lambda(x_1) - A_\lambda(x_2), x_1 - x_2 \rangle \geq \lambda \|A_\lambda(x_1) - A_\lambda(x_2)\|^2.$$

Let $\bar{x} \in S$. By (3.6), $A_\lambda(\bar{x}) = 0$ and since $\frac{1}{2}\|w^k - \bar{x}\|^2 = \frac{1}{2}\|y^k - \bar{x}\|^2 - \rho_k\lambda_k\langle y^k - \bar{x}, A_{\lambda_k}(y^k)\rangle + \frac{(\rho_k\lambda_k)^2}{2}\|A_{\lambda_k}(y^k)\|^2$, the cocoercivity property (3.7) yields

$$(3.8)\qquad \tfrac{1}{2}\|w^k - \bar{x}\|^2 \leq \tfrac{1}{2}\|y^k - \bar{x}\|^2 - \lambda_k^2 \rho_k (1 - \rho_k/2)\|A_{\lambda_k}(y^k)\|^2.$$

Define $\varphi_k := \frac{1}{2}\|x^k - \bar{x}\|^2$. Then $\varphi_{k+1} \leq \frac{1}{2}\|w^k - \bar{x}\|^2 + \|x^{k+1} - w^k\|\|w^k - \bar{x}\| + \frac{1}{2}\|x^{k+1} - w^k\|^2$. By (3.5) and (3.8),

$$(3.9)\quad \varphi_{k+1} \leq \tfrac{1}{2}\|y^k - \bar{x}\|^2 - \lambda_k^2\rho_k(1 - \rho_k/2)\|A_{\lambda_k}(y^k)\|^2 + \lambda_k\|\eta^k\|\|y^k - \bar{x}\| + \tfrac{\lambda_k^2}{2}\|\eta^k\|^2.$$

Recalling (2.8), it follows that

$$(3.10)\qquad \varphi_{k+1} \leq \varphi_k + \alpha_k(\varphi_k - \varphi_{k-1}) + \delta_k - \lambda_k^2\rho_k(1 - \rho_k/2)\|A_{\lambda_k}(y^k)\|^2,$$

where $\delta_k := \frac{\alpha_k + \alpha_k^2}{2}\|x^k - x^{k-1}\|^2 + \lambda_k\|\eta^k\|\|y^k - \bar{x}\| + \frac{\lambda_k^2}{2}\|\eta^k\|^2$. Under (1.5), (3.1), if (3.3) holds, then $\sum \delta_k < \infty$. Thus $\sum \delta_k < \infty$ and, by virtue of Lemma 2.3, (φ_k) is convergent. Moreover, we deduce that $\sum \lambda_k^2 \|A_{\lambda_k}(y^k)\|^2 < \infty$. Set $\xi^k := y^k - J^A_{\lambda_k}(y^k)$, which amounts to

$$(3.11)\qquad \xi^k/\lambda_k \in A(y^k - \xi^k).$$

Since $\sum \|\xi^k\|^2 < \infty$, in particular $\lim_{k\to\infty} \xi^k = 0$. Let $\hat{x}$ be a weak cluster point of (x^k). Since $\lim_{k\to\infty} \alpha_k \|x^k - x^{k-1}\| = 0$, $y^k \rightharpoonup \hat{x}$ and consequently $y^k - \xi^k \rightharpoonup \hat{x}$. By the weak-strong closedness of the graph of A, letting $k \to \infty$ in (3.11) gives $0 \in A(\hat{x})$. Therefore, condition (b) of Lemma 2.4 holds, which finishes the proof. ∎

Remark 3. Under (1.5) and (3.1), assume

$$\sum \alpha_k \|x^k - x^{k-1}\| < \infty. \tag{3.12}$$

From (3.9), it follows that $\|x^{k+1} - \bar{x}\| \le \|y^k - \bar{x}\| + \lambda_k \|\eta^k\| \le \|x^k - \bar{x}\| + \alpha_k \|x^k - x^{k-1}\| + \lambda_k \|\eta^k\|$. Using (3.1) and (3.12), $\|x^k - \bar{x}\|$ is convergent; in particular (y^k) is bounded. Hence in view of (3.1), condition (3.3) is realized.

Acknowledgments. The author thanks the hospitality of the Sydoco research team of INRIA-Rocquencourt (France) where part of this work was carried out. The author also wishes to thank the anonymous referee whose remarks helped him to improve the presentation of this paper.

REFERENCES

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

[2] F. Alvarez and H. Attouch, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, in Wellposedness in Optimization and Related Topics (Gargnano, 1999), Set-Valued Anal., 9 (2001), pp. 3–11.

[3] A. S. Antipin, *Minimization of convex functions on convex sets by means of differential equations*, Differential Equations, 30 (1994), pp. 1365–1375.

[4] H. Attouch, X. Goudou, and P. Redont, *The heavy ball with friction method.* I. *The continuous dynamical system*, Commun. Contemp. Math., 2 (2000), pp. 1–34.

[5] J. B. Baillon, *Un exemple concernant le comportement asymptotique de la solution du problème* $du/dt + \partial\varphi(u) \ni 0$, J. Funct. Anal., 28 (1978), pp. 369–376.

[6] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

[7] H. Brezis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, Mathematics Studies 5, North–Holland, Amsterdam, 1973.

[8] R. E. Bruck, *Asymptotic convergence of nonlinear contraction semigroups in Hilbert space*, J. Funct. Anal., 18 (1975), pp. 15–26.

[9] J. Eckstein and D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293–318.

[10] J. Eckstein and M. C. Ferris, *Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control*, INFORMS J. Comput., 10 (1998), pp. 218–235.

[11] O. R. Gárciga, A. Iusem, and B. F. Svaiter, *On the need for hybrid steps in hybrid proximal point methods*, Oper. Res. Lett., 29 (2001), pp. 217–220.

[12] E. G. Gol'shtein and N. V. Tret'yakov, *Modified Lagrangians in convex programming and their generalizations*, Math. Program. Stud., 10 (1979), pp. 86–97.

[13] O. Güler, *On the convergence of the proximal point algorithm for convex minimization*, SIAM J. Control Optim., 29 (1991), pp. 403–419.

[14] F. Jules and P. E. Maingé, *Numerical approach to a stationary solution of a second order dissipative dynamical system*, Optimization, 51 (2002), pp. 235–255.

[15] B. Martinet, *Régularisation d'inéquations variationnelles par approximations successives*, Rev. Française Informat. et Recherche Opérationnelle, 4 (1970), pp. 154–158.

[16] B. Martinet, *Determination approchée d'un point fixe d'une application pseudo-contractante*, C. R. Acad. Sci. Paris Ser. A-B, 274 (1972), pp. 163–165.

[17] G. Minty, *Monotone (nonlinear) operators in Hilbert space*, Duke Math. J., 29 (1962), pp. 341–346.

[18] J. J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.

[19] A. Moudafi, *Second-order differential proximal methods for equilibrium problems*, JIPAM J. Inequal. Pure Appl. Math., 14 (2003).

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

[21] B. T. Polyak, *Some methods of speeding up the convergence of iterative methods*, Zh. Vychisl. Mat. Mat. Fiz., 4 (1964), pp. 1–17.

[22] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898.

[23] M. V. Solodov and B. F. Svaiter, *A hybrid projection-proximal point algorithm*, J. Convex Anal., 6 (1999), pp. 59–70.

[24] M. V. Solodov and B. F. Svaiter, *Forcing strong convergence of proximal point iterations in a Hilbert space*, Math. Program., 87 (2000), pp. 189–202.

[25] F. Zirilli, F. Aluffi and V. Parisi, *DAFNE: A Differential Equations Algorithm for Nonlinear Equations*, ACM Trans. Math. Software, 10 (1984), pp. 317–324.