

M.V. Solodov · B.F. Svaiter

## Forcing strong convergence of proximal point iterations in a Hilbert space

Received January 6, 1998 / Revised version received August 9, 1999

Published online November 30, 1999

**Abstract.** This paper concerns with convergence properties of the classical proximal point algorithm for finding zeroes of maximal monotone operators in an infinite-dimensional Hilbert space. It is well known that the proximal point algorithm converges weakly to a solution under very mild assumptions. However, it was shown by Güler [11] that the iterates may fail to converge strongly in the infinite-dimensional case. We propose a new proximal-type algorithm which does converge strongly, provided the problem has a solution. Moreover, our algorithm solves proximal point subproblems inexactly, with a constructive stopping criterion introduced in [31]. Strong convergence is forced by combining proximal point iterations with simple projection steps onto intersection of two halfspaces containing the solution set. Additional cost of this extra projection step is essentially negligible since it amounts, at most, to solving a linear system of two equations in two unknowns.

**Key words.** proximal point algorithm – Hilbert spaces – weak convergence – strong convergence

### 1. Introduction

Consider the problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in T(x), \quad (1)$$

where  $\mathcal{H}$  is a real Hilbert space, and  $T(\cdot)$  is a maximal monotone operator (or a multifunction) on  $\mathcal{H}$ . We shall denote the solution set of this problem by

$$S := \{x \in \mathcal{H} \mid 0 \in T(x)\}.$$

One of the classical algorithms for solving (1) is the proximal point method, introduced by Martinet [18] and further developed by Rockafellar [24]. Some other relevant papers on the proximal point method and its applications and modifications are [20, 23, 5, 21, 17, 10, 12, 9, 4, 6, 31, 29, 27]. This list is by no means exhaustive, we refer the reader to [16] for a survey. Specifically, having  $x^k \in \mathcal{H}$ , a current approximation to the solution of (1), the proximal point method generates the next iterate  $x^{k+1}$  by solving the proximal subproblem

$$0 \in T(x) + \mu_k(x - x^k), \quad (2)$$

where  $\mu_k > 0$  is a regularization parameter. If the sequence  $\{\mu_k\}$  is chosen bounded from above, then the resulting sequence  $\{x^k\}$  of proximal point iterates converges *weakly* to

M.V. Solodov, B.F. Svaiter: Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil. e-mail: solodov@impa.br, benar@impa.br

*Mathematics Subject Classification (1991):* 49M45, 90C25, 90C33

an element of  $S$ , provided this set is nonempty, see [24]. In [24], Rockafellar also posed an open question whether (or not) the proximal point method always converges *strongly*. This question was resolved in the negative by Güler [11], who exhibited a proper closed convex function  $f$  in an infinite-dimensional Hilbert space  $l^2$ , for which the proximal point algorithm (in our framework, procedure (2) with  $T = \partial f$ ) converges weakly but not strongly. Naturally, the question arises whether the proximal point method can be modified, preferably in a simple way, so that strong convergence is guaranteed.

In this paper, we develop a strongly convergent algorithm by combining proximal point iterations with certain (computationally simple) projection steps (see Algorithm 1). Moreover, we allow the proximal point subproblems to be solved inexactly, and our tolerance requirements are less restrictive and more constructive than in the classical setting of [24] (this is discussed in detail in Sect. 2). Our approach is closely related to the hybrid projection-proximal point method presented in [31]. The principal idea of [31] is to use (approximate) solutions of proximal point subproblems to construct certain hyperplanes which separate the iterates from the solution set  $S$ , if it is nonempty. Convergence is forced by adding simple projection steps onto associated halfspaces.

Of course, weak and strong convergence are only distinguishable in the infinite-dimensional setting. On the other hand, even when we have to solve infinite-dimensional problems, numerical implementations of algorithms are certainly applied to finite-dimensional approximations of the problems. Nevertheless, it is important to have convergence theory for the infinite-dimensional case, because it guarantees robustness and stability with respect to discretization schemes employed for obtaining finite-dimensional approximations of infinite-dimensional problems. This issue is closely related to the so-called *Mesh Independence Principle* [2, 1, 15]. This principle relies on infinite-dimensional convergence to predict the convergence properties of a discretized finite-dimensional method. Furthermore, the mesh independence provides theoretical justification for the design of refinement strategies. Note that fine discretization is crucial for the obtained discrete solution to be an appropriate approximation to the true solution of the infinite-dimensional problem being solved. We note that many real-world problems in economics and engineering are modeled in the infinite-dimensional spaces. These include the optimal control and structural design problems, and the problem of minimal area surface with obstacles, among others. We refer the reader to [13, 14], where a variety of applications are described.

The rest of the paper is organized as follows. In Sect. 2, the concept of inexact solutions of proximal subproblems with associated error tolerance is formally defined, and relevant properties of such inexact solutions are established. In Sect. 3, the algorithm is formally stated, and its components are discussed. Convergence results are presented in Sect. 4. Section 5 contains some concluding remarks.

We briefly describe our notation. The inner product in  $\mathcal{H}$  is denoted by  $\langle \cdot, \cdot \rangle$ . The norm determined by the inner product is denoted by  $\| \cdot \|$ . For a nonempty closed convex set  $A \subseteq \mathcal{H}$  and an element  $x \in \mathcal{H}$ , the orthogonal projection of  $x$  onto  $A$ , i.e.,  $\arg \min\{\|y - x\| \mid y \in A\}$ , is denoted by  $P_A(x)$ . By  $\text{ri}(A)$  we denote the relative interior of  $A$ .

Finally, we state the following well known properties of the projection operator to be used in the sequel.

**Lemma 1.** [32] *Let  $A$  be any nonempty closed convex set in  $\mathcal{H}$ . For any  $x, y \in \mathcal{H}$  and any  $z \in A$ , the following properties hold :*

$$\begin{aligned} \langle x - P_A(x), z - P_A(x) \rangle &\leq 0; \\ \|P_A(x) - P_A(y)\|^2 &\leq \|x - y\|^2 - \|P_A(x) - x + y - P_A(y)\|^2. \end{aligned}$$

## 2. Inexact proximal point iterations

Because solving the proximal subproblem (2) exactly can be computationally as difficult (or almost as difficult) as solving the original problem (1) itself, of particular importance is the case when the subproblems are solved only approximately, that is we find a pair  $y^k \in \mathcal{H}$  and  $v^k \in T(y^k)$  such that

$$\varepsilon^k = v^k + \mu_k(y^k - x^k), \quad (3)$$

where  $\varepsilon^k$  is the error associated with inexact solution of subproblem (2). In the pure inexact proximal point method, one then sets

$$x^{k+1} := y^k$$

to obtain the next iterate.

The subject of appropriate approximation criteria for the inexact solution of proximal point subproblems is very important, and it has been addressed, for example, in [24, 6, 8]. Typically, the following (or conceptually similar) conditions are used to ensure convergence of the iterates:

$$\|\varepsilon^k\| \leq \sigma_k \mu_k, \quad \sum_{k=0}^{\infty} \sigma_k < \infty,$$

or

$$\|\varepsilon^k\| \leq \sigma_k \mu_k \|y^k - x^k\|, \quad \sum_{k=0}^{\infty} \sigma_k < \infty.$$

The first condition above is used to ensure global convergence, while the second condition (together with certain additional assumptions) implies local linear rate of convergence. Note that a *summable* sequence  $\{\sigma_k\}$  essentially has to be chosen a priori, which makes a constructive choice difficult in most situations. Also note that under the second condition, the *relative error* in (3) satisfies

$$\frac{\|\varepsilon^k\|}{\mu_k \|y^k - x^k\|} \leq \sigma_k, \quad \sum_{k=0}^{\infty} \sigma_k < \infty.$$

So the relative error must be summable, and hence, tend to zero. A simple example in [31] shows that if the tolerance parameters  $\sigma_k$  are fixed at a nonzero value, convergence of the inexact proximal point iterations cannot be guaranteed even in the finite-dimensional case.

A proximal-point-type method with a more realistic and constructive error tolerance has recently been proposed in [31]. In particular, for the method of [31], the relative error in (3) can be fixed. In other words, we accept  $y^k \in \mathcal{H}$  and  $v^k \in T(y^k)$  as a satisfactory approximate solution in (3) whenever *any one* of the following two conditions is met:

$$\frac{\|\varepsilon^k\|}{\mu_k \|y^k - x^k\|} \leq \sigma \quad \text{or} \quad \frac{\|\varepsilon^k\|}{\|v^k\|} \leq \sigma,$$

where  $\sigma \in [0, 1)$ . Note that here  $\sigma$  does not depend on the iteration index  $k$  (although, in principle, it certainly need not be fixed; the important issue is that it can be kept bounded away from zero). The key fact is that under this condition, the hyperplane

$$H_k := \{x \in \mathcal{H} \mid \langle v^k, x - y^k \rangle = 0\}$$

strictly separates the current iterate  $x^k$  from the solution set  $S$  (we assume here that it is nonempty). Convergence (in the weak topology) can then be guaranteed if we obtain the next iterate  $x^{k+1}$  as the orthogonal projection of  $x^k$  onto  $H_k$ :

$$x^{k+1} = P_{H_k}(x^k).$$

This hybrid method retains all the attractive convergence properties of the proximal point algorithm while allowing a fixed relative error tolerance. We refer the reader to [31] for complete analysis. An important point to note is that projection onto a hyperplane can be carried out explicitly, so it does not entail any significant additional computational cost. We also note that this projection step is indispensable in the sense that if one sets instead  $x^{k+1} := y^k$ , where  $y^k$  is computed with fixed relative error tolerance  $\sigma \in (0, 1)$ , then the iterates of the resulting *pure* inexact proximal method may fail to converge even in  $\mathfrak{R}^2$  (see an example in [31]). For an extension of the described notion of approximate solution to the setting of generalized proximal algorithms based on the Bregman functions, we refer the reader to [29]. Another extension using certain enlargements (outer approximations) of the operator defining the problem can be found in [28]. Using also the linesearch technique of [26], the framework of [31, 28] led to the development of *truly* globally convergent inexact Newton methods for monotone equations [30] and complementarity problems [25].

However, by itself, the method of [31] does not attain the goal of the present paper. Like the classical proximal method, it may converge weakly but not strongly in an infinite-dimensional space. This is easy to see because with  $\varepsilon^k = 0$ , the hybrid projection-proximal point method reduces to the standard proximal point method (again, see [31]). Hence, the counter example of Güler [11] applies. Still, the separation and projection methodology of [31] will appear very useful for devising a strongly convergent algorithm.

As already mentioned, the method proposed here will use inexact proximal iterates to set up appropriate projection steps. We proceed to study the properties of inexact solutions of proximal subproblems, which will be used in the sequel.

**Definition 1.** Let  $x \in \mathcal{H}$ ,  $\mu > 0$  and  $\sigma \in [0, 1)$ . We say that a pair  $(y, v) \in \mathcal{H} \times \mathcal{H}$  is an inexact solution with tolerance  $\sigma$  of  $0 \in T(\cdot) + \mu(\cdot - x)$ , if

$$\begin{aligned} v &\in T(y), \\ v + \mu(y - x) &= \varepsilon, \end{aligned}$$

and

$$\|\varepsilon\| \leq \sigma \max \{\|v\|, \mu\|y - x\|\}.$$

Next we establish some nice properties of inexact solutions defined above.

**Proposition 1.** *Let  $x \in \mathcal{H}$ ,  $\mu > 0$  and  $\sigma \in [0, 1)$ , and suppose that  $(y, v)$  is an inexact solution of  $0 \in T(\cdot) + \mu(\cdot - x)$  with tolerance  $\sigma$ . Then it holds that*

$$\langle x - y, v \rangle \geq (1 - \sigma) \max\{\mu\|x - y\|^2, \|v\|^2/\mu\} \geq (1 - \sigma)\|v\|\|x - y\|. \quad (4)$$

Define

$$H := \{z \in \mathcal{H} \mid \langle z - y, v \rangle \leq 0\}.$$

Then the following four statements are equivalent:

$$\begin{aligned} x &\in H; \\ y &= x; \\ v &= 0; \\ x &\text{ is a solution of (1)}. \end{aligned}$$

Furthermore,

$$\|P_H(x) - x\| \geq (1 - \sigma) \max\{\|x - y\|, \|v\|/\mu\}. \quad (5)$$

*Proof.* To prove (4), we consider the two possible cases:

$$\mu\|x - y\| \leq \|v\| \quad \text{and} \quad \mu\|x - y\| \geq \|v\|.$$

In the first case, we have that

$$\|\varepsilon\| \leq \sigma\|v\|.$$

It follows that in this case

$$\begin{aligned} \langle x - y, v \rangle &= \frac{1}{\mu} \langle v - \varepsilon, v \rangle \\ &\geq \frac{1 - \sigma}{\mu} \|v\|^2. \end{aligned} \quad (6)$$

Furthermore,

$$\begin{aligned} \langle x - y, v \rangle &\geq \frac{1 - \sigma}{\mu} \|v\|^2 \\ &\geq (1 - \sigma)\|v\|\|y - x\| \\ &\geq \mu(1 - \sigma)\|y - x\|^2. \end{aligned} \quad (7)$$

Combining (6) and (7), we conclude that (4) holds in the first case.

Consider now the second case. We have that

$$\|\varepsilon\| \leq \sigma\mu\|y - x\|.$$

Hence,

$$\begin{aligned} \langle x - y, v \rangle &= \langle x - y, \mu(x - y) + \varepsilon \rangle \\ &\geq \mu(1 - \sigma)\|x - y\|^2. \end{aligned} \quad (8)$$

Furthermore, in this case,

$$\begin{aligned} \langle x - y, v \rangle &\geq \mu(1 - \sigma)\|x - y\|^2 \\ &\geq (1 - \sigma)\|v\|\|y - x\| \\ &\geq \frac{1 - \sigma}{\mu}\|v\|^2. \end{aligned} \quad (9)$$

Again, combining (8) and (9), we obtain that (4) is satisfied in the second case also.

Next, we establish the equivalence of the four conditions. Suppose that  $x \in H$ . Then  $\langle x - y, v \rangle \leq 0$  and, by (4),  $x = y$ . If  $x = y$ , then  $\langle x - y, v \rangle = 0$  and, again by (4),  $v = 0$ . By the same reasoning, if  $v = 0$  then  $x = y$  and  $x$  is a solution of (1). Finally, if  $x$  is a solution (i.e.,  $0 \in T(x)$ ) then, by the monotonicity of  $T$ ,

$$0 \leq \langle y - x, v - 0 \rangle = \langle y - x, v \rangle,$$

and so  $x \in H$ .

Finally, to prove (5), note that if  $x \in H$ , then  $x = y$ ,  $v = 0$  and (5) holds trivially. The other case is  $x \notin H$  (and so  $v \neq 0$ ). Then

$$P_H(x) = x - \frac{\langle v, x - y \rangle}{\|v\|^2}v.$$

Hence,

$$\|P_H(x) - x\| = \frac{\langle v, x - y \rangle}{\|v\|}.$$

If  $\|v\|/\mu \geq \|x - y\|$  then  $\|v\|^2/\mu \geq \mu\|x - y\|^2$ , and (5) follows from the first inequality in (4). If  $\|v\|/\mu \leq \|x - y\|$  then (5) follows from the second inequality in (4). This completes the proof.  $\square$

### 3. The algorithm

We are now ready to formally state our algorithm.

**Algorithm 1.** Choose any  $x^0 \in \mathcal{H}$  and  $\sigma \in [0, 1)$ . At iteration  $k$ , having  $x^k$ , choose

$\mu_k > 0$  and find  $(y^k, v^k)$ , an inexact solution of

$$0 \in T(x) + \mu_k(x - x^k)$$

with tolerance  $\sigma$ . Define

$$H_k = \{z \in \mathcal{H} \mid \langle z - y^k, v^k \rangle \leq 0\},$$

and

$$W_k = \{z \in \mathcal{H} \mid \langle z - x^k, x^0 - x^k \rangle \leq 0\}.$$

Take

$$x^{k+1} = P_{H_k \cap W_k}(x^0).$$

□

Note that at each iteration, there are two subproblems to be solved: find an inexact solution of the proximal point subproblem and find the projection of  $x^0$  onto  $H_k \cap W_k$ , the intersection of two halfspaces. The proximal subproblem always has an exact solution, which is unique (this is a classical result of Minty [19]). Computing an approximate solution only makes things easier. So this part of the method is well-defined (but it is worth to note that since some error is allowed, in general the pair  $(y^k, v^k)$  is not uniquely defined). Regarding the projection step, we shall prove that the set  $H_k \cap W_k$  is never empty, even when the solution set  $S$  is empty. Therefore the whole algorithm is well-defined in the sense that it generates an infinite sequence  $\{x^k\}$  (and an associated sequence of pairs  $\{(y^k, v^k)\}$ ). Also note that starting with the same  $x^0$ , many different sequences satisfying the conditions of Algorithm 1 can be generated. Finally, note that from Proposition 1 it follows that at some iteration  $k'$ ,  $x^{k'} \in S$  if and only if  $v^{k'} = 0$ . So, in that case, for all  $k \geq k'$  we will have  $x^k = x^{k'}$  and  $v^k = 0$ .

Some remarks are in order regarding complexity of the projection step in Algorithm 1. Even though we are working in an infinite-dimensional space, projection onto an intersection of two halfspaces amounts to solving, at most, a linear system of two equations with two unknowns. Suppose that at iteration  $k$ ,  $W_k \cap H_k$  is nonempty (this fact will be formally established in Sect. 4). Then  $x^{k+1}$  is well defined and is characterized as the solution of

$$\begin{aligned} \min_z \quad & \|z - x^0\|^2 \\ \text{s.t.} \quad & \langle z - y^k, v^k \rangle \leq 0, \\ & \langle z - x^k, x^0 - x^k \rangle \leq 0. \end{aligned}$$

Let us express  $z - x^0$  as a linear combination of  $v^k$  and  $x^0 - x^k$  plus a vector orthogonal to  $v^k$  and  $x^0 - x^k$ :

$$\begin{aligned} z - x^0 &= \lambda_1 v^k + \lambda_2 (x^0 - x^k) + h, \\ \langle h, v^k \rangle &= 0, \quad \langle h, x^0 - x^k \rangle = 0. \end{aligned}$$

So, the problem above becomes

$$\begin{aligned} \min_{\lambda_1, \lambda_2, h} \quad & \|h\|^2 + \|\lambda_1 v^k + \lambda_2 (x^0 - x^k)\|^2 \\ \text{s.t.} \quad & \langle \lambda_1 v^k + \lambda_2 (x^0 - x^k) + x^0 - y^k, v^k \rangle \leq 0, \\ & \langle \lambda_1 v^k + \lambda_2 (x^0 - x^k) + x^0 - x^k, x^0 - x^k \rangle \leq 0. \end{aligned}$$

Due to the special structure of this problem, at the solution we obviously have  $h = 0$ . Therefore  $\lambda_1$  and  $\lambda_2$  are obtained solving a two-dimensional quadratic minimization problem with two linear inequality constraints.

Furthermore, it is easy to see that if the projection of  $x^0$  onto  $H_k$

$$P_{H_k}(x^0) = x^0 - \frac{\langle v^k, x^0 - y^k \rangle}{\|v^k\|^2} v^k$$

belongs to  $W_k$  then

$$P_{H_k}(x^0) = P_{H_k \cap W_k}(x^0).$$

So in this case, to obtain  $x^{k+1}$  there is no need for any further computation. Otherwise, we have

$$P_{H_k \cap W_k}(x^0) = x^0 + \lambda_1 v^k + \lambda_2 (x^0 - x^k),$$

where  $\lambda_1, \lambda_2$  is the solution of the linear system of two equations with two unknowns:

$$\begin{aligned} \lambda_1 \|v^k\|^2 + \lambda_2 \langle v^k, x^0 - x^k \rangle &= -\langle x^0 - y^k, v^k \rangle \\ \lambda_1 \langle v^k, x^0 - x^k \rangle + \lambda_2 \|x^0 - x^k\|^2 &= -\|x^0 - x^k\|^2. \end{aligned}$$

Therefore, we could actually write an explicit formulae for obtaining  $x^{k+1}$ , which means that the cost of the projection step in Algorithm 1 is essentially negligible.

The preceding analysis shows that the cost of each iteration of Algorithm 1 is about the same as that of an iteration of the inexact proximal point method (actually, of the modified projection-proximal method of [31], since we adopt the relaxed approximation criterion introduced in that reference). Thus we obtain a strongly rather than weakly convergent proximal-type algorithm at essentially no additional cost per iteration. We proceed to establish strong convergence of the method.

#### 4. Convergence analysis

We start with establishing some properties of the algorithm which hold regardless of whether the problem has solutions or not. Let  $x^k, y^k, v^k, k = 0, 1, \dots$  be variables generated by Algorithm 1. For now, they may be defined either for all iteration indices  $k$  or only up to some index  $k_0$ .

**Proposition 2.** *Suppose that Algorithm 1 reaches an iteration  $k + 1$ . Then it holds that*

$$\|x^{k+1} - x^0\|^2 \geq \|x^k - x^0\|^2 + \|x^{k+1} - x^k\|^2, \quad (10)$$

and

$$\|x^{k+1} - x^k\| \geq (1 - \sigma) \max\{\|y^k - x^k\|, \|v^k\|/\mu_k\}. \quad (11)$$



*Proof.* By the definition of  $W_k$ , it is clear that  $x^k$  is the projection of  $x^0$  onto  $W_k$ . Applying Lemma 1 with  $A = W_k$ ,  $x = x^{k+1}$  and  $y = x^0$ , we obtain

$$\|P_{W_k}(x^{k+1}) - P_{W_k}(x^0)\|^2 \leq \|x^{k+1} - x^0\|^2 - \|P_{W_k}(x^{k+1}) - x^{k+1} + x^0 - P_{W_k}(x^0)\|^2.$$

Because  $x^{k+1} \in W_k$ , we have that  $P_{W_k}(x^{k+1}) = x^{k+1}$ . Furthermore,  $P_{W_k}(x^0) = x^k$ . Hence,

$$\|x^{k+1} - x^0\|^2 \geq \|x^k - x^0\|^2 + \|x^{k+1} - x^k\|^2.$$

Since  $x^{k+1} \in H_k$ , it holds that

$$\|x^{k+1} - x^k\| \geq \|x^k - P_{H_k}(x^k)\|.$$

Now (11) follows from Proposition 1.  $\square$

As a consequence of Proposition 2, we immediately obtain the following result.

**Corollary 1.** *Suppose the sequence of regularization parameters  $\{\mu_k\}$  is bounded from above, and Algorithm 1 generates an infinite sequence  $\{x^k\}$ . Then either  $\{x^k\}$  is bounded and each of its weak accumulation points belongs to  $S \neq \emptyset$ , or  $S = \emptyset$  and  $\lim_{k \rightarrow \infty} \|x^k\| = \infty$ .*

*Proof.* Applying (10) consecutively, we obtain

$$\|x^k - x^0\|^2 \geq \sum_{j=0}^{k-1} \|x^{j+1} - x^j\|^2.$$

If  $\{x^k\}$  is bounded, letting  $k \rightarrow \infty$  yields that

$$\sum_{j=0}^{\infty} \|x^{j+1} - x^j\|^2 < \infty,$$

and hence,

$$0 = \lim_{k \rightarrow \infty} \|x^{k+1} - x^k\|.$$

By (11) and boundedness of  $\{\mu_k\}$  from above, we further obtain that

$$\lim_{k \rightarrow \infty} \|y^k - x^k\| = 0, \quad (12)$$

$$\lim_{k \rightarrow \infty} \|v^k\| = 0. \quad (13)$$

Since the sequence  $\{x^k\}$  is bounded, it must have weak accumulation points. Let  $\bar{x}$  be any weak accumulation point of  $\{x^k\}$ , and take any subsequence  $\{x^{k_j}\}$  weakly convergent to it. By (12),  $\{y^{k_j}\}$  has the same weak limit  $\bar{x}$ . Since  $v^k \in T(y^k)$  with  $v^k \rightarrow 0$  strongly, the maximal monotonicity of  $T$  implies that  $0 \in T(\bar{x})$  (the argument to show this is standard). That is,  $\bar{x} \in S$ .

Suppose now that  $S = \emptyset$ . By the preceding assertion, the sequence  $\{x^k\}$  is unbounded in this case. Since, by (10), the sequence  $\{\|x^k - x^0\|\}$  is nondecreasing, it follows that  $\|x^k - x^0\| \rightarrow \infty$  as  $k \rightarrow \infty$ , and so  $\|x^k\| \rightarrow \infty$ .  $\square$

Next, we shall prove well-definedness and strong convergence of  $\{x^k\}$  to a solution in the case  $S \neq \emptyset$ . After that, we shall establish well-definedness of  $\{x^k\}$  in the case  $S = \emptyset$ . In the latter case, we obtain that  $\|x^k\| \rightarrow \infty$ , which is somewhat stronger than simple unboundedness of  $\{x^k\}$ .

#### 4.1. The case $S \neq \emptyset$

In this subsection we assume that (1) has a solution, so  $S \neq \emptyset$ . Having chosen the initial iterate  $x^0$ , let us define the following set which will be important for the subsequent analysis:

$$U(x^0) = \{x \in \mathcal{H} \mid \forall z \in S, \langle z - x, x^0 - x \rangle \leq 0\}. \quad (14)$$

We next show that the set  $H_k \cap W_k$  always contains the solution set  $S$ . Hence, it is nonempty and the projection step in Algorithm 1 is well-defined. Furthermore, we show that the generated sequence of iterates  $\{x^k\}$  is contained in the set  $U(x^0)$ .

**Proposition 3.** *Suppose that Algorithm 1 reaches an iteration  $k$  and  $x^k \in U(x^0)$ . Then it holds that*

1.  $S \subseteq H_k \cap W_k$ .
2.  $x^{k+1}$  is well-defined and  $x^{k+1} \in U(x^0)$ .

*Proof.* First note that by the monotonicity of  $T$ ,

$$\langle v^k, y^k - x^* \rangle \geq 0$$

for any  $x^* \in S$ . Hence,  $S \subseteq H_k$ . Since  $x^k \in U(x^0)$ , we also have that

$$\langle x^* - x^k, x^0 - x^k \rangle \leq 0 \quad \forall x^* \in S.$$

By the definition of  $W_k$ , we obtain that  $S \subseteq W_k$ . We conclude that  $S \subseteq H_k \cap W_k$ .

In particular, it follows that  $H_k \cap W_k \neq \emptyset$ , and the next iterate  $x^{k+1}$  is well-defined. Because  $x^{k+1}$  is the projection of  $x^0$  onto  $H_k \cap W_k$ , by Lemma 1 we obtain that

$$\langle z - x^{k+1}, x^0 - x^{k+1} \rangle \leq 0 \quad \forall z \in H_k \cap W_k.$$

Since  $S \subseteq H_k \cap W_k$ , the above relation certainly holds for all  $z \in S$ , which implies that  $x^{k+1} \in U(x^0)$ , by the definition given in (14). □

It can be now easily verified that the whole algorithm is well-defined.

**Corollary 2.** *Algorithm 1 is well-defined and generates infinite sequences  $\{x^k\}$  and  $\{y^k\}, \{v^k\}$  such that  $x^k \in U(x^0)$  and  $S \subseteq H_k \cap W_k$  for all  $k$ . Furthermore, if  $\{\mu_k\}$  is bounded from above, then  $\{x^k\}$  is bounded and each of its weak accumulation points belongs to  $S$ .*

*Proof.* It is enough to observe that  $x^0 \in U(x^0)$  and apply Proposition 3 and induction on  $k$ . Corollary 1 further implies that  $\{x^k\}$  is bounded with weak accumulation points being in  $S$ .

□

We are now ready to prove strong convergence of any sequence  $\{x^k\}$  generated by Algorithm 1 starting from  $x^0$ , to the solution of (1) which is closest to  $x^0$ .

**Theorem 1.** *Let  $\{x^k\}$  be a sequence generated by Algorithm 1, and suppose that the sequence of regularization parameters  $\{\mu_k\}$  is bounded from above. Then  $\{x^k\}$  converges strongly to  $x^* = P_S(x^0)$ .*

*Proof.* Note that  $x^*$ , the projection of  $x^0$  onto  $S$ , exists because the solution set  $S$  is closed, convex, and we assumed it to be nonempty. By the definition of  $x^{k+1}$ , we have that

$$\|x^{k+1} - x^0\| \leq \|z - x^0\| \quad \forall z \in H_k \cap W_k.$$

Since  $x^* \in S \subseteq H_k \cap W_k$ , it follows that for all  $k$

$$\|x^k - x^0\| \leq \|x^* - x^0\|. \quad (15)$$

By Corollary 2, we already know that  $\{x^k\}$  is bounded with all of its weak accumulation points belonging to  $S$ . Let  $\{x^{k_j}\}$  be any weakly convergent subsequence of  $\{x^k\}$ , and let  $\bar{x} \in S$  be its weak limit. Observe that

$$\begin{aligned} \|x^{k_j} - x^*\|^2 &= \|x^{k_j} - x^0 - (x^* - x^0)\|^2 \\ &= \|x^{k_j} - x^0\|^2 + \|x^* - x^0\|^2 \\ &\quad - 2\langle x^{k_j} - x^0, x^* - x^0 \rangle \\ &\leq 2\|x^* - x^0\|^2 - 2\langle x^{k_j} - x^0, x^* - x^0 \rangle, \end{aligned}$$

where the inequality follows from (15). By the weak convergence of  $\{x^{k_j}\}$  to  $\bar{x}$ , we then obtain

$$\limsup_{j \rightarrow \infty} \|x^{k_j} - x^*\|^2 \leq 2(\|x^* - x^0\|^2 - \langle \bar{x} - x^0, x^* - x^0 \rangle). \quad (16)$$

Applying Lemma 1 with  $A = S$ ,  $x = x^0$  and  $z = \bar{x} \in S$ , and taking into account that  $x^*$  is the projection of  $x^0$  onto  $S$ , we have that

$$\langle x^0 - x^*, \bar{x} - x^* \rangle \leq 0.$$

Now from the latter relation it follows that

$$\|x^* - x^0\|^2 \leq \langle \bar{x} - x^0, x^* - x^0 \rangle.$$

Combining the last inequality with (16), we conclude that  $\{x^{k_j}\}$  converges *strongly* to  $x^*$ . Clearly  $\bar{x} = x^*$  because  $\bar{x}$  is a weak limit of  $\{x^{k_j}\}$ .

Since  $\bar{x}$  was taken as an *arbitrary* weak accumulation point of  $\{x^k\}$ , it follows that  $x^*$  is the *unique* weak accumulation point of this sequence. Since  $\{x^k\}$  is bounded, the whole sequence  $\{x^k\}$  weakly converges to  $x^*$ . On the other hand, we have shown that every weakly convergent subsequence of  $\{x^k\}$  converges strongly to  $x^*$ . Hence, the whole sequence  $\{x^k\}$  converges strongly to  $x^* \in S$ .

□

#### 4.2. The case $S = \emptyset$

In this subsection we assume that no solution exists. We shall prove that the sequence  $\{x^k\}$  is still defined for every  $k$ , and that it diverges.

**Theorem 2.** *If  $S = \emptyset$  then Algorithm 1 generates an infinite sequence  $\{x^k\}$ . If, in addition, the sequence of regularization parameters  $\{\mu_k\}$  is bounded from above, then  $\lim_{k \rightarrow \infty} \|x^k\| = \infty$ .*

*Proof.* In view of Corollary 1, we only have to establish well-definedness of the algorithm. We argue by induction. We first check that the initial iteration  $k = 0$  is well defined. The problem  $0 \in T(x) + \mu_0(x - x^0)$  always has the exact solution, and hence some inexact solution  $(y^0, v^0)$ . Now note that  $W_0 = \mathcal{H}$ . Since  $H_0$  cannot be empty, the next iterate  $x^1$  is generated, which is the projection of  $x^0$  onto  $H_0 = W_0 \cap H_0$ .

Note that whenever  $x^k$  is generated,  $y^k$  and  $v^k$  can further be obtained because the proximal subproblems always have (in)exact solutions. Suppose that  $x^k, (y^k, v^k)$  are defined for  $k = 0, \dots, \hat{k}$ . It is enough to prove that  $x^{\hat{k}+1}$  is well-defined. Take any

$$z^0 \in \text{ri}(D(T)),$$

where  $D(T)$  is the domain of  $T$ , and define

$$\rho = \max\{\|y^k - z^0\| \mid k = 0, \dots, \hat{k}\}$$

and

$$h(x) = \begin{cases} 0 & \text{if } \|x - z^0\| \leq \rho + 1, \\ +\infty & \text{otherwise.} \end{cases}$$

Then  $h : \mathcal{H} \rightarrow R \cup \{+\infty\}$  is a lower semicontinuous proper convex function, its subgradient  $\partial h$  is maximal monotone, and

$$T' = T + \partial h$$

is also maximal monotone [22]. Furthermore,

$$T'(z) = T(z) \quad \text{if } \|z - z^0\| < \rho + 1.$$

Therefore,  $v^k \in T'(y^k)$  for  $k = 0, \dots, \hat{k}$ . We conclude that  $x^k, (y^k, v^k)$  also satisfy the conditions of Algorithm 1 applied to the problem

$$0 \in T'(x).$$

Since  $T'$  has a bounded domain, the above problem has solutions. Using the two preceding observations and Corollary 2, it follows that  $x^{\hat{k}+1}$  is well-defined, and hence, so is the whole algorithm.

The conclusion now follows using Corollary 1. □

## 5. Concluding remarks

We presented a proximal-type algorithm for finding zeroes of maximal monotone operators, which converges strongly in an infinite-dimensional Hilbert space to a solution, provided it exists. In this sense, this is a significant theoretical improvement over the classical proximal point method which converges weakly but not strongly. Strong convergence of our method is forced by adding a simple projection step which amounts to solving, at most, a linear system of two equations with two unknowns. Furthermore, a constructive error tolerance is used in the approximate solution of proximal subproblems. Finally, we mention the recent work [7, 3], which has some interesting connections to the algorithm described in this paper.

*Acknowledgements.* We thank an anonymous referee for a constructive suggestion which helped us to somewhat simplify the analysis and improve the presentation.

Research of the first author is supported by CNPq Grant 300734/95-6 and by PRONEX–Optimization, research of the second author is supported by CNPq Grant 301200/93-9(RN) and by PRONEX–Optimization.

## References

1. Allgower, E.L., Böhmer, K. (1987): Application of the mesh-independence principle to mesh refinement strategies. *SIAM J. Numer. Anal.* **24**, 1335–1351
2. Allgower, E.L., Böhmer, K., Potra, F.-A., Rheinboldt, W.C. (1986): A mesh-independence principle for operator equations and their discretizations. *SIAM J. Numer. Anal.* **23**, 160–169
3. Bauschke, H.H., Combettes, P.L. (1999): A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces. *Math. Oper. Res.* submitted 1999
4. Bonnans, J.F., Gilbert, J.C., Lemaréchal, C., Sagastizábal, C. (1995): A family of variable metric proximal point methods. *Math. Program.* **68**, 15–47
5. Brézis, H., Lions, P.L. (1978): Produits infinis de résolvantes. *Isr. J. Math.* **29**, 329–345
6. Burke, J.V., Qian, M. (1998): A variable metric proximal point algorithm for monotone operators. *SIAM J. Control Optim.* **37**, 353–375
7. Combettes, P.L. (1999): Strong convergence of block-iterative outer approximation methods for convex optimization. *SIAM J. Control Optim.* To appear
8. Eckstein, J. (1998): Approximate iterations in Bregman-function-based proximal algorithms. *Math. Program.* **83**, 113–123
9. Eckstein, J., Bertsekas, D.P. (1992): On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.* **55**, 293–318
10. Ferris, M.C. (1991): Finite termination of the proximal point algorithm. *Math. Program.* **50**, 359–366
11. Güler, O. (1991): On the convergence of the proximal point algorithm for convex minimization. *SIAM J. Control Optim.* **29**, 403–419
12. Güler, O. (1992): New proximal point algorithms for convex minimization. *SIAM J. Optim.* **2**, 649–664
13. Henry, J., Yvon, J.-P., eds. (1994): *Lecture notes in Control and Information Sciences No. 197, System Modelling and Optimization*. Springer, Berlin
14. Kinderlehrer, D., Stampacchia, G. (1980): *An introduction to variational inequalities and their applications*. Academic Press, New York
15. Laumen, M. (1999): Newton’s mesh independence principle for a class of optimal shape design problems. *SIAM J. Control Optim.* **37**, 1070–1088
16. Lemaire, B. (1989): The proximal algorithm. In: Penot, J.P., ed., *New Methods of Optimization and Their Industrial Use*. International Series of Numerical Mathematics 87, pp. 73–87. Birkhauser, Basel
17. Luque, F.J. (1984): Asymptotic convergence analysis of the proximal point algorithm. *SIAM J. Control Optim.* **22**, 277–293
18. Martinet, B. (1970): Regularisation d’inéquations variationnelles par approximations successives. *Revue Française d’Informatique et de Recherche Opérationnelle* **4**, 154–159
19. Minty, G.J. (1962): Monotone (nonlinear) operators in Hilbert space. *Duke Math. J.* **29**, 341–346
20. Moreau, J.-J. (1965): Proximité et dualité dans un espace Hilbertien. *Bull. Soc. Math. Fr.* **93**, 273–299

21. Passty, G.B. (1979): Weak convergence theorems for nonexpansive mappings in Banach spaces. *J. Math. Anal. Appl.* **67**, 274–276
22. Rockafellar, R.T. (1970): On the maximality of sums of nonlinear monotone operators. *Trans. Am. Math. Soc.* **149**, 75–88
23. Rockafellar, R.T. (1976): Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.* **1**, 97–116
24. Rockafellar, R.T. (1976): Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **14**, 877–898
25. Solodov, M.V., Svaiter, B.F.: A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem. *SIAM J. Optim.* To appear
26. Solodov, M.V., Svaiter, B.F. (1999): A new projection method for variational inequality problems. *SIAM J. Control Optim.* **37**, 765–776
27. Solodov, M.V., Svaiter, B.F. (1999): A comparison of rates of convergence of two inexact proximal point algorithms. In: Di Pillo, G., Giannessi, F., eds., *Nonlinear Optimization and Related Topics*. Kluwer Academic Publishers. To appear
28. Solodov, M.V., Svaiter, B.F. (1998): A hybrid approximate extragradient–proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Analysis*. To appear
29. Solodov, M.V., Svaiter, B.F.: An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. *Math. Oper. Res.* To appear
30. Solodov, M.V., Svaiter, B.F. (1999): A globally convergent inexact Newton method for systems of monotone equations. In: Fukushima, M., Qi, L., eds., *Reformulation – Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, pp. 355–369. Kluwer Academic Publishers
31. Solodov, M.V., Svaiter, B.F. (1999): A hybrid projection – proximal point algorithm. *J. Convex Anal.* **6**, 59–70
32. Zarantonello, E.H. (1971): Projections on convex sets in Hilbert space and spectral theory. In: Zarantonello, E.H., ed., *Contributions to Nonlinear Functional Analysis*, pp. 237–424. Academic Press, New York