Math. Program., Ser. A 96: 513–528 (2003)

Digital Object Identifier (DOI) 10.1007/s10107-002-0369-z

M.V. Solodov

Convergence rate analysis of iteractive algorithms for solving variational inequality problems*

Received: April 17, 2001 / Accepted: December 10, 2002
Published online: April 10, 2003 –

Abstract. We present a unified convergence rate analysis of iterative methods for solving the variational inequality problem. Our results are based on certain error bounds; they subsume and extend the linear and sublinear rates of convergence established in several previous studies. We also derive a new error bound for γ-strictly monotone variational inequalities. The class of algorithms covered by our analysis in fairly broad. It includes some classical methods for variational inequalities, e.g., the extragradient, matrix splitting, and proximal point methods. For these methods, our analysis gives estimates not only for linear convergence (which had been studied extensively), but also sublinear, depending on the properties of the solution. In addition, our framework includes a number of algorithms to which previous studies are not applicable, such as the infeasible projection methods, a separation-projection method, (inexact) hybrid proximal point methods, and some splitting techniques. Finally, our analysis covers certain feasible descent methods of optimization, for which similar convergence rate estimates have been recently obtained by Luo [14].

Key words. Variational inequality – error bound – rate of convergence

1. Introduction

Given a function $F : \Re^n \to \Re^n$ and a set $C \subset \Re^n$, we consider the variational inequality problem [3, 8], VIP(F, C) for short, which 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 $\Re^n$. We assume that C is closed and convex, F is continuous, and the solution set of VIP(F, C), denoted by S, is nonempty. As is well known, this problem subsumes nonlinear equations, optimization problems (over convex feasible sets), and the nonlinear complementarity problems, among others.

Among numerous algorithms proposed for solving VIP(F, C) and its special cases, we mention projection-type methods, linearization and Newton-type methods, proximal point and splitting algorithms, and techniques based on merit functions. We refer the reader to the survey [8], and to articles and references in the more recent collections [4, 5]. Many methods for VIP(F, C) can be expressed in the following form:

$$x^0 \in D\,, \quad x^{k+1} = T(x^k) = T_2(x^k; T_1(x^k))\,, \quad k = 0, 1, 2, \ldots, \tag{1}$$

M.V. Solodov: Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil. e-mail: solodov@impa.br

Mathematics Subject Classification (2000): 90C30, 90C33, 65K05

* Research of the author is partially supported by CNPq Grant 200734/95–6, by PRONEX-Optimization, and by FAPERJ.

where D is some closed set containing C (typically, $D = C$ or $D = \Re^n$), and $T : D \to D$ is the mapping which defines each specific algorithm. Many methods for solving VIP(F, C) are based on computing a certain intermediate point, information at which is then used to obtain the next iterate, e.g., the extragradient method [12, 11, 20, 9], infeasible projection methods [33], separation-projection methods [30], etc. (See also Section 4). This makes it convenient to split the iteration (i.e., the mapping T) in two parts, as represented in (1) by the mappings $T_1 : D \to D$ and $T_2 : D \times D \to D$. This two-step framework makes it more convenient to handle a number of iterative methods considered below. (Actually, in some cases it is not clear if using a one-step scheme is at all possible for the purposes of this paper.) In any case, one-step methods can be easily recovered setting $T_2(x; y) = y$.

The purpose of this paper is to identify conditions on the data of VIP(F, C) and on the properties of the algorithm, which are sufficient for convergence of the sequence $\{x^k\}$ to the solution set S, allow one to estimate in some way the rate of convergence, and hold for a wide class of problems and algorithms. To this end, we introduce a continuous function (often called Lyapunov function)

$$f : D \to \Re_+ \text{ such that } f(x) = 0 \Leftrightarrow x \in S\,.$$

For feasible descent methods of minimization, f is usually the difference between the objective function and the optimal value of the problem [18, 14]. In the setting of general variational inequalities, f is typically the square of an appropriate distance-like function to the solution set S. Define the norm of the natural residual for VIP(F, C) as

$$R(x) = \|x - P_C[x - F(x)]\|\,, \quad x \in D\,, \tag{2}$$

where $P_C[\cdot]$ stands for the orthogonal projection operator onto C. As is well known, $R(x) = 0$ if, and only if, $x \in S$. In our algorithmic framework, we consider the following three conditions.

$\forall\, M > 0\ \exists\, c_1 > 0$ such that $\forall\, x \in \{y \in D \mid \|y\| \le M\}$ it holds that

$$f(x) - f(T(x)) \ge c_1 \|x - T_1(x)\|^\alpha\,,\ \alpha \ge 2\,. \tag{3}$$

It holds that

$$\forall\, x \in D, \quad \|x - T_1(x)\| \ge c_2 \min\{R(x), R(T(x))\}\,,\ c_2 > 0\,. \tag{4}$$

For some set Ω satisfying $\Omega \supset \{y \in D \mid R(y) \le \delta\}$, $\delta > 0$, it holds that

$$\forall\, x \in \Omega\,, \quad R(x)^{2/\beta} \ge c_3 f(x)\,,\ \beta \ge 1\,, c_3 > 0\,. \tag{5}$$

Conditions (3) and (4) above are related to the structure of the algorithm, while (5) is an error bound type condition (at least when f is some distance to the solution set S). In Section 2, assuming that the sequence $\{x^k\}$ is generated by (1) and conditions (3)-(5) hold, we give explicit estimates of the rate of convergence of $\{f(x^k)\}$ to zero. This rate is at least linear if $\beta = 1$ and $\alpha = 2$. It is sublinear in the case where $\beta > 1$ and/or $\alpha > 2$.

Our analysis is in the spirit of convergence rate results in [18, 35, 14], which are also based on error bounds. Reference [18] presents a unified analysis for linear convergence of feasible descent methods for optimization (see also [17, 15, 19, 13] for similar studies of some specific methods). Those results were subsequently extended in two different ways. In [35], the general variational inequality was considered, but the convergence results also refer to the linear rate only. In [14], the rate of convergence can be linear or sublinear, but the class of methods is the same as in [18], i.e., for optimization only. The present paper can be regarded as an extension of [35] to include possibly sublinear convergence and a larger class of algorithms for VIP(F, C), and as an extension of [14] from optimization to general variational inequalities. Compared to [35], in addition to allowing the sublinear rate of convergence, our conditions (3)–(5) also include a larger class of algorithms. First, note that (3)–(5) allow one to treat infeasible methods (i.e., methods whose iterates need not belong to the feasible set C), thus covering the modified projection-type methods of [33], for example. Additionally, the possibility of $\alpha > 2$ in (3) allows one to include algorithms for which the linear rate of convergence is not currently known even if a Lipschitzian error bound holds. The latter is the case, for example, for the separation-projection method of [30]. There are two more differences between (3) and analogous conditions in [18, 35, 14]. First, in (3) appears an intermediate point $T_1(x)$ rather than the more usual $T(x)$. As was already commented above, in the context of variational inequalities this setting appears more natural, at least for some algorithms. Furthermore, in some cases, e.g., the inexact (hybrid) proximal point schemes [31], it is not clear whether (3) holds with T_1 replaced by T (See Section 4.) The second distinctive feature of (3) is that the constant c_1 in general depends on the bounded set in consideration. The reason why this modification is useful is that for some methods the inequality in (3) does not hold uniformly for all $x \in D$. Nevertheless, typically the quantity represented by c_1 is bounded on bounded sets. For (pseudo)monotone problems, boundedness of the iterates can often be established easily and independently of the convergence rate analysis.

The rest of this paper is organized as follows. In Section 2 we present our general rate of convergence estimate. Section 3 contains a discussion of error bounds used in the convergence rate analysis, and a new error bound result for γ-strictly monotone variational inequalities. Section 4 is devoted to showing how our general results apply to a number of specific algorithms for solving variational inequalities.

2. Convergence rate estimate

We start with a general rate of convergence estimate which will be applied to specific algorithms in Section 4. Our result is related to [14, Theorem 4], but our estimate is somewhat different, and the sequence $\{f(x^k)\}$ resulting from (3)–(5) is more general than the sequence generated in [14] (in particular, the key condition (8) below is weaker than condition (10) in [14]). This relaxation is due to taking the minimum of two quantities in (4), and it is important for the analysis to be applicable to the classical proximal point method; see [35] and Section 4.

Theorem 1. *Suppose that $F : \Re^n \to \Re^n$ is continuous, $C \subset \Re^n$ is closed and convex, and the solution set S of VIP(F, C) is nonempty.*

Let $\{x^k\}$ *be any sequence generated according to* (1)*, and suppose that* f*,* T *and* R *defined in Section 1 satisfy conditions* (3)–(5)*. Suppose further that either* c_1 *in* (3) *does not depend on* M*, or that* $\{x^k\}$ *is bounded. Then the sequence* $\{f(x^k)\}$ *converges to zero.*

Furthermore, if $\alpha = 2$ *and* $\beta = 1$*, then* $\{f(x^k)\} \to 0$ *at a linear rate. Specifically, there exists an iteration index* k_0 *such that for all* $k \geq k_0$ *it holds that*

$$f(x^{k+1}) \leq (1 + c_1 c_2^{\alpha} c_3^{\alpha\beta/2})^{-1} f(x^k) \,. \tag{6}$$

If $\alpha > 2$ *and/or* $\beta > 1$*, then for any* $\tau \in (0, 1)$ *there exists some index* k_0 *such that for all* $k \geq k_0$ *it holds that*

$$f(x^k) \leq f(x^{k_0}) \, (1 + (k - k_0)\Delta)^{1/(1-\alpha\beta/2)} \,, \qquad \Delta = f(x^{k_0})^{-1+\alpha\beta/2} \tau^{\alpha\beta/2} (1+\tau)^{-1} (\alpha\beta/2 - 1) c_1 c_2^{\alpha} c_3^{\alpha\beta/2} \,. \tag{7}$$

Proof. Either by the assumption that the inequality in (3) holds with the same c_1 for all $x \in D$, or by the assumption of boundedness of $\{x^k\}$, from (1),(3) we have that $f(x^k) - f(x^{k+1}) \geq c_1 \|x^k - T_1(x^k)\|^{\alpha}$ for all $k \geq 0$, where $c_1 > 0$ is fixed from now on. Hence, the sequence $\{f(x^k)\}$ is monotonically non-increasing. Since $f(x^k) \geq 0$ for all k, it follows that $\{f(x^k)\}$ converges (to some $a \geq 0$). From the relation above, it then follows that $\|x^k - T_1(x^k)\| \to 0$. Furthermore, relation (4) implies that $\min\{R(x^k), R(x^{k+1})\} \to 0$. The latter fact clearly means that $\liminf_k R(x^k) > 0$ is not possible. Let $\{k_j\}$ be an infinite subsequence such that $R(x^{k_j}) \to 0$. Since $F(\cdot)$ is continuous and C closed and convex, $R(\cdot)$ is continuous. The continuity of $R(\cdot)$, definition of Ω in (5), and the fact that $R(x^{k_j}) \to 0$, imply that $x^{k_j} \in \Omega$ for all j large enough, say $j \geq j_0$. Hence, by (5), we conclude that $f(x^{k_j}) \to 0$. Since we already established that $\{f(x^k)\}$ converges, it must be the case that it converges to zero.

Combining (3)–(5), for each k sufficiently large, say $k \geq k_0$, we obtain that

$$f(x^k) - f(x^{k+1}) \geq c_4 f(x^{k+1})^{\gamma} \quad \text{or/and} \quad f(x^k) - f(x^{k+1}) \geq c_4 f(x^k)^{\gamma} \,, \tag{8}$$

where $c_4 = c_1 c_2^{\alpha} c_3^{\alpha\beta/2} > 0$ and $\gamma = \alpha\beta/2 \geq 1$.

Consider first the case where $\alpha = 2$, $\beta = 1$. Since $f(x^k) \geq f(x^{k+1})$, in both cases in (8) the first inequality is satisfied, and it immediately implies (6).

We next consider the case of $\alpha > 2$ and/or $\beta > 1$, in which case $\gamma > 1$. Fix any $\tau \in (0, 1)$. Suppose that the first inequality in (8) holds, but the second does not, that is,

$$f(x^k) - f(x^{k+1}) < c_4 f(x^k)^{\gamma} \,.$$

Then we have that

$$\begin{aligned} f(x^{k+1}) &> f(x^k)(1 - c_4 f(x^k)^{\gamma-1}) \\ &\geq \tau f(x^k) \,, \end{aligned}$$

where the second inequality holds for all $k \geq k_0$ (increasing the index k_0, if necessary), because $f(x^k) \to 0$ and $\gamma > 1$, while $\tau < 1$. Using the latter relation with the first one in (8), we conclude that $f(x^k) - f(x^{k+1}) \geq \tau^{\gamma} c_4 f(x^k)^{\gamma}$, and hence the two cases in (8) can be combined into $f(x^k) - f(x^{k+1}) \geq c_5 f(x^k)^{\gamma}$, where $c_5 = \tau^{\gamma} c_4 (< c_4)$. Or, equivalently,

$$f(x^{k+1}) \leq f(x^k)(1 - c_5 f(x^k)^{\gamma-1}) \,. \tag{9}$$

We further obtain

$$
\begin{aligned}
f(x^{k+1})^{1-\gamma} - f(x^k)^{1-\gamma} &= \frac{f(x^k)^{\gamma-1} - f(x^{k+1})^{\gamma-1}}{(f(x^k)f(x^{k+1}))^{\gamma-1}} \\
&\geq \frac{f(x^k)^{\gamma-1}(1-(1-c_5 f(x^k)^{\gamma-1})^{\gamma-1})}{(f(x^k)f(x^{k+1}))^{\gamma-1}} \\
&\geq \frac{1-(1-c_5 f(x^k)^{\gamma-1})^{\gamma-1}}{f(x^k)^{\gamma-1}} \\
&\geq \frac{(\gamma-1)c_5 f(x^k)^{\gamma-1}}{f(x^k)^{\gamma-1}(1+(\gamma-1)c_5 f(x^k)^{\gamma-1})} \\
&\geq c_5(\gamma-1)(1+\tau)^{-1},
\end{aligned}
$$

where the first inequality is by (9); the second inequality follows from $f(x^k) \geq f(x^{k+1})$; the third inequality is implied by the fact that $\forall\, y \in \Re_+$ sufficiently small $1-(1-y)^t \geq ty/(1+ty)$ if $t > 0$; and the last inequality is by $f(x^k) \to 0$ and $\gamma > 1$, taking also into account that $\tau \in (0, 1)$.

Using the above relation consecutively, we have that

$$
\begin{aligned}
f(x^k)^{1-\gamma} - f(x^{k_0})^{1-\gamma} &= \sum_{i=k_0}^{k-1} \left(f(x^{i+1})^{1-\gamma} - f(x^i)^{1-\gamma} \right) \\
&\geq \sum_{i=k_0}^{k-1} c_5(\gamma-1)(1+\tau)^{-1} \\
&= (k-k_0)c_5(\gamma-1)(1+\tau)^{-1}.
\end{aligned}
$$

Hence,

$$
f(x^k)^{1-\gamma} \geq f(x^{k_0})^{1-\gamma}(1+(k-k_0)c_5 f(x^{k_0})^{\gamma-1}(\gamma-1)(1+\tau)^{-1}),
$$

and therefore,

$$
f(x^k) \leq f(x^{k_0})(1+(k-k_0)f(x^{k_0})^{\gamma-1}c_5(\gamma-1)(1+\tau)^{-1})^{1/(1-\gamma)}.
$$

Now recalling the definitions of c_5 and γ gives (7). □

3. Error bounds

When $f(x) = \text{dist}(x, S)^2$, the square of the distance to the solution set of VIP(F, C), then condition (5) is a (local) error bound property:

$$
R(x)^{1/\beta} \geq c_3^{1/2}\text{dist}(x, S) \quad \forall\, x \in \Omega, \tag{10}
$$

where Ω is some set containing S and $\beta \geq 1$, as specified in (5). We refer the reader to [22] for a survey of error bounds and their applications. Here, we first list some of the known conditions which imply the desired error bound, and then derive a new condition.

(a) If F is affine and C is polyhedral, then (10) holds with $\beta = 1$, where Ω is some neighborhood of S. See [24, 16].
(b) If F is strongly monotone and Lipschitz-continuous, then (10) holds with $\beta = 1$ and $\Omega = \Re^n$. See [21].
(c) If C is polyhedral, $F(\cdot) = \nabla\varphi(\cdot)$, and $\varphi(\cdot)$ is γ-strictly convex, i.e., for some $\gamma \geq 1$ and $\mu > 0$

$$\langle \nabla\varphi(x) - \nabla\varphi(y), x - y\rangle \geq \mu\|x - y\|^{1+\gamma} \ \forall x, y \in \Re^n ,$$

then (10) holds with $\beta = \gamma$ and $\Omega = C$. See [14].
We note that in [14] this result is stated for $\Omega = \Re^n$, but the proof of [14, Theorem 3] in fact appears to require that the point under consideration be feasible, i.e., $\Omega = C$. On the other hand, the proof can be easily extended to the "asymmetric" case where F is not necessarily a gradient.
(d) If C is polyhedral and F is "monotone composite", (i.e., $F(x) = E^\top G(Ex) + q$, where E is a matrix of appropriate dimensions with no zero column, $q \in \Re^n$, and G is strongly monotone and Lipschitz-continuous), then (10) holds locally with $\beta = 1$. See [17].
(e) If $C = \{x \mid g(x) \leq 0\}$, g is convex, C satisfies some constraint qualification, and F and g are semianalytic functions, then (10) holds locally with some (unknown) β. See [14, Theorem 2], where this result is established for $F(\cdot) = \nabla\varphi(\cdot)$. The extension to the case of general F is immediate.

We proceed to prove a new error bound (see Theorem 2), and compare it to items (b) and (c) above. Condition (12) below can be thought of as (local) γ-strict monotonicity with respect to the solution set S, and condition (11) as (local) Hölder-continuity of F with respect to S. Note that (12) and (11) together imply that Ω is a sufficiently small "neighborhood" of S, and necessarily $\gamma \geq \nu$. When C is a box, then condition (12) can be replaced by the yet weaker condition that F is a γ-uniform P-function with respect to S, see (13).

Compared to item (b) above, strong monotonicity and Lipschitz-continuity of F are replaced by weaker assumptions. Of course, instead of a global error bound in (b) we obtain a local one. Compared to item (c) above, our result does not require C to be polyhedral or x to be feasible (and F to be a gradient). On the other hand, we assume Hölder-continuity of F (with respect to S), and our bound is local.

Theorem 2. *Suppose that the solution set S of VIP(F, C) is nonempty, and the set Ω, containing S, is such that*

$$\forall x \in \Omega,\ \|F(x) - F(P_S[x])\| \leq L\|x - P_S[x]\|^\nu ,\ \nu \in (0, 1] ,\ L > 0 , \qquad (11)$$

and either

$$\forall x \in \Omega,\ \langle F(x) - F(P_S[x]), x - P_S[x]\rangle \geq \mu\|x - P_S[x]\|^{1+\gamma} ,\ \gamma \geq 1 ,\ \mu > 0 , \qquad (12)$$

or C is a box (i.e, $C = \prod_{i=1}^n [l_i, u_i]$ for some $-\infty \leq l_i \leq u_i \leq +\infty$), and

$$\forall x \in \Omega,\ \max_{i\in\{1,\ldots,n\}} (F_i(x) - F_i(P_S[x])(x_i - (P_S[x])_i) \geq \mu\|x - P_S[x]\|^{1+\gamma} . \qquad (13)$$

Then the error bound (10) *holds with R given by* (2) *and $\beta = 1 + \gamma - \nu$, $c_3 = \mu^2(1 + L)^{2/(\nu-1-\gamma)}$.*

Proof. Since $S \subset \Omega$ and the error bound in question is local, we can assume that Ω is such that for all $x \in \Omega$ we have $\|x - P_S[x]\| \le 1$. Take any $x \in \Omega$, and denote $p = P_C[x - F(x)]$. By properties of the projection operator, $\langle x - F(x) - p, p - y\rangle \ge 0$ for any $y \in C$. Choosing $y = P_S[x]$ in this inequality and adding it to $\langle F(P_S[x]), p - P_S[x]\rangle \ge 0$ (which holds because $P_S[x] \in S$ and $p \in C$), we have that

$$\langle x - p + F(P_S[x]) - F(x), p - P_S[x]\rangle \ge 0\,. \tag{14}$$

We further obtain that

$$\begin{aligned}
\mu\|x - P_S[x]\|^{1+\gamma} &\le \langle F(P_S[x]) - F(x), P_S[x] - x\rangle \\
&\le \langle x - p, p - P_S[x]\rangle + \langle F(P_S[x]) - F(x), p - x\rangle \\
&\le -\|x - p\|^2 + \|x - p\|\|x - P_S[x]\| \\
&\quad + \|F(P_S[x]) - F(x)\|\|x - p\| \\
&\le \|x - p\|\left(\|x - P_S[x]\| + L\|x - P_S[x]\|^\nu\right) \\
&\le (1 + L)\|x - p\|\|x - P_S[x]\|^\nu
\end{aligned}$$

where the first inequality is by (12); the second inequality is by (14); the fourth follows from (11); and the last follows from $\|x - P_S[x]\| \le 1$ and $\nu \in (0, 1]$. Since $\|x - p\| = R(x)$, we conclude that

$$R(x)^{1/(1+\gamma-\nu)} \ge \mu(1 + L)^{1/(\nu-1-\gamma)}\mathrm{dist}(x, S)\,,$$

which establishes the claim under the assumptions (11) and (12).

The proof under the assumptions (11) and (13) is similar. As is well known and easy to see, C being a box implies that $P_C[\cdot]$ is separable. In particular, for all $i \in \{1, \ldots, n\}$ it holds that $(x_i - F_i(x) - p_i)(p_i - (P_S[x])_i) \ge 0$ and $F_i(P_S[x])(p_i - (P_S[x])_i) \ge 0$, so that

$$(x_i - p_i + F_i(P_S[x]) - F_i(x))(p_i - (P_S[x])_i) \ge 0\,.$$

Taking $j \in \{1, \ldots, n\}$ which realizes the maximum in (13), similarly to the analysis above we obtain

$$\begin{aligned}
\mu\|x - P_S[x]\|^{1+\gamma} &\le (F_j(P_S[x]) - F_j(x))((P_S[x])_j - x_j) \\
&\le |x_j - p_j||x_j - (P_S[x])_j| + |F_j(P_S[x]) - F_j(x)||x_j - p_j| \\
&\le (1 + L)\|x - p\|\|x - P_S[x]\|^\nu\,,
\end{aligned}$$

where we also use the monotonicity of the norm. □

4. Applications to specific algorithms

In this section, we outline applications of our results to some algorithms that have been previously proposed for solving variational inequalities. After some initial comments on the methods considered in [35, 14], to which our results also apply, we shall turn our attention to methods not discussed in those references.

4.1. Feasible descent methods for optimization

First, note that our framework covers the feasible descent methods studied in [14], by setting

$$F(x) = \nabla\varphi(x)\,, \quad f(x) = \varphi(x) - \min_{z\in C}\varphi(z)\,,$$

$$T_1(x) = P_C[x - \eta(x)\nabla\varphi(x) + e(x)]\,, \quad T_2(x; y) = y\,,$$

where φ is the objective function of the problem, $\eta(x) > 0$ is the (typically, stepsize) parameter, and the mapping $e : C \to \Re^n$ defines each specific algorithm. In particular, this setting includes the gradient projection, (symmetric) matrix splitting, and coordinate descent methods, among others (see [18, 14]). It is not difficult to see that conditions (3),(4) are satisfied under appropriate assumptions on $e(\cdot)$ and $\eta(\cdot)$. Indeed, suppose that

$$0 < \bar{\eta} \le \eta(x) \le \hat{\eta} \quad \text{and} \quad \|e(x)\| \le \tau\|x - T_1(x)\|\,, \tau \in [0, 1), \tag{15}$$

conditions which are known to be satisfied by the feasible descent methods mentioned above. For $x \in D = C$, by properties of the projection operator,

$$\langle x - \eta(x)\nabla\varphi(x) + e(x) - T_1(x), x - T_1(x)\rangle \le 0\,.$$

Hence,

$$\begin{aligned} -\eta(x)\langle\nabla\varphi(x), T_1(x) - x\rangle &\ge \|x - T_1(x)\|^2 - \|e(x)\|\|x - T_1(x)\| \\ &\ge (1 - \tau)\|x - T_1(x)\|^2\,, \end{aligned}$$

where the second inequality is by (15). Assuming that $\nabla\varphi(\cdot)$ is Lipschitz-continuous (with modulus $L > 0$), we have that

$$\begin{aligned} f(x) - f(T(x)) &\ge -\langle\nabla\varphi(x), T_1(x) - x\rangle - \frac{L}{2}\|x - T_1(x)\|^2 \\ &\ge \left(\frac{1-\tau}{\hat{\eta}} - \frac{L}{2}\right)\|x - T_1(x)\|^2\,. \end{aligned}$$

In particular, inequality in (3) holds uniformly for all $x \in D = C$ with $\alpha = 2$ and $c_1 = (1-\tau)/\hat{\eta} - L/2$, provided $\hat{\eta} < 2(1-\tau)/L$. Furthermore, for R given by (2) with $F(x) = \nabla\varphi(x)$, we obtain

$$\begin{aligned} \|x - T_1(x)\| &= \|x - P_C[x - \eta(x)\nabla\varphi(x) + e(x)]\| \\ &\ge \min\{1, \bar{\eta}\}R(x) \\ &\quad - \|P_C[x - \eta(x)\nabla\varphi(x)] - P_C[x - \eta(x)\nabla\varphi(x) + e(x)]\| \\ &\ge \min\{1, \bar{\eta}\}R(x) - \|e(x)\| \\ &\ge \min\{1, \bar{\eta}\}R(x) - \tau\|x - T_1(x)\|\,, \end{aligned}$$

where the first inequality is by [7, Lemma 1]; the second is by the nonexpansiveness of the projection operator; and the last follows from (15). Using the above relation, we conclude that (4) holds with $c_2 = \min\{1, \bar{\eta}\}/(1 + \tau)$. Finally, if error bound (10) is satisfied and $\varphi(\cdot)$ is Lipschitz-continuous (with modulus $L > 0$), then for $x \in \Omega$

$$\begin{aligned} R(x)^{1/\beta} &\geq c_3 \mathrm{dist}(x, S) \\ &\geq L^{-1}(\varphi(x) - \varphi(P_S[x])) \\ &= L^{-1} f(x), \end{aligned}$$

which verifies (5).

4.2. *Classical extragradient, proximal point and matrix splitting methods*

Note further that algorithms for solving VIP (F, C) considered in [35] can also be cast in our framework. This is easy to see, as the conditions used in [35] are similar in form, but less general. In [35], T is regarded as a "one-step" mapping (i.e., $T_2(x; y) = y$), in (3) one always has $\alpha = 2$ and c_1 cannot depend on M, and finally the error bound is always Lipschitzian (i.e., $\beta = 1$). Therefore, our analysis yields the rate of convergence estimates for all methods mentioned in [35]: the extragradient method, the classical proximal point method, the (asymmetric) matrix splitting, and a certain feasible descent method. The difference is that our results establish conditions not only for the linear convergence rate, but also sublinear. Furthermore, as already mentioned in the Introduction and will be exhibited below, our framework applies to some algorithms to which [35] does not apply. In particular, [35] does not cover the following situations: when the sequence $\{x^k\}$ is infeasible; if in (3) we have $\alpha > 2$; if (3) does not hold uniformly on D; and when it is not clear whether (3) holds with T_1 replaced by T.

4.3. *Infeasible projection-type methods*

Let F be monotone, and consider Algorithm 3.1 of [33] (it had been shown to be typically more efficient computationally than the classical extragradient algorithm). In terms of (1), the method is given by

$$T_1(x) = P_C[x - \eta(x)F(x)],$$

$$T_2(x; y) = x - \frac{\theta(1-\sigma)\|x-y\|^2}{\|Q^{1/2}(x-y-\eta(x)F(x)+\eta(x)F(y))\|^2} Q(x-y-\eta(x)F(x)+\eta(x)F(y)),$$

where Q is any symmetric positive definite matrix, $\theta \in (0, 2)$, and $\eta(x) > 0$ is determined by linesearch to satisfy the condition

$$\eta(x)\langle F(x) - F(T_1(x)), x - T_1(x)\rangle \leq \sigma \|x - T_1(x)\|^2, \ \sigma \in (0, 1).$$

If F is Lipschitz-continuous (with modulus $L > 0$), then it is clear that $0 < \bar{\eta} \leq \eta(x) \leq \hat{\eta} \leq \sigma/L$. With

$$f(x) = \min_{\bar{x} \in S} \|x - \bar{x}\|^2_{Q^{-1}},$$

where $\|x\|^2_{Q^{-1}} = \langle Q^{-1}x, x\rangle$, the analysis in [33, Theorem 3.1] shows that

$$f(x) - f(T(x)) \geq \frac{\theta(2-\theta)(1-\sigma)^2\|x - T_1(x)\|^4}{\|Q^{1/2}(x - T_1(x) - \eta(x)F(x) + \eta(x)F(T_1(x)))\|^2}\,.$$

By the Lipschitz-continuity of F, it is easy to see that

$$\frac{\|x - T_1(x)\|^2}{\|Q^{1/2}(x - T_1(x) - \eta(x)F(x) + \eta(x)F(T_1(x)))\|} \geq \frac{\|x - T_1(x)\|}{\|Q^{1/2}\|(1+\hat{\eta}L)}\,.$$

Combining the last two relations, we verify that condition (3) holds with $\alpha = 2$ and $c_1 = \theta(2-\theta)(1-\sigma)^2(\|Q^{1/2}\|(1+\hat{\eta}L))^{-1}$. Here, c_1 does not depend on M, i.e., inequality in (3) holds uniformly for all $x \in D = \Re^n$. Furthermore, by [7, Lemma 1],

$$\|x - T_1(x)\| = \|x - P_C[x - \eta(x)F(x)]\| \geq \min\{1, \bar{\eta}\}R(x)\,,$$

and so condition (4) is satisfied.

Finally, by the equivalence of the norms in $\Re^n$, condition (5) holds whenever we have error bound (10).

Theorem 1 applied in this context subsumes the linear rate of convergence result in [33] for the case when (10) holds with $\beta = 1$, and establishes a new rate of convergence estimate for $\beta > 1$. Note that the analysis in [35] does not cover the method considered above even for $\beta = 1$, because the generated sequence is infeasible (the range of T_2 is $\Re^n$, and not C). Finally, we note that other infeasible projection-type algorithms discussed in [33] can be analyzed similarly to what has been done above.

4.4. A separation-projection method

Let F be pseudomonotone (with respect to the solution set S), and consider the method discussed in [30] (see also [10]). In this context, we choose

$$T_1(x) = \eta(x)P_C[x - F(x)] + (1-\eta(x))x\,,$$

$$T_2(x; y) = \arg\min_z\{\|z - x\| \mid \langle F(y), z - y\rangle \leq 0\,,\ z \in C\}\,,$$

where $\eta(x) \in (0, 1]$ is chosen (by linesearch) to satisfy

$$\eta(x)\langle F(T_1(x)), x - T_1(x)\rangle \geq \sigma\|x - T_1(x)\|^2\,,\ \sigma \in (0,1)\,.$$

As is well known and easy to check, $\langle F(x), x - P_C[x - F(x)]\rangle \geq \|x - P_C[x - F(x)]\|^2$. Hence, if F is Lipschitz-continuous (with modulus $L > 0$) then

$$\begin{aligned}\langle F(T_1(x)), x - T_1(x)\rangle &\geq \langle F(x), x - T_1(x)\rangle - \|F(T_1(x)) - F(x)\|\,\|x - T_1(x)\| \\ &\geq (\eta(x)^{-1} - L)\|x - T_1(x)\|^2\,.\end{aligned}$$

Therefore, the linesearch specified above would generate stepsize values satisfying $0 < \bar{\eta} \leq \eta(x) \leq \hat{\eta} \leq (1-\sigma)/L$. Choosing

$$f(x) = \mathrm{dist}(x, S)^2\,,$$

the analysis in [30, Theorem 2.1] shows that

$$f(x) - f(T(x)) \geq \frac{\sigma^2 \|x - T_1(x)\|^4}{\hat{\eta}^2 \|F(T_1(x))\|^2} .$$

If x belongs to some bounded set B, obviously the set $\{T_1(x), x \in B\}$ is bounded, and hence also $\{F(T_1(x)), x \in B\}$. It then follows from the inequality above that condition (3) holds with $\alpha = 4$ and appropriate c_1 which depends on M defining each bounded set.

Furthermore, condition (4) is also satisfied:

$$\|x - T_1(x)\| = \eta(x) R(x) \geq \bar{\eta} R(x) .$$

Now, whenever an error bound (10) holds, so does condition (5). Finally, the proof of [30, Theorem 2.1] shows that $\{x^k\}$ generated by the method is bounded, because for any fixed $\bar{x} \in S$ it holds that $\|x^k - \bar{x}\| \geq \|x^{k+1} - \bar{x}\|$. Therefore, Theorem 1 is applicable and gives a rate of convergence estimate for the algorithm. Note that no rate of convergence was given in [30]. Note also that the framework of [35] is not applicable here, because (3) does not hold uniformly, and also $\alpha = 4$ rather than $\alpha = 2$.

This algorithm does not have a proven linear rate of convergence. Nevertheless, it has some advantages over extragradient and related methods when the projection onto the feasible set is computationally expensive (note that the linesearch procedure here does not require any projections for computing or testing each trial point). Also, this algorithm appears useful for globalizing certain Newton-type algorithms for solving VIP(F, C) [27, 32, 26].

4.5. Hybrid inexact proximal point algorithms

Denoting by $N_C(\cdot)$ the normal operator for the feasible set C, the exact proximal point method [25] for solving VIP(F, C) can be written in the form of (1) by choosing

$$T_1(x) = (I + \eta(x)(F + N_C))^{-1}(x) , \quad T_2(x; y) = y ,$$

where $\eta(x) \geq \bar{\eta} > 0$ is the regularization parameter. As is well known [25], for

$$f(x) = \mathrm{dist}(x, S)^2$$

condition (3) holds with $\alpha = 2$ and $c_1 = 1$ for all $x \in D = C$. Also, as shown in [35],

$$\|x - T_1(x)\| \geq \min\{1, \eta(x)\} R(T(x)) ,$$

so that condition (4) is satisfied. The analysis in [35] establishes the linear rate of convergence of the *exact* proximal point method, provided error bound (10) holds with $\beta = 1$ (note that this is already weaker than the more usual condition of Lipschitz-continuity of $(F + N_C)^{-1}$ at zero [25], which implies uniqueness of the solution). In what follows, we extend those results to a class of inexact proximal-type algorithms, and beyond the case of linear convergence rate.

We consider here the inexact proximal-based algorithm of [31] (see also conceptually related methods in [29, 28]). To this end, let

$$F_x(y) = \eta(x)F(y) + y - x\,,$$

i.e., $F_x(\cdot)$ is the proximal regularization of $F(\cdot)$ with respect to x with parameter $\eta(x) > 0$. We take

$$T_1(x) = \left\{ y \in C \mid \Psi_x(y) \leq \frac{\sigma}{2}\|y - x\|^2 \right\}\,,\ \sigma \in [0, 1)\,,$$
$$T_2(x; y) = P_C[x - \eta(x)F(y)]\,,$$

where

$$\Psi_x(y) = \langle F_x(y), y - P_C[y - F_x(y)]\rangle - \frac{1}{2}\|y - P_C[y - F_x(y)]\|^2$$

is the regularized gap function [2, 6] for VIP(F_x, C). Note that formally, T_1 is not single-valued here. On one hand, the extension of conditions (3), (4) to the multi-valued case is straightforward. Alternatively, T_1 above can be considered single-valued (as a function of given x) if we are to specify explicitly how y with the given property is computed (for example, by applying a specific feasible descent method to the problem $\min_{y\in C} \Psi_x(y)$ with x as a starting point, and iterating until the tolerance prescribed in the definition of T_1 is achieved). One advantage of the approximation criterion represented by T_1, is that it is directly and constructively related to the subproblem VIP(F_x, C), see [27, 28, 32, 26] for some applications of hybrid proximal-based strategies.

Choosing

$$f(x) = \operatorname{dist}(x, S)^2\,,$$

by the analysis in [31, Theorem 3], it holds that

$$f(x) - f(T(x)) \geq (1 - \sigma^2)\|x - T_1(x)\|^2\,,$$

so that (3) holds with $\alpha = 2$ and $c_1 = 1 - \sigma^2$ for all $x \in D = C$. To verify (4), we argue as follows. It is well known that the regularized gap function provides an upper bound for the natural residual. Specifically for VIP(F_x, C), we have that

$$\Psi_x(y) \geq \frac{1}{2}\|y - P_C[y - F_x(y)]\|^2 \ \ \forall\, y \in C\,.$$

Using this relation and the definition of T_1, we have that

$$\begin{aligned}
\sigma^{1/2}\|x - T_1(x)\| &\geq \sqrt{2\Psi_x(T_1(x))} \\
&\geq \|T_1(x) - P_C[T_1(x) - F_x(T_1(x))]\| \\
&= \|T_1(x) - P_C[x - \eta(x)F(T_1(x))]\| \\
&\geq \|x - P_C[x - \eta(x)F(T_1(x))]\| - \|x - T_1(x)\|\,.
\end{aligned}$$

If F is Lipschitz-continuous (with modulus $L > 0$), we further obtain

$$\begin{aligned}
(1 + \sigma^{1/2})\|x - T_1(x)\| &\geq \|x - P_C[x - \eta(x)F(T_1(x))]\| \\
&\geq \min\{1, \bar{\eta}\}R(x) \\
&\quad - \|P_C[x - \eta(x)F(T_1(x))] - P_C[x - \eta(x)F(x)]\| \\
&\geq \min\{1, \bar{\eta}\}R(x) - L\hat{\eta}\|x - T_1(x)\|\,,
\end{aligned}$$

where the the second inequality is by [7, Lemma 1] assuming $\eta(x) \geq \bar{\eta}$, and the last follows from the nonexpansiveness of the projection operator, assuming also that $\eta(x) \leq \hat{\eta}$. The relation above implies (4) with $c_2 = \min\{1, \bar{\eta}\}(1 + \sigma^{1/2} + L\hat{\eta})^{-1}$.

Now, if the error bound (10) holds with some $\beta \geq 1$, then Theorem 1 gives a rate of convergence estimate for the inexact proximal point algorithm, including linear and sublinear rates. We remark that it is not clear whether the framework of [35] applies to the inexact hybrid scheme, even in the case of linear convergence. Specifically, it is not clear how to establish (3) with T_1 replaced by T. Regarding T as a two-step mapping appears to be very natural in this setting.

Finally, we note that in the special case of optimization other convergence estimates are possible [37, 1]; these can be in terms of the objective function rather than the distance to the solution set.

4.6. A splitting algorithm

Suppose now that the function F defining VIP(F, C) has the structure

$$F(x) = A(x) + B(x)\,.$$

Suppose further that, for each x fixed, VIP($A(\cdot) + B(x), C$) is in some sense easier to solve than the original problem. In this setting splitting-type methods are often useful. Assume that F is continuous and monotone, and $-B$ (and hence also A) strongly monotone with modulus $\mu > 0$. The following splitting method was considered in [26]:

$$T_1(x) = -(A + N_C)^{-1}(B(x) + \rho(x))\,,$$

$$T_2(x; y) = \arg\min_z\{\|z - x\| \mid \langle F(T_1(x)), z - T_1(x)\rangle \leq 0\}\,,$$

where $\rho(x)$ measures approximation to the solution of VIP($A(\cdot)+B(x), C$) and satisfies

$$\|\rho(x)\| \leq \sigma\mu\|x - T_1(x)\|\,,\ \ \sigma \in [0, 1)\,.$$

For

$$f(x) = \text{dist}(x, S)^2\,,$$

the analysis in [26] shows that

$$f(x) - f(T(x)) \geq c_6 \frac{\|x - T_1(x)\|^4}{\|F(T_1(x))\|^2}\,,\quad c_6 > 0\,.$$

It is not difficult to see that if x is contained in some bounded set B, then $\{F(T_1(x)), x \in B\}$ is bounded, and so (3) is satisfied with $\alpha = 4$ and some c_1 which depends on M. Furthermore, assuming that A is Lipschitz-continuous (with modulus $L > 0$), we have

$$\begin{aligned}\|x - T_1(x)\| &= \|x - P_C[T_1(x) - A(T_1(x)) - B(x)]\| \\ &\geq R(x) - \|P_C[x - F(x)] - P_C[T_1(x) - A(T_1(x)) - B(x)]\| \\ &\geq R(x) - \|x - T_1(x)\| - \|A(x) - A(T_1(x))\| \\ &\geq R(x) - (1 + L)\|x - T_1(x)\|\,,\end{aligned}$$

which implies condition (4) with $c_2 = (2 + L)^{-1}$.

Since [26] also shows convergence of the sequence, the sequence is certainly bounded, and hence Theorem 1 provides a convergence rate estimate for this splitting method under the given assumptions.

4.7. Minimizing the D-gap function

As one example of an (infeasible) descent algorithm for solving VIP(F, C), we shall consider minimization of the D-gap function [23, 36, 34]

$$h_{t,s}(x) := g_t(x) - g_s(x) \quad \forall x \in \Re^n ,$$

where $t > s > 0$ are parameters and g is the regularized gap function [2, 6] for VIP(F, C):

$$g_t(x) = \langle F(x), x - P_C[x - tF(x)]\rangle - \frac{1}{2t}\|x - P_C[x - tF(x)]\|^2 .$$

The gradient method for minimizing $h_{t,s}$ is given by

$$T_1(x) = x - \eta(x)\nabla h_{t,s}(x) , \quad T_2(x; y) = y .$$

If $\nabla h_{t,s}$ is Lipschitz-continuous (which holds, for example, if F' is Lipschitz-continuous and C is bounded), then any standard linesearch ensures that $\eta \geq \bar{\eta} > 0$ and condition (3) is satisfied with $\alpha = 2$ for all $x \in D = \Re^n$. Furthermore, it holds that $\nabla h_{t,s}(x^k) \to 0$. Note however that linear convergence does not follow from the classical analysis of gradient descent, because $h_{t,s}$ is not strongly convex, even locally.

To verify (4), we assume that F' is uniformly positive definite on the set of stationary points of $h_{t,s}$. In that case,

$$\begin{aligned}\|x - T_1(x)\| &= \eta(x)\|\nabla h_{t,s}(x)\| \\ &\geq c_7(1/s - 1/t)h_{t,s}(x) \\ &\geq c_7 2^{-1}(1/s - 1/t)^2 \min\{1, s\}R(x) ,\end{aligned}$$

where the second inequality is by $\eta(x) \geq \bar{\eta}$ and [34, Lemma 5], and the third is well known, e.g., [34, Lemma 1]. Since F' is positive definite (on the set of interest), error bound (10) holds with $\beta = 1$ (locally). Then Theorem 1 establishes the linear rate of convergence for gradient algorithm based on the D-gap function. Alternatively, the linear convergence could also be obtained applying [18], once the error bound in terms of $\|\nabla h_{t,s}(x)\|$ is established.

5. Concluding remarks

We have presented a unified analysis of some iterative algorithms for solving variational inequalities. Our framework includes the class of feasible descent methods of optimization, various popular projection schemes for variational inequalities, and proximal point methods, among others. This unifies convergence rate analysis for optimization and variational inequalities, including both linear are sublinear estimates. In addition, the framework has been extended to include a number of methods which do not appear to fit in previous studies. Finally, a new error bound result for γ-strictly monotone problems was presented.

Acknowledgements. I thank the three anonymous referees and the editor for constructive suggestions which led to considerable improvement of the paper.

The author is supported in part by CNPq Grant 300734/95-6, by PRONEX–Optimization, and by FAPERJ.

References

1. Alber, Ya.I., Burachik, R., Iusem, A.N.: A proximal point method for nonsmooth convex optimization problems in Banach spaces. Abstract and Appl. Anal. **2**, 97–120 (1997)
2. Auchmuty, G.: Variational principles for variational inequalities. Numer. Func. Anal. Optim. **10**, 863–874 (1989)
3. Cottle, R.W., Giannessi, F., Lions, J.-L.: Variational Inequalities and Complementarity Problems: Theory and Applications. Wiley, New York, 1980
4. Ferris, M.C., Pang, J.-S.: (editors). Complementarity and variational problems: State of the Art. SIAM Publications, 1997
5. Ferris, M.C., Mangasarian, O.L. Pang, J.-S.: (editors). Complementarity: Applications, Algorithms and Extensions. Kluwer Academic Publishers, 2001
6. Fukushima, M.: Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. Math. Prog. **53**, 99–110 (1992)
7. Gafni, E.M., Bertsekas, D.P.: Two–metric projection methods for constrained optimization. SIAM J. Control and Optim. **22**, 936–964 (1984)
8. Harker, P.T., Pang, J.-S.: Finite-dimensional variational inequality problems: A survey of theory, algorithms and applications. Math. Prog. **48**, 161–220 (1990)
9. Iusem, A.N.: An iterative algorithm for the variational inequality problem. Comput. Appl. Math. **13**, 103–114 (1994)
10. Iusem, A.N., Svaiter, B.F.: A variant of Korpelevich's method for variational inequalities with a new search strategy. Optim. **42**, 309–321 (1997)
11. Khobotov, E.N.: A modification of the extragradient method for the solution of variational inequalities and some optimization problems. USSR Comput.Math. Mathematical Phys. **27**, 1462–1473 (1987)
12. Korpelevich, G.M.: The extragradient method for finding saddle points and other problems. Matecon **12**, 747–756 (1976)
13. Li, W.: Remarks on matrix splitting algorithms for symmetric linear complementarity problems. SIAM J. Optim. **3**, 155–163 (1993)
14. Luo, Z.-Q.: New error bounds and their applications to convergence analysis of iterative algorithms. Math. Prog. **88**, 341–355 (2000)
15. Luo, Z.-Q., Tseng, P.: On the convergence of a matrix splitting algorithm for the symmetric monotone linear complementarity problem. SIAM J. Cont. Optim. **29**, 1037–1060 (1991)
16. Luo, Z.-Q., Tseng, P.: Error bound and convergence analysis of matrix splitting algorithms for the affine variational inequality problem. SIAM J. Optim. **2**, 43–54 (1992)
17. Luo, Z.-Q., Tseng, P.: On the linear convergence of descent methods for convex essentially smooth minimization. SIAM J. Cont. Optim. **30**, 408–425 (1992)
18. Luo, Z.-Q., Tseng, P.: Error bounds and convergence analysis of feasible descent methods: A general approach. Ann. Oper. Res. **46**, 157–178 (1993)
19. Luo, Z.-Q., Tseng, P.: Error bounds and reduced gradient projection algorithms for convex minimization over a polyhedral set. SIAM J. Optim. **3**, 43–59 (1993)
20. Marcotte, P.: Application of Khobotov's algorithm to variational inequalities and network equilibrium problems. Inf. Sys. Operational Res. **29**, 258–270 (1991)
21. Pang, J.-S.: A posteriori error bounds for the linearly–constrained variational inequality problem. Math. Operations Res. **12**, 474–484 (1987)
22. Pang, J.-S.: Error bounds in mathematical programming. Math. Prog. **79**, 299–332 (1997)
23. Peng, J.-M.: Equivalence of variational inequality problems to unconstrained optimization. Math. Prog. **78**, 347–356 (1997)
24. Robinson, S.M.: Some continuity properties of polyhedral multifunctions. Math. Prog. Study **14**, 206–214 (1981)

25. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control and Optim. **14**, 877–898 (1976)
26. Solodov, M.V.: A class of globally convergent algorithms for pseudomonotone variational inequalities. In: M.C. Ferris, O.L. Mangasarian, and J.-S. Pang, editors, Complementarity: Applications, Algorithms and Extensions, pages 297–315. Kluwer Academic Publishers, 2001
27. Solodov, M.V., Svaiter, B.F.: 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
28. Solodov, M.V., Svaiter, B.F.: A hybrid approximate extragradient–proximal point algorithm using the enlargement of a maximal monotone operator. Set-Valued Analysis **7**, 323–345 (1999)
29. Solodov, M.V., Svaiter, B.F.: A hybrid projection–proximal point algorithm. J. Convex Anal. **6**, 59–70 (1999)
30. Solodov, M.V., Svaiter, B.F.: A new projection method for variational inequality problems. SIAM J. Control and Optim. **37**, 765–776 (1999)
31. Solodov, M.V., Svaiter, B.F.: Error bounds for proximal point subproblems and associated inexact proximal point algorithms. Math. Prog. **88**, 371–389 (2000)
32. Solodov, M.V., Svaiter, B.F.: A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem. SIAM J. Optim. **10**, 605–625 (2000)
33. Solodov, M.V., Tseng, P.: Modified projection-type methods for monotone variational inequalities. SIAM J. Control and Optim. **34**, 1814–1830 (1996)
34. Solodov, M.V., Tseng, P.: Some methods based on the the D-gap function for solving monotone variational inequalities. Comput. Optim. Appl. **17**, 255–277 (2000)
35. Tseng, P.: On linear convergence of iterative methods for the variational inequality problem. J. Comput. Appl. Math. **60**, 237–252 (1995)
36. Yamashita, N., Taji, K., Fukushima, M.: Unconstrained optimization reformulations of variational inequality problems. J. Optim. Theory and Appl. **92**, 439–456 (1997)
37. Zhu, C.: Asymptotic convergence analysis of some inexact proximal point algorithms for minimization. SIAM J. Optim. **6**, 626–637 (1996)