Numerical Functional Analysis and Optimization
Vol. 22, pp. 1013-1035, 2001.

A UNIFIED FRAMEWORK FOR SOME INEXACT PROXIMAL POINT ALGORITHMS*

M. V. Solodov† and B. F. Svaiter†

ABSTRACT

We present a unified framework for the design and convergence analysis of a class of algorithms based on approximate solution of proximal point subproblems. Our development further enhances the constructive approximation approach of the recently proposed hybrid projection–proximal and extragradient–proximal methods. Specifically, we introduce an even more flexible error tolerance criterion, as well as provide a unified view of these two algorithms. Our general method possesses global convergence and local (super)linear rate of convergence under standard assumptions, while using a constructive approximation criterion suitable for a number of specific implementations. For example, we show that close to a regular solution of a monotone system of semismooth equations, two Newton iterations are sufficient to solve the proximal subproblem within the required error tolerance. Such systems of equations arise naturally when reformulating the nonlinear complementarity problem.

Key words. Maximal monotone operator, proximal point algorithm, approximation criteria.

---

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

† Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil.
Email : `solodov@impa.br` and `benar@impa.br` .

1 Introduction and Motivation

We consider the classical problem

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

where $\mathcal{H}$ is a real Hilbert space, and T is a maximal monotone operator (or a multifunction) on $\mathcal{H}$, i.e., $T : \mathcal{H} \to \mathcal{P}(\mathcal{H})$, where $\mathcal{P}(\mathcal{H})$ stands for the family of subsets of $\mathcal{H}$. It is well known that many problems in applied mathematics, economics and engineering can be cast in this general form.

One of the fundamental regularization techniques in this setting is the proximal point method. Using the current approximation to the solution of (1), $x^k \in \mathcal{H}$, this method generates the next iterate x^{k+1} by solving the subproblem

$$0 \in c_k T(x) + (x - x^k)\,, \tag{2}$$

where $c_k > 0$ is a regularization parameter. Equivalently this can be written as

$$x^{k+1} = (c_k T + I)^{-1}(x^k)\,. \tag{3}$$

Theoretically, the proximal point method has quite nice global and local convergence properties [30], see also [19]. Its main practical drawback is that the subproblems are structurally as difficult to solve as the original problem. This makes straightforward application of the method impractical in most situations. Nevertheless, the proximal point methodology is important for the development of a variety of successful numerical techniques, such as operator splitting and bundle methods, to name a few. It is therefore of importance to develop proximal-type algorithms with improved practical characteristics. One possibility in this sense is inexact solution of subproblems. In particular, approximation criteria should be as constructive and realistic as possible. We next discuss previous work in this direction.

The first inexact proximal scheme was introduced in [30], and until very recently criteria of the same (summability) type remained the only choice available (for example, see [41, 7, 11, 9]). Specifically, in [30] equation (3) is relaxed, and the approximation rule is the following:

$$\|x^{k+1} - (c_k T + I)^{-1}(x^k)\| \leq e_k \,, \quad \sum_{k=0}^{\infty} e_k < \infty \,. \tag{4}$$

This condition, together with $\{c_k\}$ being bounded away from zero, ensures convergence of the iterates to some $\bar{x} \in T^{-1}(0)$, provided the latter set is nonempty [30, Theorem 1]. Convergence here is understood in the weak topology. In general, proximal point iterates may fail to converge strongly [17]. However, strong convergence can be forced by some simple modifications [39], see also [1]. Regarding the rate of convergence, the classical result is the following. If the iterates are bounded, $c_k \geq c > 0$, and T^{-1} is Lipschitz-continuous at zero, then condition

$$\|x^{k+1} - (c_k T + I)^{-1}(x^k)\| \leq d_k \|x^{k+1} - x^k\| \,, \quad \sum_{k=0}^{\infty} d_k < \infty \tag{5}$$

implies convergence (in the strong sense) at the linear rate [30, Theorem 2]. It is interesting to note that by itself, (5) does not guarantee convergence of the sequence. Only if T^{-1} is globally Lipschitz-continuous (which is a rather strong assumption), then the boundedness assumption on iterates is superfluous [30, Proposition 5].

Regarding the error tolerance rules (4) and (5), it is important to keep in mind that $(I + c_k T)^{-1}(x^k)$ is the point to be approximated, so it is not available. Therefore, the above two criteria usually cannot be directly used in implementations of the algorithm. In [30, Proposition 3] it was established that (4), (5) are implied, respectively, by the following two conditions:

$$\mathrm{dist}(0, c_k T(x^{k+1}) + x^{k+1} - x^k) \leq e_k \,, \quad \sum_{k=0}^{\infty} e_k < \infty \,, \tag{6}$$

and

$$\mathrm{dist}(0, c_k T(x^{k+1}) + x^{k+1} - x^k) \leq d_k \|x^{k+1} - x^k\| \,, \quad \sum_{k=0}^{\infty} d_k < \infty \,. \tag{7}$$

These two conditions are usually easier to verify in practice, because they only require evaluation of T at x^{k+1}. Note that since $T(x)$ is closed, the inexact proximal point algorithm managed by criteria (6) or (7) can be restated as follows. Having a current iterate x^k, find $(x^{k+1}, v^{k+1}) \in \mathcal{H} \times \mathcal{H}$ such that

$$\begin{cases} v^{k+1} \in T(x^{k+1}), \\ c_k v^{k+1} + x^{k+1} - x^k = r^k, \end{cases} \tag{8}$$

where $r^k \in \mathcal{H}$ is the error associated with approximation. With this notation, criterion (6) becomes

$$\|r^k\| \leq e_k\,, \quad \sum_{k=0}^{\infty} e_k < \infty\,, \tag{9}$$

and criterion (7) becomes

$$\|r^k\| \leq d_k \|x^{k+1} - x^k\|\,, \quad \sum_{k=0}^{\infty} d_k < \infty\,. \tag{10}$$

One question which arises immediately if these rules are to be used, is how to choose the value of e_k or d_k for each k when solving a specific problem. Obviously, the number of choices at every step is infinite, and it is not quite clear how the choice should be related to the behavior of the method on the given problem. This goes to say that, in this sense, these rules are not of constructive nature. Developing constructive approximation criteria is one of the subjects of the present paper (following [37, 36]). Another important question is how to solve the subproblems efficiently within the approximation rule, once it is chosen. This question can be addressed most effectively under some further assumptions on the structure of the operator T, e.g., T is the subdifferential of a convex function, T is smooth, T is a sum of two operators with certain structure, T represents a variational inequality, etc. Developments for some of these cases can be found in [35, 43, 36, 34, 38], where some Newton-type and splitting-type methods were considered. We note that the use of constructive approximation criteria of [37, 36] was crucial in those references. The case of a general maximal monotone operator was discussed in [31], where subproblems are solved by bundle techniques, similar to [4]. In this paper we consider one more application, which has to do with solving a system of semismooth (but not in general differentiable) systems of equations. Monotone systems of this type arise, for example, in reformulation of complementarity problems.

The next question we shall discuss is whether and how (9), (10) can be relaxed. A natural rule to consider in this setting would be

$$\|r^k\| \leq \sigma \|x^{k+1} - x^k\| , \quad 0 \leq \sigma < 1 , \tag{11}$$

where, for simplicity, the relaxation parameter σ is fixed (the important point is not to have it necessarily fixed, but rather to keep it bounded away from zero). The above condition is obviously weaker than (10). It is also less restrictive than (9), because a sequence $\{x^k\}$ generated by the proximal point method based on it may not converge (unlike when (9) is used). We refer the reader to [37], where an example of divergence in a finite-dimensional space R^2 is given. Hence, condition (11) cannot be used without modifications to the algorithm. Fortunately, modifications that are needed are quite simple and do not increase the computational burden per iteration of the algorithm. This development is another goal of this paper, following [37, 36]. We note that conditions in the spirit of (11) are very natural and computationally realistic. We mention classical inexact Newton methods, e.g. [10], as one example where similar approximations are used.

We now turn our attention to the more recent research in the area of inexact proximal point methods. Let us consider the "proximal system"

$$\begin{cases} v \in T(x), \\ c_k v + x - x^k = 0, \end{cases} \tag{12}$$

which is clearly equivalent to the subproblem (2). This splitting of equation and inclusion seems to be a natural view of subproblems, the advantages of which will be further evident later. In [37], the following notion of inexact solutions was introduced. A pair $\tilde{x}^{k+1}, \tilde{v}^{k+1}$ is an *approximate* solution of (12) if

$$\begin{cases} \tilde{v}^{k+1} \in T(\tilde{x}^{k+1}) \\ c_k \tilde{v}^{k+1} + \tilde{x}^{k+1} - x^k = r^k \end{cases} \tag{13}$$

and

$$\|r^k\| \leq \sigma \max\{c_k \|\tilde{v}^{k+1}\|, \|\tilde{x}^{k+1} - x^k\|\}, \tag{14}$$

where $\sigma \in [0,1)$ is *fixed*. Note that this condition is even more general than (11). The algorithm of [37], called the Hybrid Projection–Proximal Point Method (HPPPM), proceeds as follows. Instead of taking $\tilde{x}^{k+1}$ as the next iterate x^{k+1} (recall that this may not work even in R^2), x^{k+1} is defined as the projection of x^k onto the halfspace

$$H_{k+1} := \{x \in \mathcal{H} \mid \langle \tilde{v}^{k+1}, x - \tilde{x}^{k+1} \rangle \leq 0\}.$$

Note that since $\tilde{v}^{k+1} \in T(\tilde{x}^{k+1})$, any solution of (1) belongs to H_{k+1}, by the monotonicity of T. On the other hand, using (14), it can be verified that if x^k is not a solution then

$$\langle \tilde{v}^{k+1}, x^k - \tilde{x}^{k+1} \rangle > 0,$$

and so $x^k \notin H_{k+1}$. Hence, the projected point x^{k+1} is closer to the solution set $T^{-1}(0)$ than x^k, which essentially ensures global convergence of the method. Recall that projection onto a halfspace is explicit (so it does not entail any nontrivial computational cost). Specifically, it is given by

$$x^{k+1} := x^k - \frac{\langle \tilde{v}^{k+1}, x^k - \tilde{x}^{k+1} \rangle}{\|\tilde{v}^{k+1}\|^2} \tilde{v}^{k+1}. \tag{15}$$

It is also worth to mention that the same rule (14) used for global convergence also gives local linear rate of convergence under standard assumptions. Observe that if $(\tilde{x}^{k+1}, \tilde{v}^{k+1})$ happens to be the *exact* solution of (12) (for example, this will necessarily be the case if one takes $\sigma = 0$ in (14)), then

x^{k+1} will be equal to $\tilde{x}^{k+1}$, and we retrieve the exact proximal iteration. Of course, the case of interest is when $\sigma \neq 0$. Convergence properties of HPPPM are precisely the same as of the standard proximal point method, see [37]. The advantage of HPPPM is that the approximation rule (14) is constructive and less restrictive than the classical one. This is important not only theoretically, but even more so in applications. For example, tolerance rule (14) proved crucial in the design of *truly* globally convergent Newton-type algorithms [35, 34], which combine the proximal and projection methods of [37, 33] with the linearization methodology. An extension of (14) to proximal methods with Bregman-distance regularization can be found in [40].

Let us go back to the proximal system (12). Up to now, we discussed relaxing the equation part of this system. The inclusion $v \in T(x)$ can also be relaxed using some "outer approximation" of T. In [3] T^{ε}, an enlargement of a maximal monotone operator T, was introduced. Specifically, given $\varepsilon \geq 0$, define $T^{\varepsilon} : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ as

$$T^{\varepsilon}(x) = \{v \in \mathcal{H} \mid \langle w - v, z - x\rangle \geq -\varepsilon \text{ for all } z \in \mathcal{H},\, w \in T(z)\}\,. \tag{16}$$

Since T is maximal monotone, $T^0(x) = T(x)$ for all $x \in \mathcal{H}$. Furthermore, the relation

$$T(x) \subseteq T^{\varepsilon}(x)$$

holds for any $\varepsilon \geq 0$, $x \in \mathcal{H}$. So, T^{ε} may be seen as a certain outer approximation of T. In the particular case $T = \partial f$, where f is a proper closed convex function, it holds that $T^{\varepsilon}(x) \supseteq \partial_{\varepsilon} f(x)$, where $\partial_{\varepsilon} f$ stands for the standard ε-subdifferential as defined in Convex Analysis. Some further properties and applications of T^{ε} can be found in [3, 5, 4, 6, 38, 31].

In [36], a variation of the proximal point method was proposed, where both the equation and the inclusion were relaxed in system (12). In addition, the projection step of [37] was replaced by a more simple extragradient-like step. This algorithm is called the Hybrid Extragradient–Proximal Point Method (HEPPM). Specifically, HEPPM works as follows. A pair $(\tilde{x}^{k+1}, \tilde{v}^{k+1})$

is considered an acceptable approximate solution of the proximal system (12) if for some $\varepsilon_{k+1} \geq 0$

$$\begin{cases} \tilde{v}^{k+1} \in T^{\varepsilon_{k+1}}(\tilde{x}^{k+1}), \\ c_k\tilde{v}^{k+1} + \tilde{x}^{k+1} - x^k = r^k, \end{cases} \tag{17}$$

and

$$\|r^k\|^2 + 2c_k\varepsilon_{k+1} \leq \sigma^2\|\tilde{x}^{k+1} - x^k\|^2, \tag{18}$$

where $\sigma \in [0, 1)$ is fixed. Having in hand those objects, the next iterate is

$$x^{k+1} := x^k - c_k\tilde{v}^{k+1}. \tag{19}$$

Convergence properties of this method are again the same as of the original proximal point algorithm. We note that for $\varepsilon_{k+1} = 0$, condition (18) for HEPPM is somewhat stronger than condition (14) for HPPPM, although it should not be much more difficult to satisfy in practice. It has to be stressed, however, that the introduction of $\varepsilon_{k+1} \neq 0$ is not merely formal or artificial here. It is important for several reasons, see [36, 38] for a more detailed discussion. Here, we shall mention that T^ε arises naturally in the context of variational inequalities where subproblems are solved using the regularized gap function [14]. Another situation where T^ε is useful are bundle techniques for general maximal monotone inclusions, see [4, 31] (unlike T, T^ε has certain continuity properties [5], much like the ε-subdifferential in nonsmooth convex optimization).

The main goal of the present paper is to unify the methods of [37] and [36] within a more general framework which, in addition, utilizes an even better approximation criterion for solution of subproblems than the ones studied previously. A new application to solving monotone systems of semismooth equations is also discussed.

2 The general framework

Given any $x \in \mathcal{H}$ and $c > 0$, consider the proximal point problem of solving the system

$$\begin{cases} v \in T(y), \\ cv + y - x = 0. \end{cases} \tag{20}$$

We start directly with our new notion of approximate solutions of (20).

Definition 1 *We say that a triple $(y, v, \varepsilon) \in \mathcal{H} \times \mathcal{H} \times R_+$ is an approximate solution of the proximal system* (20) *with error tolerance $\sigma \in [0, 1)$, if*

$$\begin{aligned} & v \in T^{\varepsilon}(y)\,, \\ & \|cv + y - x\|^2 + 2c\varepsilon \le \sigma^2 \left(\|cv\|^2 + \|y - x\|^2 \right)\,. \end{aligned} \tag{21}$$

First note that condition (21) is more general than either (14) or (18). This is easy to see because the right-hand side in (21) is larger. In addition, (14) works only with the exact values of the operator (i.e., with $\varepsilon = 0$).

Observe that if (y, v) is the *exact* solution of (20) then, taking $\varepsilon = 0$, we conclude that (y, v, ε) satisfy the approximation criteria (21) for any $\sigma \in [0, 1)$. Conversely, if $\sigma = 0$, then only the exact solution of (20) (with $\varepsilon = 0$) will satisfy this approximation criterion. So this definition is quite natural. For $\sigma \in (0, 1)$, system (20) has at least one, and typically many, approximate solutions in the sense of Definition 1.

We next establish some basic properties of approximate solutions defined above.

Lemma 2 *Take $x \in \mathcal{H}$, $c > 0$. The triple $(y, v, \varepsilon) \in \mathcal{H} \times \mathcal{H} \times R_+$ being an approximate solution of the proximal system* (20) *with error tolerance $\sigma \in [0, 1)$ is equivalent to the following relations:*

$$v \in T^{\varepsilon}(y)\,, \quad \langle v, x - y \rangle - \varepsilon \ge \frac{1 - \sigma^2}{2c} \left(\|cv\|^2 + \|y - x\|^2 \right)\,. \tag{22}$$

In addition, it holds that

$$\frac{1 - \rho}{1 - \sigma^2} c\|v\| \le \|y - x\| \le \frac{1 + \rho}{1 - \sigma^2} c\|v\|\,, \tag{23}$$

where

$$\rho := \sqrt{1-(1-\sigma^2)^2}\,.$$

Furthermore, the three conditions

1. $0 \in T(x)$,

2. $v = 0$,

3. $y = x$

are equivalent and imply $\varepsilon = 0$.

Proof. Re-writing (21), we have

$$\begin{aligned}\sigma^2\left(\|cv\|^2 + \|y-x\|^2\right) &\geq \|cv + y - x\|^2 + 2c\varepsilon \\ &= 2c\varepsilon + \|cv\|^2 + \|y-x\|^2 + 2c\langle v, y-x\rangle\,,\end{aligned}$$

which is further equivalent to (22), by a simple rearrangement of terms.

Furthermore, by the Cauchy-Schwarz inequality and the nonnegativity of ε, using (22), we have that

$$\|cv\|\|y-x\| \geq c\langle v, y-x\rangle - c\varepsilon \geq \frac{1-\sigma^2}{2}\left(\|cv\|^2 + \|y-x\|^2\right)\,.$$

Denoting $t := \|y-x\|$ and resolving the quadratic inequality in t

$$t^2 - \frac{2\|cv\|}{1-\sigma^2}\,t + \|cv\|^2 \leq 0\,,$$

proves (23).

Suppose now that $0 \in T(x)$. Since $v \in T^{\varepsilon}(y)$, we have that

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

and it follows that the right-hand side of (22) is zero. Hence, $v = 0$, and $x = y$. If we assume that $v = 0$, then (22) implies that $x = y$, and vice versa.

And obviously, these conditions imply that $0 \in T(x)$. It is also clear from (22) that in those cases $\varepsilon = 0$. ∎

The following lemma is the key to developing convergent iterative schemes based on approximate solutions of proximal subproblems defined above.

Lemma 3 *Take any $x \in \mathcal{H}$. Suppose that*

$$\langle v, x - y \rangle - \varepsilon > 0\,,$$

where $\varepsilon \geq 0$ and $v \in T^{\varepsilon}(y)$. Then for any $x^* \in T^{-1}(0)$ and any $\tau \geq 0$, it holds that

$$\|x^* - x^+\|^2 \leq \|x^* - x\|^2 - (1 - (1 - \tau)^2)\|av\|^2\,,$$

where

$$\begin{aligned} x^+ &:= x - \tau a v\,, \\ a &:= \frac{\langle v, x - y \rangle - \varepsilon}{\|v\|^2}\,. \end{aligned}$$

Proof. Define the closed halfspace H as

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

By the assumption, $x \notin H$. Let $\bar{x}$ be the projection of x onto H. It is well known, and easy to check, that

$$\bar{x} = x - av\,,$$

where the quantity a is defined above. For any $z \in H$, it holds that $\langle z - \bar{x}, v \rangle \leq 0$, and so

$$\langle z - \bar{x}, x^+ - x \rangle \geq 0\,.$$

Using the definition of x^+, we obtain

$$\begin{aligned}
&\|z-x\|^2 = \|(z-x^+)+(x^+-x)\|^2 \\
&= \|z-x^+\|^2 + \|x^+-x\|^2 + 2\langle z-x^+, x^+-x\rangle \\
&= \|z-x^+\|^2 + \|x^+-x\|^2 + 2\langle \bar{x}-x^+, x^+-x\rangle \\
&\quad +2\langle z-\bar{x}, x^+-x\rangle \\
&\geq \|z-x^+\|^2 + \|x^+-x\|^2 + 2\langle \bar{x}-x^+, x^+-x\rangle \\
&= \|z-x^+\|^2 + (1-(1-\tau)^2)a^2\|v\|^2 .
\end{aligned}$$

Suppose $x^* \in T^{-1}(0)$. By the definition of T^ε,

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

Therefore, $\langle v, x^*-y\rangle - \varepsilon \leq 0$, which means that $x^* \in H$. Setting $z = x^*$ in the chain of inequalities above completes the proof. ■

Lemma 3 shows that if we take $\tau \in (0,2)$ then the point x^+ is closer to the solution set $T^{-1}(0)$ than the point x. Using further Lemma 2, this can serve as a basis for constructing a convergent iterative algorithm. Allowing the parameter τ to vary within the interval $(0,2)$ will make it possible to unify the algorithms of [37] and [36].

Algorithm 2.1 *Choose any $x^0 \in \mathcal{H}$, $0 \leq \bar{\sigma} < 1$, $\bar{c} > 0$, and $\theta \in (0,1)$.*

1. *Choose $c_k \geq \bar{c}$ and find $(v^k, y^k, \varepsilon_k) \in \mathcal{H} \times \mathcal{H} \times R_+$, an approximate solution with error tolerance $\sigma_k \leq \bar{\sigma}$ of the proximal system* (12), *i.e.,*

$$\begin{aligned}
&v^k \in T^{\varepsilon_k}(y^k) \\
&\|c_k v^k + y^k - x^k\|^2 + 2c_k\varepsilon_k \leq \sigma_k^2(\|c_k v^k\|^2 + \|y^k - x^k\|^2) .
\end{aligned}$$

2. *If $y^k = x^k$,* STOP. *Otherwise,*

3. *Choose $\tau_k \in [1-\theta, 1+\theta]$, and set*

$$\begin{aligned}
a_k &:= \frac{\langle v^k, x^k - y^k\rangle - \varepsilon_k}{\|v^k\|^2} , \\
x^{k+1} &:= x^k - \tau_k a_k v^k ,
\end{aligned}$$

and go to Step 1.

Note that the algorithm stops when $y^k = x^k$ which, by Lemma 2, means that $x^k \in T^{-1}(0)$. In our convergence analysis we shall assume that this does not occur, and so an infinite sequence of iterates is generated.

We next show that Algorithm 2.1 contains the methods proposed in [37] and [36] as its special cases. First, as was already remarked, any approximate solution of the proximal point subproblem satisfying (14) or (18) (used in [37] and [36], respectively), certainly satisfies approximation criterion (21) employed in Algorithm 2.1. It remains to show that the ways x^{k+1} is obtained in (15) and (19) (used in [37] and [36], respectively), are also two special cases of the update rule in Algorithm 2.1. With respect to [37] this is completely obvious, as (15) corresponds to taking $\tau_k = 1$ in Algorithm 2.1 (with $\varepsilon_k = 0$, as in the setting of [37]). For the method of [36], on the other hand, this issue is not immediately clear. Essentially, one has to show that under the approximation criterion (18), there exists $\tau_k \in [1-\theta, 1+\theta]$ such that $\tau_k a_k = c_k$, or equivalently, there exists $\theta \in (0,1)$ such that $(1-\theta)a_k \le c_k \le (1+\theta)a_k$. The following proposition establishes this fact.

Proposition 4 *Take* $x, y, v \in \mathcal{H}$, $\varepsilon \ge 0$, $c > 0$ *and* $\sigma \in [0,1)$ *such that* $0 \notin T(x)$ *and*

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

Then

$$(1-\sigma)a \le c \le (1+\sigma)a \,,$$

where

$$a = \frac{\langle v, x - y\rangle - \varepsilon}{\|v\|^2} \,.$$

Proof. First note that by the hypotheses of the proposition, Lemma 2 is applicable, and we have that $v \ne 0$ and $a > 0$.

Furthermore, by our assumptions it holds that $\|cv + y - x\| \le \sigma\|y - x\|$, from which by the triangle inequality, it follows that

$$(1-\sigma)\|y-x\| \le c\|v\| \le (1+\sigma)\|y-x\|\,. \tag{24}$$

By the nonnegativity of ε and the Cauchy-Schwarz inequality, we have that

$$a \le \langle v, x-y\rangle/\|v\|^2 \le \|y-x\|/\|v\|\,.$$

Combining the latter relation with (24), we conclude that

$$(1-\sigma)a \le (1-\sigma)\|y-x\|/\|v\| \le c\,,$$

which proves one part of the assertion.

To prove the other part, note that under our assumptions

$$\begin{aligned} a &= \frac{\|cv\|^2 + \|y-x\|^2 - (\|cv+y-x\|^2 + 2c\varepsilon)}{2c\|v\|^2} \\ &\ge \frac{\|cv\|^2 + (1-\sigma^2)\|y-x\|^2}{2c\|v\|^2} \\ &= \frac{c}{2}\left(1 + (1-\sigma^2)\frac{\|x-y\|^2}{\|cv\|^2}\right)\,. \end{aligned}$$

Using further (24), we obtain

$$a \ge (c/2)\left(1 + \frac{1-\sigma^2}{(1+\sigma)^2}\right) = c/(1+\sigma)\,,$$

which concludes the proof. ∎

Proposition 4 implies that if we choose $\theta \ge \bar{\sigma}$ in Algorithm 2.1, then for each k there exists $\tau_k \in [1-\theta, 1+\theta]$ such that $\tau_k a_k = c_k$. Hence, HEPPM falls within the presented framework of Algorithm 2.1.

3 Convergence analysis

As already remarked, if Algorithm 2.1 terminates finitely, it does so at a solution of the problem. From now on, we assume that infinite sequences

$\{x^k\}$, $\{v^k\}$, $\{y^k\}$ and $\{\varepsilon_k\}$ are generated. Using Lemma 2 we conclude that for all k, $v^k \neq 0$, $y^k \neq x^k$ and

$$\langle v^k, x^k - y^k \rangle - \varepsilon_k \geq \frac{1-\sigma_k^2}{2c_k} \left(\left\| c_k v^k \right\|^2 + \left\| y^k - x^k \right\|^2 \right) > 0 \,. \tag{25}$$

Therefore, by the definition of a_k,

$$a_k \geq \frac{1-\sigma_k^2}{2} \, \frac{\|c_k v^k\|^2 + \|y^k - x^k\|^2}{c_k \|v^k\|^2} \,. \tag{26}$$

Furthermore, using the fact that $\|c_k v^k\|^2 + \|y^k - x^k\|^2 \geq 2c_k \|v^k\| \|y^k - x^k\|$, we obtain that

$$a_k \|v^k\| \geq (1-\sigma_k^2) \, \|y^k - x^k\| \,. \tag{27}$$

Proposition 5 *If the solution set of problem* (1) *is nonempty, then*

1. *The sequence* $\{x^k\}$ *is bounded;*
2. $\sum_{k=0}^{\infty} \|a_k v^k\|^2 < \infty$*;*
3. $\lim_{k\to\infty} \|y^k - x^k\| = \lim_{k\to\infty} \|v^k\| = \lim_{k\to\infty} \varepsilon_k = 0$.

Proof. Applying (25) and Lemma 3, we have that for any $x^* \in T^{-1}(0)$ it holds that

$$\begin{aligned} \|x^* - x^{k+1}\|^2 &\leq \|x^* - x^k\|^2 - (1-(1-\tau_k)^2) a_k^2 \|v^k\|^2 \\ &\leq \|x^* - x^k\|^2 - (1-\theta^2) \|a_k v^k\|^2 \,. \end{aligned}$$

It immediately follows that the sequence $\{x^k\}$ is bounded, and that the second item of the proposition holds. Therefore we also have that $0 = \lim_{k\to\infty} a_k \|v^k\|$. From (26) it easily follows that

$$a_k \|v^k\| \geq \frac{1-\sigma_k^2}{2} \, \|c_k v^k\| \,.$$

Using this relation and (27), we immediately conclude that $0 = \lim_{k\to\infty} \|v^k\| = \lim_{k\to\infty} \|y^k - x^k\|$. Since $\varepsilon_k \geq 0$, relation (25) also implies that $0 = \lim_{k\to\infty} \varepsilon_k$.

■

We are now in position to prove convergence of our algorithm.

Theorem 6 *If the solution set of problem* (1) *is nonempty, then the sequence* $\{x^k\}$ *converges weakly to a solution.*

Proof. By Proposition 5, the sequence $\{x^k\}$ is bounded, and so it has at least one weak accumulation point, say $\bar{x}$. Let $\{x^{k_j}\}$ be some subsequence weakly converging to $\bar{x}$. Since $\|y^k - x^k\| \to 0$ (by Proposition 5), it follows that the subsequence $\{y^{k_j}\}$ also converges weakly to $\bar{x}$. Moreover, we know that $v^{k_j} \to 0$ (strongly) and $\varepsilon_{k_j} \to 0$. Take any $x \in \mathcal{H}$ and $u \in T(x)$. Because $v^k \in T^{\varepsilon_k}(y^k)$, for any index j it holds that

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

Therefore

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

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

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

But (x, u) was taken as an arbitrary element in the graph of T. From maximality of T it follows that $0 \in T(\bar{x})$, i.e., $\bar{x}$ is a solution of (1). We have thus proved that every weak accumulation point of $\{x^k\}$ is a solution.

The proof of uniqueness of weak accumulation point in this setting is standard. For example, uniqueness follows easily by applying Opial's Lemma [22] (observe that $\{\|x^k - \bar{x}\|\}$ is nonincreasing), or by using the analysis of [30]. ■

When $T^{-1}(0) = \emptyset$, i.e., no solution exists, the sequence $\{x^k\}$ can be shown to be unbounded. Because one has to work with an enlargement of the operator, the proof of this fact is somewhat different from the classical (e.g., [30]). We refer the reader to [36] where the required analysis is similar.

We next turn our attention to the study of convergence rate of Algorithm 2.1. This will be done under the standard assumption that T^{-1} is Lipschitz-continuous at zero, i.e., there exist the unique $x^* \in T^{-1}(0)$ and some $\delta, L > 0$ such that

$$v \in T(x)\,,\ \|v\| \le \delta \quad \Rightarrow \quad \|x - x^*\| \le L\|v\|\,. \tag{28}$$

We shall assume, for the sake of simplicity, that $\tau_k = 1$ for all k, i.e., there is no relaxation in the "projection" step of Algorithm 2.1.

The following *error bound*, established in [38], will be the key to the analysis.

Theorem 7 *[38, Corollary 2.1] Let y^*, v^* be the exact solution of the proximal system* (20)*. Then for any $y \in \mathcal{H}, v \in T^{\varepsilon}(y)$, it holds that*

$$\|y - y^*\|^2 + c^2\|v - v^*\|^2 \le \|cv + y - x\|^2 + 2c\varepsilon\,.$$

A lower bound for a_k will also be needed. Using (26), we have

$$\begin{aligned} a_k &\ge \frac{1-\sigma_k^2}{2}(1 + (\|y^k - x^k\|/\|c_k v^k\|)^2)c_k \\ &\ge \frac{1-\bar{\sigma}^2}{2}(1 + (\|y^k - x^k\|/\|c_k v^k\|)^2)\bar{c}\,. \end{aligned}$$

Using also (23), after some algebraic manipulations, we obtain

$$a_k \ge \left[\frac{1 + \sqrt{1 - (1-\bar{\sigma}^2)^2}}{1 - \bar{\sigma}^2}\right]^{-1} \bar{c}\,. \tag{29}$$

We are now ready to establish the linear rate of convergence of our algorithm.

Theorem 8 *Suppose that T^{-1} is Lipschitz-continuous at zero, and $\tau_k = 1$ for all k. Then for k sufficiently large,*

$$\|x^* - x^{k+1}\| \le \frac{\lambda}{\sqrt{1+\lambda^2}}\|x^* - x^k\|\,,$$

where

$$\lambda := \sqrt{\alpha^2+1}\sqrt{\beta^2-1}+\alpha\beta$$

with

$$\alpha := (L/\bar{c})\frac{1+\sqrt{1-(1-\bar{\sigma}^2)^2}}{1-\bar{\sigma}^2}\,, \quad \beta := 1/(1-\bar{\sigma}^2)\,.$$

Proof. Define, for each k, $\hat{x}^k$, $\hat{v}^k \in T(\hat{x}^k)$ as the exact solution of the proximal problem $0 \in a_k T(x) + x - x^k$, where a_k was defined in Algorithm 2.1 (recall that $a_k > 0$). Since $v^k \in T^{\varepsilon_k}(y^k)$, by Theorem 7 it follows that

$$\begin{aligned}
&\|\hat{x}^k - y^k\|^2 + a_k^2\|\hat{v}^k - v^k\|^2 \le \|a_k v^k + y^k - x^k\|^2 + 2a_k\varepsilon_k \\
&= \;\; \|a_k v^k + y^k - x^k\|^2 + 2a_k(\langle v^k, x^k - y^k\rangle - a_k\|v^k\|^2) \\
&= \;\; \|y^k - x^k\|^2 - \|a_k v^k\|^2.
\end{aligned}$$

Using also (27) we get

$$\|\hat{x}^k - y^k\|^2 + a_k^2\|\hat{v}^k - v^k\|^2 \le ((1-\sigma_k^2)^{-2} - 1)\|a_k v^k\|^2. \tag{30}$$

Therefore,

$$\|\hat{v}^k - v^k\|^2 \le ((1-\bar{\sigma}^2)^{-2}) - 1)\|v^k\|^2.$$

Now, since v_k converges to zero (strongly) we conclude that $\hat{v}^k$ also converges to zero. Hence, for k large enough, it holds that $\|\hat{v}^k\| \le \delta$, and, by (28),

$$\|x^* - \hat{x}^k\| \le L\|\hat{v}^k\| = (L/a_k)\|\hat{x}^k - x^k\|\,.$$

In the sequel, we assume that k is large enough, so that the above bound holds. By the triangle inequality, we further obtain

$$\begin{aligned}
\|x^* - x^{k+1}\| &\le \|x^* - \hat{x}^k\| + \|\hat{x}^k - x^{k+1}\| \\
&\le (L/a_k)\|\hat{x}^k - x^k\| + \|\hat{x}^k - x^{k+1}\| \\
&\le (L/a_k)\|\hat{x}^k - y^k\| + (L/a_k)\|y^k - x^k\| + \|\hat{x}^k - x^{k+1}\|\,.
\end{aligned}$$

Note that because $\hat{x}^k = x^k - a_k \hat{v}^k$ and $x^{k+1} = x^k - a_k v^k$, we have that $a_k \|\hat{v}^k - v^k\| = \|\hat{x}^k - x^{k+1}\|$. Using this relation and (30), we obtain that

$$\|\hat{x}^k - y^k\|^2 + \|\hat{x}^k - x^{k+1}\|^2 \leq [(1 - \sigma_k^2)^{-2} - 1]\|a_k v^k\|^2 .$$

Using the Cauchy-Schwarz inequality, it now follows that

$$\begin{array}{rl} (L/a_k)\|\hat{x}^k - y^k\| + \|\hat{x}^k - x^{k+1}\| & \\ \leq & \sqrt{(L/a_k)^2 + 1}\sqrt{\|\hat{x}^k - y^k\|^2 + \|\hat{x}^k - x^{k+1}\|^2} \\ \leq & \sqrt{(L/a_k)^2 + 1}\sqrt{(1 - \sigma_k^2)^{-2} - 1}\|a_k v^k\| . \end{array}$$

Hence,

$$\begin{array}{rl} \|x^* - x^{k+1}\| & \\ \leq & \sqrt{(L/a_k)^2 + 1}\sqrt{(1 - \sigma_k^2)^{-2} - 1}\|a_k v^k\| + (L/a_k)\|y^k - x^k\| \\ \leq & \left[\sqrt{(L/a_k)^2 + 1}\sqrt{(1 - \sigma_k^2)^{-2} - 1} + (L/a_k)(1 - \sigma_k^2)^{-1}\right] \|a_k v^k\| , \end{array}$$

where (27) was used again. Recalling the definition of λ, and using (29), we get

$$\|x^* - x^{k+1}\| \leq \lambda \|a_k v^k\| .$$

Using also Lemma 3, we now have that

$$\begin{array}{rcl} \|x^* - x^k\|^2 & \geq & \|x^* - x^{k+1}\|^2 + \|a_k v^k\|^2 \\ & \geq & \|x^* - x^{k+1}\|^2 + (1/\lambda)^2 \|x^* - x^{k+1}\|^2 , \end{array}$$

and it follows that

$$\|x^* - x^{k+1}\| \leq \frac{\lambda}{\sqrt{1 + \lambda^2}} \|x^* - x^k\| ,$$

which establishes the claim. ∎

Remark 9 *It is easy to see that we could repeat the analysis of Theorem 8 with α_k, β_k and λ_k defined for each k as functions of c_k and σ_k (instead of their bounds $\bar{c}$ and $\bar{\sigma}$). Then the analysis above shows that if $c_k \to \infty$ and*

$\sigma_k \to 0$, *then* $\alpha_k \to 0\,, \beta_k \to 1$, *and hence,* $\lambda_k \to 0$. *We conclude that under the additional assumptions that* $c_k \to \infty$ *and* $\sigma_k \to 0$, *in the setting of this section* $\{x^k\}$ *converges to* x^* superlinearly.

In the more general case where $\tau_k \in [1-\theta, 1+\theta]$, under the same condition (28), the following estimate is valid:

$$\|x^* - x^{k+1}\| \le \frac{\lambda+\theta}{\sqrt{(1-\theta^2)+(\lambda+\theta)^2}}\|x^* - x^k\|\,.$$

4 Systems of semismooth equations and reformulations of monotone complementarity problems

In this section, we consider a specific situation where

$$\mathcal{H} = R^n\,,\ T = F\,,\ F : R^n \to R^n\,,$$

and F is a locally Lipschitz-continuous *semismooth* monotone function. As will be discussed below, one interesting example of problems having this structure are equation-based reformulations [26, 15] of monotone nonlinear complementarity problems [24, 12]. A globally convergent Newton method for solving systems of monotone equations was proposed in [35]. The attractive feature of this algorithm is that given an arbitrary starting point, it is guaranteed to generate a sequence of iterates convergent to a solution of the problem without any regularity-type assumptions. This is a very desirable property, not shared by standard globalizations strategies based on merit functions. We refer the reader to [35, 34] for a detailed discussion of this issue. We next state a related method based on the more general framework of Algorithm 2.1.

Algorithm 4.1 *Choose any* $x^0 \in R^n$, $0 \le \bar\sigma < 1$, $\bar c > 0$, *and* $\beta\,, \theta \in (0, 1)$.

1. **(Newton Step)**
   Choose $c_k \geq \bar{c}$, $\sigma_k \leq \bar{\sigma}$, and a positive semidefinite matrix H_k. Compute d^k, the solution of the linear system

   $$F(x^k) + (H_k + c_k^{-1} I) d = 0 \,.$$

   Stop if $d^k = 0$. If

   $$\|c_k F(x^k + d^k) + d^k\|^2 \leq \sigma_k^2 (\|c_k F(x^k + d^k)\|^2 + \|d^k\|^2) \,,$$

   then set $y^k := x^k + d^k$, and go to **Projection Step**. *Otherwise,*

2. **(Linesearch Step)**
   Find m_k, the smallest nonnegative integer m, such that

   $$-c_k \langle F(x^k + \beta^m d^k), d^k \rangle \geq \sigma_k^2 \|d^k\|^2 \,.$$

   Set $y^k := x^k + \beta^{m_k} d^k$.

3. **(Projection Step)**
   Choose $\tau_k \in [1 - \theta, 1 + \theta]$, and set

   $$x^{k+1} := x^k + \tau_k \frac{\langle F(y^k), d^k \rangle}{\|F(y^k)\|^2} \, F(y^k) \,.$$

   Set $k := k + 1$, and go to **Newton Step**.

The key idea in Algorithm 4.1 is to try solving the proximal point subproblem by just one Newton-type iteration. When this is not possible, the Newton point is refined by means of a linesearch. The following properties for Algorithm 4.1 can be established essentially using the line of analysis developed in [35].

Theorem 10 *Suppose that F is continuous and monotone and let $\{x^k\}$ be any sequence generated by Algorithm 4.1. Then $\{x^k\}$ is bounded. Furthermore, if $\|H_k\| \leq C_1$ and $c_k \leq C_2 \|F(x^k)\|^{-1}$ for some $C_1, C_2 > 0$, then $\{x^k\}$ converges to some $\bar{x}$ such that $F(\bar{x}) = 0$.*

In addition, if $F(\cdot)$ is differentiable in a neighborhood of $\bar{x}$ with $F'(\cdot)$ Lipschitz-continuous, $F'(\bar{x})$ is positive definite, $H_k = F'(x^k)$ for all $k \geq k_0$, and $0 = \lim_{k\to\infty} c_k^{-1} = \lim_{k\to\infty} c_k \|F(x^k)\|$, then the convergence rate is superlinear.

Consider the classical nonlinear complementarity problem [24, 12], which is to find an $x \in R^n$ such that

$$g(x) \geq 0, \quad x \geq 0, \quad \langle g(x), x \rangle = 0\,, \tag{31}$$

where $g : R^n \to R^n$ is differentiable. One of the most useful approaches to numerical and theoretical treatment of the NCP consists in reformulating it as a system of (nonsmooth) equations [26, 15]. One popular choice is given by the so-called *natural residual* [23]

$$F_i(x) = \min\{x_i, \alpha g_i(x)\}\,,\ i = 1, \dots, n\,, \tag{32}$$

where $\alpha > 0$ is a parameter. It is easy to check that for this mapping the solution set of the system of equations $F(x) = 0$ coincides with the solution set of the NCP (31). Furthermore, it is known [44] (see also [32, 16]) that F given by (32) is monotone for all α sufficiently small if g is *co-coercive* (for the definition of co-coerciveness and its role for variational inequalities, see [2, 42, 45]). Since F is obviously continuous, it immediately follows that for co-coercive NCP Algorithm 4.1 applied to the natural residual F generates a sequence of iterates globally converging to some solution x^* of the NCP, provided the parameters are chosen as specified. However, it is clear that this F is in general not differentiable. In particular, it is not differentiable at a solution x^* of the NCP, unless the strict complementarity condition $x_i^* + g_i(x^*) > 0\,, i = 1, \dots, n$, holds. This condition, however, is known to be very restrictive. Hence, the superlinear rate of convergence stated in Theorem 10 does not apply in this context.

On the other hand, by Remark 9, Algorithm 2.1 does converge locally superlinearly under appropriate assumptions. Of course, it is important to study computational cost involved in finding acceptable approximate solutions for the proximal point subproblems in Algorithm 2.1. We now turn our attention to this issue. Specifically, we shall show that when x^k is close to a solution x^* with certain properties, two steps of the generalized Newton method applied to solve the k-th proximal subproblem

$$0 = F_k(x) := F(x) + c_k^{-1}(x - x^k) \tag{33}$$

are sufficient to guarantee our error tolerance criterion. As a consequence, local superlinear rate of convergence of Algorithm 2.1 is ensured by solving at most two systems of linear equations at each iteration, and by taking $c_k \to +\infty\,, \sigma_k \to 0$ in an appropriate way.

Let $\partial F(x)$ denote the Clarke's *generalized Jacobian* [8] of F at $x \in R^n$. Specifically, $\partial F(x)$ is the convex hull of the *B-subdifferential* [29] of F at x, which is the set

$$\partial_B F(x) = \{H \in R^{n\times n} \mid \exists \{x^k\} \subset D_F \ : \ x^k \to x \text{ and } F'(x^k) \to H\},$$

with D_F being the set of points at which F is differentiable (by the Rademacher's Theorem, a locally Lipschitz-continuous function is differentiable almost everywhere). Recall that F is semismooth [21, 27, 28] at $x \in R^n$ if it is directionally differentiable at x, and

$$Hd - F'(x;d) = o(\|d\|)\,, \quad \forall\, d \to 0\,,\ \forall\, H \in \partial F(x+d)\,,$$

where $F'(x;d)$ stands for the usual directional derivative (this is one of a number of equivalent ways to define semismoothness). Similarly, F is said to be strongly semismooth at $x \in R^n$ if

$$Hd - F'(x;d) = O(\|d\|^2)\,, \quad \forall\, d \to 0\,,\ \forall\, H \in \partial F(x+d)\,.$$

When we say that F is (strongly) semismooth, we mean that this property holds for all points $x \in R^n$. By the calculus for semismooth mappings [13], if g is (strongly) semismooth, then so is F given by (32). In particular, if

g is continuously differentiable, then F is semismooth; and F is strongly semismooth if the derivative of g is Lipschitz-continuous (see also [20]). Furthermore, at any point $x \in R^n$ elements in $\partial F(x)$ or $\partial_B F(x)$ are easily computable by explicit formulas, e.g., see [18]. It is known [18, 20] that if x^* is a b-regular [25] solution of NCP (which is one of the weakest regularity conditions for NCP), then all elements in $\partial_B F(x^*)$ are nonsingular (the so-called BD-regularity of F at x^*).

Consider $\{x^k\} \to x^*$, and assume that x^* is a BD-regular solution of $F(x) = 0$. The latter implies that x^* is an isolated solution, and the following *error bound* holds [26, Proposition 3]: there exist $M_1 > 0$ and $\delta > 0$ such that

$$M_1 \|x - x^*\| \le \|F(x)\| \quad \forall\, x \in R^n \text{ such that } \|x - x^*\| \le \delta\,. \tag{34}$$

Denote by z^k the exact solution of (33), $\{z^k\} \to x^*$. By the well-known properties of the (exact) proximal point method [30],

$$\|z^k - x^*\|^2 \le \|x^k - x^*\|^2 - \|z^k - x^k\|^2\,,$$

from which using $x^k - z^k = c_k F(z^k)$ and (34), we obtain

$$\|z^k - x^*\| = O(c_k^{-1} \|x^k - x^*\|)\,. \tag{35}$$

Suppose F is strongly semismooth. Clearly, F_k is strongly semismooth, and z^k is a BD-regular solution of (33) (elements in $\partial_B F_k(z^k)$ are nonsingular, once x^k is sufficiently close to x^*, so that z^k is also close to x^*).

Let $\hat{y}^k \in R^n$ be the point obtained after one iteration of the generalized Newton method applied to (33):

$$\hat{y}^k = x^k - P_k^{-1} F_k(x^k)\,, \quad P_k \in \partial_B F_k(x^k)\,.$$

Since $P_k = H_k + c_k^{-1} I\,, H_k \in \partial_B F(x^k)$, we have that

$$\begin{aligned} \|\hat{y}^k - z^k\| &= \|x^k - z^k - (H_k + c_k^{-1} I)^{-1} F(x^k)\| \\ &\le \|x^k - H_k^{-1} F(x^k) - x^*\| + \|z^k - x^*\| \end{aligned}$$

$$
\begin{aligned}
& +\|(H_k^{-1} - (H_k + c_k^{-1} I)^{-1}) F(x^k)\| \\
= \;& O(\|x^k - x^*\|^2) + O(c_k^{-1}\|x^k - x^*\|) + O(c_k^{-1}\|F(x^k)\|) \\
= \;& O(c_k^{-1}\|F(x^k)\|)\,, \qquad (36)
\end{aligned}
$$

where the second equality follows from the quadratic convergence [27, 28] of the generalized Newton method applied to $F(x) = 0$, (35), and the classical fact of linear algebra about the inverse of a perturbed nonsingular matrix; and the last equality follows from (34). By the Lipschitz-continuity of F, the triangle inequality, and (35), we also have

$$
\begin{aligned}
\|F(x^k)\| &\le M_2\|x^k - x^*\| \\
&\le M_2(\|z^k - x^k\| + \|z^k - x^*\|) \\
&\le M_2\|z^k - x^k\| + O(c_k^{-1}\|x^k - x^*\|)\,.
\end{aligned}
$$

Using (34) in the last relation, it follows that $\|F(x^k)\| \le O(\|z^k - x^k\|)$. From (36), we then obtain that

$$
\|\hat{y}^k - z^k\| = O(c_k^{-1}\|x^k - z^k\|)\,.
$$

Denoting by y^k the next Newton step for (33),

$$
y^k = \hat{y}^k - \hat{P}_k^{-1} F_k(\hat{y}^k)\,, \quad \hat{P}_k \in \partial_B F_k(\hat{y}^k)\,,
$$

we have

$$
\|y^k - z^k\| = O(c_k^{-1}\|\hat{y}^k - z^k\|)\,,
$$

and, using (36), we conclude that

$$
\|y^k - z^k\| = O(c_k^{-2}\|F(x^k)\|)\,. \qquad (37)
$$

The left-hand side of the tolerance criterion (21) for proximal subproblem (33) can be bounded as follows:

$$
\begin{aligned}
\|c_k F(y^k) + y^k - x^k\| &\le c_k\|F(y^k) - F(z^k)\| + \|y^k - z^k\| \\
&\le 2M_2 c_k\|y^k - z^k\| \\
&= O(c_k^{-1}\|F(x^k)\|)\,, \qquad (38)
\end{aligned}
$$

where the first inequality follows from the definition of z^k and the Cauchy-Schwarz inequality; the second inequality follows from the Lipschitz-continuity of F; and the last follows from (37). On the other hand, we have that

$$\begin{aligned} \|y^k - x^k\| &\geq \|x^k - x^*\| - \|y^k - z^k\| - \|z^k - x^*\| \\ &\geq \|x^k - x^*\| - O(c_k^{-2}\|F(x^k)\|) - O(c_k^{-1}\|x^k - x^*\|) \\ &\geq M_3\|F(x^k)\|, \end{aligned} \tag{39}$$

where the second inequality follows from (37) and (35); and the last inequality follows with some $M_3 > 0$, by the Lipschitz-continuity of F. Using (39) and (38), we conclude that the error tolerance condition (21) is guaranteed to be satisfied for y^k, the second Newton iterate for solving (33), if we choose parameters σ_k and c_k so that

$$c_k^{-1} = o(\sigma_k).$$

With this choice, in the framework of Algorithm 2.1 y^k is an acceptable approximate solution of the proximal point subproblem. By Remark 9, convergence rate of $\{x^k\}$ generated by Algorithm 2.1 is superlinear.

In fact, comparing (36) and (38), we can see that for the first Newton point the left and right-hand sides of the tolerance criterion are already of the same order. Since these are merely rough upper and lower estimates, as a practical matter, one might expect the error tolerance rule to be satisfied already after the first Newton step.

References

[1] H.H. Bauschke and P.L. Combettes. A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces. *Mathematics of Operations Research*, 26:248–264, 2001.

[2] R.E. Bruck. An iterative solution of a variational inequality for certain monotone operators in a Hilbert space. *Bulletin of the American Mathematical Society*, 81:890–892, 1975.

[3] R.S. Burachik, A.N. Iusem, and B.F. Svaiter. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Analysis*, 5:159–180, 1997.

[4] R.S. Burachik, C.A. Sagastizábal, and B. F. Svaiter. Bundle methods for maximal monotone operators. In R. Tichatschke and M. Théra, editors, *Ill-posed variational problems and regularization techniques*, Lecture Notes in Economics and Mathematical Systems, No. 477, pages 49–64. Springer–Verlag, 1999.

[5] R.S. Burachik, C.A. Sagastizábal, and B.F. Svaiter. ε-Enlargements of maximal monotone operators: Theory and applications. In M. Fukushima and L. Qi, editors, *Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, pages 25–44. Kluwer Academic Publishers, 1999.

[6] R.S. Burachik and B.F. Svaiter. ε-Enlargements of maximal monotone operators in Banach spaces. *Set-Valued Analysis*, 7:117–132, 1999.

[7] J.V. Burke and M. Qian. A variable metric proximal point algorithm for monotone operators. *SIAM Journal on Control and Optimization*, 37:353–375, 1998.

[8] F.H. Clarke. *Optimization and Nonsmooth Analysis.* John Wiley & Sons, New York, 1983.

[9] R. Cominetti. Coupling the proximal point algorithm with approximation methods. *Journal of Optimization Theory and Applications*, 95:581–600, 1997.

[10] R.S. Dembo, S.C Eisenstat, and T. Steihaug. Inexact Newton methods. *SIAM Journal on Numerical Analysis*, 19:400–408, 1982.

[11] J. Eckstein. Approximate iterations in Bregman-function-based proximal algorithms. *Mathematical Programming*, 83:113–123, 1998.

[12] M.C. Ferris and C. Kanzow. Complementarity and related problems. In P.M. Pardalos and M.G.C. Resende, editors, *Handbook of Applied Optimization*. Oxford University Press, 2000.

[13] A. Fischer. Solution of monotone complementarity problems with locally Lipschitzian functions. *Mathematical Programming*, 76:513–532, 1997.

[14] M. Fukushima. Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. *Mathematical Programming*, 53:99–110, 1992.

[15] M. Fukushima and L. Qi (editors). *Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*. Kluwer Academic Publishers, 1999.

[16] D. Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods : Applications to the Solution of Boundary-Value Problems*, pages 299–331. North–Holland, 1983.

[17] O. Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM Journal on Control and Optimization*, 29:403–419, 1991.

[18] C. Kanzow and M. Fukushima. Solving box constrained variational inequality problems by using the natural residual with D-gap function globalization. *Operations Research Letters*, 23:45–51, 1998.

[19] B. Lemaire. The proximal algorithm. In J.-P. Penot, editor, *New Methods of Optimization and Their Industrial Use. International Series of Numerical Mathematics 87*, pages 73–87. Birkhauser, Basel, 1989.

[20] T. De Luca, F. Facchinei, and C. Kanzow. A theoretical and numerical comparison of some semismooth algorithms for complementarity problems. *Computational Optimization and Applications*, 16:173–205, 2000.

[21] R. Mifflin. Semismooth and semiconvex functions in constrained optimization. *SIAM Journal on Control and Optimization*, 15:957–972, 1977.

[22] Z. Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bulletin of the American Mathematical Society*, 73:591–597, 1967.

[23] J.-S. Pang. Inexact Newton methods for the nonlinear complementarity problem. *Mathematical Programming*, 36:54–71, 1986.

[24] J.-S. Pang. Complementarity problems. In R. Horst and P. Pardalos, editors, *Handbook of Global Optimization*, pages 271–338. Kluwer Academic Publishers, Boston, Massachusetts, 1995.

[25] J.-S. Pang and S.A. Gabriel. NE/SQP: A robust algorithm for the nonlinear complementarity problem. *Mathematical Programming*, 60:295–337, 1993.

[26] J.-S. Pang and L. Qi. Nonsmooth equations: Motivation and algorithms. *SIAM Journal on Optimization*, 3:443–465, 1993.

[27] L. Qi. Convergence analysis of some algorithms for solving nonsmooth equations. *Mathematics of Operations Research*, 18:227–244, 1993.

[28] L. Qi and J. Sun. A nonsmooth version of Newton's method. *Mathematical Programming*, 58:353–367, 1993.

[29] S. M. Robinson. Local structure of feasible sets in nonlinear programming, Part III: Stability and sensitivity. *Mathematical Programming Study*, 30:45–66, 1987.

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

[31] C.A. Sagastizábal and M.V. Solodov. On the relation between bundle methods for maximal monotone inclusions and hybrid proximal point algorithms. In D. Butnariu, Y. Censor, and S. Reich, editors, *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, volume 8 of *Studies in Computational Mathematics*, pages 441–455. Elsevier Science B.V., 2001.

[32] M. Sibony. Méthodes itératives pour les équations et inéquations aux dérivées partielles nonlinéares de type monotone. *Calcolo*, 7:65–183, 1970.

[33] M. V. Solodov and B. F. Svaiter. A new projection method for variational inequality problems. *SIAM Journal on Control and Optimization*, 37:765–776, 1999.

[34] M. V. Solodov and B. F. Svaiter. A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem. *SIAM Journal on Optimization*, 10:605–625, 2000.

[35] M.V. Solodov and B.F. Svaiter. A globally convergent inexact Newton method for systems of monotone equations. In M. Fukushima and L. Qi, editors, *Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, pages 355–369. Kluwer Academic Publishers, 1999.

[36] M.V. Solodov and B.F. Svaiter. A hybrid approximate extragradient–proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Analysis*, 7:323–345, 1999.

[37] M.V. Solodov and B.F. Svaiter. A hybrid projection–proximal point algorithm. *Journal of Convex Analysis*, 6:59–70, 1999.

[38] M.V. Solodov and B.F. Svaiter. Error bounds for proximal point subproblems and associated inexact proximal point algorithms. *Mathematical Programming*, 88:371–389, 2000.

[39] M.V. Solodov and B.F. Svaiter. Forcing strong convergence of proximal point iterations in a Hilbert space. *Mathematical Programming*, 87:189–202, 2000.

[40] M.V. Solodov and B.F. Svaiter. An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. *Mathematics of Operations Research*, 25:214–230, 2000.

[41] P. Tossings. The perturbed proximal point algorithm and some of its applications. *Applied Mathematics and Optimization*, 29:125–159, 1994.

[42] P. Tseng. Further applications of a splitting algorithm to decomposition in variational inequalities and convex programming. *Mathematical Programming*, 48:249–263, 1990.

[43] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38:431–446, 2000.

[44] Y.-B. Zhao and D. Li. Monotonicity of fixed point and normal mappings associated with variational inequality and its application. *SIAM Journal on Optimization*, 11:962–973, 2001.

[45] D.L. Zhu and P. Marcotte. Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities. *SIAM Journal on Optimization*, 6:714–726, 1996.