Chapter 14

A CLASS OF GLOBALLY CONVERGENT ALGORITHMS FOR PSEUDOMONOTONE VARIATIONAL INEQUALITIES

M. V. Solodov

Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil *

solodov@impa.br

Abstract We describe a fairly broad class of algorithms for solving variational inequalities, global convergence of which is based on the strategy of generating a hyperplane separating the current iterate from the solution set. The methods are shown to converge under very mild assumptions. Specifically, the problem mapping is only assumed to be continuous and pseudomonotone with respect to at least one solution. The strategy to obtain (super)linear rate of convergence is also discussed. The algorithms in this class differ in the tools which are used to construct the separating hyperplane. Our general scheme subsumes an extragradient-type projection method, a globally and locally superlinearly convergent Josephy-Newton-type method, a certain minimization-based method, and a splitting technique.

1 INTRODUCTION

Given a function $F : \boldsymbol{R}^n \to \boldsymbol{R}^n$ and a set $C \subset \boldsymbol{R}^n$, the classical variational inequality problem [1, 3, 8, 9, 5], abbreviated VIP(F, C), is to find a point x such that

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

where $\langle \cdot, \cdot \rangle$ denotes the usual inner product in $\boldsymbol{R}^n$. Throughout this paper we assume that C is closed convex, F is continuous, and the solution

*This research is supported in part by CNPq Grant 300734/95-6 and by PRONEX–Optimization.

set of VIP(F, C), denoted by S, is nonempty. When C is the nonnegative orthant $\boldsymbol{R}^n_+$, VIP(F, C) reduces to the nonlinear complementarity problem NCP(F), which is to find a point $x \in \boldsymbol{R}^n$ such that

$$x \geq 0\,, \quad F(x) \geq 0\,, \quad \langle F(x), x \rangle = 0\,.$$

More generally, when C is a box (i.e., a Cartesian product of intervals), VIP(F, C) is the mixed complementarity problem [6].

For the convergence of our general algorithm, we further assume that there exists $\bar{x}$, some element of the solution set S, such that

$$\langle F(y), y - \bar{x} \rangle \geq 0 \quad \text{for all } y \in C\,. \tag{14.1}$$

Under this condition, we prove that the sequence of iterates generated by our method is bounded and each of its accumulation points is a solution of VIP(F, C). Furthermore, if condition (14.1) holds for all $\bar{x} \in S$ then the accumulation point is unique, i.e., the whole sequence of iterates converges to a solution.

Condition (14.1) is very mild. It can be thought of as pseudomonotonicity of F on the feasible set C with respect to *one* solution of VIP(F, C). Note that pseudomonotonicity of F with respect to the whole solution set S is a stronger, but still fairly mild assumption. It is clear that the latter is satisfied if F is monotone, i.e.,

$$\langle F(x) - F(y), x - y \rangle \geq 0 \quad \text{for all } x, y \in \boldsymbol{R}^n\,.$$

More generally, it also holds if F is pseudomonotone (as defined in [12]), i.e., for all $x, y \in \boldsymbol{R}^n$

$$\langle F(y), x - y \rangle \geq 0 \quad \Longrightarrow \quad \langle F(x), x - y \rangle \geq 0\,.$$

Moreover, it is possible to construct examples where (14.1) is satisfied but F is not monotone or pseudomonotone everywhere. Of course, since this condition involves (unknown!) solution of the problem, in general it is not possible to verify it a priori. Still, one might expect (14.1) to hold under some kind of generalized monotonicity assumptions on F, some of which are not difficult to check (see [20, 4]). And at any rate, (14.1) obviously holds on the class of (pseudo)monotone problems.

The general framework to be presented is essentially based on simple geometric considerations. If given a point $x \in \boldsymbol{R}^n$ we could find (by whichever means) some point $y \in C$ such that

$$\langle F(y), x - y \rangle > 0\,, \tag{14.2}$$

then condition (14.1) implies that the hyperplane

$$H_y := \{w \in \boldsymbol{R}^n \mid \langle F(y), w - y \rangle = 0\}$$

separates the point x from any solution $\bar{x}$ satisfying (14.1). It then follows that moving from x in the direction of this hyperplane decreases the distance to the solution point $\bar{x}$. Formally, this can be stated as follows.

Lemma 14.1 *Let $x \in \boldsymbol{R}^n$ and $y \in C$ be any two points satisfying (14.2), and let*

$$\hat{x}(\beta) := x - \beta \frac{\langle F(y), x - y \rangle}{\|F(y)\|^2} F(y)\,,$$

where $\beta \geq 0$. Then for any $\bar{x} \in S$ satisfying (14.1) it holds that

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

Proof Denote

$$t := \frac{\langle F(y), x - y \rangle}{\|F(y)\|^2} > 0\,,$$

so that $\hat{x}(\beta) = x - \beta t F(y)$. With this notation, we obtain

$$\begin{aligned} \|\hat{x}(\beta) - \bar{x}\|^2 &= \|x - \bar{x}\|^2 + t^2\beta^2\|F(y)\|^2 - 2t\beta \langle F(y), x - \bar{x} \rangle \\ &\leq \|x - \bar{x}\|^2 + t^2\beta^2\|F(y)\|^2 - 2t\beta \langle F(y), x - y \rangle \\ &= \|x - \bar{x}\|^2 - \beta(2 - \beta)t \langle F(y), x - y \rangle\,, \end{aligned}$$

where the inequality follows from (14.1), and the last equality follows from the relation $t\|F(y)\|^2 = \langle F(y), x - y \rangle$. Using now the definition of t concludes the proof. □

Lemma 14.1 implies that if we choose $\beta \in (0, 2)$ then the point $\hat{x}(\beta)$ is closer to $\bar{x} \in S$ than the point x. Note that taking $\beta = 1$ corresponds to $\hat{x}(\beta)$ being the projection of x onto the separating hyperplane H_y, while choosing $\beta \in (1, 2)$ or $\beta \in (0, 1)$ corresponds to a step with over-relaxation or under-relaxation, respectively. Alternatively, $\hat{x}(\beta)$ can also be viewed as a projection of x onto a shifted hyperplane $H_y(\beta)$, with $H_y(1) = H_y$. If furthermore the point y is obtained in such a way that the quantity $\langle F(y), x - y \rangle$ has a meaningful connection to "optimality" of x, then Lemma 14.1 can be used as the basis for devising algorithms for solving VIP(F, C) under the assumption (14.1).

Of course, the ways to compute (given some x) an acceptable point y, is the central question for this algorithmic development. In this paper, after presenting a general framework (Section 2), we shall give a number of examples of how this can be done (Section 3). These include a projection algorithm, a (Josephy-)Newton method, a splitting method, and a certain minimization-based algorithm. The separation-projection methodology appears to be useful for devising some new computational techniques, forcing convergence of iterative schemes that otherwise might diverge, and for relaxing error tolerance requirements in solving subproblems within some convergent algorithms. Each of those issues will be illustrated by the given specific instances of the general method.

2 GENERAL ALGORITHMIC FRAMEWORK

A common approach to solving VIP(F, C) consists of solving a sequence of subproblems VIP(F_k, C) approximating the original problem (see, for example, [9]). This is a natural view, which we shall adopt also for our development. For the resulting algorithm to be potentially practical, it is clear that F_k must be such that subproblems are structurally easier to solve than the original problem. It is also reasonable to say that F_k should provide some kind of approximation for F around the current iterate x^k. It is therefore natural to write

$$F_k(x) := F(x^k) + A_k(x^k, x)\,, \tag{14.3}$$

where $A_k : \boldsymbol{R}^n \times \boldsymbol{R}^n \to \boldsymbol{R}^n$ is some mapping whose choice will define each specific algorithm.

For $x \in \boldsymbol{R}^n$ define

$$r(x) := \|x - P_C[x - F(x)]\|\,,$$

where $P_C[\cdot]$ stands for the orthogonal projection operator onto the set C. It is well known that some x solves VIP(F, C) if, and only if, $r(x) = 0$. We are now ready to formally state our general method.

Algorithm 14.2 Choose parameters $t > 0$, $0 \le \rho_u < 1$, $0 < \sigma_l \le \sigma_u < 1$, $0 < \beta_l \le \beta_u < 2$, $0 < \gamma < 1$ and $0 < \mu_l \le \mu_u$. Choose any $x^0 \in \boldsymbol{R}^n$. Set $k := 0$.

Variational inequality subproblem. If $r(x^k) = 0$, stop. Otherwise, choose a mapping A_k such that $A_k(x^k, \cdot)$ is strongly monotone with modulus $\mu_k > 0$, where

$$\min\{\mu_l\,,\, r(x^k)^t\} \le \mu_k \le \mu_u\,. \tag{14.4}$$

Find $z^k \in C$ as an approximate solution of VIP(F_k, C) in the sense that

$$\langle F_k(z^k) + e^k, x - z^k\rangle \ge 0 \quad \forall\, x \in C\,, \tag{14.5}$$

where

$$\|e^k\| \le \rho_k \mu_k \|z^k - x^k\| , \quad \rho_k \in [0, \rho_u] . \tag{14.6}$$

Linesearch step. Choose $\sigma_k \in [\sigma_l, \sigma_u]$. Find $y^k = x^k + \alpha_k(z^k - x^k)$, where $\alpha_k = \gamma^{m_k}$ with m_k being the smallest nonnegative integer m such that

$$\langle F(x^k + \gamma^m(z^k - x^k)), x^k - z^k \rangle \ge \sigma_k \mu_k (1 - \rho_k) \|z^k - x^k\|^2 . \tag{14.7}$$

Projection step. Choose $\beta_k \in [\beta_l, \beta_u]$. Compute

$$x^{k+1} = P_{C_k}[\hat{x}^k] , \tag{14.8}$$

where

$$\hat{x}^k = x^k - \beta_k \frac{\langle F(y^k), x^k - y^k \rangle}{\|F(y^k)\|^2} F(y^k) , \tag{14.9}$$

and

$$C_k = C \quad \text{or} \quad C_k = C \cap \{w \in \boldsymbol{R}^n \mid \langle F(y^k), w - y^k \rangle \le 0\} . \tag{14.10}$$

Set $k := k + 1$; and repeat.

Before proceeding with the convergence analysis, some comments are in order. The requirement that $A_k(x^k, \cdot)$ be strongly monotone is natural and not restrictive. We can always satisfy it by adding a regularization term with a suitable parameter (for monotone problems, this parameter can be exactly μ_k). Note that the regularization parameter is allowed to go to zero (but no faster than some power of the problem residual $r(x^k)$). This feature is important for ensuring fast local convergence of Newton-type methods.

Our condition (14.5) for inexact solution of subproblems is in the spirit of conditions used in [16, 15, 14]. For implementation purposes, it might be preferable to to re-write it in terms of some computable subproblem residual, as in [21]. Here, we shall keep it in the form of (14.5) because it simplifies somewhat the subsequent analysis.

The role of the linesearch step is to ensure the separation property (i.e., inequality (14.2)), so that Lemma 14.1 comes into play, and global convergence of the method can be obtained. Observe that our linesearch step is computationally inexpensive, as it only requires evaluation of F at the trial points. In particular, in contrast to many projection algorithms for solving VIP(F, C), it does not require a projection operation for each trial point. This is an advantage when C does not have a special structure and making projections is costly.

Note that if one is to use the second choice for the set C_k in (14.10), then $P_{C_k}[\hat{x}^k] = P_{C_k}[x^k]$ (see [22, Lemma 2.2]). That is, in that case x^{k+1} is the projection of x^k onto the intersection of the feasible set C with the halfspace containing the solution $\bar{x} \in S$. In particular, one does not have to compute $\hat{x}^k$. Whether it is better to use the first or second choice for C_k in (14.10) depends mostly on the structure of the set C. Theoretically, it is clear that projecting onto the intersection yields faster convergence (because we get closer to the solution). This also seems to have been confirmed by some numerical experiments in [22]. However, if projection onto C is explicit then it might be better to choose $C_k = C$, because in that case x^{k+1} would also be computed explicitly. Introducing an additional constraint, even though it is linear, may ruin this structure.

Finally, we formally specify requirements regarding the choice of the algorithmic mapping A_k.

1. For each k, $A_k(x^k, \cdot)$ is strongly monotone with modulus $\mu_k > 0$:

$$\langle A_k(x^k, z_1) - A_k(x^k, z_2), z_1 - z_2 \rangle \geq \mu_k \|z_1 - z_2\|^2 .$$

2. For each k, $A_k(x^k, x^k) = 0$.

3. For each k, there exists $M_k > 0$ such that

$$\|A_k(x, z)\| \leq M_k \|x - z\| \quad \forall\, x, z \in \boldsymbol{R}^n .$$

   Furthermore, for any bounded set $D \subset \boldsymbol{R}^n$ there exists some $M > 0$ (independent of k) such that

$$\|A_k(x, z)\| \leq M \|x - z\| \quad \forall\, x, z \in D .$$

As we shall see in Section 3, these conditions are natural and are satisfied in many cases of interest.

We next establish convergence properties of Algorithm 14.2.

Theorem 1 *Suppose F is continuous, and let $\{x^k\}$ be any sequence generated by Algorithm 14.2.*

If condition (14.1) holds for some $\bar{x} \in S$ then the sequence $\{x^k\}$ is bounded and each of its accumulation points is a solution of VIP(F, C).

If (14.1) holds for all $\bar{x} \in S$ then the sequence $\{x^k\}$ converges to a solution of VIP(F, C).

Proof We start with showing that the method is well-defined. First note that by strong monotonicity of $A_k(x^k, \cdot)$, VIP(F_k, C) always has (unique) solution. Hence, it also has approximate solutions satisfying (14.5), (14.6) (this is because the exact solution always satisfies these conditions, for any value of ρ_k).

We next show that the linesearch procedure (14.7) always terminates with a positive stepsize α_k. Suppose that for some iteration index k this is not the case. That is, for all integers m we have

$$\langle F(x^k + \gamma^m(z^k - x^k)), x^k - z^k\rangle < \sigma_k\mu_k(1-\rho_k)\|z^k - x^k\|^2 .$$

Since F is continuous, passing onto the limit as $m \to \infty$, we obtain

$$\langle F(x^k), x^k - z^k\rangle \le \sigma_k\mu_k(1-\rho_k)\|z^k - x^k\|^2 . \tag{14.11}$$

On the other hand, by (14.5), we obtain that

$$\begin{aligned} \langle F(x^k), x^k - z^k\rangle &\ge \langle A_k(x^k, z^k) + e^k, z^k - x^k\rangle \\ &\ge \langle A_k(x^k, z^k) - A_k(x^k, x^k), z^k - x^k\rangle - \|e^k\|\|z^k - x^k\| \\ &\ge \mu_k(1-\rho_k)\|z^k - x^k\|^2 , \end{aligned} \tag{14.12}$$

where the assumptions on A_k and (14.6) were used in the last two relations. From the assumptions on the parameters of the method, it follows that (14.11) and (14.12) can be consistent only if $z^k = x^k$.

Note that by (14.5),

$$z^k - P_C[z^k - F_k(z^k) - e^k] = 0 .$$

By the triangle and Cauchy-Schwarz inequalities, the nonexpansiveness of the projection operator, the assumption on A_k, and (14.6), we have

$$\begin{aligned} \|z^k - x^k\| &= \|P_C[z^k - F_k(z^k) - e^k] - x^k\| \\ &\ge r(x^k) - \|P_C[x^k - F(x^k)] - P_C[z^k - F_k(z^k) - e^k]\| \\ &\ge r(x^k) - \|x^k - z^k + A_k(x^k, z^k) + e^k\| \\ &\ge r(x^k) - (1 + M_k + \rho_k\mu_k)\|x^k - z^k\| . \end{aligned}$$

Hence,

$$(2 + M_k + \rho_k\mu_k)\|z^k - x^k\| \ge r(x^k) . \tag{14.13}$$

Now, if the linesearch is activated then $r(x^k) > 0$ (otherwise the method would have stopped), and so necessarily $z^k \ne x^k$, by (14.13). The latter contradicts the conclusion above, obtained assuming that the linesearch procedure does not terminate in a finite number of steps. Hence, that assumption was invalid, and so the linesearch always terminates with some positive stepsize α_k.

It also follows immediately from $z^k \neq x^k$ and (14.7) that $F(y^k) \neq 0$, and so the point $\hat{x}^k$ is well-defined. Furthermore, C_k is a closed convex set which is nonempty (for either choice in (14.10), $\bar{x} \in C_k$), which implies that the projection step is also well-defined. This concludes the proof that the whole algorithm is well-defined.

We assume, from now on, that $r(x^k) > 0$ for each k, so that an infinite sequence $\{x^k\}$ is generated. Take any $\bar{x} \in S$ satisfying (14.1). Since $\bar{x} \in C_k$, (14.8) clearly implies that

$$\|x^{k+1} - \bar{x}\| \leq \|\hat{x}^k - \bar{x}\| .$$

By (14.7) we have that

$$\langle F(y^k), x^k - y^k\rangle = \alpha_k \langle F(y^k), x^k - z^k\rangle \geq \sigma_k \mu_k (1-\rho_k)\alpha_k \|z^k - x^k\|^2 ,$$

and so Lemma 14.1 is readily applicable. In particular,

$$\begin{aligned}
\|x^{k+1} - \bar{x}\|^2 &\leq \|\hat{x}^k - \bar{x}\|^2 \\
&\leq \|x^k - \bar{x}\|^2 - \beta_k(2-\beta_k)\left(\frac{\langle F(y^k), x^k - y^k\rangle}{\|F(y^k)\|}\right)^2 \\
&\leq \|x^k - \bar{x}\|^2 - \beta_k(2-\beta_k)\left(\sigma_k\mu_k(1-\rho_k)\alpha_k \frac{\|z^k - x^k\|^2}{\|F(y^k)\|}\right)^2
\end{aligned}$$

and it follows that the sequence $\{\|x^k - \bar{x}\|\}$ converges, and so the sequence $\{x^k\}$ is bounded. Taking further into account assumptions on the parameters, the last relation implies that

$$0 = \lim_{k\to\infty} \mu_k \alpha_k \frac{\|z^k - x^k\|^2}{\|F(y^k)\|} . \tag{14.14}$$

Suppose there exists $\tilde{x}$, an accumulation point of the sequence $\{x^k\}$, which is not a solution of VIP(F, C). In that case, $r(\tilde{x}) > 0$. Let $\{x^{k_i}\}$ be the subsequence of $\{x^k\}$ converging to $\tilde{x}$. By (14.4) we have that in this case $\liminf_{i\to\infty} \mu_{k_i} > 0$. Note that, by (14.12),

$$\|F(x^{k_i})\| \geq \mu_{k_i}(1-\rho_{k_i})\|z^{k_i} - x^{k_i}\| ,$$

and, from boundedness of $\{x^k\}$, it follows that $\{z^{k_i}\}$ is bounded. From (14.13) and the assumption on A_k, we now have that

$$(2 + M + \rho_u\mu_u)\|z^{k_i} - x^{k_i}\| \geq r(x^{k_i}) ,$$

where M is a positive constant (depending on $\{x^{k_i}\}, \{z^{k_i}\}$). Hence, $\liminf_{i\to\infty} \|z^{k_i} - x^{k_i}\| > 0$. Furthermore, since $\{y^{k_i}\}$ is obviously bounded, so is $\{F(y^{k_i})\}$. Using all these facts, (14.14) implies that

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

The latter is equivalent to saying that $m_{k_i} \to \infty$. It follows that for every sufficiently large i, the stepsize is decreased at least twice, i.e., $m_{k_i} \geq 2$. This means that the stepsize rule (14.7) is not satisfied for the value of $\gamma^{m_{k_i}-1}$, i.e.

$$\langle F(x^{k_i} + \gamma^{m_{k_i}-1}(z^{k_i} - x^{k_i})), x^{k_i} - z^{k_i}\rangle < \sigma_{k_i}(1-\rho_{k_i})\mu_{k_i}\|z^{k_i} - x^{k_i}\|^2 .$$

Taking into account boundedness of all the sequences involved, and passing onto a subsequence if necessary, as $i \to \infty$ we obtain

$$\langle F(\tilde{x}), \tilde{x} - \tilde{z}\rangle \leq \tilde{\sigma}(1-\tilde{\rho})\tilde{\mu}\|\tilde{z} - \tilde{x}\|^2 ,$$

where $\tilde{x}$, $\tilde{\mu}$, $\tilde{\rho}$ and $\tilde{z}$ are limits of corresponding subsequences. On the other hand, passing onto the limit along the same subsequences in (14.12), we have that

$$\langle F(\tilde{x}), \tilde{x} - \tilde{z}\rangle \geq (1-\tilde{\rho})\tilde{\mu}\|\tilde{z} - \tilde{x}\|^2 .$$

But the last two relations are a contradiction because $\tilde{\sigma} \in (0,1)$, $\tilde{\rho} < 1$, while $\tilde{z} \neq \tilde{x}$, $\tilde{\mu} > 0$ by the assumption that $r(\tilde{x}) > 0$. Hence, this case is not possible. It follows that every accumulation point of $\{x^k\}$ solves VIP(F, C).

Now, if condition (14.1) holds for all $\bar{x} \in S$, then we can choose above as $\bar{x}$ any accumulation point of $\{x^k\}$. As before, we then have that the sequence $\{\|x^k - \bar{x}\|\}$ converges. Since $\bar{x}$ is actually an accumulation point of $\{x^k\}$, it must then be the case that $\{x^k\}$ converges to $\bar{x} \in S$. This completes the proof. $\square$

Note that strong monotonicity of $A_k(x^k, \cdot)$ was used mainly to ensure that subproblems VIP(F_k, C) are solvable. In the proof, we essentially need the following property:

$$\langle A_k(x^k, z^k), z^k - x^k\rangle \geq \mu_k\|z^k - x^k\|^2 , \tag{14.15}$$

which may hold without strong monotonicity of $A_k(x, \cdot)$ (more on this in Section 3). So for the convergence result to hold, we could alternatively *assume* that subproblems have (approximate) solutions, and that (14.15) holds.

3 SOME SPECIFIC ALGORITHMS

We next turn our attention to some specific choices of the algorithmic mapping A_k. The list of possibilities discussed in this section is by no means exhaustive. Our objective here is to highlight only some of the situations where our general strategy can be useful.

3.1 A PROJECTION ALGORITHM

Projection algorithm corresponds to the most simple choice of

$$A_k(x, z) = z - x \quad \text{for all } k\,.$$

Obviously, this A_k satisfies all the assumptions imposed in Section 2. It is easy to see that the (exact) solution of VIP(F_k, C) in this case is given by

$$z^k = P_C[x^k - F(x^k)]\,,$$

and the point defining the separating hyperplane has the form

$$y^k = \alpha_k P_C[x^k - F(x^k)] + (1 - \alpha_k)x^k\,, \quad \alpha_k \in (0, 1]\,.$$

The next iterate x^{k+1} is obtained by projecting x^k either on this hyperplane and then onto C, or directly onto the intersection of C with the halfspace containing a solution point. We thus recover some methods studied in [10, 22]. Note that in [10] F was assumed to be monotone, and in [22] pseudomonotonicity with respect to the whole solution set was used. In Section 2, our assumption is weaker. Furthermore, inexact computation of z^k is allowed, which might be a useful feature in practice. [In principle, we could also analyze inexact computing of x^{k+1}, so that to make things consistent in this projection method. This is certainly possible. We shall not pursue this issue, mainly for the sake of clarity of exposition.]

One advantage of this algorithm compared to most other projection techniques (for example, see [23] and references therein), is that no projections are necessary within the linesearch procedure. Thus only two projections per iteration are required.

Note that the strong monotonicity constant in the simplest implementation is $\mu_k = 1$ for all k. In principle, no adaptive regularization is needed in the projection method. However, as a practical matter, it is important to use an adaptive stepsize (or regularization, in the terminology of the present paper) to compute the first projection point, i.e., $z^k = P_C[x^k - \mu_k^{-1} F(x^k)]$, which corresponds to taking $A_k(x, z) = \mu_k(z - x)$. In [22] some ways to coordinate the two stepsizes μ_k and α_k were proposed. In the numerical experiments of [22] this issue appeared to be of importance.

3.2 A JOSEPHY-NEWTON METHOD

The Josephy-Newton (or linearization) method consists of solving successive approximations VIP(F_k, C), where $F_k(x) = F(x^k) + F'(x^k)(x -$

x^k), assuming F is differentiable. While this scheme possesses fast local convergence [11, 17, 2] in a neighbourhood of a regular [18] solution, there are serious problems with ensuring global convergence. In particular, far from a regular solution of the problem, there is no guarantee that the linearized subproblems are solvable. And even if the subproblem solution exists, there is no guarantee that it actually constitutes some progress towards solving the original problem, VIP(F, C).

A modified Josephy-Newton-type method was developed in [21] to address these two problems, (this was done in the context of monotone NCP(F), but global convergence analysis readily extends to the more general setting). In the context of the present paper, this algorithm corresponds to taking

$$A_k(x, z) = (F'(x) + \mu_k I)(z - x)\,,$$

where $\mu_k > 0$ is a suitable regularization parameter. It is easy to see that this choice satisfies all the assumptions of Section 2 if F is monotone (so that $F'(x)$ is positive semidefinite). If $F'(x)$ is not positive semidefinite, one could add an extra regularization term, and then the global convergence analysis still goes through.

By applying the framework of Section 2 to a monotone VIP(F, C), we obtain a Newton-type method which is globally well-defined and convergent to a solution without making any regularity assumptions. Note that the solution set S may even be unbounded. To see why this is a very strong property (in the context of Newtonian methods), it is enough to think of the more typical globalization strategies based on merit functions. For methods based on those strategies, when S is unbounded, even boundedness of iterates cannot be established in general (this is because the level sets of merit functions are necessarily unbounded in this case). Furthermore, global convergence of the whole sequence to a solution typically requires assumptions which imply (global) uniqueness of the solution point. We refer the reader to [21] for a more detailed discussion.

The issue of the local rate of convergence of this method will be addressed in Section 4.

3.3 A MINIMIZATION-BASED METHOD

Consider, for the moment, a monotone affine VIP(F, C), i.e., C is a polyhedral set, and $F(x) = Mx + q$ where M is a positive semidefinite matrix. An interesting iterative procedure was suggested in [7] for solving VIP(F, C) in this special case. Given $x^k \in C$, compute

$$x^{k+1} = \arg\min_{z \in C} \langle Mz + q, z - x^k \rangle + \frac{\mu_k}{2} \|z - x^k\|^2 , \tag{14.16}$$

where $\mu_k > 0$. One advantage of this approach is the induced symmetry of the quadratic term which permits the use of reliable and efficient quadratic programming techniques for solving the subproblems, such as the successive overrelaxation methods [13]. This is indeed an appealing feature. Unfortunately, a sequence defined by (14.16) need not converge (we note that no convergence proof was given in [7]). To see this, consider the rotation operator in $\boldsymbol{R}^2$ given by

$$M = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} ,$$

and let $C = \boldsymbol{R}^2$, $q = 0$. In this case, VIP(F, C) reduces to the system of linear equations $Mx = 0$ which has the unique solution $\bar{x} = (0, 0)^\top$. Take any $x^k \neq \bar{x}$. It is easy to check that (14.16) yields

$$x^{k+1} = \left(I + \frac{1}{\mu_k} M^\top \right) x^k ,$$

and

$$\|x^{k+1}\|^2 = \left(1 + \frac{1}{\mu_k^2} \right) \|x^k\|^2 .$$

Therefore for any choice of $\mu_k > 0$, $\|x^{k+1}\| > \|x^k\|$, and the sequence $\{x^k\}$ generated this way does not converge to $\bar{x} = 0 \in \boldsymbol{R}^2$. In fact, it diverges (at least if $\{\mu_k\}$ is bounded above).

We next show that if, instead of accepting the solution of (14.16) as our next iterate, we use it within the framework of Algorithm 14.2, then a convergent algorithm is obtained. In fact, a solution of (14.16) immediately gives a separating hyperplane, so no linesearch is necessary. Furthermore, we extend the idea of (14.16) to the general VIP(F, C) setting. Let F be monotone and differentiable, and consider the mapping

$$A_k(x, z) = F(z) - F(x) + (F'(z)^\top + \mu_k I)(z - x) .$$

It is easy to see that in this case VIP(F_k, C) is equivalent to the first-order necessary optimality conditions of the optimization problem

$$\min_{z \in C} \langle F(z), z - x^k \rangle + \frac{\mu_k}{2} \|z - x^k\|^2 , \tag{14.17}$$

which is a formal extension of (14.16). If F is affine, then A_k satisfies the assumptions of Section 2 for any $\mu_k > 0$. If F is not affine, then making $A_k(x, \cdot)$ strongly monotone may not be so easy (when C is compact, this can still be done by taking the regularization parameter large enough). Note however that we could use the weaker condition discussed in the end of Section 2 (see (14.15)). Indeed, we have that

$$\begin{aligned}
&\langle A_k(x^k, z^k), z^k - x^k\rangle \\
&= \langle F(z^k) - F(x^k) + (F'(z^k)^\top + \mu_k I)(z^k - x^k), z^k - x^k\rangle \\
&\geq \mu_k \|z^k - x^k\|^2 ,
\end{aligned}$$

by monotonicity of F, and so (14.15) holds. In that case, we have to make a separate assumption that the optimization problem (14.17) has stationary points.

By the first-order necessary optimality condition, z^k, an approximate solution (or stationary point) of (14.17), satisfies

$$\begin{aligned}
\langle F(z^k), x^k - z^k\rangle &\geq \langle (F'(z^k)^\top + \mu_k I)(z^k - x^k) - e^k, z^k - x^k\rangle \\
&\geq \mu_k \|z^k - x^k\|^2 - \langle e^k, z^k - x^k\rangle \\
&\geq \mu_k(1 - \rho_k)\|z^k - x^k\|^2 ,
\end{aligned}$$

where the second inequality follows from positive semidefiniteness of $F'(z^k)$, and the third from (14.6). In particular, we conclude that the linesearch rule (14.7) is always satisfied with $m = 0$ for any $\sigma_k \leq 1$. Hence, $\alpha_k = 1$ and $y^k = z^k$ for all k (no linesearch is necessary).

Theorem 1 implies that any sequence based on solving (14.17) and modified using the strategy of Section 2 converges to a solution of $\mathrm{VIP}(F, C)$. Let us go back to the example with the rotation operator considered above. For simplicity, in Algorithm 14.2 set $\mu_k = \beta_k = 1$ and $\rho_k = 0$ for all k. It is easy to see that with this choice of parameters $y^k = z^k = (I + M^\top)x^k$, $F(y^k) = (I + M)x^k$, and $x^k - y^k = -M^\top x^k$. Furthermore,

$$\langle F(y^k), x^k - y^k\rangle = \|x^k\|^2 , \quad \|F(y^k)\|^2 = 2\|x^k\|^2 ,$$

where we have used the facts that $M = -M^\top$, and $\|(I + M)x^k\| = \|x^k\| = \|Mx^k\|$. Hence,

$$x^{k+1} = x^k - \frac{\langle F(y^k), x^k - y^k\rangle}{\|F(y^k)\|^2} F(y^k) = \frac{1}{2}(I - M)x^k .$$

Therefore,

$$\|x^{k+1}\|^2 = \frac{1}{2}\|x^k\|^2 ,$$

which means that now $\{x^k\}$ converges to $\bar{x} = 0$. In fact, the rate of convergence is linear. The question of convergence rate will be further discussed in Section 4.

3.4 A SPLITTING ALGORITHM

Suppose now that the function F defining $\mathrm{VIP}(F, C)$ is monotone and has the structure

$$F(x) = F^1(x) + F^2(x)\,,$$

where the part $-F^2$ is strongly monotone. Note that this implies that F^1 is also strongly monotone (and we can take the same modulus). Suppose further that $\mathrm{VIP}(F_k, C)$ with

$$F_k(x) = F^1(x) + F^2(x^k)\,,$$

where x^k is fixed, and $-F^2$ is strongly monotone with modulus $\mu_k > 0$, is in some sense easier to solve than the original problem. In this setting the splitting method often proves to be useful. This method consists of solving a sequence of subproblems $\mathrm{VIP}(F_k, C)$ with the structure defined above. In our framework, this corresponds to taking

$$A_k(x, z) = F^1(z) - F^1(x^k)\,.$$

Obviously, this choice satisfies the assumptions of Section 2. Hence, we have a convergent splitting-type algorithm, with the error tolerance criterion given by

$$\|e^k\| \leq \mu_k \rho_k \|z^k - x^k\|\,,$$

where the relaxation quantity $\mu_k \rho_k$ can be kept bounded away from zero. This criterion compares favorably to conditions often used in this setting, which are of the type

$$\sum_{k=0}^{\infty} \|e^k\|^2 < \infty\,.$$

Issues related to constructive approximation rules are discussed in more detail in [26, 24] (the proximal point techniques studied there are in this sense quite close to the splitting methods).

Finally, we note that no linesearch is necessary when the splitting method is used within our Algorithm 14.2. Indeed,

$$\begin{aligned}
\langle F(z^k), x^k - z^k \rangle &= \langle F^1(z^k) + F^2(z^k), x^k - z^k \rangle \\
&= \langle F^1(z^k) + F^2(x^k) + F^2(z^k) - F^2(x^k), x^k - z^k \rangle \\
&\geq \langle F^2(z^k) - F^2(x^k), x^k - z^k \rangle - \langle e^k, x^k - z^k \rangle \\
&\geq \mu_k (1 - \rho_k) \|x^k - z^k\|^2\,,
\end{aligned}$$

where strong monotonicity of $-F^2$ was used. This shows that the unit stepsize always satisfies the linesearch rule (14.7).

Finally, note that we could easily consider dynamic splitting, where the choice of F^1 and F^2 depends on the iteration index k.

4 SPEEDING UP CONVERGENCE

In this section, we briefly outline a technique that can be applied to speed up convergence of Algorithm 14.2 (here, F has to be monotone). It is based on a hybrid inexact proximal point method of [26]. In essence, this approach consists of checking whether z^k, solution of subproblem VIP(F_k, C), provides an acceptable (in the sense of [26]) approximate solution to the proximal point subproblem defined at x^k:

$$0 \in (F + N_C)(z) + \mu_k(z - x^k)\,, \tag{14.18}$$

where N_C stands for the normal cone to the set C. If this is the case, a hybrid proximal step is performed. This proximal step certainly preserves global convergence, but it is also (locally) a linear or superlinear step under standard assumptions. More specifically, Algorithm 14.2 can be modified the following way.

For simplicity, we shall assume here that z^k is the exact solution of VIP(F_k, C), i.e., $e^k = 0$. We then have that $0 \in (F_k + N_C)(z^k)$, and hence

$$v^k = F(z^k) - F_k(z^k) \in (F + N_C)(z^k)\,.$$

According to [26], a pair $z^k, v^k \in (F + N_C)(z^k)$ is an acceptable approximate solution of the proximal point subproblem (14.18), if

$$\|v^k + \mu_k(z^k - x^k)\| \le \theta \max\{\|v^k\|, \mu_k\|z^k - x^k\|\}\,, \quad \theta \in [0,1)\,. \tag{14.19}$$

Note that all information needed for checking this condition is readily available after solving VIP(F_k, C), and so no additional work is required. If this condition is satisfied, then the next iterate is obtained by the same projection step as in Algorithm 14.2, but with $F(y^k)$ replaced by v^k (and also setting $y^k = z^k$). Now, if one can show that (14.19) is always satisfied in a neighbourhood of a solution with certain properties, then the resulting iterates can be viewed as iterates generated by the method of [26]. This method has the same rate of convergence properties as the classical exact/inexact proximal point method [19]. In particular, if $(F + N_C)^{-1}$ is Lipschitz continuous at zero then convergence is linear. If, in addition, $\mu_k \to 0$ then it is superlinear. We next show that this indeed happens for some methods considered in Section 3.

For the Josephy-Newton method,

$$\begin{aligned}\varepsilon^k &= v^k + \mu_k(z^k - x^k)\\ &= F(z^k) - F(x^k) - A_k(x^k, z^k) + \mu_k(z^k - x^k)\\ &= F(z^k) - F(x^k) - F'(x^k)(z^k - x^k)\,.\end{aligned}$$

Assuming F' is locally Lipschitz continuous with constant $L > 0$, for k large enough we have that

$$\|\varepsilon^k\| \le L\|z^k - x^k\|^2 \le L\|z^k - x^k\|^{1+t'}$$

for any $t' \in (t, 1)$. Since in Algorithm 14.2 we have $\mu_k \ge r(x^k)^t$,

$$\theta \max\{\|v^k\|\,, \mu_k\|z^k - x^k\|\} \ge \theta r(x^k)^t \|z^k - x^k\|\,,$$

and to verify (14.19) it remains to show that

$$\|z^k - x^k\| = O(r(x^k))\,. \tag{14.20}$$

We already know that the sequence converges to some $\bar{x}$, a solution of VIP(F, C). If this solution satisfies conditions for the superlinear convergence of the (pure) Josephy-Newton method, it is easy to see that

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

In this setting also

$$\|x^k - \bar{x}\| = O(r(x^k))$$

(the projection residual provides a local error bound on the distance to the solution), and so (14.20) holds. This essentially implies that our Newton method (modified this way) converges superlinearly if we choose $\mu_k \to 0$. We refer the reader to [25, 21], where the case of systems of equations ($C = \boldsymbol{R}^n$) and nonlinear complementarity problems ($C = \boldsymbol{R}^n_+$) are studied in detail, including the effect of inexact subproblems solution.

Finally, consider the sequential minimization algorithm given by (14.17). In this case,

$$\begin{aligned}\varepsilon^k &= v^k + \mu_k(z^k - x^k)\\ &= -F'(z^k)^\top (z^k - x^k)\,.\end{aligned}$$

It is easy to see that

$$\|\varepsilon^k\| \le \|F'(z^k)^\top\|\|z^k - x^k\| \le \theta\mu_k\|z^k - x^k\|\,,$$

if $\mu_k \ge \|F'(z^k)^\top\|/\theta$. Since $\{z^k\}$ is bounded, μ_k can be kept bounded above, and the rate of convergence is linear (under the standard assumption that $(F + N_C)^{-1}$ is Lipschitz continuous at zero). Note that it does not seem possible to drive the regularization parameter to zero in this case, and so superlinear convergence cannot be guaranteed.

5 CONCLUDING REMARKS

We presented a class of algorithms for solving variational inequalities under the mild assumption that the problem mapping is continuous and pseudomonotone with respect to at least one solution. The algorithms are based on generating a certain hyperplane separating the current iterate from the solution set. Different ways of constructing this hyperplane give rise to different methods within our general framework. A projection method, a Josephy-Newton method, a certain minimization-based method, and a splitting technique, were considered. Obtaining (super)linear convergence rate was also discussed.

Acknowledgments

I wish to thank Patrice Marcotte whose question after my talk at ICCP99 led me to consider the condition of pseudomonotonicity with respect to only one solution point. I also thank the two anonymous referees for their helpful comments.

References

[1] A. A. Auslender. *Optimisation Méthodes Numériques.* Masson, Paris, 1976.

[2] J. F. Bonnans. Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Applied Mathematics and Optimization*, 29:161–186, 1994.

[3] R.W. Cottle, F. Giannessi, and J.-L. Lions. *Variational Inequalities and Complementarity Problems: Theory and Applications.* Wiley, New York, 1980.

[4] J.-P. Crouzeix. Characterizations of generalized convexity and generalized monotonicity, A survey. In J.-P. Crouzeix et al., editor, *Generalized convexity, generalized monotonicity: Recent results*, pages 237–256. Kluwer Academic Publishers, 1998.

[5] M. C. Ferris and J.-S. Pang (editors). *Complementarity and variational problems: State of the Art.* SIAM Publications, 1997.

[6] M. C. Ferris and J.-S. Pang. Engineering and economic applications of complementarity problems. *SIAM Review*, 39:669–713, 1997.

[7] M.C. Ferris and O.L. Mangasarian. Error bounds and strong upper semicontinuity for monotone affine variational inequalities. *Annals of Operations Research*, 47:293–305, 1993.

[8] R. Glowinski, J.-L. Lions, and R. Trémolières. *Numerical Analysis of Variational Inequalities.* North-Holland, Amsterdam, 1981.

[9] P.T. Harker and J.-S. Pang. Finite-dimensional variational inequality problems : A survey of theory, algorithms and applications. *Mathematical Programming*, 48:161–220, 1990.

[10] A.N. Iusem and B.F. Svaiter. A variant of Korpelevich's method for variational inequalities with a new search strategy. *Optimization*, 42:309–321, 1997.

[11] N.H. Josephy. Newton's method for generalized equations. Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin, 1979.

[12] S. Karamardian. Complementarity problems over cones with monotone and pseudomonotone maps. *Journal of Optimization Theory and Applications*, 18:445–455, 1976.

[13] R. De Leone, O.L. Mangasarian, and T.-H. Shiau. Multi-sweep asynchronous parallel successive overrelaxation for the nonsymmetric linear complementarity problem. *Annals of Operations Research*, 22:43–54, 1990.

[14] W. Li. Remarks on matrix splitting algorithms for symmetric linear complementarity problems. *SIAM Journal on Optimization*, 3:155–163, 1993.

[15] Z.-Q. Luo and P. Tseng. Error bound and convergence analysis of matrix splitting algorithms for the affine variational inequality problem. *SIAM Journal on Optimization*, 2:43–54, 1992.

[16] O.L. Mangasarian. Convergence of iterates of an inexact matrix splitting algorithm for the symmetric monotone linear complementarity problem. *SIAM Journal on Optimization*, 1:114–122, 1991.

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

[18] S. M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research*, 5:43–62, 1980.

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

[20] S. Schaible, S. Karamardian, and J.-P. Crouzeix. Characterizations of generalized monotone maps. *Journal of Optimization Theory and Applications*, 76:399–413, 1993.

[21] M. V. Solodov and B. F. Svaiter. A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem, 1998. *SIAM Journal on Optimization*, to appear.

[22] 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.

[23] M. V. Solodov and P. Tseng. Modified projection-type methods for monotone variational inequalities. *SIAM Journal on Control and Optimization*, 34:1814–1830, 1996.

[24] M.V. Solodov and B.F. Svaiter. A hybrid approximate extragradient–proximal point algorithm using the enlargement of a maximal monotone operator, 1998. *Set-Valued Analysis*, to appear.

[25] 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.

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