Journal of Convex Analysis
Volume 6 (1999), No. 1, 59–70

A Hybrid Projection-Proximal Point Algorithm*

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,benar*}*@impa.br*

Received February 6, 1997

We propose a modification of the classical proximal point algorithm for finding zeroes of a maximal monotone operator in a Hilbert space. In particular, an approximate proximal point iteration is used to construct a hyperplane which strictly separates the current iterate from the solution set of the problem. This step is then followed by a projection of the current iterate onto the separating hyperplane. All information required for this projection operation is readily available at the end of the approximate proximal step, and therefore this projection entails no additional computational cost. The new algorithm allows significant relaxation of tolerance requirements imposed on the solution of proximal point subproblems, which yields a more practical framework. Weak global convergence and local linear rate of convergence are established under suitable assumptions. Additionally, presented analysis yields an alternative proof of convergence for the exact proximal point method, which allows a nice geometric interpretation, and is somewhat more intuitive than the classical proof.

Keywords: Maximal monotone operators, proximal point methods, projection methods

1991 Mathematics Subject Classification: 90C25, 49J45, 49M45

1. Introduction

We consider the problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in T(x), \tag{1.1}$$

where $\mathcal{H}$ is a real Hilbert space, and $T(\cdot)$ is a maximal monotone operator (or a multifunction) on $\mathcal{H}$. A wide variety of problems, such as optimization and mini-max problems, complementarity problems and variational inequalities, fall within this general framework.

Having $x^i \in \mathcal{H}$, a current approximation to the solution of (1.1), the proximal point algorithm generates the next iterate by solving the subproblem

$$0 \in T(x) + \mu_i(x - x^i), \tag{1.2}$$

where $\mu_i > 0$ is a regularization parameter. Because solving (1.2) exactly can be as difficult (or almost as difficult) as solving the original problem (1.1) itself, of practical relevance is the case when the subproblems are solved approximately, that is

$$\text{find } x^{i+1} \in \mathcal{H} \text{ such that } 0 = v^{i+1} + \mu_i(x^{i+1} - x^i) + \varepsilon^i, \quad v^{i+1} \in T(x^{i+1}), \tag{1.3}$$

*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.

ISSN 0944-6532 / $ 2.50

where $\varepsilon^i \in \mathcal{H}$ is an error associated with inexact solution of subproblem (1.2). In [19], the following approximation criteria for the solution of proximal point subproblems (1.2) were used:

$$\|\varepsilon^i\| \le \sigma_i \mu_i, \quad \sum_{i=0}^{\infty} \sigma_i < \infty, \tag{1.4}$$

and

$$\|\varepsilon^i\| \le \sigma_i \mu_i \|x^{i+1} - x^i\|, \quad \sum_{i=0}^{\infty} \sigma_i < \infty. \tag{1.5}$$

The first criterion was needed to establish global convergence of the proximal point algorithm, while the second was required for local linear rate of convergence result [19, Theorems 1 and 2] (it was also assumed that the sequence $\{\mu_i\}$ is uniformly bounded from above). Similar criteria were used, for example, in [2]. Proximal point methods have been studied extensively (e.g. [14, 18, 19, 15, 12, 11, 6, 7, 5], see [8] for a survey). It has long been realized however that in many applications, proximal point methods in the classical form are not very efficient. Developments aimed at speeding up convergence of proximal and related methods can be found in [1, 4, 9, 10, 13, 17, 2, 3]. Most of the work just cited is focused on the variable metric approach and other ways of incorporating second order information to achieve faster convergence. As it has been remarked in [19] and [2], for a proximal point method to be practical, it is also important that it should work with approximate solutions of subproblems. It is therefore worthwhile to develop new algorithms which admit less stringent requirements on solving the subproblems. There have been some advances in this direction in the context of optimization problems (for example, [1, 25]). In the general case of solving operator equations, the situation seems to be more complicated, and we are not aware of any previous work in this direction. In this paper, we propose one such algorithm. In particular, we show that the tolerance requirements for solving the proximal point subproblems can be significantly relaxed if solving each subproblem is followed by a projection onto a certain hyperplane which separates the current iterate from the solution set of the problem. We emphasize that the new method retains all the attractive convergence properties of the classical proximal point algorithm.

Our approach is based on the interpretation of the exact proximal point algorithm, given by (1.3) with $\varepsilon^i = 0$, as a certain projection-type method. In particular, an exact proximal point step can be shown to be equivalent to projecting the current iterate x^i onto a certain hyperplane which separates x^i from the solution set of (1.1). Indeed, by monotonicity of $T(\cdot)$, it follows that

$$0 \ge \langle v^{i+1}, \bar{x} - x^{i+1} \rangle$$

for any $\bar{x}$ such that $0 \in T(\bar{x})$. On the other hand, setting $\varepsilon^i = 0$ in (1.3), we have

$$v^{i+1} = -\mu_i (x^{i+1} - x^i) \in T(x^{i+1}),$$

and hence

$$\langle v^{i+1}, x^i - x^{i+1} \rangle = \mu_i \|x^i - x^{i+1}\|^2 > 0$$

(the right-hand-side is zero only when $x^i = x^{i+1}$, in which case x^i is a solution of the problem). It follows that if x^i is not a solution, then the hyperplane

$$H := \{x \in \mathcal{H} \mid \langle v^{i+1}, x - x^{i+1} \rangle = 0\}$$

strictly separates x^i from the solution set of (1.1). Moreover, it is not difficult to check that

$$P_H[x^i] = x^i - \frac{\langle v^{i+1}, x^i - x^{i+1} \rangle}{\|v^{i+1}\|^2} v^{i+1} = x^{i+1},$$

that is x^{i+1} is the projection of x^i onto the separating hyperplane H.

Our principal idea is the following. We will use *inexact* proximal iteration (1.3) to construct an appropriate hyperplane which strictly separates the current iterate from the solutions of the problem. Then the next iterate is obtained as a (computationally trivial) projection of the current iterate onto this hyperplane. By separation arguments, it can be shown that the new iterate is closer to the solution set than the previous one, which essentially entails global convergence of the algorithm. A similar idea has also been used in [20] to devise new projection methods for solving variational inequality problems with a continuous (pseudo)monotone operator. It turns out that conditions needed to be imposed on tolerances ε^i in (1.3) in order to construct an appropriate separating hyperplane, are significantly weaker than conditions needed for the inexact proximal point iterations to converge (see also an example below). Fortunately, whenever such a separating hyperplane is available, convergence can be forced by adding a simple projection step.

Specifically, we propose the following method.

Algorithm 1.1. *Choose any $x^0 \in \mathcal{H}$ and $\sigma \in [0, 1)$. Having x^i, choose $\mu_i > 0$ and*

$$\text{find } y^i \in \mathcal{H} \text{ such that } 0 = v^i + \mu_i(y^i - x^i) + \varepsilon^i, \quad v^i \in T(y^i), \tag{1.6}$$

where

$$\|\varepsilon^i\| \le \sigma \max\{\|v^i\|, \mu_i\|y^i - x^i\|\}. \tag{1.7}$$

Stop if $v^i = 0$ or $y^i = x^i$. Otherwise, let

$$x^{i+1} = x^i - \frac{\langle v^i, x^i - y^i \rangle}{\|v^i\|^2} v^i. \tag{1.8}$$

Note that (1.8) is equivalent to

$$x^{i+1} = P_{H_i}[x^i], \tag{1.9}$$

where

$$H_i := \{x \in \mathcal{H} \mid \langle v^i, x - y^i \rangle = 0\}. \tag{1.10}$$

Observe that if $\sigma = 0$, Algorithm 1.1 reduces to the classical *exact* proximal point method. This is because $\sigma = 0$ implies that $\varepsilon^i = 0$ (i.e., the subproblems are solved exactly), and consequently $x^{i+1} = y^i$ (see the interpretation of the exact proximal point algorithm as a projection method discussed above). For $\sigma \neq 0$ however, Algorithm 1.1 is different from other inexact proximal-based methods in the literature.

We point out that our tolerance condition (1.7) is less restrictive than the corresponding tolerance requirements in [19, 2] given by (1.4) or (1.5), which are standard in the literature. For one thing, the right-hand-side of (1.7) involves the *largest* of two quantities, $\|v^i\|$ and $\mu_i\|y^i - x^i\|$. And, more importantly, the relaxation parameter σ can be fixed, while in the classical setting it has to be summable, and, hence, tend to zero. Condition (1.7) is sufficient for both, global convergence and local linear rate of convergence results (see Theorems 2.2 and 2.4).

We next show, by exhibiting an example, that the projection step (1.8) is essential to the convergence of Algorithm 1.1. In particular, we show that the iterates generated by inexact proximal point algorithm, which can be viewed as Algorithm 1.1 with (1.8) replaced by $x^{i+1} = y^i$, need not converge.

Let $\mathcal{H} = \Re^2$ and $T(x) = Ax$, where

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Then (1.1) becomes a system of linear equations which has the unique solution $\bar{x} = (0,0)^\top$. Suppose $x^i = (1,0)^\top$ and $\mu_i = 1$. It is easy to check that the point $y^i = (0,1)^\top$ is admissible according to (1.6)–(1.7). Indeed,

$$0 = T(y^i) + \mu_i(y^i - x^i) + \varepsilon^i = (1,0)^\top + (-1,1)^\top + \varepsilon^i$$

with $\varepsilon^i = (0,-1)^\top$. So we have that $\|\varepsilon^i\| = 1$ while $\|y^i - x^i\| = \sqrt{2}$. So (1.7) is satisfied with, say, $\sigma = 1/\sqrt{2}$. But $\|y^i\| = \|x^i\| = 1$ so no progress is being made towards $\bar{x}$, the solution of the problem, if we accept y^i as the next iterate x^{i+1}. In fact, taking $\sigma \in (1/\sqrt{2}, 1)$, it is even possible to construct a sequence satisfying (1.6)–(1.7) and $x^{i+1} := y^i$, such that $\|x^i\| \to \infty$ as $i \to \infty$. On the other hand, if the projection step (1.8) is added after y^i is computed, we obtain the solution in the example under consideration in just one step, because $0 = P_{H_i}[x^i]$.

2. Convergence Analysis

We start with a simple lemma.

Lemma 2.1. *Let $x, y, v, \bar{x}$ be any elements of $\mathcal{H}$ such that*

$$\langle v, x - y\rangle > 0, \quad \textit{and} \quad \langle v, \bar{x} - y\rangle \leq 0.$$

Let $z = P_H[x]$, where

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

Then

$$\|z - \bar{x}\|^2 \leq \|x - \bar{x}\|^2 - \left(\frac{\langle v, x - y\rangle}{\|v\|}\right)^2.$$

Proof. It follows from the hypothesis that H separates x from $\bar{x}$. Moreover, $z = P_H[x]$ is also the projection of x onto the halfspace $\{s \in \mathcal{H} \mid \langle v, s - y\rangle \le 0\}$. Since $\bar{x}$ belongs to this halfspace, it follows from the basic properties of the projection operator (see, for example, [16, p.121]) that $\langle x - z, z - \bar{x}\rangle \ge 0$. Therefore

$$\begin{aligned} \|x - \bar{x}\|^2 &= \|x - z\|^2 + \|z - \bar{x}\|^2 + 2\langle x - z, z - \bar{x}\rangle \\ &\ge \|x - z\|^2 + \|z - \bar{x}\|^2. \end{aligned}$$

Now the result follows from the relation

$$z = P_H[x] = x - \frac{\langle v, x - y\rangle}{\|v\|^2} v.$$

□

We are now ready to prove our main global convergence result. Throughout we assume that the solution set of the problem is nonempty.

Theorem 2.2. *Let $\{x^i\}$ be any sequence generated by* Algorithm 1.1. *Then the following statements hold:*

(i) *The sequence $\{x^i\}$ is bounded, and the sequence $\{x^i - y^i\}$ converges strongly to zero.*
(ii) *If $\sum_{i=0}^{\infty} \mu_i^{-2} = \infty$, then there exists a subsequence of $\{v^i\}$ which converges strongly to zero. Moreover, some weak accumulation point of $\{x^i\}$ solves* (1.1).
(iii) *If $\mu_i \le \bar{\mu} < \infty$, then the sequence $\{v^i\}$ converges strongly to zero, and the sequence $\{x^i\}$ converges weakly to a solution of* (1.1).

Proof. Suppose that the algorithm terminates at some iteration i. If $v^i = 0$ then $0 \in T(y^i)$, so that y^i is a solution. If $y^i = x^i$ then, by (1.6), $v^i = -\varepsilon^i$. On the other hand, from (1.7) we have that $\|\varepsilon^i\| \le \sigma\|v^i\|$. Because $\sigma < 1$, it again follows that $v^i = 0$, i.e., y^i is a solution of the problem. From now on, we assume that an infinite sequence of iterates is generated.

We first show that the hyperplane H_i, given by (1.10), strictly separates x^i from any solution $\bar{x}$ of the problem.

By monotonicity of $T(\cdot)$,

$$\langle v^i, \bar{x} - y^i\rangle \le 0.$$

Consider the two possible cases that may occur in (1.7). Suppose first that

$$\|\varepsilon^i\| \le \sigma\mu_i\|x^i - y^i\|. \tag{2.1}$$

Then, using (1.6) and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} \langle v^i, x^i - y^i\rangle &= \mu_i\|x^i - y^i\|^2 - \langle \varepsilon^i, x^i - y^i\rangle \\ &\ge \mu_i\|x^i - y^i\|^2 - \|\varepsilon^i\|\|x^i - y^i\| \\ &\ge \mu_i(1 - \sigma)\|x^i - y^i\|^2. \end{aligned} \tag{2.2}$$

The right-hand-side of the above inequality is positive unless $x^i = y^i$, in which case the algorithm terminates at a solution of the problem. Suppose now that

$$\|\varepsilon^i\| \le \sigma\|v^i\|. \tag{2.3}$$

In that case, by (1.6) and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\langle v^i, x^i - y^i\rangle &= \frac{1}{\mu_i}\langle v^i, v^i + \varepsilon^i\rangle \\ &\geq \frac{1}{\mu_i}\left(\|v^i\|^2 - \|v^i\|\|\varepsilon^i\|\right) \\ &\geq \frac{1-\sigma}{\mu_i}\|v^i\|^2, \end{aligned} \tag{2.4}$$

which again is positive unless y^i is a solution of the problem.

In either case, we are in position to apply Lemma 2.1, which gives

$$\|x^{i+1} - \bar{x}\|^2 \leq \|x^i - \bar{x}\|^2 - \left(\frac{\langle v^i, x^i - y^i\rangle}{\|v^i\|}\right)^2. \tag{2.5}$$

It follows that the sequence $\{\|x^i - \bar{x}\|\}$ is nonincreasing. Hence it converges and, in particular, the sequence $\{x^i\}$ is bounded.

If (2.1) holds then, using (2.2), we have

$$\frac{\langle v^i, x^i - y^i\rangle}{\|v^i\|} \geq \frac{\mu_i(1-\sigma)\|x^i - y^i\|^2}{\|v^i\|}. \tag{2.6}$$

Furthermore, by (1.6),

$$\begin{aligned}\mu_i\|x^i - y^i\| &= \|v^i + \varepsilon^i\| \\ &\geq \|v^i\| - \sigma\mu_i\|x^i - y^i\|.\end{aligned}$$

Hence

$$\mu_i\|x^i - y^i\| \geq \frac{1}{1+\sigma}\|v^i\|. \tag{2.7}$$

Combining the latter inequality with (2.6), we obtain that

$$\frac{\langle v^i, x^i - y^i\rangle}{\|v^i\|} \geq \frac{1-\sigma}{1+\sigma}\|x^i - y^i\|. \tag{2.8}$$

If (2.3) holds then, by (2.4), we have

$$\frac{\langle v^i, x^i - y^i\rangle}{\|v^i\|} \geq \frac{1-\sigma}{\mu_i}\|v^i\|. \tag{2.9}$$

Since, by (1.6),

$$\frac{1}{\mu_i}(v^i + \varepsilon^i) = x^i - y^i,$$

we also have that

$$\frac{1+\sigma}{\mu_i}\|v^i\| \geq \|x^i - y^i\|. \tag{2.10}$$

Using (2.9) and (2.10), we have that (2.8) holds in the second case also.

Combining (2.8) with (2.5), we obtain

$$\|x^{i+1} - \bar{x}\|^2 \leq \|x^i - \bar{x}\|^2 - \left(\frac{1-\sigma}{1+\sigma}\right)^2 \|x^i - y^i\|^2.$$

It immediately follows that the sequence $\{x^i - y^i\}$ converges strongly to zero. This completes the proof of the first assertion of the theorem.

Suppose now that $\sum_{i=0}^{\infty} \mu_i^{-2} = \infty$. When (2.1) holds, (2.6) and (2.7) imply that

$$\frac{\langle v^i, x^i - y^i \rangle}{\|v^i\|} \geq \frac{1-\sigma}{\mu_i(1+\sigma)^2}\|v^i\|.$$

By (2.9), the same relation also holds in the case of (2.3). Hence, from (2.5) it follows that

$$\|x^{i+1} - \bar{x}\|^2 \leq \|x^i - \bar{x}\|^2 - \left(\frac{1-\sigma}{(1+\sigma)^2\mu_i}\right)^2 \|v^i\|^2. \tag{2.11}$$

Suppose that $\liminf_{i\to\infty} \|v^i\| = 2\delta > 0$. It follows that there exists some index i_0 such that $\|v^i\| \geq \delta$ for all $i \geq i_0$. By (2.11), we have that

$$\begin{aligned} \|x^{i_0} - \bar{x}\|^2 - \|x^i - \bar{x}\|^2 &\geq c\sum_{j=i_0}^{i-1} \|v^j\|^2 \mu_j^{-2} \\ &\geq c\delta^2 \sum_{j=i_0}^{i-1} \mu_j^{-2}, \end{aligned}$$

where $c := (1-\sigma)^2(1+\sigma)^{-4}$. Now letting $i \to \infty$ gives a contradiction with the fact that $\sum_{i=0}^{\infty} \mu_i^{-2} = \infty$. Hence the assumption is invalid and $\liminf_{i\to\infty} \|v^i\| = 0$, i.e. there exists a subsequence $\{v^{i_k}\}$ of $\{v^i\}$ which strongly converges to zero.

Since the sequence $\{x^i\}$ is bounded, so is $\{x^{i_k}\}$. Let $\hat{x}$ be a weak accumulation point of $\{x^{i_k}\}$. Since $\|x^{i_k} - y^{i_k}\| \to 0$, $\hat{x}$ is also a weak accumulation point of $\{y^{i_k}\}$. Passing onto a subsequence, if necessary, we can assume that $\{y^{i_k}\}$ converges weakly to $\hat{x}$. By monotonicity of $T(\cdot)$, for any $z \in \mathcal{H}$ and any $w \in T(z)$, we have

$$0 \leq \langle z - y^{i_k}, w - v^{i_k} \rangle = \langle z - y^{i_k}, w \rangle - \langle z - y^{i_k}, v^{i_k} \rangle.$$

Passing onto the limit as $k \to \infty$, and taking into account weak convergence of $\{y^{i_k}\}$ to $\hat{x}$, and strong convergence of $\{v^{i_k}\}$ to zero, we obtain that

$$0 \leq \langle z - \hat{x}, w \rangle.$$

Now maximality of $T(\cdot)$ implies that $0 \in T(\hat{x})$. This proves the second assertion of the theorem.

Finally, let $\mu_i \leq \bar{\mu} < \infty$ for all i. It follows from (2.11) that the whole sequence $\{v^i\}$ converges strongly to zero. Therefore we can repeat the preceding argument for any weak accumulation point of $\{x^i\}$ to show that each of them is a solution of the problem. The proof of uniqueness of a weak accumulation point is standard.

Let $\tilde{x}$, $\hat{x}$ be two weak accumulation points of $\{x^i\}$, and $\{x^{i_m}\}$, $\{x^{i_k}\}$ be two subsequences of $\{x^i\}$ weakly convergent to $\tilde{x}$ and $\hat{x}$, respectively. Let $\pi = \lim_{i\to\infty} \|x^i - \tilde{x}\|^2$, $\zeta = \lim_{i\to\infty} \|x^i - \hat{x}\|^2$. Those limits exist by (2.5), since $\tilde{x}$ and $\hat{x}$ are solutions of the problem. We have that

$$\|x^{i_k} - \tilde{x}\|^2 = \|x^{i_k} - \hat{x}\|^2 + \|\hat{x} - \tilde{x}\|^2 + 2\langle x^{i_k} - \hat{x}, \hat{x} - \tilde{x}\rangle,$$

and

$$\|x^{i_m} - \hat{x}\|^2 = \|x^{i_m} - \tilde{x}\|^2 + \|\tilde{x} - \hat{x}\|^2 + 2\langle x^{i_m} - \tilde{x}, \tilde{x} - \hat{x}\rangle.$$

Take limits in the above two relations as $k \to \infty$ and $m \to \infty$. The inner products in the right-hand-sides converge to zero because $\tilde{x}$, $\hat{x}$ are the weak limits of $\{x^{i_m}\}$, $\{x^{i_k}\}$ respectively. Therefore, using the definitions of π and ζ we obtain

$$\pi = \zeta + \|\hat{x} - \tilde{x}\|^2,$$

and

$$\zeta = \pi + \|\hat{x} - \tilde{x}\|^2.$$

It follows that $\pi - \zeta = \|\hat{x} - \tilde{x}\|^2 = \zeta - \pi$, which implies that $\tilde{x} = \hat{x}$, i.e. all weak accumulation ponts of $\{x^i\}$ coincide. So $\{x^i\}$ weakly converges to a solution of the problem.

The proof is complete. $\square$

Remark 2.3. In the finite-dimensional case, condition $\sum_{i=0}^{\infty} \mu_i^{-2} = \infty$ is sufficient to ensure that the sequence of iterates converges to a solution of the problem.

Indeed, by the second assertion of Theorem 2.2, the sequence $\{x^i\}$ has some accumulation point $\hat{x}$ such that $0 \in T(\hat{x})$. We can then choose $\bar{x} = \hat{x}$ in (2.5) (or (2.11)). It follows that $\{x^i - \hat{x}\}$ converges, and since $\hat{x}$ is an accumulation point of $\{x^i\}$, it must be the case that $\{x^i\}$ converges to $\hat{x}$.

Since in the special case of $\sigma = 0$ the iterates generated by Algorithm 1.1 coincide with iterates of the exact proximal point method, Theorem 2.2 provides an alternative convergence proof for this classical method. In some ways, the given proof is more intuitive, as it allows for a nice and simple geometric interpretation using separation arguments.

We next show that the iterates converge at a linear rate to the solution of the problem, provided T^{-1} is Lipschitz continuous at zero (note that, in that case, the solution is unique) [19].

Theorem 2.4. *Suppose there exist positive constants L and δ such that*

$$\|y - \bar{x}\| \le L\|v\| \quad \text{for all } y \in T^{-1}(v),\ \|v\| \le \delta.$$

Then any sequence $\{x^i\}$ generated by Algorithm 1.1 *with $\mu_i \le \bar{\mu} < \infty$ converges strongly to $\bar{x}$, the solution of* (1.1)*, and the rate of convergence is Q-linear.*

Proof. Let $\bar{x}$ be any solution of (1.1). By (2.11), we have

$$\|x^{i+1} - \bar{x}\|^2 \leq \|x^i - \bar{x}\|^2 - (1-\sigma)^2(1+\sigma)^{-4}\mu_i^{-2}\|v^i\|^2. \tag{2.12}$$

By Theorem 2.2, it follows that $\|v^i\| \leq \delta$ holds for all i sufficiently large, say $i \geq i_0$. Again consider the two cases that may occur in (1.7). In the case of (2.1), applying the Cauchy-Schwarz inequality to the left-hand-side of (2.2), we obtain

$$\|x^i - y^i\| \leq \frac{1}{\mu_i(1-\sigma)}\|v^i\|. \tag{2.13}$$

In the case of (2.3), we have that (2.10) holds, which in turn implies (2.13) because $(1-\sigma)^{-1} \geq 1+\sigma$ for $\sigma < 1$. So in both cases (2.13) holds. Using (2.13), the hypothesis of the theorem, and the triangle inequality, we obtain, for $i \geq i_0$, that

$$\begin{aligned} L\|v^i\| &\geq \|y^i - \bar{x}\| \\ &\geq \|x^i - \bar{x}\| - \|x^i - y^i\| \\ &\geq \|x^i - \bar{x}\| - (\mu_i(1-\sigma))^{-1}\|v^i\|. \end{aligned}$$

Rearranging terms, we obtain

$$\|v^i\| \geq \frac{\mu_i(1-\sigma)}{L\mu_i(1-\sigma)+1}\|x^i - \bar{x}\|.$$

Combining the latter relation with (2.12), we have that

$$\begin{aligned} \|x^{i+1} - \bar{x}\|^2 &\leq \left(1 - \frac{(1-\sigma)^4}{(1+\sigma)^4(L\mu_i(1-\sigma)+1)^2}\right)\|x^i - \bar{x}\|^2 \\ &\leq \left(1 - \frac{(1-\sigma)^4}{(1+\sigma)^4(L\bar{\mu}(1-\sigma)+1)^2}\right)\|x^i - \bar{x}\|^2, \end{aligned} \tag{2.14}$$

which establishes the claim. □

Observe that the relation (2.14) also implies parametric superlinear convergence, which is obtained if the tolerance parameter σ and the regularization parameter μ tend to zero.

Remark 2.5. In the finite-dimensional case, it is possible to relax the hypothesis of Theorem 2.4. In particular, if $\mu_i \leq \bar{\mu} < \infty$, then the following condition [24] is sufficient for linear convergence of the algorithm :

$$\|y - P_S[y]\| \leq L\|v\| \quad \text{for all } y \in T^{-1}(v),\ \|v\| \leq \delta,$$

where $P_S[\cdot]$ denotes the projection map onto the solution set S (which need not be a singleton).

By Remark 2.3, we have that $\{x^i\} \to \hat{x}$, where $\hat{x} \in S$. In the same way as before, we obtain the relation

$$\|v^i\| \geq \frac{\mu_i(1-\sigma)}{L\mu_i(1-\sigma)+1}\|x^i - P_S[x^i]\|.$$

In (2.12), let $\bar{x} = P_S[x^i]$. Combining the latter inequality above with (2.12), we have

$$\begin{aligned} \|x^{i+1} - P_S[x^{i+1}]\|^2 &\leq \|x^{i+1} - \bar{x}\|^2 \\ &\leq \|x^i - \bar{x}\|^2 - (1-\sigma)^2(1+\sigma)^{-4}\mu_i^{-2}\|v^i\|^2 \\ &\leq \left(1 - \frac{(1-\sigma)^4}{(1+\sigma)^4(L\bar{\mu}(1-\sigma)+1)^2}\right)\|x^i - P_S[x^i]\|^2. \end{aligned}$$

It follows that the sequence $\{\|x^i - P_S[x^i]\|\}$ converges Q-linearly to zero. Letting $\bar{x} = P_S[x^i]$ in (2.12), for any $j \geq 1$, we have

$$\|x^{i+j} - P_S[x^i]\| \leq \|x^{i+j-1} - P_S[x^i]\| \leq \cdots \leq \|x^i - P_S[x^i]\|.$$

Letting $j \to \infty$, we obtain that

$$\|\hat{x} - P_S[x^i]\| \leq \|x^i - P_S[x^i]\|.$$

Therefore

$$\begin{aligned} \|x^i - \hat{x}\| &\leq \|x^i - P_S[x^i]\| + \|P_S[x^i] - \hat{x}\| \\ &\leq 2\|x^i - P_S[x^i]\|. \end{aligned}$$

Since $\{\|x^i - P_S[x^i]\|\}$ converges Q-linearly to zero, it follows that $\{x^i\}$ converges R-linearly to $\hat{x} \in S$.

3. Concluding Remarks

We presented a hybrid algorithm which combines some projection and proximal point ideas. In particular, an approximate proximal point iteration is used to construct a hyperplane which strictly separates the current iterate from the solutions of the problem. This step is then followed by a projection of the current iterate onto the separating hyperplane. The new method relaxes tolerance requirements for solving the subproblems of the proximal point algorithm, while preserving all of its convergence properties. Thus a more practical framework is obtained.

The general algorithm presented here has already proved to be useful for devising *truly* globally convergent (i.e. the *whole* sequence of iterates is globally convergent to a solution without any regularity assumptions) and locally superlinearly convergent inexact Newton methods for solving systems of monotone equations [21] and monotone nonlinear complementarity problems [23]. An extention to inexact Bregman-function-based proximal algorithms is given in [22].

Using the presented approach in conjunction with variable metric proximal point methods and decomposition (operator splitting) algorithms can be an interesting subject for future research.

References

[1] J.F. Bonnans, J. C. Gilbert, C. Lemaréchal, C. Sagastizábal: A family of variable metric proximal point methods, Mathematical Programming 68 (1995) 15–47.

[2] J. V. Burke, Maijian Qian: The variable metric proximal point algorithm, I: Basic convergence theory, Department of Mathematics, University of Washington, Seattle, WA (1996).

[3] J. V. Burke, Maijian Qian: The variable metric proximal point algorithm, II: Quasi-Newton updating, Department of Mathematics, University of Washington, Seattle, WA (1996).

[4] X. Chen, M. Fukushima: Proximal quasi-Newton methods for nondifferentiable convex optimization, Mathematical Programming, To appear.

[5] J. Eckstein, D. P. Bertsekas: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, Mathematical Programming 55 (1992) 293–318.

[6] M. C. Ferris: Finite termination of the proximal point algorithm, Mathematical Programming 50 (1991) 359–366.

[7] O. Güler: New proximal point algorithms for convex minimization, SIAM Journal on Optimization 2 (1992) 649–664.

[8] B. Lemaire: The proximal algorithm, In: J. P. Penot (ed.), International Series of Numerical Mathematics, Birkhäuser, Basel, 1989, 73–87.

[9] C. Lemaréchal, C. Sagastizábal: An approach to variable metric bundle methods, In: J. Henry, J.-P. Yvon (eds.), Lecture Notes in Control and Information Sciences No. 197, System Modelling and Optimization, Springer-Verlag, Berlin et al., 1994, 144–162.

[10] C. Lemaréchal, C. Sagastizábal: Variable metric bundle methods: From conceptual to implementable forms, Mathematical Programming 76 (1997) 393–410.

[11] F. J. Luque: Asymptotic convergence analysis of the proximal point algorithm, SIAM Journal on Control and Optimization 22 (1984) 277–293.

[12] B. Martinet: Regularisation d'inequations variationelles par approximations successives, Revue Française d'Informatique et de Recherche Opérationelle 4 (1970) 154–159.

[13] R. Mifflin: A quasi-second-order proximal bundle algorithm, Mathematical Programming 73 (1996) 51–72.

[14] J.-J. Moreau: Proximité et dualité dans un espace Hilbertien, Bulletin de la Société Mathématique de France 93 (1965) 273–299.

[15] G. B. Passty: Weak convergence theorems for nonexpansive mappings in Banach spaces, Journal of Mathematical Analysis and Applications 67 (1979) 274–276.

[16] B. T. Polyak: Introduction to Optimization, Optimization Software, Inc., Publications Division, New York, 1987.

[17] L. Qi, X. Chen: A preconditioning proximal Newton method for nondifferentiable convex optimization, Technical Report AMR 95/20, Department of Applied Mathematics, University of New South Wales, Sydney, SW, Australia, 1995.

[18] R. T. Rockafellar: Augmented Lagrangians and applications of the proximal point algorithm in convex programming, Mathematics of Operations Research 1 (1976) 97–116.

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

[20] M. V. Solodov, B. F. Svaiter: A new projection method for variational inequality problems, SIAM Journal on Control and Optimization, To appear.

[21] M. V. Solodov, B. F. Svaiter: A globally convergent inexact Newton method for systems of monotone equations, In: M. Fukushima, L. Qi (eds.), Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, Kluwer Academic Publishers, 1998, 355–369.

[22] M. V. Solodov, B. F. Svaiter: An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions, Mathematics of Operations Research, submitted.

[23] M. V. Solodov, B. F. Svaiter: A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem, SIAM Journal on Optimization, submitted.

[24] P. Tseng: On linear convergence of iterative methods for the variational inequality problem, Journal of Computational and Applied Mathematics 60 (1995) 237–252.

[25] C. Zhu: Asymptotic convergence analysis of some inexact proximal point algorithms for minimization, SIAM Journal on Optimization 6 (1996) 626–637.