A NEW PROXIMAL-BASED GLOBALIZATION STRATEGY FOR THE JOSEPHY-NEWTON METHOD FOR VARIATIONAL INEQUALITIES*

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.
Email : solodov@impa.br *and* benar@impa.br.

(Received 6 March 2002; in final form 25 July 2002)

We propose a new approach to globalizing the Josephy-Newton algorithm for solving the monotone variational inequality problem. Known globalization strategies rely either on minimization of a suitable merit function, or on a projection-type approach. The technique proposed here is based on a linesearch in the regularized Josephy-Newton direction which finds a trial point and a proximal point subproblem (i.e., subproblem with suitable parameters), for which this trial point is an acceptable approximate solution. We emphasize that this requires only checking a certain approximation criterion, and in particular, does not entail actually solving any nonlinear proximal point subproblems. The method converges globally under very mild assumptions. Furthermore, an easy modification of the method secures the local superlinear rate of convergence under standard conditions.

KEY WORDS: Variational inequality, Josephy-Newton method, proximal point method, globalization

1 INTRODUCTION

Given a function $F : \Re^n \to \Re^n$ and a set $C \subset \Re^n$, the classical variational inequality problem [8, 15], $\mathrm{VIP}(F, C)$, is to find a point x such that

$$x \in C, \quad \langle F(x), u - x\rangle \geq 0 \quad \text{ for all } u \in C\,,$$

where $\langle \cdot, \cdot\rangle$ denotes the usual inner product in $\Re^n$. When C is the nonnegative orthant $\Re^n_+$, $\mathrm{VIP}(F, C)$ reduces to the nonlinear complementarity problem $\mathrm{NCP}(F)$, which is to find a point $x \in \Re^n$ such that $x \geq 0,\ F(x) \geq 0,\ \langle F(x), x\rangle = 0$. When C is the whole space $\Re^n$, $\mathrm{VIP}(F, C)$ becomes a system of nonlinear equations $F(x) = 0$. Throughout this paper we assume that C is closed convex, F is continuously differentiable and monotone (i.e., $\langle F(x) - F(y), x - y\rangle \geq 0$ for all $x, y \in \Re^n$), and the solution set of $\mathrm{VIP}(F, C)$ is nonempty.

* 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.

One of the successful and widely used approaches to solving VIP(F,C) is the Josephy-Newton method [17, 20, 3], which consists of solving successive linearizations of the problem. Given a point x^k, this method generates the next iterate as follows:

$$x^{k+1} \text{ is a solution of VIP}(F_k, C)\,,$$

where $F_k(\cdot)$ is the first-order approximation of $F(\cdot)$ at x^k :

$$F_k(x) = F(x^k) + F'(x^k)(x - x^k)\,.$$

If the starting point is sufficiently close to some *regular* solution $\bar{x}$ of VIP(F,C) the sequence generated by the Josephy-Newton method is well-defined and converges to $\bar{x}$ superlinearly. Regularity in this context is meant in the sense of Robinson [24] (see also [3]).

There are two key difficulties with using the Josephy-Newton method for solving VIP(F,C). First of all, in the absence of regularity even local convergence cannot be ensured. Second, even if regularity holds at a solution, ensuring global convergence is often difficult. In particular, far from a solution of the problem (even if it is regular), there is no guarantee that subproblems are solvable and the method is well-defined. Even if the subproblem solution exists, there is no guarantee that it is "useful", i.e., that it constitutes some progress towards solving VIP(F,C). To enlarge the domain of convergence of the Josephy-Newton method, some globalization strategy has to be used.

One possibility is adopting a linesearch procedure in the obtained Newton direction (if it exists) aimed at decreasing the value of a suitable *merit function* (see [14, 28, 12] for surveys of merit functions for variational inequality and complementarity problems). Globalizations based on this approach have been proposed in [19, 37, 22, 23, 10]. The method of [19] is based on the use of the *gap function* [2]. For global convergence, F has to be monotone, C compact, and the linesearch has to be the exact minimization along the direction. Globalization developed in [37] employs the *regularized gap function* [1, 13]. For convergence, F has to be strongly monotone. In both [19] and [37], a strict complementarity condition is needed for the superlinear convergence rate. Algorithms proposed in [22, 23, 10] are based on unconstrained merit functions. Specifically, the first two use the *D-gap function* [21, 39, 36], and the last uses the Fischer-Burmeister function [11, 9]. These globalizations are similar in spirit. Due to the "safeguard" possibility of performing a standard gradient descent step for the merit function whenever the Newton direction does not exist or is not satisfactory, the methods of [22, 23, 10] are well-defined for any F. It typically holds that every accumulation point of the generated sequence of iterates is a stationary point of the merit function employed in the algorithm. However, the existence of such accumulation points, and the equivalence of stationary points of merit functions to solutions of VIP(F,C), cannot be guaranteed without further assumptions. For example, in [22] F is assumed to be strongly monotone, and in [23] F is a uniform P-function and C is a box. Either of those assumptions implies that solution of VIP(F,C) is in fact globally unique. If F' is further Lipschitz continuous around the solution, local superlinear rate of convergence is obtained.

An alternative globalization strategy for monotone problems had been proposed in [29, 34] for the two special cases of VIP(F, C): the nonlinear equation $F(x) = 0$ and the NCP(F), respectively (see also some remarks in [27] concerning the general case). A distinctive feature of this approach is that it is not based on minimizing any merit function. Global convergence is ensured by a certain separation and projection procedure similar to the projection methods for variational inequalities [16, 32]. Local superlinear rate of convergence, on the other hand, is based on the inexact proximal point scheme of [31]. The key fact here is that under natural assumptions, close to a solution the Newton step satisfies the approximation criterion of [31], and so the method converges superlinearly due to its relation to the proximal point algorithm with an appropriate control of parameters. We note that the methods of [29, 34] generate a sequence of iterates which converges globally (from any starting point) to a solution of the problem even if it is not locally unique. This compares favorably with globalizations using merit functions.

In this paper, we propose a new way of globalizing the Josephy-Newton method, which is based on the following interesting fact. Performing a linesearch in the Newton direction, we can find a point and an associated proximal subproblem for which this point is an acceptable (in some sense) approximate solution. We point out that this is done by checking a certain approximation criterion, without actually solving any nonlinear proximal point subproblems. Thus the computational cost of this procedure is comparable to any typical linesearch. Once this point and the corresponding values of the proximal subproblem parameters are obtained, the next iterate is computed in the spirit of the hybrid inexact proximal point methods [31, 30, 33, 35]. This will be made precise in the following section. Note that this strategy is different from the methods in [29, 34, 27], where hybrid proximal steps are taken only locally (essentially, when conditions for superlinear convergence are satisfied), while globalization is based on a projection procedure.

Our notation is quite standard. For a differentiable function $F : \Re^n \to \Re^n$, its Jacobian at a point $x \in \Re^n$ is denoted by $F'(x)$. By $P_C(x)$ we denote the orthogonal projection of the point $x \in \Re^n$ onto the closed convex set $C \subset \Re^n$. The normal operator of C is denoted by N_C, that is $N_C(x) = \emptyset$ if $x \notin C$, and $N_C(x) = \{w \in \Re^n \mid \langle w, u - x\rangle \le 0 \ \forall u \in C\}$ if $x \in C$. Finally, I stands for the identity matrix (of appropriate dimension).

2 PROXIMAL-BASED GLOBALIZATION

In the context of this paper, VIP(F, C) is equivalent to solving the inclusion

$$0 \in T(x)\,, \tag{1}$$

where the operator $T = F + N_C$ is maximal monotone. Recall that the (exact) proximal point scheme [25, 18] for (1) can be stated as follows: given the current iterate x, choose the regularization parameter $c > 0$, and find $(y, v) \in \Re^n \times \Re^n$ such that

$$v \in T(y)\,, \quad cv + y - x = 0\,, \tag{2}$$

then set y to be the next iterate, and repeat. Of course, the exact proximal point scheme is not a practical computational method. The utility of any proximal-based framework depends critically on the ability to handle approximate solutions of proximal subproblems. In this respect, one may try to relax the inclusion in (2), or the equation in (2), or both. As discussed in [33], in the context of VIP(F, C) it is natural and important to relax both. To this end, recall that given some $\varepsilon \geq 0$, the ε-enlargement of a maximal monotone operator T, introduced in [4] and denoted T^ε, is given by

$$T^\varepsilon(x) = \{v \in \Re^n \mid \langle w - v, z - x\rangle \geq -\varepsilon \text{ for all } z \in \Re^n,\ w \in T(z)\}\,. \tag{3}$$

We refer the reader to [4, 6, 5, 7, 33] for properties and applications of ε-enlargements of maximal monotone operators. The following is a unification [35] of the approximation criteria considered in [31, 30]: we say that (y, v) is an acceptable *approximate* solution of the proximal point subproblem (2) if it holds that

$$\begin{array}{l} v \in T^\varepsilon(y)\,, \\ \|cv + y - x\|^2 + 2c\varepsilon \leq \sigma^2(\|cv\|^2 + \|y - x\|^2)\,, \\ c > 0\,, \varepsilon \geq 0\,, \sigma \in [0, 1)\,. \end{array} \tag{4}$$

Calling such a pair acceptable is justified, because convergence of the resulting approximate proximal point iterations can be ensured by adding simple explicit "correction" steps of projection [31] or extragradient [30] type. The advantage of approximation criteria of this kind is that they are constructive and suitable for a number of applications (for example, Newton methods [29, 34], forward-backward splitting [38, 30], and bundle techniques [5, 26], to name some). In what follows, we show how (4) can be used for globalization of the Josephy-Newton method for VIP(F, C).

Let x be the current iterate and consider the regularized (with regularization parameter $\lambda > 0$) variational inequality: find w such that

$$w \in C\,, \quad \langle \lambda F(w) + w - x, u - w\rangle \geq 0 \quad \forall\, u \in C\,. \tag{5}$$

Let z be its Josephy-Newton point, i.e., the solution of the linearization of (5):

$$z \in C\,, \quad \langle \lambda F(x) + (\lambda F'(x) + I)(z - x), u - z\rangle \geq 0 \quad \forall\, u \in C\,. \tag{6}$$

Note that one advantage of considering the regularized subproblems is that the Newton direction always exists, due to the strong monotonicity of the (affine) variational inequality (6) [15]. In what follows, we show that for all $t > 0$ sufficiently small, the point $y(t) = x + t(z - x)$, the modified regularization parameter $c(t) = t\lambda > 0$, and certain associated $v(t) \in \Re^n$, $\varepsilon(t) \geq 0$, $\sigma(t) \in [0, 1)$ satisfy the approximation criterion (4). In other words, performing an Armijo-type linesearch in the Josephy-Newton direction we can find a proximal point subproblem for which an acceptable approximate solution is readily available. This solution can then be used in the hybrid proximal point framework to obtain a globally convergent algorithm.

It is convenient to start with the following auxiliary result.

Lemma 2.1. *Let F be monotone and continuously differentiable, and let z be the point given by* (6)*, where $x \in C$ is not a solution of VIP(F, C). Then for all $t \in (0, 1]$ sufficiently small and all $\sigma \in [0, 1)$ close enough to one, $y(t) = x + t(z - x)$ satisfies*

$$\|t\lambda F(y(t)) + y(t) - x\|^2 \le \sigma^2 (\|t\lambda F(y(t))\|^2 + \|y(t) - x\|^2) . \tag{7}$$

Proof. Using (6) with $u = x \in C$, we have that

$$\begin{aligned} \lambda\langle F(x), z - x\rangle &\le \langle(\lambda F'(x) + I)(z - x), x - z\rangle \\ &\le -\|z - x\|^2 , \end{aligned}$$

where the second inequality follows from positive semidefiniteness of $F'(x)$ (since F is monotone). By the Cauchy-Schwarz inequality, we further obtain

$$\begin{aligned} \lambda\langle F(y(t)), y(t) - x\rangle &= t\lambda\langle F(x), z - x\rangle + t\lambda\langle F(y(t)) - F(x), z - x\rangle \\ &\le -t\|z - x\|^2 + t\lambda\|F(y(t)) - F(x)\|\|z - x\| \\ &= -t(1 - t\lambda)\|z - x\|(\|z - x\| + \|F(y(t)) - F(x)\|) . \end{aligned} \tag{8}$$

Hence,

$$\begin{aligned} &\|t\lambda F(y(t)) + y(t) - x\|^2 - \sigma^2(\|t\lambda F(y(t))\|^2 + \|y(t) - x\|^2) \\ = \; & t^2(1 - \sigma^2)(\|\lambda F(y(t))\|^2 + \|z - x\|^2) + 2t\lambda\langle F(y(t)), y(t) - x\rangle \\ \le \; & t^2((1 - \sigma^2)\|\lambda F(y(t))\|^2 - \|z - x\|((1 + \sigma^2)\|z - x\| - 2\lambda\|F(y(t)) - F(x)\|)) , \end{aligned}$$

where the inequality follows from (8). Hence, (7) will be satisfied whenever the right-hand side in the relation above is non-positive, i.e.,

$$(1 - \sigma^2)\|\lambda F(y(t))\|^2 \le \|z - x\|((1 + \sigma^2)\|z - x\| - 2\lambda\|F(y(t)) - F(x)\|) .$$

As $t \to 0$ and $\sigma \to 1$, the left-hand side in the above inequality tends to zero, while the right-hand side tends to $(1 + \sigma^2)\|z - x\|^2 > 0$, provided $z \ne x$ (observe that if $z = x$, then (6) implies that x solves VIP(F, C)). This proves the assertion. ∎

It can be seen that if F is further Lipschitz-continuous (with modulus $L > 0$), then the preceeding analysis shows that condition (7) is certainly satisfied whenever

$$1 > \sigma \ge \left(\max\left\{0\, ,\, 1 - \frac{\|z - x\|^2}{\|\lambda F(y(t))\|^2}\right\}\right)^{1/2} , \quad 0 < t \le \frac{\sigma^2}{2\lambda L} . \tag{9}$$

We are now in position to exhibit the pair (y, v) which satisfies the conditions specified in (4).

Proposition 2.2. *Suppose F is monotone and condition* (7) *holds. Then for $c = t\lambda$ and*

$$v(t) = \frac{1}{c(1 - \sigma^2)}\left(x - P_C(x - c(1 - \sigma^2)F(y(t)))\right) ,$$

the pair $(y(t), v(t))$ *satisfies the proximal point approximation criterion* (4).

Proof. In what follows, in our notation we shall omit the explicit dependence of y and v on t, as it clear from the context.

We first verify the inclusion in (4), i.e., we show that $v \in T^{\varepsilon}(y)$ for a certain $\varepsilon \geq 0$, where $T = F + N_C$. Denote $q = P_C(x - c(1-\sigma^2)F(y))$. By properties of the normal operator,

$$q = x - c(1-\sigma^2)F(y) - \nu\,, \quad \nu \in N_C(q)\,. \tag{10}$$

With this notation,

$$v = \frac{1}{c(1-\sigma^2)}(x-q) = F(y) + a\,, \quad a = \frac{1}{c(1-\sigma^2)}\nu \in N_C(q)\,. \tag{11}$$

We have to verify that for some $\varepsilon \geq 0$,

$$\langle v - w, y - u\rangle \geq -\varepsilon \quad \forall\, u \in \Re^n, w \in T(u)\,.$$

Since $N_C(u) = \emptyset$ for $u \notin C$, $w \in T(u)$ implies that $u \in C$, and $w = F(u) + b, b \in N_C(u)$. Using also (11), we have that

$$\begin{array}{rcl} \langle v - w, y - u\rangle & = & \langle F(y) - F(u), y - u\rangle + \langle a - b, y - u\rangle \\ & \geq & \langle a, y - u\rangle + \langle b, u - y\rangle \\ & \geq & \langle a, y - u\rangle \\ & = & \langle a, y - q\rangle + \langle a, q - u\rangle \\ & \geq & \langle a, y - q\rangle\,, \end{array}$$

where the first inequality follows from the monotonicity of F, the second follows from $y \in C$, $b \in N_C(u)$, and the last from $u \in C$, $a \in N_C(q)$. We conclude that $v \in T^{\varepsilon}(y)$ with

$$\varepsilon = -\langle a, y - q\rangle \geq 0\,, \tag{12}$$

where the nonnegativity of ε is due to $y \in C$, $a \in N_C(q)$.

Next, we prove the inequality in (4).

$$\begin{array}{rl} & \|cv + y - x\|^2 + 2c\varepsilon - \sigma^2\|cv\|^2 - \sigma^2\|y - x\|^2 \\ = & \|cF(y) + y - x\|^2 + \|ca\|^2 + 2c\langle a, cF(y) + y - x\rangle - 2c\langle a, y - q\rangle \\ & -\sigma^2\|cF(y)\|^2 - \sigma^2\|ca\|^2 - 2\sigma^2\langle cF(y), ca\rangle - \sigma^2\|y - x\|^2 \\ \leq & (1-\sigma^2)\|ca\|^2 + 2c\langle a, c(1-\sigma^2)F(y) + q - x\rangle \\ = & (1-\sigma^2)\|ca\|^2 - 2c\langle a, \nu\rangle \\ = & -c^2(1-\sigma^2)\|a\|^2 \leq 0\,, \end{array}$$

where the first inequality follows from the hypothesis (7), the next to last equality follows from (10), and the last from (11). This completes the proof. ■

We have therefore demonstrated how by performing an Armijo-type linesearch in the Josephy-Newton direction we can obtain all the necessary objects to satisfy

(4). Those objects can then be used in the framework of hybrid proximal point algorithms. However, convergence of such a method would not follow directly from the properties of hybrid proximal algorithms. Note that in the setting of Proposition 2.2, c and σ are not user-chosen parameters. Their values are obtained by linesearch. Thus standard assumptions on those parameters, such as $0 < \liminf_{k\to\infty} c_k$ and $1 > \limsup_{k\to\infty} \sigma_k$, are not automatic in the context of this paper. For this reason, an independent convergence analysis will be required. But first, we formally state the algorithm that we propose.

Let

$$r(x) := x - P_C(x - F(x))$$

be the *natural residual* of VIP(F, C). As is well known, some x solves VIP(F, C) if, and only if, $r(x) = 0$.

Algorithm 1. *Choose parameters* $\hat{\lambda}, \tilde{\lambda}, s > 0$; $\gamma, \theta \in (0, 1)$, *and any* $x^0 \in C$. *Set* $k := 0$.

1. Stopping test. *Compute* $r(x^k)$, *and stop if it is zero.*

2. Josephy-Newton subproblem.
Find z^k, *the solution of VIP*(F_k, C), *where*

$$F_k(z) = \lambda_k F(x^k) + (\lambda_k F'(x^k) + I)(z - x^k)\,,$$

$$\min\{\hat{\lambda},\, \tilde{\lambda}\|r(x^k)\|^{-s}\} \leq \lambda_k \leq \tilde{\lambda}\|r(x^k)\|^{-s}\,.$$

3. Linesearch.
Find

$$y^k = x^k + t_k(z^k - x^k)\,, \quad c_k = t_k\lambda_k\,, \quad \sigma_k = 1 - t_k\theta\,,$$

where $t_k = \gamma^{m_k}$ *with* m_k *being the smallest nonnegative integer* m *such that the above objects, together with*

$$v^k = \frac{1}{c_k(1-\sigma_k^2)}\left(x^k - P_C(x^k - c_k(1-\sigma_k^2)F(y^k))\right)\,,$$

$$\varepsilon_k = \langle F(y^k) - v^k, y^k - P_C(x^k - c_k(1-\sigma_k^2)F(y^k))\rangle\,,$$

satisfy

$$\begin{array}{l} v^k \in T^{\varepsilon_k}(y^k) \\ \|c_k v^k + y^k - x^k\|^2 + 2c_k\varepsilon_k \leq \sigma_k^2(\|c_k v^k\|^2 + \|y^k - x^k\|^2)\,. \end{array} \tag{13}$$

4. Variable update.

$$x^{k+1} := P_C(x^k - \alpha_k v^k)\,, \ \textit{where}\ \alpha_k = \|v^k\|^{-2}\left(\langle v^k, x^k - y^k\rangle - \varepsilon_k\right)\,.$$

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

Note that there are two possibilities in how the linesearch step can be executed. One is to compute y^k, c_k and σ_k such that the condition (7) of Lemma 2.1 is satisfied (although, in that case the part that "m_k is the smallest nonnegative integer" may not apply). By Proposition 2.2, it then immediately follows that with the given definition of v^k and ε_k, the approximation criterion (13) is satisfied. The advantage of this approach is that no projections onto the feasible set C are performed during the linesearch (see (7)). Indeed, the projection is needed only once, to compute the point $P_C(x^k - c_k(1-\sigma_k^2)F(y^k))$ when linesearch terminates. This is a useful feature when projection onto C is computationally expensive. On the other hand, what we really need is to satisfy (13) rather than (7). In that sense, the latter condition might be too conservative in some situations and it can result in smaller stepsize values than really necessary. Note that if C has a simple structure, say is a box, then performing projections to compute each trial vector v does not entail any nontrivial computational burden. In that case, it might be preferable to check (13) directly. In the general case, a combination of conditions (7) and (13) can be used, in order to balance cheaper computation of possibly smaller stepsizes with more costly computation of larger ones.

Theorem 2.3. *Let F be monotone and continuously differentiable. Then any sequence $\{x^k\}$ generated by Algorithm 1 converges to a solution of VIP(F, C), provided one exists.*

Proof. Consider any iteration index k. We can assume that $r(x^k) \neq 0$, as otherwise the method terminates at a solution. By the monotonicity of F, F_k is strongly monotone and so z^k exists (and is unique) for each k. Note also that $x^k \in C$. By Lemma 2.1 and Proposition 2.2, the linesearch procedure is well-defined and terminates finitely with some $t_k = \gamma^{m_k} > 0$. Next, the variable update rule is well-defined if $v^k \neq 0$. Now, if $v^k = 0$ then (13) reduces to $(1-\sigma_k^2)\|y^k - x^k\|^2 + 2c_k\varepsilon_k \leq 0$, which implies that $y^k = x^k$ and $\varepsilon_k = 0$, so that $0 = v^k \in T(x^k)$. But then x^k is a solution of VIP(F, C) and the algorithm would have terminated at the stage of the stopping test. We have therefore established that the method is well-defined, and either terminates at a solution or generates an infinite sequence. We next consider the latter case.

By (13), we obtain

$$2c_k\langle v^k, y^k - x^k\rangle + 2c_k\varepsilon_k \leq (\sigma_k^2 - 1)(\|c_k v^k\|^2 + \|y^k - x^k\|^2)\,,$$

and hence,

$$\langle v^k, x^k - y^k\rangle - \varepsilon_k \geq \frac{1-\sigma_k^2}{2c_k}(\|c_k v^k\|^2 + \|y^k - x^k\|^2)\,. \tag{14}$$

On the other hand, for any solution $\bar{x}$ of VIP(F, C), $0 \in T(\bar{x})$ and $v^k \in T^{\varepsilon_k}(y^k)$ imply that

$$\langle v^k, \bar{x} - y^k\rangle - \varepsilon_k \leq 0\,.$$

It follows from the above relation and (14) that the hyperplane $H_k := \{u \in \Re^n \mid \langle v^k, u - y^k\rangle - \varepsilon_k = 0\}$ separates x^k from $\bar{x}$, and in fact, $x^k - \alpha_k v^k = P_{H_k}(x^k)$. By the well-known (and geometrically obvious) properties of the orthogonal projections, it holds that

$$\begin{aligned} \|x^{k+1} - \bar{x}\|^2 &\leq \|P_{H_k}(x^k) - \bar{x}\|^2 \\ &\leq \|x^k - \bar{x}\|^2 - \|P_{H_k}(x^k) - x^k\|^2 \\ &= \|x^k - \bar{x}\|^2 - \|\alpha_k v^k\|^2\,. \end{aligned} \tag{15}$$

We conclude that $\{\|x^k - \bar{x}\|\}$ converges, $\{x^k\}$ is bounded, and

$$0 = \lim_{k\to\infty} \alpha_k \|v^k\| = \lim_{k\to\infty} \|v^k\|^{-1} \left(\langle v^k, x^k - y^k\rangle - \varepsilon_k\right).$$

Combining the latter relation with (14), we conclude that

$$0 = \lim_{k\to\infty} (1-\sigma_k^2) c_k \|v^k\|. \tag{16}$$

We next consider the two possible cases:

$$0 < \liminf_{k\to\infty} t_k \quad \text{and} \quad 0 = \liminf_{k\to\infty} t_k. \tag{17}$$

In the first case we have that $0 < \liminf_k (1-\sigma_k^2)c_k$, and (16) implies that

$$0 = \lim_{k\to\infty} v^k. \tag{18}$$

Passing further onto the limit in (13), we have that $0 \geq \limsup_k (1-\sigma_k^2)\|y^k - x^k\|^2 + 2c_k\varepsilon_k$, and hence,

$$0 = \lim_{k\to\infty} \|y^k - x^k\|, \quad 0 = \lim_{k\to\infty} \varepsilon_k. \tag{19}$$

Now, let x^* be any accumulation point of $\{x^k\}$, and $\{x^{k_i}\}$ be some subsequence converging to x^*. By (19), $\{y^{k_i}\}$ also converges to x^*. Since $v^k \in T^{\varepsilon_k}(y^k)$ for all k, for any $u \in \Re^n$ and $w \in T(u)$ we have that

$$\langle w - v^{k_i}, u - y^{k_i}\rangle \geq -\varepsilon_{k_i}.$$

Now passing onto the limit in the above relation, and taking into account (18) and (19), we conclude that

$$\langle w - 0, u - x^*\rangle \geq 0.$$

By the maximal monotonicity of T, it now follows that $0 \in T(x^*)$, i.e., x^* solves $\mathrm{VIP}(F,C)$. Choosing $\bar{x} = x^*$ in (15), we conclude that $\{\|x^k - x^*\|\}$ converges to some number, and since x^* is an accumulation point of $\{x^k\}$, this number must be zero. In other words, $\{x^k\}$ converges to x^*, which is a solution of $\mathrm{VIP}(F,C)$.

Consider now the second case in (17). Let $\{k_i\}$ be a subsequence of iteration indices such that

$$0 = \lim_{i\to\infty} t_{k_i}.$$

Since F is continuously differentiable and $\{x^k\}$ is bounded, F is Lipschitz-continuous on some set containing $\{x^k\}$. By (9), since the stepsize $t = \gamma^{m_{k_i}-1} = \gamma^{-1}t_{k_i}$ was rejected, it must have been the case that

$$1 - \gamma^{-1}t_{k_i}\theta < \left(\max\left\{0\,,\, 1 - \frac{\|z^{k_i} - x^{k_i}\|^2}{\|\lambda_{k_i} F(y^{k_i})\|^2}\right\}\right)^{1/2} \tag{20}$$

or/and

$$\gamma^{-1}t_{k_i} > \frac{(1-\gamma^{-1}t_{k_i}\theta)^2}{2\lambda_{k_i} L}. \tag{21}$$

At least one of the inequalities above must hold an infinite number of times. If the second one holds an infinite number of times, then passing onto the limit along the corresponding indices, (21) yields

$$0 = \liminf_{i\to\infty} \lambda_{k_i}^{-1} \,. \qquad (22)$$

By the choice of λ_k, (22) implies that

$$0 = \liminf_{i\to\infty} \|r(x^{k_i})\|^s \,,$$

which means that $\{x^{k_i}\}$ has an accumulation point x^* such that $r(x^*) = 0$. Hence, this x^* solves VIP(F, C). The proof that the whole sequence $\{x^k\}$ converges to a solution can now follow the same pattern as above.

If (21) holds a finite number of times, then (20) must hold for all indices i sufficiently large. Passing onto the limit in (20) as $i \to \infty$, we obtain that

$$0 = \lim_{i\to\infty} \frac{\|z^{k_i} - x^{k_i}\|^2}{\|\lambda_{k_i} F(y^{k_i})\|^2} \,,$$

which means that

$$0 = \lim_{i\to\infty} \|z^{k_i} - x^{k_i}\| \quad \text{or/and} \quad +\infty = \lim_{i\to\infty} \lambda_{k_i} \|F(y^{k_i})\| \,. \qquad (23)$$

If $+\infty = \limsup_i \lambda_{k_i}$ then we are again in the setting of (22), and the argument above shows the existence of an accumulation point x^* which solves VIP(F, C), and ultimately the convergence of $\{x^k\}$ to this x^*. Hence, suppose that $\lambda_{k_i} \leq M$ for some $M > 0$ and all i. By the definition of z^{k_i}, for each $u \in C$

$$\lambda_{k_i} \langle F(x^{k_i}), u - z^{k_i}\rangle \geq \langle (\lambda_{k_i} F'(x^{k_i}) + I)(x^{k_i} - z^{k_i}), u - z^{k_i}\rangle \,. \qquad (24)$$

Choosing $u = x^{k_i}$, using the Cauchy-Schwarz inequality in the left-hand side of (24) and positive semidefiniteness of $F'(x^{k_i})$ in the right-hand side, we obtain

$$\lambda_{k_i} \|F(x^{k_i})\| \geq \|z^{k_i} - x^{k_i}\| \,.$$

Since $\{x^k\}$ is bounded, so is $\{F(x^k)\}$. Using further boundedness of $\{\lambda_{k_i}\}$, we conclude that $\{z^{k_i}\}$ is bounded. It then also follows that $\{y^{k_i}\}$ and $\{F(y^{k_i})\}$ are bounded. Hence, in (23) the first equality must hold. Let x^* be a limit of some subsequence of $\{x^{k_i}\}$. By (23), the corresponding subsequence of $\{z^{k_i}\}$ converges to the same limit x^*. Passing onto the limit in (24) along the subsequences of $\{x^{k_i}\}$ and $\{z^{k_i}\}$ converging to x^*, and taking into account boundedness of $\{\lambda_{k_i}\}$, for each $u \in C$ we obtain

$$\langle F(x^*), u - x^*\rangle \geq 0 \,,$$

which means that x^* solves VIP(F, C). Convergence of the whole sequence $\{x^k\}$ to x^* follows as before. ∎

3 SUPERLINEAR CONVERGENCE

Given the nice global convergence properties of Algorithm 1 (for monotone problems, the only requirement is existence of solutions), ideally one would have liked to show that in a neighbourhood of a solution satisfying some conditions, the unit stepsize $t_k = 1$ is admissible, thus implying the local superlinear rate. At this time, we do not have a proof establishing this fact. However, there exists a simple way to ensure superlinear convergence under reasonable assumptions. Essentially, it consists of checking, before performing the linesearch, whether the Josephy-Newton point z^k solves the proximal subproblem

$$z \in C\,, \quad \langle \lambda_k F(z) + z - x^k, u - z\rangle \geq 0\,, \quad \forall\, u \in C \tag{25}$$

within the tolerance required by the hybrid proximal point method. If this is so (as we show below, this would be always the case in a neighbourhood of a solution with certain properties), a readily computable hybrid proximal step is performed. This step ensures local superlinear rate of convergence. It can be checked that Algorithm 1 modified in this way also retains its global convergence properties, because hybrid proximal steps always decrease the distance to the solution set. This strategy had been used in [34] in the setting of NCP(F). We next provide some details.

Since z^k solves VIP(F_k, C), we have that $0 \in (F_k + N_C)(z^k)$. Hence, $-F_k(z^k) \in N_C(z^k)$, so that

$$F(z^k) - \lambda_k^{-1} F_k(z^k) \in T(z^k) = (F + N_C)(z^k)\,.$$

In this situation, the approximation criterion of [35] for subproblem (25) (recall also (13)) reduces to the following: z^k is acceptable if

$$\|\lambda_k F(z^k) - F_k(z^k) + z^k - x^k\| \leq (1-\theta)(\|\lambda_k F(z^k) - F_k(z^k)\|^2 + \|z^k - x^k\|^2)^{1/2}\,. \tag{26}$$

If this condition is satisfied, x^{k+1} is obtained by the same update formula as in Step 4 of Algorithm 1, but with $v^k = F(z^k) - \lambda_k^{-1} F_k(z^k)$, $y^k = z^k$ (and $\varepsilon^k = 0$). We point out that if $(F + N_C)^{-1}$ is Lipschitz-continuous at zero, then this is a superlinear step [35], provided $\lambda_k \to +\infty$ (e.g., if $\lambda_k = \tilde{\lambda}\|r(x^k)\|^{-s}$, as allowed in Algorithm 1).

We next show that (26) is satisfied when close to a solution $\bar{x}$ with $F'(\bar{x})$ positive definite. Let $F'(\cdot)$ be locally Lipschitz-continuous around $\bar{x}$ with modulus $L > 0$. As is easy to see,

$$\begin{aligned}\|\lambda_k F(z^k) - F_k(z^k) + z^k - x^k\| &= \lambda_k \|F(z^k) - F(x^k) - F'(x^k)(z^k - x^k)\| \\ &\leq L\lambda_k \|z^k - x^k\|^2 \\ &\leq L\tilde{\lambda}\|r(x^k)\|^{-s}\|z^k - x^k\|^2\,,\end{aligned}$$

where the last inequality is by the choice of λ_k in Algorithm 1. It is now evident that (26) is guaranteed to hold for all k sufficiently large if

$$0 = \lim_{k\to\infty} \|r(x^k)\|^{-s}\|z^k - x^k\|\,. \tag{27}$$

Positive definiteness of $F'(\bar{x})$ implies that $\bar{x}$ is regular [22], and thus the Josephy-Newton step is superlinear:

$$\|z^k - \bar{x}\| = o(\|x^k - \bar{x}\|) .$$

The above relation easily leads to

$$\|z^k - x^k\| \leq 2\|x^k - \bar{x}\| .$$

Also, from positive definiteness of $F'(\bar{x})$ it follows [15] that for all k sufficiently large,

$$\|x^k - \bar{x}\| \leq M\|r(x^k)\| , \quad M > 0 .$$

Combining the last two relations, we have that

$$\|r(x^k)\|^{-s}\|z^k - x^k\| \leq 2M\|r(x^k)\|^{1-s} ,$$

implying (27) if $s \in (0,1)$ is chosen in Algorithm 1.

4 CONCLUDING REMARKS

We have proposed a new globalization strategy for the Josephy-Newton method for solving monotone variational inequalities. Our approach is based on checking an approximation criterion for a proximal point subproblem, without making a direct attempt to solve this (nonlinear) subproblem. The novel idea consists of performing a linesearch not only in the problem variables, but also for the regularization and relaxation parameters involved in the inexact proximal point framework.

Global convergence properties of the resulting algorithm are very satisfactory. In particular, the whole sequence of iterates converges to a solution of the variational inequality, provided a solution exists and the underlying mapping is monotone and continuously differentiable.

One area where some improvements would be desirable, are local convergence properties. First, in its current form, the method requires checking an additional criterion in order to ensure local superlinear rate of convergence. In some sense, this is more of an aesthetical drawback, since no significant extra computational work is involved. In fact, globalization strategies based on merit functions also require an extra criterion to ensure that the full Newton step is asymptotically accepted. Nevertheless, it would be nice to come up with a “one-piece” linesearch rule which guarantees both the global convergence and the local superlinear rate. A more important issue, however, is obtaining fast local convergence in the case of nonisolated solutions, i.e., under more general regularity-type assumptions. This issue certainly deserves further investigation.

REFERENCES

[1] G. Auchmuty (1989). Variational principles for variational inequalities. *Numerical Functional Analysis and Optimization*, **10**, 863–874.

[2] A. A. Auslender (1976). *Optimisation Méthodes Numériques.* Masson, Paris.

[3] J. F. Bonnans (1994). Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Applied Mathematics and Optimization*, **29**, 161–186.
[4] R.S. Burachik, A.N. Iusem, and B.F. Svaiter (1997). Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Analysis*, **5**, 159–180.
[5] R.S. Burachik, C.A. Sagastizábal, and B. F. Svaiter (1999). Bundle methods for maximal monotone operators. In: R. Tichatschke and M. Théra (Eds.), *Ill-posed variational problems and regularization techniques*, Lecture Notes in Economics and Mathematical Systems, No. 477, pp. 49–64. Springer–Verlag, Berlin.
[6] R.S. Burachik, C.A. Sagastizábal, and B.F. Svaiter (1999). $\mathcal{E}$-Enlargements of maximal monotone operators: Theory and applications. In: M. Fukushima and L. Qi (Eds.), *Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, pp. 25–44. Kluwer Academic Publishers, Dordrecht.
[7] R.S. Burachik and B.F. Svaiter (1999). $\mathcal{E}$-Enlargements of maximal monotone operators in Banach spaces. *Set-Valued Analysis*, **7**, 117–132.
[8] R.W. Cottle, F. Giannessi, and J.-L. Lions (1980). *Variational Inequalities and Complementarity Problems: Theory and Applications*. Wiley, New York.
[9] F. Facchinei and J. Soares (1997). A new merit function for nonlinear complementarity problems and a related algorithm. *SIAM Journal on Optimization*, **7**, 225–247.
[10] M.C. Ferris, C. Kanzow, and T.S. Munson (1999). Feasible descent algorithms for mixed complementarity problems. *Mathematical Programming*, **86**, 475–497.
[11] A. Fischer (1992). A special Newton-type optimization method. *Optimization*, **24**, 269–284.
[12] A. Fischer and H. Jiang (2000). Merit functions for complementarity and related problems: A survey. *Computational Optimization and Applications*, **17**, 159–182.
[13] M. Fukushima (1992). Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. *Mathematical Programming*, **53**, 99–110.
[14] M. Fukushima (1996). Merit functions for variational inequality and complementarity problems. In: G. Di Pillo and F. Giannessi (Eds.), *Nonlinear Optimization and Applications*, pp. 155–170. Plenum Publishing Corporation, New York.
[15] P.T. Harker and J.-S. Pang (1990). Finite-dimensional variational inequality problems: A survey of theory, algorithms and applications. *Mathematical Programming*, **48**, 161–220.
[16] A.N. Iusem and B.F. Svaiter (1997). A variant of Korpelevich's method for variational inequalities with a new search strategy. *Optimization*, **42**, 309–321.
[17] N.H. Josephy (1979). Newton's method for generalized equations. Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin.
[18] B. Lemaire (1989). The proximal algorithm. In: J.-P. Penot (Ed.), *New Methods of Optimization and Their Industrial Use. International Series of Numerical Mathematics 87*, pp. 73–87. Birkhauser, Basel.
[19] P. Marcotte and J.-P. Dussault (1987). A note on a globally convergent Newton method for solving monotone variational inequalities. *Operations Research Letters*, **6**, 35–42.
[20] J.-S. Pang (1986). Inexact Newton methods for the nonlinear complementarity problem. *Mathematical Programming*, **36**, 54–71.
[21] J.-M. Peng (1997). Equivalence of variational inequality problems to unconstrained optimization. *Mathematical Programming*, **78**, 347–356.
[22] J.-M. Peng and M. Fukushima (1999). A hybrid Newton method for solving the variational inequality problem via the D-gap function. *Mathematical Programming*, **86**, 367–386.
[23] J.-M. Peng, C. Kanzow, and M. Fukushima (1999). A hybrid Josephy-Newton method for solving box constrained variational inequality problem via the D-gap function. *Optimization Methods and Software*, **10**, 687–710.
[24] S.M. Robinson (1980). Strongly regular generalized equations. *Mathematics of Operations Research*, **5**, 43–62.
[25] R.T. Rockafellar (1976). Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, **14**, 877–898.
[26] C.A. Sagastizábal and M.V. Solodov (2001). On the relation between bundle methods for maximal monotone inclusions and hybrid proximal point algorithms. In: D. Butnariu,

Y. Censor, and S. Reich (Eds.), *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, volume 8 of *Studies in Computational Mathematics*, pp. 441–455. Elsevier Science B.V., Amsterdam.

[27] M.V. Solodov (2001). A class of globally convergent algorithms for pseudomonotone variational inequalities. In: M.C. Ferris, O.L. Mangasarian, and J.-S. Pang (Eds.), *Complementarity: Applications, Algorithms and Extensions*, pp. 297–315. Kluwer Academic Publishers, Dordrecht.

[28] M.V. Solodov (2001). Implicit Lagrangian. In: C. Floudas and P. Pardalos (Eds.), *Encyclopedia of Optimization*. Kluwer Academic Publishers, Dordrecht.

[29] M.V. Solodov and B.F. Svaiter (1999). 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*, pp. 355–369. Kluwer Academic Publishers, Dordrecht.

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

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

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

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

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

[35] M.V. Solodov and B.F. Svaiter (2001). A unified framework for some inexact proximal point algorithms. *Numerical Functional Analysis and Optimization*, **22**, 1013–1035.

[36] M.V. Solodov and P. Tseng (2000). Some methods based on the the D-gap function for solving monotone variational inequalities. *Computational Optimization and Applications*, **17**, 255–277.

[37] K. Taji, M. Fukushima, and T. Ibaraki (1993). A globally convergent Newton method for solving strongly monotone variational inequalities. *Mathematical Programming*, **58**, 369–383.

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

[39] N. Yamashita, K. Taji, and M. Fukushima (1997). Unconstrained optimization reformulations of variational inequality problems. *Journal of Optimization Theory and Applications*, **92**, 439–456.