Set-Valued Analysis **7:** 323–345, 1999.
© 2000 *Kluwer Academic Publishers. Printed in the Netherlands.*

A Hybrid Approximate Extragradient – Proximal Point Algorithm Using the Enlargement of a Maximal Monotone Operator ⋆

M. V. SOLODOV and 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: 28 May 1998; in final form: 10 June 1999)

Abstract. We propose a modification of the classical extragradient and proximal point algorithms for finding a zero of a maximal monotone operator in a Hilbert space. At each iteration of the method, an approximate extragradient-type step is performed using information obtained from an approximate solution of a proximal point subproblem. The algorithm is of a hybrid type, as it combines steps of the extragradient and proximal methods. Furthermore, the algorithm uses elements in the enlargement (proposed by Burachik, Iusem and Svaiter) of the operator defining the problem. One of the important features of our approach is that it allows significant relaxation of tolerance requirements imposed on the solution of proximal point subproblems. This yields a more practical proximal-algorithm-based framework. Weak global convergence and local linear rate of convergence are established under suitable assumptions. It is further demonstrated that the modified forward-backward splitting algorithm of Tseng falls within the presented general framework.

Mathematics Subject Classifications (1991): 90C25, 49J45, 49M45.

Key words: maximal monotone operator, enlargement of a maximal monotone operator, proximal point method, extragradient method.

1. Introduction

We consider the classical 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 $\mathcal{H}$, i.e., $T\colon \mathcal{H} \to \mathcal{P}(\mathcal{H})$, where $\mathcal{P}(\mathcal{H})$ stands for the family of subsets of $\mathcal{H}$. A wide variety of problems, such as optimization and min-max problems, complementarity problems and variational inequalities, fall within this general framework.

⋆ 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.

Having $x^k \in \mathcal{H}$, a current approximation to the solution of (1.1), Rockafellar's proximal point algorithm generates the next iterate by the approximate rule [25, expression (1.7)]:

$$x^{k+1} \approx (I + c_k T)^{-1} x^k, \tag{1.2}$$

where $\{c_k\}$ is some sequence of positive real numbers. Note that $(I + c_k T)^{-1} x^k$ is the *exact* solution of the 'proximal subproblem'

$$0 \in c_k T(x) + (x - x^k). \tag{1.3}$$

Since the exact computation of $(I + c_k T)^{-1} x^k$ (or equivalently, the exact solution of (1.3)) can be quite difficult or even impossible in practice, the use of approximate solutions is essential for devising implementable algorithms. In [25], the following approximation criteria were introduced to manage iterations given by (1.2):

$$\|x^{k+1} - (I + c_k T)^{-1} x^k\| \leqslant e_k, \quad \sum_{k=0}^{\infty} e_k < \infty, \tag{1.4}$$

and

$$\|x^{k+1} - (I + c_k T)^{-1} x^k\| \leqslant d_k \|x^{k+1} - x^k\|, \quad \sum_{k=0}^{\infty} d_k < \infty. \tag{1.5}$$

Since $(I + c_k T)^{-1} x^k$ is the point to be approximated, in general it is not available. Therefore, the above two criteria usually cannot be directly used in practical implementations. In [25, Proposition 6] it was established that (1.4), (1.5) are implied, respectively, by the following two conditions:

$$\operatorname{dist}(0, c_k T(x^{k+1}) + x^{k+1} - x^k) \leqslant e_k, \quad \sum_{k=0}^{\infty} e_k < \infty, \tag{1.6}$$

and

$$\operatorname{dist}(0, c_k T(x^{k+1}) + x^{k+1} - x^k) \leqslant d_k \|x^{k+1} - x^k\|, \quad \sum_{k=0}^{\infty} d_k < \infty. \tag{1.7}$$

The latter two conditions have the advantage that they are usually easier to verify in practice, because they only require evaluation of T at x^{k+1}.

A sequence generated by the proximal point algorithm under criterion (1.4) (or (1.6)), with

$$c_k \geqslant \bar{c} > 0, \tag{1.8}$$

converges weakly to a solution of (1.1), provided this problem has solutions [25, Theorem 1]. If T^{-1} is Lipschitz continuous around 0, then $\{x^k\}$ generated using criterion (1.5) (or (1.7)) converges linearly to the solution, provided $\{x^k\}$ is bounded and regularization parameters satisfy (1.8) [25, Theorem 2]. Observe that criterion (1.5) (or (1.7)) plus Lipschitz continuity of T^{-1} and (1.8) do not guarantee boundedness of $\{x^k\}$ (see also an example in Section 3).

We emphasize that criteria (1.6), (1.7) are more suitable for applications than (1.4), (1.5). We next discuss them in more detail. Observe that since $T(x)$ is closed, for any $\lambda \geqslant 0$

$$\operatorname{dist}(0, c_k T(x^{k+1}) + x^{k+1} - x^k) \leqslant \lambda$$

is equivalent to

$$\exists\, v^{k+1} \in T(x^{k+1}) \quad \text{such that} \quad \|c_k v^{k+1} + x^{k+1} - x^k\| \leqslant \lambda.$$

Hence, the inexact proximal point algorithm managed by criteria (1.6) or (1.7) can be restated as follows. Having a current iterate x^k,

$$\begin{aligned}&\text{find } x^{k+1} \in \mathcal{H} \text{ and } v^{k+1} \in T(x^{k+1}) \text{ such that}\\ &0 = c_k v^{k+1} + (x^{k+1} - x^k) - r^k,\end{aligned} \tag{1.9}$$

where $r^k \in \mathcal{H}$ is an error associated with the approximation. With this notation, criterion (1.6) becomes

$$\|r^k\| \leqslant e_k, \quad \sum_{k=0}^{\infty} e_k < \infty, \tag{1.10}$$

and criterion (1.7) becomes

$$\|r^k\| \leqslant d_k \|x^{k+1} - x^k\|, \quad \sum_{k=0}^{\infty} d_k < \infty. \tag{1.11}$$

Error tolerance criteria of this kind (i.e., similar to (1.10), (1.11)) are standard in the study of proximal point and related methods (e.g., see also [1, 32, 4, 8, 7]). The condition $c_k \geqslant \bar{c} > 0$ is also a common requirement (in the finite-dimensional setting, it can be relaxed to the condition that $\sum c_k^2 = \infty$, e.g., see [30]). Proximal point methods have been studied extensively. The literature on this subject is vast and includes, but is not limited to, [24, 25, 21, 18, 17, 10, 13, 9] (see [15] for a survey).

In [2], Burachik, Iusem and Svaiter introduced an enlargement T^ε of a maximal monotone operator T. Given $\varepsilon \geqslant 0$ define $T^\varepsilon\colon \mathcal{H} \to \mathcal{P}(\mathcal{H})$ as

$$T^\varepsilon(x) = \{v \in \mathcal{H} \mid \langle w - v, z - x\rangle \geqslant -\varepsilon \text{ for all } z \in \mathcal{H},\ w \in T(z)\}. \tag{1.12}$$

Some of the properties of T^ε will be discussed in Section 2. Specifically, our emphasis will be on the relations between T^ε and approximate solutions of proximal subproblems. In [2], the enlargement T^ε was further used to devise an approximate generalized proximal point method. The exact version of this method can be stated as follows: having x^k, the next iterate x^{k+1} is the solution of

$$0 \in c_k T(\cdot) + \nabla f(\cdot) - \nabla f(x^k), \tag{1.13}$$

where f is a suitable regularization function. Note that for $f(\cdot) = (1/2)\|\cdot\|^2$, the above method becomes the classical proximal point algorithm (1.3). Approximate solutions of (1.13) are treated in [2] via T^ε, an (outer) approximation to T. Specifically, an approximate solution of (1.13) is regarded as an *exact* solution of

$$0 \in c_k T^{\varepsilon_k}(\cdot) + \nabla f(\cdot) - \nabla f(x^k)$$

for an appropriate value of ε_k. Note that for $f(\cdot) = (1/2)\|\cdot\|^2$, the latter relation is equivalent to

$$\begin{aligned} &\text{find } x^{k+1} \in \mathcal{H},\ \varepsilon_k \geqslant 0 \text{ and } v^{k+1} \in T^{\varepsilon_k}(x^{k+1}) \text{ such that} \\ &\quad 0 = c_k v^{k+1} + (x^{k+1} - x^k). \end{aligned} \tag{1.14}$$

To ensure convergence (assuming additional conditions usual in the context of generalized proximal point methods), the following requirement was imposed in [2]:

$$\sum_{k=0}^{\infty} \varepsilon_k < \infty. \tag{1.15}$$

We note that the latter requirement is of the same nature as (1.10). One drawback of conditions of this kind is that quite often there is no *constructive* way to enforce them. Indeed, there exist infinitely many summable sequences, and it is not specified how to choose the value of e_k in (1.10) or the value of ε_k in (1.15) at a specific iteration k and for the given problem, so as to ensure convergence. From the algorithmic standpoint, one would prefer to have a *computable* error tolerance condition which is related to the progress of the algorithm at every given step when applied to the given problem. This is one of the main motivations of [30], as well as the present paper.

It has long been realized that in many applications, proximal point methods in the classical forms are not very efficient. Developments aimed at speeding up convergence of proximal point methods for solving operator equations can be found in [6, 22, 4] (the situation for the special case of optimization, i.e., when T is a subdifferential mapping, is somewhat easier, so more strategies are available; however, the area of optimization is not our main focus). 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 [25] and [4], 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 and/or more constructive requirements on solving the proximal subproblems. In [30], Solodov and Svaiter proposed a new criterion for the approximate solution of proximal subproblems. Specifically, a pair y^k, v^k is admissible if

$$v^k \in T(y^k), \quad 0 = c_k v^k + (y^k - x^k) - r^k,$$

and the error r^k satisfies

$$\|r^k\| \leqslant \sigma \max\{c_k \|v^k\|, \|y^k - x^k\|\}, \tag{1.16}$$

where σ is a *fixed* scalar in $[0, 1)$. The next iterate x^{k+1} is obtained by projecting x^k onto the hyperplane

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

which can be shown to strictly separate x^k from any zero of T. This hybrid method preserves all the desirable convergence properties of the exact algorithm (1.3) (or its approximate version given by (1.9) with error criteria (1.10), (1.11)), see [30]. It is important to note that projection onto a hyperplane is explicit, and thus computationally negligible. The advantage of the new error criterion (1.16) is that it is constructive, and the relaxation parameter σ need not tend to zero to ensure convergence (it is fixed). Let us give an interpretation of (1.16) as a *relative error* tolerance. An *exact* solution of $0 \in c_k T(x) + x - x^k$ may be regarded as a pair y, v satisfying

$$v \in T(y), \qquad c_k v + y - x^k = 0.$$

For an approximate solution,

$$c_k v + (y - x^k) = r \approx 0.$$

To estimate the relative error in the above relation one has to look at the ratios between $\|r\|$ and $\|c_k v\|$, $\|r\|$ and $\|y - x^k\|$, i.e., the quantities

$$\frac{\|r\|}{\|c_k v\|}, \qquad \frac{\|r\|}{\|y - x^k\|}.$$

In this sense, (1.16) is equivalent to saying that the bound for the *relative error* in solving the proximal subproblem can be fixed (at σ), and need not tend to zero. We emphasize that computationally, the requirement of bounded relative error is realistic. Some further applications of the hybrid strategy described above can be found in [27–29].

In this paper, we combine the ideas of [30, 2], and propose an even simpler method, in which no projection is performed. We obtain a rather broad framework in which the approximate solutions are handled both through T^{ε_k}, a current approximation of T, and the error term r^k appearing in the subproblem equation. An approximate solution is regarded as a pair y^k, v^k such that

$$v^k \in T^{\varepsilon_k}(y^k), \quad 0 = c_k v^k + (y^k - x^k) - r^k, \tag{1.17}$$

where ε_k, r^k are 'relatively small' in comparison with $\|y^k - x^k\|$ (this will be made precise later). Our new method uses this information to get the next iterate by the rule

$$x^{k+1} = x^k - c_k v^k.$$

Because v^k, the direction of change, is obtained by an (approximate) evaluation of the operator T at an intermediate point y^k, the algorithm shares some features with the extragradient method [14]. This is reflected in the name we give to our algorithm. It is worth to point out that the notion of approximate solutions introduced in (1.17) is conceptually broader than that of [2] given in (1.14). This is because in general, the relation $v \in T(y),\ cv + y - x = r\ (r \neq 0)$ does not imply that there exists some $\varepsilon \geqslant 0$ for which $0 \in cT^\varepsilon(y) + y - x$ (see the example in Section 3).

This paper is organized as follows. In Section 2, the Burachik–Iusem–Svaiter (BIS, for short) enlargement of a maximal monotone operator is reviewed and the error tolerance in solving proximal subproblems is formally defined. Some properties of such approximate solutions are derived. In Section 3, the hybrid approximate extragradient–proximal method is introduced and the convergence theorems are stated. Section 4 contains convergence analysis. In Section 5 we give some useful applications of the general approach. Specifically, we show that the recently proposed modified forward-backward splitting method of Tseng [35] can be viewed as a realization of our framework. We further describe how our approximate proximal scheme can be used for constructing a globally convergent Newton-type algorithm.

2. BIS Enlargement of a Maximal Monotone Operator and the Proximal Subproblem Error Tolerance

For convenience, we shall restate the definition of T^ε: for $\varepsilon \geqslant 0$ and $x \in \mathcal{H}$,

$$T^\varepsilon(x) = \{v \in \mathcal{H} \mid \langle w - v, z - x\rangle \geqslant -\varepsilon \text{ for all } z \in \mathcal{H},\ w \in T(z)\}. \tag{2.1}$$

It follows directly from the definition that

$$0 \leqslant \varepsilon_1 \leqslant \varepsilon_2 \Rightarrow T^{\varepsilon_1}(x) \subseteq T^{\varepsilon_2}(x) \quad \forall x \in \mathcal{H}.$$

Also, from the maximal monotonicity of T, it follows that

$$T^0 = T.$$

Hence, for any $\varepsilon \geqslant 0$ and all $x \in \mathcal{H}$,

$$T(x) \subseteq T^\varepsilon(x).$$

Therefore, it is reasonable to call T^ε an 'enlargement' of T. The term 'submonotone enlargement' seems to be most suitable for the enlargement defined above. Unfortunately, 'submonotone' has been already used in [31, 12], so we call the above enlargement BIS, after its authors, Burachik, Iusem and Svaiter. Some theoretical properties of the BIS enlargement, together with applications, are presented in [2, 3].

DEFINITION 2.1. Let $x \in \mathcal{H}$, $c > 0$ and $\sigma \in [0, 1)$. We say that a pair (y, v) is an approximate solution with error tolerance σ of $0 \in cT(\cdot) + (\cdot - x)$ if for some $\varepsilon \geqslant 0$ it holds that

$$v \in T^{\varepsilon}(y),$$
$$cv + (y - x) = r$$

and

$$\|r\|^2 + 2c\varepsilon \leqslant \sigma^2 \|y - x\|^2.$$

Note that since we are working with a maximal monotone operator T, the proximal problem $0 \in cT(\cdot) + (\cdot - x)$ has always a unique solution (this is a classical result by Minty [19]). This exact solution is always an approximate solution, for any $\sigma \in [0, 1)$. Therefore, the proximal subproblem has, a fortiori, approximate solutions, possibly many. For the case $\sigma = 0$ (no error tolerance), only the exact solution of the proximal subproblem satisfies Definition 2.1. So this view of approximate solutions is quite natural.

Observe also that imposing $\varepsilon = 0$, the above error condition gives

$$v \in T(y), \quad cv + (y - x) = r$$

and

$$\|r\| \leqslant \sigma \|y - x\|,$$

which implies the error tolerance (1.16) used in [30]. Even though (*for the case* $\varepsilon = 0$!) the new error condition is stronger, it should not be much harder to satisfy in practice, and it allows to omit the extra projection used in [30]. From now on, approximate solutions of proximal subproblems are understood in the sense of Definition 2.1. An important question is: How close is an approximate solution of a proximal subproblem to the exact solution? An answer is furnished by the following lemma.

LEMMA 2.2. *Take any $x \in \mathcal{H}$ and $c > 0$. Suppose that (y, v) is an approximate solution, with error tolerance $\sigma \in [0, 1)$, of*

$$0 \in cT(\cdot) + (\cdot - x).$$

Let z be the exact solution of the above problem. Then it holds that

$$\|z - y\| \leqslant \sigma \|y - x\|.$$

Proof. There exists some $\varepsilon \geqslant 0$ such that

$$\begin{aligned} & v \in T^{\varepsilon}(y), \\ & \|r\|^2 + 2c\varepsilon \leqslant \sigma^2 \|y - x\|^2, \end{aligned} \tag{2.2}$$

where

$$r = cv + (y - x).$$

Since $0 \in cT(z) + (z - x)$, it follows that

$$u := -(1/c)(z - x) \in T(z).$$

Using (2.1), we obtain

$$\langle u - v, z - y \rangle \geqslant -\varepsilon.$$

Observe that

$$\begin{aligned} u - v &= -(1/c)(z - x) - (1/c)(r - (y - x)) \\ &= -(1/c)r - (1/c)(z - y). \end{aligned}$$

Combining the last two relations, we have that

$$(1/c)\|z - y\|^2 + (1/c)\langle r, z - y \rangle - \varepsilon \leqslant 0.$$

Multiplying by c and using the Cauchy–Schwarz inequality, we obtain

$$\|z - y\|^2 - \|r\| \, \|z - y\| - c\varepsilon \leqslant 0.$$

Resolving the latter quadratic inequality in $\|z - y\|$, it follows that

$$\|z - y\| \leqslant \frac{\|r\| + \sqrt{\|r\|^2 + 4c\varepsilon}}{2} \leqslant \sqrt{\|r\|^2 + 2c\varepsilon},$$

where the concavity of the square root was used in the second inequality. Now the desired result follows from (2.2). □

Next we prove the key property of approximate solutions of proximal subproblems, which reveals their usefulness and is the basis for devising our convergent algorithm.

LEMMA 2.3. *Take any $x \in \mathcal{H}$ and $c > 0$. Suppose that (y, v) is an approximate solution, with error tolerance $\sigma \in [0, 1)$, of*

$$0 \in cT(\cdot) + (\cdot - x).$$

Define

$$x^+ = x - cv.$$

Then, for any solution x^* *of* (1.1), *it holds that*

$$\|x^* - x\|^2 - \|x^* - x^+\|^2 \geqslant (1-\sigma^2)\|y - x\|^2.$$

Proof. By direct algebraic manipulations, we obtain

$$\begin{aligned}
&\|x^* - x\|^2 - \|x^* - x^+\|^2 \\
&\quad = \|x^* - y\|^2 + 2\langle x^* - y, y - x\rangle + \|y - x\|^2 - \\
&\qquad - \left[\|x^* - y\|^2 + 2\langle x^* - y, y - x^+\rangle + \|y - x^+\|^2\right] \\
&\quad = 2\langle x^* - y, x^+ - x\rangle + \|y - x\|^2 - \|y - x^+\|^2.
\end{aligned} \tag{2.3}$$

By assumption, for some $\varepsilon \geqslant 0$ it holds that

$$\begin{aligned}
&v \in T^\varepsilon(y), \\
&\|r\|^2 + 2c\varepsilon \leqslant \sigma^2\|y - x\|^2,
\end{aligned}$$

where

$$r = cv + (y - x).$$

Since $0 \in T(x^*)$ and $v \in T^\varepsilon(y)$, by (2.1) we have

$$\begin{aligned}
\langle x^* - y, x^+ - x\rangle &= \langle x^* - y, -cv\rangle \\
&= c\langle x^* - y, 0 - v\rangle \geqslant -c\varepsilon.
\end{aligned}$$

Furthermore, since $r = cv + (y - x)$,

$$y - x^+ = r.$$

Therefore, combining the above relations with (2.3), we obtain

$$\begin{aligned}
\|x^* - x\|^2 - \|x^* - x^+\|^2 &\geqslant -2c\varepsilon + \|y - x\|^2 - \|r\|^2 \\
&= \|y - x\|^2 - \left(\|r\|^2 + 2c\varepsilon\right) \\
&\geqslant (1-\sigma^2)||y - x||^2.
\end{aligned}$$

□

3. The Hybrid Approximate Extragradient–Proximal Point Method

We are now ready to formally state our algorithm.

ALGORITHM 3.1 (Hybrid Approximate Extragradient–Proximal Point Algorithm).

Initialization: Choose any $x^0 \in \mathcal{H}$ and $\sigma \in [0, 1)$

Iteration: For $k = 0, 1, \ldots$

(1) Choose some $c_k > 0$, and find (y^k, v^k) as an approximate solution with error tolerance σ of

$$0 \in c_k T(\cdot) + (\cdot - x^k),$$

i.e., for some $\varepsilon_k \geqslant 0$,

$$v^k \in T^{\varepsilon_k}(y^k) \quad \text{and}$$
$$\|c_k v^k + (y^k - x^k)\|^2 + 2c_k\varepsilon_k \leqslant \sigma^2\|y^k - x^k\|^2.$$

(2) Define

$$x^{k+1} = x^k - c_k v^k.$$

To guarantee convergence of the algorithm, two standard assumptions will be needed: problem (1.1) has a solution, and regularization parameters c_k are bounded away from zero.

We have already discussed the possibility of solving approximately the proximal subproblems. Observe that if the kth proximal point subproblem is solved exactly, then the kth iteration of the above algorithm coincides with the classical proximal point iteration (this is because $c_k v^k + (y^k - x^k) = 0$ and $\varepsilon_k = 0$ imply that $x^{k+1} = y^k$ and $v^k \in T(y^k)$). Therefore, in the special case $\sigma = 0$ we retrieve the classical (exact) proximal point method. Of course, the case of interest is $\sigma > 0$, i.e., the use of approximate proximal iterations. In that case, Algorithm 3.1 is different from other approximate/inexact proximal schemes in the literature.

We next give an example which highlights two important issues. The first one is that the approximation framework of Algorithm 3 is significantly more general than the framework of [2]. The second one is that the 'extragradient' step of Algorithm 3.1 is essential for the convergence of the method. Consider the case where $\mathcal{H} = \Re^2$ and T is the $\pi/2$ rotation operator, i.e.,

$$T(x) := Mx,$$

where

$$M := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

It is easy to check that such T is (maximal) monotone, and that the origin is the unique solution of $T(x) = 0$.

It can be further checked that $T^\varepsilon \equiv T$, i.e., for every x it holds that $T^\varepsilon(x) = T(x)$ for all $\varepsilon \geqslant 0$ (see [2]). Hence, the relation

$$v \in T(y),$$
$$cv + y - x = r \quad (r \neq 0)$$

does not imply that there exists some $\varepsilon \geqslant 0$ for which

$$0 \in cT^\varepsilon(y) + y - x.$$

This means that actually only *exact* solutions can be handled by the method of [2] in this, albeit very special, case.

We next show that the 'extragradient' step of Algorithm 3.1 is essential, i.e., the approximate proximal method with the same error tolerance criterion may fail to converge. Note that the latter method can be viewed as Algorithm 3.1 where we replace the last step with $x^{k+1} = y^k$. Recalling that in our example $\varepsilon_k > 0$ does not give anything more than having $\varepsilon_k = 0$, we consider a sequence defined by the following relations:

$$\begin{aligned} & w^{k+1} \in T(x^{k+1}), \\ & \|c_k w^{k+1} + x^{k+1} - x^k\| \leqslant \sigma \|x^{k+1} - x^k\|. \end{aligned} \tag{3.1}$$

We shall prove that in this simple case the sequence satisfying (3.1) may diverge. Take

$$x^0 \neq (0,0), \quad \sigma \in (1/\sqrt{2}, 1), \quad c_k = 1/2 \quad \text{for all } k.$$

Define

$$Q := \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix},$$

and let $Q^{(k)}$ be the kth power of Q. Define further the two sequences $\{x^k\}$ and $\{w^k\}$ by

$$x^k := Q^{(k)} x^0, \quad w^k := M x^k \in T(x^k).$$

It is easy to see that $\|x^k\| \to +\infty$. However, we shall show that these sequences satisfy (3.1). Since $x^{k+1} = Q x^k$, we obtain

$$\begin{aligned} x^{k+1} - x^k &= (Q - I) x^k, \\ c_k w^{k+1} + x^{k+1} - x^k &= ((1/2) M Q + Q - I) x^k, \end{aligned}$$

where I stands for the identity matrix. Observe that

$$(1/2) M Q + Q - I = \begin{bmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix}.$$

It is now easy to check that

$$\|c_k w^{k+1} + x^{k+1} - x^k\| = (1/\sqrt{2}) \|x^k\|,$$

and

$$\|x^{k+1} - x^k\| = \|x^k\|.$$

So, we have

$$\|c_k w^{k+1} + x^{k+1} - x^k\| < \sigma \|x^{k+1} - x^k\|,$$

which means that the sequences $\{x^k\}$ and $\{w^k\}$ satisfy relations (3.1). However, $\{x^k\}$ not only does not converge to the origin (the unique solution of $T(x) = 0$), it is in fact unbounded! This illustrates that the 'extragradient' step in Algorithm 3.1 is indispensable, if we are to use our relaxed requirements on the error tolerance in solving the proximal subproblems.

For the rest of this paper, let $\{c_k\}$ be a particular sequence of regularization parameters and $\{x^k\}$, $\{(y^k, v^k)\}$, $\{\varepsilon_k\}$ be sequences generated by Algorithm 3.1. Observe that for $r^k := c_k v^k + y^k - x^k$ we have

$$x^{k+1} = y^k - r^k,$$

and so the sequences $\{y^k\}$ and $\{v^k\}$ can be treated implicitly in Algorithm 3.1. Specifically, one can rewrite the algorithm in terms of the sequences $\{x^k\}$, $\{r^k\}$ (in $\mathcal{H}$) and $\{\varepsilon_k\}$ ($\varepsilon_k \geqslant 0$), which satisfy the conditions:

$$\frac{x^k - x^{k+1}}{c_k} \in T^{\varepsilon_k}(x^{k+1} + r^k),$$

and

$$\|r^k\|^2 + 2c_k\varepsilon_k \leqslant \sigma^2\|x^{k+1} + r^k - x^k\|^2.$$

However, from the implementation point of view, the setting of Algorithm 3.1 is more direct. Indeed, what is typically available at every inner iteration of some method applied to solving the kth proximal subproblem $0 \in c_k T(\cdot) + (\cdot - x^k)$, are some values of y and v (which are current approximations to the solution of the subproblem). So it is natural to state the stopping (error tolerance) criterion for the subproblem in terms of y and v, as it is done in Algorithm 3.1. Furthermore, this formulation is more suitable for applications, as can be seen from Section 5.

Our main convergence results are the following.

THEOREM 3.1. *Suppose $c_k \geqslant \bar{c} > 0$ for all k. If problem* (1.1) *has solutions then $\{x^k\}$ converges weakly to a solution. If problem* (1.1) *has no solution, the sequence $\{x^k\}$ is unbounded.*

THEOREM 3.2. *If in addition to the assumption of Theorem* 3.1 *it holds that T^{-1} is Lipschitz continuous around zero, then $\{x^k\}$ converges linearly to the solution.*

4. Convergence Analysis

To prove Theorems 3.1 and 3.2 we shall need some intermediate results on the properties of sequences $\{x^k\}$, $\{(y^k, v^k)\}$, $\{\varepsilon_k\}$ generated by Algorithm 3.1.

LEMMA 4.1. *If x^* is some solution of problem* (1.1) *then*

$$\|x^* - x^k\|^2 - \|x^* - x^{k+1}\|^2 \geqslant (1 - \sigma^2)\|y^k - x^k\|^2$$

for every iteration index k.

Proof. The assertion follows directly from Lemma 2.3. □

COROLLARY 4.2. *If problem* (1.1) *has a solution then*

(1) $\{x^k\}$ *is bounded,*
(2) $\sum_{k=0}^{\infty} ||y^k - x^k||^2 < \infty$,
(3) $\{y^k\}$ *is bounded.*

If, in addition, $c_k \geqslant \bar{c} > 0$ *then*

(4) $\sum_{k=0}^{\infty} \|v^k\|^2 \leqslant (1+\sigma)^2 \sum_{k=0}^{\infty} c_k^{-2} \|y^k - x^k\|^2 < \infty$,
(5) $\sum_{k=0}^{\infty} \varepsilon_k < \infty$.

Proof. Let x^* be any solution of (1.1). By Lemma 4.1, the sequence $\{\|x^k - x^*\|\}$ is nonincreasing. Hence, the sequence $\{x^k\}$ is bounded.

Applying again Lemma 4.1 for $k = 0, 1, \ldots, q$, we conclude that

$$(1-\sigma^2) \sum_{k=0}^{q} \|y^k - x^k\|^2 \leqslant \|x^* - x^0\|^2 - \|x^* - x^{q+1}\|^2 \leqslant \|x^* - x^0\|^2.$$

Taking into account that $1 > \sigma$, and passing to the limit as $q \to \infty$, we obtain item (2).

Item (3) is a direct consequence of items (1) and (2).

To prove item (4), note that for each k,

$$\|c_k v^k + (y^k - x^k)\| \leqslant \sigma \|y^k - x^k\|.$$

By the triangle inequality we further obtain

$$\begin{aligned} c_k ||v^k|| &\leqslant \|c_k v^k + (y^k - x^k)\| + \|y^k - x^k\| \\ &\leqslant (1+\sigma)\|y^k - x^k\|. \end{aligned}$$

Using the condition $c_k \geqslant \bar{c} > 0$ and item (2), item (4) follows.

The proof of item (5) is similar: for each k it holds that

$$2 c_k \varepsilon_k \leqslant \sigma^2 \|y^k - x^k\|^2,$$

so the condition $c_k \geqslant \bar{c} > 0$ and item (2) imply item (5). □

COROLLARY 4.3. *If problem* (1.1) *has a solution and* $c_k \geqslant \bar{c} > 0$ *for all* k*, then all weak accumulation points of* $\{x^k\}$ *are solutions of problem* (1.1).

Proof. By Corollary 4.2, the sequence $\{x^k\}$ is bounded. Hence, it has at least one weak accumulation point. Let $\bar{x}$ be any weak accumulation point of $\{x^k\}$ and $\{x^{k_j}\}$ be some subsequence weakly converging to $\bar{x}$:

$$x^{k_j} \stackrel{w}{\to} \bar{x}.$$

By item (2) of Corollary 4.2 it follows that $\|y^k - x^k\|$ converges to zero, hence

$$y^{k_j} \xrightarrow{w} \bar{x}.$$

By item (4) of Corollary 4.2 it follows also that $\{\|v^k\|\}$ converges to zero, so we have that $\{v^{k_j}\}$ converges strongly to zero:

$$v^{k_j} \to 0.$$

Note that by the construction of the algorithm,

$$v^k \in T^{\varepsilon_k}(y^k).$$

Furthermore, by item (5) of Corollary 4.2 we also have that $\{\varepsilon_k\}$ converges to zero, so

$$\varepsilon_{k_j} \to 0.$$

Take any $x \in \mathcal{H}$ and $u \in T(x)$. Then for any index j it holds that

$$\langle u - v^{k_j}, x - y^{k_j} \rangle \geqslant -\varepsilon_{k_j}.$$

Therefore

$$\langle u - 0, x - y^{k_j} \rangle \geqslant \langle v^{k_j}, x - y^{k_j} \rangle - \varepsilon_{k_j}.$$

Since $\{y^{k_j}\}$ converges weakly to $\bar{x}$, $\{v^{k_j}\}$ converges *strongly* to zero and $\{y^{k_j}\}$ is bounded, and $\{\varepsilon_{k_j}\}$ converges to zero, taking the limit as $j \to \infty$ in the above inequality we obtain that

$$\langle u - 0, x - \bar{x} \rangle \geqslant 0.$$

Note that (x, u) was taken as an arbitrary element in the graph of T. Since T is maximal monotone, it follows that $0 \in T(\bar{x})$, i.e., $\bar{x}$ is a solution of (1.1). □

We are now in position to prove Theorem 3.1.

Proof of Theorem 3.1. First suppose that problem (1.1) has a solution. By Corollary 4.2, the sequence $\{x^k\}$ is bounded. Therefore it has some weak accumulation point, say $\bar{x}$. Corollary 4.3 ensures that $\bar{x}$ is a solution of (1.1). Furthermore, from Lemma 4.1 it follows that $\{\|x^k - \bar{x}\|\}$ is nonincreasing. The proof of uniqueness of weak accumulation point in this setting is standard. For example, uniqueness follows easily by applying Opial's lemma [20], or by using the analysis in [25]. So, if problem (1.1) has a solution, the sequence $\{x^k\}$ converges weakly to a solution.

Now suppose that problem (1.1) *does not* have solutions. For contradictory purposes, suppose that $\{x^k\}$ is bounded. Since

$$v^k = (1/c_k)(x^k - x^{k+1})$$

and $c_k \geqslant \bar{c} > 0$, it follows that $\{c_k v^k\}$ and $\{v^k\}$ are bounded. Note that

$$\begin{aligned} \|c_k v^k\| &\geqslant (\|y^k - x^k\| - \|c_k v^k + (y^k - x^k)\|) \\ &\geqslant (1-\sigma)\|y^k - x^k\|, \end{aligned}$$

so $\{y^k\}$ is also bounded. Take some $\bar{z} \in \mathrm{Dom}(T)$ (since T is maximal monotone, $\mathrm{Dom}(T) \neq \emptyset$), and some R such that

$$R > \max\left\{\|\bar{z}\|, \sup_k \|x^k\|, \sup_k \|y^k\|\right\}.$$

Furthermore, define

$$B := \{z \in \mathcal{H} \mid \|z\| \leqslant 2R\},$$

and let N_B be the normal cone operator for the set B:

$$N_B = \partial I_B,$$

where $I_B\colon \mathcal{H} \to \Re \cup \{+\infty\}$ is the indicator function for B: it is zero on B and $+\infty$ elsewhere. Then N_B is maximal monotone and $\bar{z} \in \mathrm{Dom}(T) \cap \mathrm{int}(\mathrm{Dom}(N_B)) \neq \emptyset$. Therefore, the operator

$$T' = T + N_B$$

is maximal monotone [23]. Moreover, T' has a bounded domain and, hence, it has a zero ([25, Proposition 2]). So the problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in T'(x), \tag{4.1}$$

has a solution.

Now we claim that for every index k,

$$v^k \in (T')^{\varepsilon_k}(y^k). \tag{4.2}$$

Take any $z \in \mathcal{H}$ and $w' \in T'(z)$. Then there exist $w, \nu \in \mathcal{H}$ such that

$$\begin{aligned} &w' = w + \nu, \\ &w \in T(z), \ \nu \in N_B(z). \end{aligned}$$

Note that $y^k \in B$ for all k. Therefore, by the definition of the normal operator, it holds that

$$\langle \nu, z - y^k \rangle \geqslant 0.$$

Taking into account that $v^k \in T^{\varepsilon_k}(y^k)$, we further obtain

$$\begin{aligned} \langle w' - v^k, z - y^k \rangle &= \langle w - v^k, z - y^k \rangle + \langle \nu, z - y^k \rangle \\ &\geqslant -\varepsilon_k, \end{aligned}$$

which proves (4.2). Therefore $\{x^k\}$, $\{(y^k, v^k)\}$, $\{\varepsilon_k\}$ can be considered as sequences generated by Algorithm 3.1 applied to problem (4.1) which has solutions. Then, according to the first part of Theorem 3.1, $\{x^k\}$ converges weakly to some x', a solution of (4.1). We have that

$$0 \in T'(x') = T(x') + N_B(x'),$$

and $\|x'\| \leqslant R$, so that x' is in the interior of the set B. Therefore $N_B(x') = \{0\}$ and $0 \in T(x')$, in contradiction with the hypothesis that problem (1.1) has no solution. It follows that in the latter case, $\{x^k\}$ must be unbounded. □

It is interesting to note that we could also use in our Algorithm 3.1 the more standard tolerance criterion of summability of error terms. Specifically, we could require that

$$\sqrt{\|r^k\|^2 + 2c_k\varepsilon_k} \leqslant \sigma_k, \quad \sum_{k=0}^{\infty} \sigma_k^2 < \infty. \tag{4.3}$$

After making an obvious appropriate change in the assertion of Lemma 2.3, Lemma 4.1 becomes

$$\|x^{k+1} - x^*\|^2 \leqslant \|x^k - x^*\|^2 + \sigma_k^2 - \|y^k - x^k\|^2.$$

Using the condition that $\sum \sigma_k^2 < \infty$ and standard arguments, one concludes that the sequence $\{x^k\}$ is bounded, $\|y^k - x^k\| \to 0$, etc. We note that even if we use this requirement of summability of *squares* of error terms (4.3), it is still weaker than summability of error terms ($\sum \sigma_k < \infty$) needed in the classical method. In any case, we emphasize once again that such conditions of *a priori* summability are not constructive.

To prove Theorem 3.2, recall that its hypotheses imply that problem (1.1) has a unique solution, say x^*, and that there exist some constants $L > 0$ and $\delta > 0$ such that

$$\|v\| \leqslant \delta,\ v \in T(y) \Rightarrow \|y - x^*\| \leqslant L||v||.$$

Proof of Theorem 3.2. By Corollary 4.2 (items (1) and (4)), we already know that

$$\lim_{k\to\infty} ||v^k|| = 0, \qquad \lim_{k\to\infty} ||y^k - x^k|| = 0.$$

For each k, let z^k be the *exact* solution of the proximal subproblem $0 \in c_k T(\cdot) + (\cdot - x^k)$. Then

$$u_k := -(1/c_k)(z^k - x^k) \in T(z^k).$$

Applying Lemma 2.2, we have that

$$\|z^k - y^k\| \leqslant \sigma\|y^k - x^k\|.$$

Using the triangle inequality, we obtain

$$\|z^k - x^k\| \leqslant \|z^k - y^k\| + \|y^k - x^k\| \leqslant (1+\sigma)\|y^k - x^k\|. \tag{4.4}$$

Therefore,

$$\|u_k\| \leqslant (1/c_k)(1+\sigma)\|y^k - x^k\| \leqslant (1/\bar{c})(1+\sigma)\|y^k - x^k\|.$$

Since $\{\|y^k - x^k\|\}$ tends to zero, so does $\{\|u^k\|\}$. Because $u^k \in T(z^k)$, by the Lipschitz continuity of T^{-1} around zero, we conclude that for k large enough

$$\|z^k - x^*\| \leqslant L\|u^k\| = L(1/c_k)\|z^k - x^k\|.$$

Using further the triangle inequality, we obtain

$$\begin{aligned} \|x^* - x^k\| &\leqslant \|x^* - z^k\| + \|z^k - x^k\| \\ &\leqslant L(1/c_k)\|z^k - x^k\| + \|z^k - x^k\| \\ &= (1 + L/c_k)\|z^k - x^k\|. \end{aligned}$$

Now using (4.4), we have

$$\begin{aligned} \|x^* - x^k\| &\leqslant (1 + L/c_k)(1+\sigma)\|y^k - x^k\| \\ &\leqslant (1 + L/\bar{c})(1+\sigma)\|y^k - x^k\|. \end{aligned}$$

By Lemma 4.1 and the above inequality, we conclude that for k large enough it holds that

$$\begin{aligned} \|x^* - x^{k+1}\|^2 &\leqslant \|x^* - x^k\|^2 - (1-\sigma^2)\|y^k - x^k\|^2 \\ &\leqslant \left[1 - \frac{1-\sigma^2}{((1+L/\bar{c})(1+\sigma))^2}\right]\|x^* - x^k\|^2 \\ &= \left[1 - \frac{1-\sigma}{(1+L/\bar{c})^2(1+\sigma)}\right]\|x^* - x^k\|^2, \end{aligned}$$

which means that $\{x^k\}$ converges to x^* at a linear rate. □

Observe that the last relation also implies that if we allow the error tolerance σ_k tend to zero, and the regularization parameters c_k tend to infinity, then the rate of convergence is superlinear.

5. Some Applications

Finally, we give two specific applications, where our general framework appears to be helpful. The first one was developed independently by Tseng [35].

5.1. MODIFIED FORWARD-BACKWARD SPLITTING

Suppose that the operator T has the structure

$$T = A + B,$$

where A is point-to-point, continuous and monotone (with $\mathrm{Dom}(A) = \mathcal{H}$), while B is only maximal monotone. When the operator $(cB + I),\ c > 0$, is easily invertible, the forward-backward splitting method [16, 5, 33, 34] for finding a zero of $A + B$ is of special relevance:

$$x^{k+1} := J(x^k, c_k),$$

where

$$J(x, c) := (I + cB)^{-1}(I - cA)(x), \quad c > 0.$$

This method is known to converge when the inverse of the forward mapping, i.e., the mapping A^{-1}, is strongly monotone [11] (note that this implies Lipschitz continuity of A). The forward-backward splitting is very useful for decomposition, but the requirement of strong monotonicity of A^{-1} is somewhat limiting. In [35] a useful modification is proposed which removes this requirement. The method of [35] has the following structure:

$$\begin{aligned} y^k &:= J(x^k, c_k), \\ x^{k+1} &:= P_X[y^k - c_k(A(y^k) - A(x^k))], \end{aligned} \tag{5.1}$$

where X is any closed convex set such that $X \cap T^{-1}(0) \neq \emptyset$, $P_X[\cdot]$ is the orthogonal projection operator onto the set X, and c_k is chosen to satisfy

$$c_k \| A(J(x^k, c_k)) - A(x^k)) \| \leqslant \theta \| J(x^k, c_k) - x^k \|, \tag{5.2}$$

where $\theta \in (0, 1)$. An Armijo-type linesearch procedure can be used to determine a suitable value of c_k (the continuity of A guarantees that this procedure is well defined, see [35]). We note that the choice of the set X, and the resulting projection in (5.1), are not conceptually important here. To simplify the demonstration, let us assume $\mathrm{Dom}(A) = \mathrm{Dom}(B) = \mathcal{H} = X$.

We next show that the modified forward-backward splitting method (5.1)–(5.2) falls within the framework of Algorithm 3.1. The definition of y^k in (5.1) is equivalent to

$$0 \in c_k(B(y^k) + A(x^k)) + (y^k - x^k).$$

Hence,

$$b^k := (1/c_k)(x^k - y^k) - A(x^k) \in B(y^k). \tag{5.3}$$

Define

$$v^k := A(y^k) + b^k, \qquad r^k := c_k v^k + y^k - x^k.$$

Note that $v^k \in T(y^k)$, so that in the terminology of the present paper, r^k is the error in the solution of the proximal subproblem

$$0 \in c_k T(\cdot) + (\cdot - x^k).$$

Let us evaluate this error:

$$\begin{aligned} r^k &= c_k(A(y^k) + b^k) + (y^k - x^k) \\ &= c_k(A(x^k) + b^k) + (y^k - x^k) + c_k(A(y^k) - A(x^k)) \\ &= c_k(A(y^k) - A(x^k)), \end{aligned}$$

where the last equality follows from (5.3). Using again the definition of y^k, we obtain

$$r^k = c_k\big(A(J(x^k, c_k)) - A(x^k)\big). \tag{5.4}$$

If in the setting of our Algorithm 3.1 we take the operator enlargement parameter $\varepsilon_k = 0$ (the exact operator T is used) and the error tolerance parameter $\sigma = \theta$, then y^k, v^k is an acceptable approximate solution if

$$\|r^k\| \leqslant \theta \|y^k - x^k\|.$$

In view of (5.4) and of the definition of y^k, the above condition is exactly the Armijo-type condition (5.2) used in [35]. Furthermore, the last step in Algorithm 3.1 is

$$\begin{aligned} x^{k+1} &= x^k - c^k v^k \\ &= y^k - r^k \\ &= y^k - c_k(A(y^k) - A(x^k)), \end{aligned}$$

which is again the same as (5.1) (recall that we assumed $X = \mathcal{H}$).

It is now clear that the modified forward-backward splitting of [35] can be regarded as a specific implementation of our general Algorithm 3.1 (the two algorithms have been developed independently). Note that according to Corollary 4.2, the sequences $\{x^k\}$ and $\{y^k\}$ are bounded independently of the choice of $\{c_k\}$. When A is Lipschitz continuous on the bounded set containing those sequences (with modulus $L > 0$, say), it is easy to see that any $c_k \in [0, \theta/L]$ satisfies condition (5.2). In particular, if an Armijo-type search is used to compute c_k, then the condition $c_k \geqslant \bar{c}$ will be satisfied. In that case, convergence properties of the modified forward-backward splitting method follow immediately from Theorems 3.1 and 3.2. We note that convergence of Tseng's method can also be studied if $\liminf_k c_k = 0$, but in that case additional assumptions would be required [35, Theorem 3.1].

In the case when X is a proper subset of $\mathcal{H}$, each iteration of Tseng's algorithm may be regarded as a step of Algorithm 3.1, followed by a projection onto X. Convergence properties will be preserved because projection onto X and the "hybrid" steps decrease the distance from the iterates to the set $X \cap T^{-1}(0)$.

5.2. REGULARIZED NEWTON-TYPE METHOD

Suppose now that $\mathcal{H} = \Re^n$, and $T = F$, where $F\colon \Re^n \to \Re^n$ is monotone and continuously differentiable, with its Jacobian ∇F being Lipschitz continuous, i.e., there exists some $\gamma > 0$ such that

$$\|\nabla F(x) - \nabla F(y)\| \leqslant \gamma \|x - y\|, \quad \forall x, y \in \Re^n.$$

So the problem under consideration is that of solving a system of smooth equations

$$F(x) = 0.$$

The fundamental algorithm in this setting is the Newton method. It is well known that while possessing fast superlinear/quadratic convergence close to *regular* solutions, there are serious difficulties with ensuring global convergence of Newton methods, even if merit functions and linesearch techniques are used for globalization. In [28], a truly globally convergent (on the class of monotone equations) regularized Newton method was developed based on the approximate proximal point scheme of [30]. For globalization, a linesearch technique which ensures a certain separation property (motivated by [26]) was used.

We next outline a different Newton-type method based on Algorithm 3.1. The following fact is the key to the construction: if x^k is *not* a solution, then a *single* step of the Newton method applied to the proximal subproblem

$$0 = c_k F(x) + (x - x^k)$$

is enough to ensure the error tolerance requirements at the k-th step of Algorithm 3.1, provided we choose

$$c_k \in \left[0.1\big(\gamma \|F(x^k)\|\big)^{-1/2}, \big(\gamma \|F(x^k)\|\big)^{-1/2}\right] \quad \text{and} \quad \sigma \in [1/2, 1)$$

(other choices of parameters are also possible). In other words, if

$$y^k := x^k + s^k,$$

where

$$s^k := -(c_k \nabla F(x^k) + I)^{-1} c_k F(x^k),$$

then it holds that

$$\|c_k F(y^k) + y^k - x^k\| \leqslant \sigma \|y^k - x^k\|. \tag{5.5}$$

We next demonstrate this fact. By the Lipschitz continuity of ∇F, it holds that

$$\|F(y^k) - F(x^k) - \nabla F(x^k)(y^k - x^k)\| \leqslant \frac{\gamma}{2} \|y^k - x^k\|^2. \tag{5.6}$$

Since

$$y^k - x^k = s^k = -c_k F(x^k) - c_k \nabla F(x^k) s^k,$$

combining the latter equations with (5.6), we further obtain

$$\frac{1}{c_k}\|c_k F(y^k) + y^k - x^k\| \leqslant \frac{\gamma}{2}\|s^k\|\|y^k - x^k\|.$$

Since $\nabla F(x)$ is positive semidefinite, it can be easily verified that

$$\|s^k\| \leqslant c_k\|F(x^k)\|.$$

Therefore,

$$\|c_k F(y^k) + y^k - x^k\| \leqslant \frac{c_k^2\gamma\|F(x^k)\|}{2}\|y^k - x^k\|.$$

Now (5.5) follows from the choice of σ and c_k.

We conclude that in this setting, Algorithm 3.1 reduces to a Newton-type method, where each iteration consists of one regularized Newton step, followed by an extragradient step. Observe that $\{x^k\}$ is bounded independently of the choice of $\{c_k\}$ (due to Corollary 4.2.1), and so $\{\|F(x^k)\|\}$ is also bounded. Our choice of c_k implies now that $c_k \geqslant \bar{c} > 0$. Applying Theorem 3.1, we immediately deduce that the *whole* sequence generated by this Newton-type method converges *globally* to a solution of the system of equations, provided one exists, without any regularity assumptions. This compares favorably with globalized Newton methods based on merit functions, where even to prove convergence of *subsequences* to stationary points of these functions, one needs to assume the boundedness of their level sets (note that this implies the boundedness of the solution set). Furthermore, in general these stationary points need not be solutions of $F(x) = 0$. And to guarantee even local convergence of the whole sequence, uniqueness of the solution has to be imposed. We refer the reader to [28] for a more detailed discussion.

We note that if the value (or a bound for it) of γ is not available, a suitable c_k can be obtained by an Armijo-type linesearch procedure. Finally, assuming nonsingularity of ∇F at the solution, the local superlinear rate of convergence can be established. Furthermore, under this condition, for k large enough one can take

$$c_k = \frac{1}{\sqrt{\|F(x^k)\|}},$$

and no linesearch is necessary. In this paper, we will not pursue technical details. We refer the reader to [28], where the required analysis is similar.

6. Concluding Remarks

A new method for solving the inclusion problem (or generalized equation) for a maximal monotone operator was presented. One of the key parts of the presented framework is a new concept of approximate solutions of proximal subproblems, where the errors are handled both through the equation residual, and through an enlargement of the operator. This new method gives a very general framework in which algorithms can be designed. To illustrate this, two applications are described. One is the modified forward-backward splitting algorithm proposed by Tseng [35], and the other is the regularized Newton method.

An interesting subject for future research is the use of algorithmically generated elements in T^{ε}, similar to the bundle methodology in nonsmooth convex optimization.

Acknowledgement

We thank the two anonymous referees for careful reading and constructive suggestions which led to the improvement of the paper.

References

1. Auslender, A.: Numerical methods for nondifferentiable convex optimization, *Math. Programming Study* **30** (1987), 102–126.
2. Burachik, R. S., Iusem, A. N. and Svaiter, B. F.: Enlargement of monotone operators with applications to variational inequalities, *Set-Valued Anal.* **5** (1997), 159–180.
3. Burachik, R. S., Sagastizábal, C. A. and Svaiter, B. F.: ε-Enlargements of maximal monotone operators: Theory and applications, in: M. Fukushima and L. Qi (eds), *Reformulation – Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, Kluwer Acad. Publ., 1999, pp. 25–44.
4. Burke, J. V. and Qian, M.: A variable metric proximal point algorithm for monotone operators, *SIAM J. Control Optim.* **37** (1998), 353–375.
5. Chen, H.-G.: Forward-backward splitting techniques: Theory and applications, PhD Thesis, University of Washington, Seattle, Washington, December 1994.
6. Chen, X. and Fukushima, M.: Proximal quasi-Newton methods for nondifferentiable convex optimization, *Math. Programming* (1999), to appear.
7. Cominetti, R.: Coupling the proximal point algorithm with approximation methods, *J. Optim. Theory Appl.* **95** (1997), 581–600.
8. Eckstein, J.: Approximate iterations in Bregman-function-based proximal algorithms, *Math. Programming* **83** (1998), 113–123.
9. Eckstein, J. and Bertsekas, D. P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Math. Programming* **55** (1992), 293–318.
10. Ferris, M. C.: Finite termination of the proximal point algorithm, *Math. Programming* **50** (1991), 359–366.
11. Gabay, D.: Applications of the method of multipliers to variational inequalities, in: M. Fortin and R. Glowinski (eds), *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*, North-Holland, 1983, pp. 299–331.
12. Georgiev, P. Gr.: Submonotone mappings in Banach spaces and applications, *Set-Valued Anal.* **5** (1997), 1–35.
13. Güler, O.: New proximal point algorithms for convex minimization, *SIAM J. Optim.* **2** (1992), 649–664.
14. Korpelevich, G. M.: The extragradient method for finding saddle points and other problems, *Matecon* **12** (1976), 747–756.
15. Lemaire, B.: The proximal algorithm, in: J. P. Penot (ed.), *New Methods of Optimization and Their Industrial Use*, Internat. Ser. Numer. Math. 87, Birkhäuser, Basel, 1989, pp. 73–87.

16. Lions, P. L. and Mercier, B.: Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.* **16** (1979), 964–979.
17. Luque, F. J.: Asymptotic convergence analysis of the proximal point algorithm, *SIAM J. Control Optim.* **22** (1984), 277–293.
18. Martinet, B.: Regularisation d'inequations variationelles par approximations successives, *Revue Française d'Informatique et de Recherche Opérationelle* **4** (1970), 154–159.
19. Minty, G. J.: Monotone (nonlinear) operators in Hilbert space, *Duke Math. J.* **29** (1962), 341–346.
20. Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings, *Bull. Amer. Math. Soc.* **73** (1967), 591–597.
21. Passty, G. B.: Weak convergence theorems for nonexpansive mappings in Banach spaces, *J. Math. Anal. Appl.* **67** (1979), 274–276.
22. Qi, L. and Chen, X.: 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.
23. Rockafellar, R. T.: On the maximality of sums of nonlinear monotone operators, *Trans. Amer. Math. Soc.* **149** (1970), 75–88.
24. Rockafellar, R. T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming, *Math. Oper. Res.* **1** (1976), 97–116.
25. Rockafellar, R. T.: Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.* **14** (1976), 877–898.
26. Solodov, M. V. and Svaiter, B. F.: A new projection method for variational inequality problems, *SIAM J. Control Optim.* **37** (1999), 765–776.
27. Solodov, M. V. and Svaiter, B. F.: Forcing strong convergence of proximal point iterations in a Hilbert space, *Math. Programming* (1997), to appear.
28. Solodov, M. V. and Svaiter, B. F.: A globally convergent inexact Newton method for systems of monotone equations, in: M. Fukushima and L. Qi (eds), *Reformulation – Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, Kluwer Acad. Publ., Dordrecht, 1999, pp. 355–369.
29. Solodov, M. V. and Svaiter, B. F.: An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions, *Math. Oper. Res.* (1998), to appear.
30. Solodov, M. V. and Svaiter, B. F.: A hybrid projection – proximal point algorithm, *J. Convex Anal.* **6**(1) (1999).
31. Spingarn, J. E.: Submonotone subdifferentials of Lipschitz functions, *Trans. Amer. Math. Soc.* **264** (1981), 77–89.
32. Tossings, P.: The perturbed proximal point algorithm and some of its applications, *Appl. Math. Optim.* **29** (1994), 125–159.
33. Tseng, P.: Further applications of a splitting algorithm to decomposition in variational inequalities and convex programming, *Math. Programming* **48** (1990), 249–263.
34. Tseng, P.: Applications of a splitting algorithm to decomposition in convex programming and variational inequalities, *SIAM J. Control Optim.* **29** (1991), 119–138.
35. Tseng, P.: A modified forward-backward splitting method for maximal monotone mappings, Department of Mathematics, University of Washington, Seattle, WA, 1998.