On the Convergence of Descent Methods for Monotone Variational Inequalities

M. PATRIKSSON[1]

Abstract. Recently, Zhu and Marcotte [15] established the convergence of a modified descent algorithm for monotone variational inequalities. Using algorithmic equivalence results due to Patriksson [10, 11] and Larsson and Patriksson [7], we show that this convergence result may be used to establish the convergence of slightly modified versions of the classical successive approximation algorithms of Dafermos [2] and Cohen [1], and of the descent algorithms of Wu *et al.* [14], Patriksson [10, 11], and Larsson and Patriksson [7], under assumptions that are both much milder and much easier to verify than those for their original statements.

Key Words. Monotone variational inequalities, successive approximation algorithms, merit functions, descent methods, algorithmic equivalence, convergence

1 Introduction

Let X be a nonempty, closed and convex set in $\Re^n$, and $F : X \mapsto \Re^n$ a continuous and monotone mapping on X. Consider the *variational inequality problem* of finding an $x^* \in X$ such that

$$[\mathrm{VIP}(F,X)] \qquad F(x^*)^{\mathrm{T}}(x - x^*) \geq 0, \qquad \forall x \in X.$$

The variational inequality problem (and its special cases) has a large variety of applications in the mathematical and engineering sciences, for example in partial differential equations, equilibrium problems in games, economics and transportation analysis, and nonlinear programming.

It is a well-known fact that the problem [VIP(F, X)] describes the first-order optimality conditions of a convex optimization problem if F is the gradient of a convex function $f : X \mapsto \Re$, and of a convex-concave saddle point problem if $F(z,y)^{\mathrm{T}} = [\nabla_z L(z,y)^{\mathrm{T}}, -\nabla_y L(z,y)^{\mathrm{T}}]$ for some convex-concave function $L : Z \times Y \mapsto \Re$. (Here, $x^{\mathrm{T}} = (z^{\mathrm{T}}, y^{\mathrm{T}})$ and $X = Z \times Y$.) In these cases, many iterative algorithms are immediately available for the solution of [VIP(F, X)] where, in each iteration, a sufficient decrease in a merit function is obtained through a line search in a feasible descent direction; natural choices of merit functions are f and $f(\cdot) = \sup_{y \in Y} L(\cdot, y)$, respectively. In other cases (in which the term *asymmetric* variational inequality is frequently used) there is no merit function immediately available for monitoring the convergence of an iterative algorithm.

[1]Department of Mathematics, Linköping Institute of Technology, Linköping, Sweden.

The classical approach to solving an asymmetric variational inequality is to replace the mapping F with a sequence $\{F^k\}$ of simpler mappings, and define the iterate x^{k+1} by the solution to [VIP(F^k, X)]. Such *successive approximation* approaches include linear approximation methods [$F^k(x) = F(x^k) + A^k(x - x^k)$, where A^k is a positive definite matrix], among which we find Newton-type approaches, linearized Jacobi methods and the projection algorithm, and nonlinear approximation algorithms, such as the Jacobi method [$F_i^k(x_i) = F_i(x_{i\_}^k, x_i, x_{i_+}^k)$, $i \in \{1, \ldots, n\}$, where $x_{i\_}$ denotes the subvector of x with indices lower than i, and x_{i_+}, correspondingly, the subvector with higher indices], the Gauss–Seidel approach [$F_i^k(x_i) = F_i(x_{i\_}, x_i, x_{i_+}^k)$, $i \in \{1, \ldots, n\}$], and regularization methods [$F^k(x) = F(x) + \gamma_k R(x, x^k)$, where $\{\gamma_k\}$ is a sequence of positive parameters and $x \mapsto R(x, x^k)$ is a strongly monotone mapping]; see [9, 4] for surveys of successive approximation methods for [VIP(F, X)].

The convergence of such classical approaches is often based on a guaranteed monotone decrease of a merit function which is *artificial* in the sense that it is not possible to evaluate and therefore possible to utilize neither to monitor the convergence of the algorithm nor in line searches for accelerating its convergence. One example of an artificial merit function is the (unknown) Euclidean distance from a given iterate to the solution set Ω of [VIP(F, X)].

Recently, reformulations of asymmetric variational inequalities as, in general nonconvex, optimization problems have been shown to naturally lead to convergent descent algorithms (e.g., [3, 7]). In such reformulations, the problem [VIP(F, X)] is supplied with a specially constructed (calculable) merit function, ψ, whose (local) minima coincide with the solution set Ω of [VIP(F, X)]. In algorithms based on the direct minimization of ψ, the solution y^k to an approximate variational inequality [VIP(F^k, X)] defines a descent direction, $d^k \stackrel{\text{def}}{=} y^k - x^k$, for ψ; the next iterate is defined as the (approximate) solution to the problem

$$\min_{t \geq 0:\ x^k + td^k \in X} \psi(x^k + td^k).$$

The first known descent algorithm of this type is given by Zuhovickiĭ *et al.* [16]; see [10, 11, 7, 12] for surveys of descent methods for variational inequalities. Some of these algorithms are convergent under weaker conditions on the problem data than the classical successive approximation approaches, and the introduction of the line search may also lead to a higher practical convergence rate.

In general, however, in order to establish the convergence of either one of the two types of algorithms a strict or strong monotonicity assumption on F is required. [Examples of algorithms that require only monotonicity of F are the extragradient method [6] (a modification of the projection method), the descent method of Smith [13], and descent algorithms based on Newton's method (e.g., [8]).] In addition, the convergence conditions frequently include the required knowledge of certain problem parameters, such as the strong monotonicity and Lipschitz constants of F, which may be difficult to estimate, and further conditions which are difficult to verify in practice. This is in contrast to the mild (and in many cases easily verified) conditions that guarantee the convergence of these types of algorithms in the case of convex programming [11].

The need for these strong assumptions to hold in order to guarantee the convergence of iterative algorithms for [VIP(F, X)] is very unfortunate, in view of the fact that important applications of variational inequalities may fail to satisfy them; one example is traffic equilibrium models with detailed representations of junctions which, according to Heydecker [5], can not satisfy a strict monotonicity assumption.

Recently, Zhu and Marcotte [15] established the convergence of a general descent algorithm for variational inequalities, which requires only monotonicity of F. The purpose of this note is to show that this algorithm may be shown to be a slight modification of classical successive approximation schemes as well as descent algorithms recently presented for the solution of [VIP(F,X)], which all require strict or strong monotonicity of F in their original statements. Thus, by introducing the proper modifications of these algorithms, convergence may be established under much milder assumptions on F. Furthermore, the additional technical conditions present for the original algorithms may be removed, thus enabling a much easier verification of the convergence conditions.

In the next Section, we give a short description of the algorithm of Zhu and Marcotte, and the requirements for its convergence. In Section 3, we relate the algorithm classes of Dafermos [2], Cohen [1], Wu *et al.* [14], Patriksson [10, 11], and Larsson and Patriksson [7] to each other and to the algorithm of Zhu and Marcotte. Utilizing these relationships we establish the modifications to these algorithms that are necessary in order to yield convergence under only a monotonicity assumption on F.

2 The algorithm of Zhu and Marcotte

Henceforth, we shall assume that X is bounded. (This ensures that the solution set Ω is nonempty.)

Let $\phi : X \times X \mapsto \Re$ be a continuously differentiable function on $X \times X$ of the form $\phi(x,y)$, strictly convex with respect to y. Further, let ϕ be such that $\nabla_y \phi(x,x) = 0$, for all $x \in X$.

Let $\alpha > 0$, define

$$G_\alpha(x) = \max_{y \in X} \left\{ F(x)^{\mathrm{T}}(x-y) - \frac{1}{\alpha}\phi(x,y) \right\}, \tag{1}$$

and let $y_\alpha(x)$ denote the (unique) solution to the inner problem of (1).

The algorithm of Zhu and Marcotte proceeds as follows.

Step 0 (*Initialization*) Choose $x^0 \in X$, $\varepsilon, \alpha_0, \Delta_\alpha > 0$, and $0 < \beta, \gamma, \sigma < 1$. Let $k = 0$.

Step 1 (*Search direction generation*) Evaluate $G_{\alpha_k}(x^k)$, and let $d^k = y_{\alpha_k}(x^k) - x^k$.

Step 2 (*Stopping criterion*) If $G_{\alpha_k}(x^k) < \varepsilon$, then stop. If

$$G_{\alpha_k}(x^k) \le \frac{1}{(1-\gamma)\alpha} \left\{ \nabla_x \phi(x^k, y_{\alpha_k}(x^k))^{\mathrm{T}}(y_{\alpha_k}(x^k) - x^k) - \phi(x^k, y_{\alpha_k}(x^k)) \right\},$$

then let $\alpha_{k+1} = \alpha_k + \Delta_\alpha$, $x^{k+1} = x^k$, and go to Step 1; otherwise, let $\alpha_{k+1} = \alpha_k$, and go to Step 3.

Step 3 (*Armijo-type line search*) If $G_{\alpha_k}(x^k + d^k) \le G_{\alpha_k}(x^k)$, then let $t_k = 1$; otherwise, let m be the smallest positive integer such that

$$G_{\alpha_k}(x^k) - G_{\alpha_k}(x^k + \beta^m d^k) \ge -\sigma\beta^m \nabla G_{\alpha_k}(x^k)^{\mathrm{T}} d^k, \tag{2}$$

and $t_k = \beta^m$. Let $x^{k+1} = x^k + t_k d^k$, $k := k + 1$, and go to Step 1.

The convergence of the algorithm is given by the below theorem.

Theorem 1 [15] (Convergence of the algorithm of Zhu and Marcotte) *Let F be continuously differentiable, monotone and Lipschitz continuous on X. Let x^0 be arbitrary in X. Then either the algorithm terminates at a solution to* [VIP(F, X)] *or it generates an infinite sequence $\{x^k\}$ such that any limit point is a solution to* [VIP(F, X)].

In the next section we relate some earlier proposed iterative schemes for the solution of [VIP(F, X)] to each other and to the algorithm of Zhu and Marcotte. We finally describe the modifications to these algorithms that are necessary in order to ensure convergence under only a monotonicity assumption on F.

3 Algorithmic equivalence results and modified descent algorithms

We first state the generic cost approximation algorithm for [VIP(F, X)], introduced by Patriksson [10, 11] and Larsson and Patriksson [7].

Choose a sequence $\{\varphi^k\}$ of strictly convex and continuously differentiable functions $\varphi^k : X \mapsto \Re^n$, and a merit function ψ.

Step 0 (*Initialization*) Choose $x^0 \in X$, and let $k = 0$.

Step 1 (*Search direction generation*) Let $y(x^k)$ be the solution to

$$\max_{y \in X} \left\{ [F(x^k) - \nabla\varphi^k(x^k)]^{\mathrm{T}}(x^k - y) - \varphi^k(y) \right\}. \tag{3}$$

Let $d^k = y(x^k) - x^k$.

Step 2 (*Stopping criterion*) If $y(x^k) = x^k$, then stop.

Step 3 (*Line search*) Find a step length, t_k, which (approximately) solves the problem

$$\min_{t \ge 0:\ x^k + td^k \in X} \psi(x^k + td^k). \tag{4}$$

Let $x^{k+1} = x^k + t_k d^k$, $k := k + 1$, and go to Step 1.

Below, we state an algorithmic equivalence result, which establishes the close relationships among iterative methods for [VIP(F, X)], including the one of Zhu and Marcotte.

Theorem 2 (Algorithmic equivalence results)

(a) *If the sequence* $\{\varphi^k\}$ *is given by* $\varphi^k = \varphi(\cdot, x^k)$*, where* $\varphi : X \times X \mapsto \Re$ *is continuous on* $X \times X$*, then the resulting subproblem (3) is equivalent to the subproblems of Dafermos [2] and Wu et al. [14].*

(b) *If the sequence* $\{\varphi^k\}$ *is given by* $\varphi^k = \alpha_k \varphi$*, where* $\alpha_k > 0$ *and* $\varphi : X \mapsto \Re$ *is strictly convex, then the resulting subproblem (3) is equivalent to the subproblem of Cohen [1].*

(c) *If the sequence* $\{\varphi^k\}$ *is given by* $\varphi^k = \frac{1}{\alpha_k}\phi(x^k, \cdot)$*, where* $\alpha_k > 0$ *and* $\phi : X \times X \mapsto \Re$ *is given as in Section 2, then the resulting subproblem (3) is equivalent to the inner problem of (1).*

Proof.

(a) See [10, 11, 7].

(b) See [11, 7].

(c) Let $\varphi^k = \frac{1}{\alpha_k}\phi(x^k, \cdot)$. Then $\nabla\varphi^k(x^k) = 0$, and (3) equals (1). □

To describe the above mentioned algorithms as instances of the generic cost approximation algorithm, we need only to specify the choices of step lengths in Step 3. The successive approximation algorithms of Dafermos and Cohen correspond to taking unit steps, while the descent algorithms of Wu *et al.* and Larsson and Patriksson include line searches with respect to merit functions of the form (3). The generic algorithm includes a wide variety of classical algorithms for variational inequalities; for instance, with $t_k \equiv 1$, the projection algorithm results from the choice $\varphi^k(x) = \frac{1}{2}x^{\mathrm{T}}Bx$, where $B \in \Re^{n\times n}$ is symmetric and positive definite, and the symmetrized Newton method from the choice $\varphi^k(x) = \frac{1}{2}x^{\mathrm{T}}\nabla F(x^k)x$. Further examples are found in [10, 11, 7].

The convergence of the algorithms of Cohen, Dafermos, Wu *et al.*, and Larsson and Patriksson all require F to be at least strictly monotone. (In the case of Cohen, F must be strongly monotone.) Further, the convergence conditions of all these algorithms include additional assumptions on the relationships among problem parameters, such as the strong monotonicity and Lipschitz constants of F and of the respective choices of φ^k. By introducing the proper modifications of these algorithms, so that they fall into the framework of Zhu and Marcotte, all these assumptions may be replaced by only monotonicity, according to Theorem 1.

In the case of Cohen's [1] algorithm (the *auxiliary problem principle*), the sequence $\{\alpha_k\}$ of parameters would be governed by the scheme of Section 2. In the case of Dafermos [2], Wu *et al.* [14], Patriksson [10, 11], and Larsson and Patriksson [7], the function φ would be replaced by a sequence $\{\frac{1}{\alpha_k}\varphi\}$ of functions. In all these algorithms, a line search of the type given in Step 3 of the algorithm of Zhu and Marcotte would replace a line search of the form (4) or a unit step.

Disregarding the line search step of the algorithm of Section 2, which is only considered when it is necessary in order to guarantee global convergence, the algorithm contains the same steps as the successive approximation version of the generic cost approximation algorithm; an implementation of, for instance, a modified version of an instance of the algorithm class of Dafermos would therefore amount to only minor adjustments of an existing implementation.

4 Conclusion

In this note, we have established that with only minor adjustments, classical successive approximation algorithms and descent algorithms for variational inequalities, which require the problem defining mapping to be at least strictly monotone, can be made globally convergent under only a monotonicity assumption. Furthermore, with these minor adjustments, the additional technical assumptions and the required knowledge of estimates of certain problem parameters, which are present in the convergence conditions of their original statements, may be removed; thus, a much easier verification of the convergence conditions is enabled.

References

[1] G. Cohen, *Auxiliary problem principle extended to variational inequalities*, Journal of Optimization Theory and Applications, 59 (1988), pp. 325–333.

[2] S. Dafermos, *An iterative scheme for variational inequalities*, Mathematical Programming, 26 (1983), pp. 40–47.

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

[4] P. T. Harker and J.-S. Pang, *Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications*, Mathematical Programming, 48 (1990), pp. 161–220.

[5] B. G. Heydecker, *Some consequences of detailed junction modeling in road traffic assignment*, Transportation Science, 17 (1983), pp. 263–281.

[6] G. M. Korpelevich, *The extragradient method for finding saddle points and other problems*, Matekon, 13 (1977), pp. 35–49.

[7] T. Larsson and M. Patriksson, *A class of gap functions for variational inequalities*, Mathematical Programming, 64 (1994), pp. 53–79.

[8] P. Marcotte and J.-P. Dussault, *A modified Newton method for solving variational inequalities*, in Proceedings of the 24th IEEE Conference on Decision and Control, Fort Lauderdale, FL, 1985, pp. 1433–1436.

[9] J.-S. Pang and D. Chan, *Iterative methods for variational and complementarity problems*, Mathematical Programming, 24 (1982), pp. 284–313.

[10] M. Patriksson, *A unified description of iterative algorithms for traffic equilibria*, European Journal of Operational Research, 71 (1993), pp. 154–176.

[11] M. Patriksson, *A unified framework of descent algorithms for nonlinear programs and variational inequalities*, PhD thesis, Department of Mathematics, Linköping Institute of Technology, Linköping, Sweden, 1993.

[12] M. Patriksson, *The Traffic Assignment Problem: Models and Methods*, VSP, Utrecht, The Netherlands, 1995.

[13] M. J. Smith, *A descent algorithm for solving monotone variational inequalities and monotone complementarity problems*, Journal of Optimization Theory and Applications, 44 (1984), pp. 485–496.

[14] J. H. Wu, M. Florian, and P. Marcotte, *A general descent framework for the monotone variational inequality problem*, Mathematical Programming, 61 (1993), pp. 281–300.

[15] D. L. Zhu and P. Marcotte, *Modified descent methods for solving the monotone variational inequality problem*, Operations Research Letters, 14 (1993), pp. 111–120.

[16] S. I. Zuhovickiĭ, R. A. Polyak, and M. E. Primak, *Two methods of search for equilibrium points of n-person concave games*, Soviet Mathematics Doklady, 10 (1969), pp. 279–282.