A QP-FREE CONSTRAINED NEWTON-TYPE METHOD FOR VARIATIONAL INEQUALITY PROBLEMS

Christian Kanzow[1] and Hou-Duo Qi[2]

[1] University of Hamburg
Institute of Applied Mathematics
Bundesstrasse 55, D-20146 Hamburg, Germany
e-mail: kanzow@math.uni-hamburg.de

[2] Chinese Academy of Sciences
Institute of Computational Mathematics
and Scientific/Engineering Computing
P.O. Box 2719, Beijing, China
e-mail: qhd@lsec.cc.ac.cn

March 27, 1997 (revised November 10, 1997)

Abstract: We consider a simply constrained optimization reformulation of the Karush-Kuhn-Tucker conditions arising from variational inequalities. Based on this reformulation, we present a new Newton-type method for the solution of variational inequalities. The main properties of this method are: (a) it is well-defined for an arbitrary variational inequality problem, (b) it is globally convergent at least to a stationary point of the constrained reformulation, (c) it is locally superlinearly/quadratically convergent under a certain regularity condition, (d) all iterates remain feasible with respect to the constrained optimization reformulation, and (e) it has to solve just one linear system of equations at each iteration. Some preliminary numerical results indicate that this method is quite promising.

Key Words: Variational inequality problem, Newton's method, semismoothness, global convergence, quadratic convergence, strong regularity.

1 Introduction

Consider the *variational inequality problem* VIP(X, F) which is to find a vector $x^* \in X$ such that
$$F(x^*)^T(x - x^*) \geq 0 \quad \text{for all } x \in X,$$
where $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable and the feasible set $X \subseteq \mathbb{R}^n$ is described by some equality and inequality constraints:
$$X := \{x \in \mathbb{R}^n \,|\, h(x) = 0, g(x) \geq 0\};$$
here, the constraint functions $h : \mathbb{R}^n \to \mathbb{R}^p$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are assumed to be twice continuously differentiable.

The variational inequality problem has a number of important applications in operations research, engineering problems and economic equilibrium problems. Constrained optimization problems with a pseudoconvex objective function, saddlepoint problems, complementarity and mixed complementarity problems are special cases of a variational inequality. We refer the reader to the survey paper [15] by Harker and Pang, to the book [24] by Nagurney as well as to the recent work [9] by Ferris and Pang.

Many, if not most, algorithms for the solution of the variational inequality problem do not really try to solve this problem directly; instead, they often try to solve the corresponding *Karush-Kuhn-Tucker (KKT) system.* In order to state this KKT system, let us denote by
$$L(x, y, z) := F(x) + \nabla h(x)y - \nabla g(x)z$$
the *Lagrangian* of VIP(X, F). The KKT system of VIP(X, F) can now be written as follows:
$$\begin{aligned} L(x, y, z) &= 0, \\ h(x) &= 0, \\ g(x) \geq 0, z \geq 0, z^T g(x) &= 0. \end{aligned}$$
Any vector $w^* = (x^*, y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ satisfying this system is called a *KKT point* of VIP(X, F).

The relationship between the variational inequality problem VIP(X, F) and its KKT conditions is very close and, in fact, much closer than for optimization problems. For example, if x^* denotes a (local) solution of VIP(X, F) and if any of the standard constraint qualifications (like the Mangasarian-Fromovitz or the linear independence constraint qualification or the linearity of the constraint functions) holds at this solution, then there exist multipliers $y^* \in \mathbb{R}^p$ and $z^* \in \mathbb{R}^m$ such that the triple $w^* := (x^*, y^*, z^*)$ is a KKT point. Conversely, if $w^* = (x^*, y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ is a KKT point, h is an affine mapping and g has concave component functions, then x^* is a solution of VIP(X, F) (note that, in this case, we do not have to assume monotonicity or anything else for the mapping F). In particular, if both h and g are affine, then the variational inequality VIP(X, F) and the corresponding KKT system are fully equivalent. The reader is referred to [15] for some more details.

The algorithm to be described in this paper is based on the following reformulation of the KKT system: Let $\varphi : \mathbb{R}^2 \to \mathbb{R}$ denote the Fischer function [10]
$$\varphi(a, b) := \sqrt{a^2 + b^2} - a - b,$$

define

$$\phi(g(x), z) := (\varphi(g_1(x), z_1), \ldots, \varphi(g_m(x), z_m))^T \in \mathbb{R}^m,$$

and let $\Phi : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ be the equation operator

$$\Phi(w) := \Phi(x, y, z) := \begin{pmatrix} L(x, y, z) \\ h(x) \\ \phi(g(x), z) \end{pmatrix}.$$

Then it is easy to see that $w^* = (x^*, y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ is a KKT point of VIP(X, F) if and only if this vector solves the nonlinear system of equations

$$\Phi(w) = 0. \tag{1}$$

Alternatively, $w^* = (x^*, y^*, z^*)$ is a KKT point of VIP(X, F) if and only if it is a global minimizer of the problem

$$\min \;\; \Psi(w)$$

with $\Psi(w^*) = 0$, where

$$\Psi(w) := \frac{1}{2}\Phi(w)^T\Phi(w) = \frac{1}{2}\|\Phi(w)\|^2$$

denotes the natural merit function of the equation operator Φ. This approach has been used in [3, 4, 5] in order to develop some (unconstrained) Newton-type methods for the solution of problem (1) as well as in [30] for the local solution of the KKT system arising from constrained optimization problems.

Despite the quite strong theoretical and numerical properties of these unconstrained Newton-type methods, however, Peng [28] was able to present an example which indicates that these methods may fail even for strongly monotone variational inequalities.

In order to overcome this problem and motivated by the fact that we know a priori that $z^* \geq 0$ at any KKT point $w^* = (x^*, y^*, z^*)$ of VIP(X, F), Facchinei et al. [7] investigated a QP-based method for the solution of the constrained minimization problem

$$\min \;\; \Psi(w) \quad \text{s.t.} \quad z \geq 0. \tag{2}$$

The main motivation for our paper is to describe an algorithm for the solution of problem (2) which just solves one linear system of equations instead of a quadratic program at each iteration without destroying the global and local convergence properties of the method from [7]. In particular, our new algorithm has the following nice features:

(a) it is well-defined for an arbitrary variational inequality problem;

(b) every accumulation point of a sequence generated by this algorithm is at least a stationary point for the constrained reformulation (2);

(c) all iterates remain feasible with respect to the constraints in (2);

(d) it has to solve just one linear system of equations at each iteration; this system is actually of reduced dimension;

(e) it is locally superlinearly/quadratically convergent under Robinson's [32] strong regularity condition.

In particular, the combination of points (c), (d) and (e) seems to be a difficult task. As to the knowledge of the authors, the current paper is the first one which has all these desirable features. For some other approaches which are also based on a suitable reformulation of the KKT conditions, we refer the reader to [12, 19, 21, 25, 26, 33, 35, 36].

The paper is organized as follows: In Section 2, we restate some background material on the functions Φ and Ψ as well as on the constrained reformulation (2). The algorithm itself is stated in Section 3 where we also show that it is well-defined for an arbitrary variational inequality problem. The global convergence of this algorithm to a stationary point of (2) is established in Section 4, whereas we prove local fast convergence in Section 5. Section 6 is devoted to some preliminary numerical results. We conclude with some final remarks in Section 7.

Some words about our notation: A function $G : \mathbb{R}^t \to \mathbb{R}^t$ is called a C^k function if it is k times continuously differentiable, and an LC^k function if it is a C^k function such that the kth derivative is locally Lipschitz continuous everywhere. The Jacobian of a C^1 function $G : \mathbb{R}^t \to \mathbb{R}^t$ at a point $w \in \mathbb{R}^t$ is denoted by $G'(w)$, whereas $\nabla G(w)$ is the transposed Jacobian. This notation is consistent with our notion of a gradient vector $\nabla g(w)$ for a real-valued function $g : \mathbb{R}^t \to \mathbb{R}$ since we view $\nabla g(w)$ as a column vector.

If $M \in \mathbb{R}^{t\times t}, M = (m_{ij})$, is any given matrix and $I, J \subseteq \{1, \ldots, t\}$ are two subsets, then M_{IJ} denotes the $|I| \times |J|$ submatrix with elements $m_{ij}, i \in I, j \in J$. Similarly, $M_{\cdot J}$ indicates the submatrix with elements $m_{ij}, i \in \{1, \ldots, t\}, j \in J$, i.e., we obtain $M_{\cdot J}$ from M by removing all columns belonging to indices $j \notin J$. A similar notation is used for subvectors.

If $w = (x^T, y^T, z^T)^T \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, we often simplify our notation and write $w = (x, y, z)$. All vector norms used in this paper are Euclidian norms, and matrix norms are assumed to be consistent with this vector norm and to satisfy the inequality $\|AB\| \leq \|A\| \, \|B\|$ whenever the matrix product AB is defined.

Finally, we make extensive use of the Landau symbols $o(\cdot)$ and $O(\cdot)$: If $\{\alpha_k\}$ and $\{\beta_k\}$ are two sequences of positive numbers, then $\alpha_k = O(\beta_k)$ if $\limsup_{k\to\infty} \alpha_k/\beta_k < +\infty$, i.e., if there exists a constant $c > 0$ such that $\alpha_k \leq c\beta_k$ for all k, and $\alpha_k = o(\beta_k)$ if $\lim_{k\to\infty} \alpha_k/\beta_k = 0$ for $\beta_k \to 0$.

2 Mathematical Background

In this section, we restate a couple of properties of the equation operator Φ and the corresponding merit function Ψ. We start by noting that Φ is locally Lipschitz continuous everywhere, so that Clarke's [1] generalized Jacobian $\partial\Phi(w)$ is well-defined at any point $w = (x, y, z) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$. The structure of this set is given in the following result whose proof can be found in [4].

Proposition 2.1 *Let $w = (x, y, z) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$. Then, each element $H \in \partial\Phi(w)$ can* *be represented as follows:*

$$H = \begin{pmatrix} \nabla_x L(w) & \nabla h(x) & \nabla g(x) D_a(w) \\ \nabla h(x)^T & 0 & 0 \\ -\nabla g(x)^T & 0 & D_b(w) \end{pmatrix}^T,$$

where $D_a(w) = diag(a_1(w), \dots, a_m(w)), D_b(w) := diag(b_1(w), \dots, b_m(w)) \in \mathbb{R}^{m \times m}$ *are diagonal matrices whose* j*th diagonal elements are given by*

$$a_j(w) = \frac{g_j(x)}{\sqrt{g_j(x)^2 + z_j^2}} - 1, \quad b_j(w) = \frac{z_j}{\sqrt{g_j(x)^2 + z_j^2}} - 1$$

if $(g_j(x), z_j) \neq (0,0)$*, and by*

$$a_j(w) = \xi_j - 1, \quad b_j(w) = \zeta_j - 1 \quad \text{for any } (\xi_j, \zeta_j) \text{ with } \|(\xi_j, \zeta_j)\| \leq 1$$

if $(g_j(x), z_j) = (0,0)$.

The next property follows from the fact that Φ is a (strongly) semismooth operator under certain smoothness assumptions for F, h and g, see, e.g., [29, 31, 27, 11].

Proposition 2.2 *For any* $w = (x, y, z) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$*, we have*

$$\|\Phi(w + d) - \Phi(w) - Hd\| = o(\|d\|) \quad \text{for } d \to 0 \text{ and } H \in \partial\Phi(w + d).$$

If F *is an* LC^1 *mapping and* h, g *are* LC^2 *mappings, then*

$$\|\Phi(w + d) - \Phi(w) - Hd\| = O(\|d\|^2) \quad \text{for } d \to 0 \text{ and } H \in \partial\Phi(w + d).$$

The following result gives a characterization of Robinson's [32] strong regularity condition in terms of an important property of the equation operator Φ. A proof of this result can be found in [4]. For the precise definition, some further characterizations and sufficient conditions for strong regularity, we refer the reader to Robinson [32] as well as to the recent article by Liu [20].

Proposition 2.3 *A KKT point* $w^* = (x^*, y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ *of VIP(*X, F*) is a strongly regular KKT point if and only if all elements in the generalized Jacobian* $\partial\Phi(w^*)$ *are nonsingular.*

An immediate consequence of Proposition 2.3 and the above-mentioned semismoothness of Φ is the following result, see, e.g., [31, 27].

Proposition 2.4 *Assume that* $w^* = (x^*, y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ *is a strongly regular KKT point of VIP(*X, F*). Then the following statements hold:*

(a) *There are constants* $c_1 > 0$ *and* $\delta_1 > 0$ *such that the matrices* $H \in \partial\Phi(w)$ *are nonsingular with*

$$\|H^{-1}\| \leq c_1$$

for all w *with* $\|w - w^*\| \leq \delta_1$.

(b) There are constants $c_2 > 0$ and $\delta_2 > 0$ such that

$$\|\Phi(w)\| \geq c_2 \|w - w^*\|$$

for all w with $\|w - w^*\| \leq \delta_2$.

The next result is essential for the design of our algorithm. Its proof can also be found in [4].

Proposition 2.5 *The merit function Ψ is continuously differentiable with $\nabla\Psi(w) = H^T\Phi(w)$ for an arbitrary element $H \in \partial\Phi(w)$.*

We finally restate a result from [7] which shows that, under reasonable conditions, a stationary point of (2) is already a KKT point of VIP(X, F).

Proposition 2.6 *Let $w^* = (x^*, y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ be a stationary point of (2). Assume that*

(a) $\nabla_x L(w^*)$ is positive semidefinite on $\mathbb{R}^n$;

(b) $\nabla_x L(w^*)$ is positive definite on a certain cone $\mathcal{C}(x^*) \subseteq \mathbb{R}^n$ (see [7] for more details);

and either of the following two conditions holds:

(c1) $\nabla h(x^*)$ has full column rank;

(c2) h is an affine function, and the system $h(x) = 0$ is consistent.

Then w^* is a KKT point of VIP(X, F).

Since $z^* \geq 0$ for any stationary point $w^* = (x^*, y^*, z^*)$ of (2), condition (a) in Proposition 2.6 is obviously satisfied for monotone variational inequalities where F is a monotone function, each g_i is concave and h is affine; on the other hand, the cone $\mathcal{C}(x^*)$ in condition (b) is usually quite small, but even in the unlikely situation where $\mathcal{C}(x^*) = \mathbb{R}^n$, condition (b) is satisfied for strongly monotone variational inequalities where F is strongly monotone and h, g satisfy the assumptions for monotone problems. Finally, we note that, if h is an affine function, then the second condition in (c2) is quite obvious since otherwise the feasible set X would be empty.

3 Statement of Algorithm

In this section, we give a detailed description of our algorithm and show that it is well-defined for an arbitrary variational inequality problem VIP(X, F).

From now on, we will often abbreviate the gradient vector $\nabla\Psi(w^k)$ by g^k (in contrast go $g(x^k)$ which denotes the function value of the inequality constraints at the current point x^k, so there should be no ambiguity). Moreover, we will use the index sets

$$\begin{aligned} \mathcal{I} &:= \{1, \ldots, n\}, \\ \mathcal{P} &:= \{n+1, \ldots, n+p\}, \\ \mathcal{J} &:= \{n+p+1, \ldots, n+p+m\}, \end{aligned}$$

i.e., $\mathcal{I}$ denotes the index set for the variables x, $\mathcal{P}$ is the index set for the equality constraints and the variables y, and $\mathcal{J}$ is the index set for the inequality constraints and the variables z. For example, if $w = (x, y, z) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ is any given vector, then $w_{\mathcal{I}} = x, w_{\mathcal{P}} = y$ and $w_{\mathcal{J}} = z$. We also stress that, if $j \in \mathcal{J}$ or $J \subseteq \mathcal{J}$, then w_j is a component of the z-part of the vector w and w_J is a subvector of the z-part of w.

Finally, we will denote by $\rho : \mathbb{R} \to \mathbb{R}$ a *forcing function*, i.e., a continuous function with the following two properties:

(a) $\rho(s) \geq 0$ for all $s \in \mathbb{R}$;

(b) $\rho(s) = 0 \Longleftrightarrow s = 0$.

We are now in the position to give a precise statement of our algorithm for the solution of the constrained reformulation (2) of a KKT system arising from variational inequalities.

Algorithm 3.1 *(QP-free Constrained Newton-type Algorithm)*

(S.0) (Initialization)
Choose $w^0 = (x^0, y^0, z^0) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ *with* $z^0 \geq 0, \sigma \in (0,1), \beta \in (0,1), \gamma \in (0,1), c > 0, \delta > 0, \varepsilon \geq 0,$ *and set* $k := 0$.

(S.1) (Active Set Strategy)
Let

$$\delta_k := \min\{\delta, c\sqrt{\|\Phi(w^k)\|}\}$$

and define

$$J_k := \{j \in \mathcal{J} | \, w_j^k \leq \delta_k\}.$$

(S.2) (Termination Criterion)
Let

$$v^k := \begin{pmatrix} v^k_{J_k} \\ v^k_{\bar{J}_k} \end{pmatrix}$$

be defined by

$$v^k_{J_k} = \min\{w^k_{J_k}, g^k_{J_k}\} \quad and \quad v^k_{\bar{J}_k} = g^k_{\bar{J}_k},$$

where $\bar{J}_k := \mathcal{I} \cup \mathcal{P} \cup (\mathcal{J} \setminus J_k)$. *If* $\|v^k\| \leq \varepsilon$, *stop.*

(S.3) (Subproblem Solution)
Select $H_k \in \partial\Phi(w^k)$ *and write* $H_k = (H^k_{.J_k}, H^k_{.\bar{J}_k})$. *Let* $d^k_{\bar{J}_k}$ *be the unique solution of the reduced linear system*

$$\left((H^k_{.\bar{J}_k})^T H^k_{.\bar{J}_k} + \rho(\Psi(w^k))I\right) d_{\bar{J}_k} = -v^k_{\bar{J}_k}. \tag{3}$$

(S.4) (Feasibility Preserving Safeguard)
Compute the search directions

$$d^k := \begin{pmatrix} -w^k_{J_k} \\ d^k_{\bar{J}_k} \end{pmatrix} \quad and \quad \tilde{d}^k := \begin{pmatrix} -v^k_{J_k} \\ d^k_{\bar{J}_k} \end{pmatrix},$$

and define

$$\begin{aligned} \bar{\tau}_k &:= \sup\{\tau \,|\, w^k_{\mathcal{J}\setminus J_k} + \tau d^k_{\mathcal{J}\setminus J_k} \geq 0\}, \\ \tau_k &:= \min\{1, \bar{\tau}_k\}. \end{aligned}$$

(S.5) (Computation of New Iterate)
If

$$\Psi(w^k + \tau_k d^k) \leq \gamma \Psi(w^k) \tag{4}$$

then *(*** **fast step** ***)*
set $w^{k+1} := w^k + \tau_k d^k$
else *(*** **safe step** ***)*
compute $t_k := \max\{\beta^\ell \,|\, \ell = 0, 1, 2, \dots\}$ *such that*

$$\Psi(w^k + \tau_k t_k \tilde{d}^k) \leq (1 - \sigma \tau_k t_k^2) \Psi(w^k) \tag{5}$$

and set $w^{k+1} := w^k + \tau_k t_k \tilde{d}^k$.

(S.6) (Update)
Set $k \leftarrow k + 1$ *and go to (S.1).*

Before analysing the properties of Algorithm 3.1, let us give some explanations and motivational remarks: First of all, the set J_k defined in Step (S.1) is used as an approximation for the set of active constraints

$$J_* := \{j \in \mathcal{J} \,|\, w^*_j = 0\}$$

for the constrained reformulation (2). In fact, we will show in Section 5 that J_k eventually coincides with J_* under reasonable assumptions. The main motivation for the precise definition of J_k originates from the recent paper [6] to which we refer for some further details.

The definition of v^k in Step (S.2) plays a crucial role in the definition of the search direction $\tilde{d}^k$ in Step (S.4) and therefore in the safe step branch of Step (S.5). In (S.2), it is only used in order to terminate the iteration. The termination criterion will be motivated by Lemma 3.2 below which says that $v^k = 0$ can only happen if w^k is already a stationary point of (2).

The basic motivation behind the linear system (3) in Step (S.3) is to take the standard Newton equation

$$H_k d = -\Phi(w^k)$$

for the equation operator Φ and to use a Levenberg-Marquardt-type regularization of it:

$$\left(H_k^T H_k + \rho(\Psi(w^k)) I\right) d = -H_k^T \Phi(w^k) = -g^k, \tag{6}$$

where the second equality comes from Proposition 2.5. The linear system (3) can now be derived from (6) by taking into account that J_k is an approximation of J_* so that we view the variables w^k_j corresponding to the indices $j \in J_k$ as being fixed.

The main idea of Step (S.4) is to compute the largest possible stepsize which guarantees that the next iterate $w^{k+1} = (x^{k+1}, y^{k+1}, z^{k+1})$ also satisfies the feasibility condition $z^{k+1} \geq 0$.

Note that the value of τ_k is always positive since $w_j^k > 0$ for all $j \in \mathcal{J} \setminus J_k$, and that there is a simple expression for τ_k, namely

$$\tau_k = \min \left\{ 1, \min_{\{j \in \mathcal{J} \setminus J_k \mid d_j^k < 0\}} \{ -w_j^k / d_j^k \} \right\}.$$

In Step (S.5), we utilize the two search directions d^k and $\tilde{d}^k$ computed in the previous step. Note, however, that there is not a big difference in these two directions; both depend on the solution $d_{\bar{J}_k}^k$ of the linear system (3) and, in fact, coincide completely if $w_j^k \leq g_j^k$ for all $j \in J_k$. In general, however, these two vectors differ from each other in the components $j \in J_k$, and this difference is essential in our convergence analysis for Algorithm 3.1: The direction d^k is used in order to prove local fast convergence under Robinson's [32] strong regularity condition. Hence, if the descent test (4) for our merit function Ψ is satisfied, we call the iteration k a fast step (and the search direction d^k itself a fast search direction), whereas in the safe step branch in Step (S.5) we take the direction $\tilde{d}^k$ and perform a line search for the merit function Ψ along this direction. This will guarantee global convergence to stationary points of (2).

We now start our analysis of Algorithm 3.1. We will always assume implicitly that the termination parameter ε in Algorithm 3.1 is equal to zero, and that the algorithm does not terminate after finitely many iterations. This is a reasonable assumption since the next result shows that otherwise the current iterate $w^k = (x^k, y^k, z^k)$ would already be a stationary point of (2).

Lemma 3.2 *Let $w^k = (x^k, y^k, z^k) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ be any given point with $z^k \geq 0$. Then the following statements are equivalent:*

(a) w^k is a stationary point of (2).

(b) $(g^k)^T v^k = 0$.

(c) $v^k = 0$.

Proof. (a) $\Rightarrow$ (b): Assume that w^k is a stationary point of (2). Then

$$g_{\mathcal{I}}^k = 0,\ g_{\mathcal{P}}^k = 0,\ z^k \geq 0,\ g_{\mathcal{J}}^k \geq 0,\ (z^k)^T g_{\mathcal{J}}^k = 0. \tag{7}$$

In order to verify statement (b), let us partition the index set J_k into the following two subsets:

$$J_k^> := \{ j \in J_k \mid w_j^k > g_j^k \} \quad \text{and} \quad J_k^\leq := \{ j \in J_k \mid w_j^k \leq g_j^k \}.$$

Now the definition of v^k in Step (S.2) of Algorithm 3.1 implies

$$\begin{array}{rcl} (g^k)^T v^k & = & \sum_{j \in \bar{J}_k} (g_j^k)^2 + \sum_{j \in J_k} g_j^k \min\{w_j^k, g_j^k\} \\ & = & \sum_{j \in \bar{J}_k} (g_j^k)^2 + \sum_{j \in J_k^>} (g_j^k)^2 + \sum_{j \in J_k^\leq} g_j^k w_j^k. \end{array} \tag{8}$$

From (7) and the definitions of the index sets J_k, $J_k^>$ and $J_k^\leq$, we obtain

$$g_j^k = 0 \quad \forall j \in \mathcal{I} \cup \mathcal{P} \cup (\mathcal{J} \setminus J_k) \cup J_k^>$$

and

$$w_j^k = 0 \quad \forall j \in J_k^{\leq}.$$

Hence (8) implies $(g^k)^T v^k = 0$.

(b) $\Rightarrow$ (c): Assume that statement (b) holds. We then obtain from (8) that

$$g_j^k = 0 \quad \forall j \in \bar{J}_k \cup J_k^{>}$$

and

$$g_j^k w_j^k = 0 \quad \forall j \in J_k^{\leq}$$

since $g_j^k w_j^k \geq 0$ for all $j \in J_k^{\leq}$ (recall the definition of $J_k^{\leq}$ and take into account that $z^k \geq 0$ by assumption). Hence the definition of v^k immediately gives $v^k = 0$.

(c) $\Rightarrow$ (a): If $v^k = 0$, we have

$$g_{\bar{J}_k}^k = 0 \quad \text{and} \quad \min\{w_{J_k}^k, g_{J_k}^k\} = 0$$

which is equivalent to

$$g_j^k = 0 \ \ \forall j \in \bar{J}_k \quad \text{and} \quad w_j^k \geq 0, g_j^k \geq 0, w_j^k g_j^k = 0 \ \ \forall j \in J_k.$$

Since $z^k \geq 0$ by our general assumption, this implies that (7) holds, i.e., w^k is a stationary point of (2). $\square$

The following result shows that the direction vector $\tilde{d}^k$ used in the safe step branch of Algorithm 3.1 is a descent direction for the merit function Ψ.

Lemma 3.3 *Let $w^k = (x^k, y^k, z^k) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ be any given point with $z^k \geq 0$. Then*

$$\nabla\Psi(w^k)^T \tilde{d}^k \leq 0.$$

Moreover, if w^k is not a stationary point of (2), then

$$\nabla\Psi(w^k)^T \tilde{d}^k < 0.$$

Proof. Using (3), we have

$$\begin{aligned} \nabla\Psi(w^k)^T \tilde{d}^k &= (g_{J_k}^k)^T \tilde{d}_{J_k}^k + (g_{\bar{J}_k}^k)^T d_{\bar{J}_k}^k \\ &= -(g_{J_k}^k)^T v_{J_k}^k - (d_{\bar{J}_k}^k)^T \left((H_{\cdot\bar{J}_k}^k)^T H_{\cdot\bar{J}_k}^k + \rho(\Psi(w^k)) I \right) d_{\bar{J}_k}^k. \end{aligned}$$

Now the first part on the right-hand side is nonpositive by the proof of Lemma 3.2, cf. (8). Moreover, the second part is nonpositive since the matrix $(H_{\cdot\bar{J}_k}^k)^T H_{\cdot\bar{J}_k}^k + \rho(\Psi(w^k)) I$ is positive semidefinite. This proves the first statement.

Now assume that w^k is not a stationary point of (2). Then $\Psi(w^k) > 0$ so that the matrix $(H_{\cdot\bar{J}_k}^k)^T H_{\cdot\bar{J}_k}^k + \rho(\Psi(w^k)) I$ is positive definite. Assume that $\nabla\Psi(w^k)^T \tilde{d}^k = 0$. Then we must have $(g_{J_k}^k)^T v_{J_k}^k = 0$ and $d_{\bar{J}_k}^k = 0$. This implies $g_{\bar{J}_k}^k = 0$ and therefore $(g^k)^T v^k = 0$. Hence w^k is a stationary point of (2) by Lemma 3.2, a contradiction to our assumption. $\square$

As a simple consequence of Lemma 3.3, we now show that the line search in (5) is always well-defined.

Corollary 3.4 *Let $w^k = (x^k, y^k, z^k) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ be any given point with $z^k \geq 0$, and assume that w^k is not a stationary point of (2). Then there is a finite exponent ℓ_k such that the stepsize $t_k = \beta^{\ell_k}$ satisfies the line search test (5) in Algorithm 3.1.*

Proof. Assume that condition (5) is not satisfied for any integer $\ell \geq 0$. Then

$$\Psi(w^k + \tau_k \beta^\ell \tilde{d}^k) > (1 - \sigma \tau_k \beta^{2\ell}) \Psi(w^k)$$

for all $\ell \geq 0$. Rearranging terms gives

$$\frac{\Psi(w^k + \tau_k \beta^\ell \tilde{d}^k) - \Psi(w^k)}{\tau_k \beta^\ell} > -\sigma \beta^\ell \Psi(w^k)$$

for all $\ell \geq 0$. Taking the limit $\ell \to \infty$, we obtain from the continuous differentiability of Ψ (cf. Proposition 2.5) that

$$\nabla \Psi(w^k)^T \tilde{d}^k \geq 0.$$

Hence, by Lemma 3.3, w^k must be a stationary point of (2) in contrast to our assumption. $\Box$

The following result shows that the iterates $\{w^k\}$ generated by Algorithm 3.1 stay feasible with respect to the constrained reformulation (2).

Lemma 3.5 *Let $w^k = (x^k, y^k, z^k) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ be any given point with $z^k \geq 0$, and assume that w^k is not a stationary point of (2). Then the next iterate w^{k+1} can be computed by Algorithm 3.1, and it holds that $z^{k+1} \geq 0$.*

Proof. The fact that w^{k+1} can be computed by Algorithm 3.1 follows immediately from Corollary 3.4. In order to prove the nonnegativity of z^{k+1}, first assume that we take a fast step at iteration k, i.e.,

$$z^{k+1} = z^k + \tau_k d^k_{\mathcal{J}}.$$

Then $w_j^{k+1} \geq 0$ for all $j \in \mathcal{J} \setminus J_k$ by definition of τ_k, whereas we have

$$w_j^{k+1} = w_j^k + \tau_k d_j^k = (1 - \tau_k) w_j^k \geq 0$$

for $j \in J_k$ since $\tau_k \leq 1$ and $w_j^k \geq 0$ by assumption.

If, on the other hand, we take a safe step at iteration k, we have

$$z^{k+1} = z^k + \tau_k t_k \tilde{d}^k_{\mathcal{J}}.$$

Using the fact that $\tau_k t_k \leq \bar{\tau}_k$ and the definition of $\bar{\tau}_k$, we obtain $w_j^{k+1} = w_j^k + \tau_k t_k d_j^k \geq 0$ for all $j \in \mathcal{J} \setminus J_k$. So assume $j \in J_k$. Then

$$\begin{aligned} w_j^{k+1} &= w_j^k + \tau_k t_k \tilde{d}_j^k \\ &= w_j^k - \tau_k t_k \min\{w_j^k, g_j^k\} \\ &\geq \begin{cases} w_j^k \geq 0 & \text{if } \min\{w_j^k, g_j^k\} \leq 0, \\ w_j^k - w_j^k = 0 & \text{if } \min\{w_j^k, g_j^k\} > 0, \end{cases} \end{aligned}$$

where we used the inequality $\tau_k t_k \leq 1$. This completes the proof. $\square$

Since the matrix of the linear system (3) is positive definite as long as w^k is not a global solution of (2), we can now apply an induction argument and summarize the previous results into the following

Theorem 3.6 *Algorithm 3.1 is well-defined and generates a sequence* $\{w^k\} = \{(x^k, y^k, z^k)\}$ *with* $z^k \geq 0$ *for all* k.

4 Global Convergence

In this section, we show that any accumulation point of a sequence generated by Algorithm 3.1 is a stationary point of our constrained reformulation (2). We note that this result holds without making any further assumptions apart from the standard differentiability conditions stated in the beginning of Section 1.

We start our global convergence analysis with a simple perturbation result.

Lemma 4.1 *Let* $\{A_k\} \subseteq \mathbb{R}^{t \times t}$ *be any sequence of symmetric matrices converging to a symmetric and positive definite matrix* $A_* \in \mathbb{R}^{t \times t}$. *Assume further that* $\{b^k\} \subseteq \mathbb{R}^t$ *is a sequence converging to a vector* $b^* \in \mathbb{R}^t$. *Let* $d^* \in \mathbb{R}^t$ *be the unique solution of the linear system* $A_* d = b^*$. *Then the linear systems* $A_k d = b^k$ *also have a unique solution* $d^k \in \mathbb{R}^t$ *for all* k *sufficiently large, and* $d^k \to d^*$.

Proof. Since A_* is symmetric positive definite and $A_k \to A_*$, it is easy to see that the symmetric matrices A_k are also positive definite for all k sufficiently large. Hence the linear systems $A_k d = b^k$ are uniquely solvable for those k. Moreover, it follows from a well-known perturbation result (see, e.g., [2, Theorem 3.1.4]) that $A_k^{-1} \to A_*^{-1}$. Hence we obtain for k sufficiently large:

$$\begin{aligned} \|d^k - d^*\| &= \|A_k^{-1} b^k - A_*^{-1} b^*\| \\ &\leq \|A_k^{-1} b^k - A_*^{-1} b^k\| + \|A_*^{-1} b^k - A_*^{-1} b^*\| \\ &\leq \|A_k^{-1} - A_*^{-1}\| \, \|b^k\| + \|A_*^{-1}\| \|b^k - b^*\| \\ &\to 0, \end{aligned}$$

as desired. $\square$

Lemma 4.1 allows us to prove the following global convergence result.

Theorem 4.2 *Let* $\{w^k\} \subseteq \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ *be a sequence generated by Algorithm 3.1. Then every accumulation point of* $\{w^k\}$ *is a stationary point of (2).*

Proof. We first note that $\{\Psi(w^k)\}$ is a strictly decreasing sequence which is bounded from below by zero. Hence $\{\Psi(w^k)\}$ converges to an element $\Psi^* \geq 0$. If $\Psi^* = 0$, then every accumulation point of $\{w^k\}$ is a global minimizer and hence a KKT point of (2). So assume $\Psi^* > 0$.

This implies that we eventually take only safe steps in Algorithm 3.1 since otherwise we would have

$$\Psi(w^{k+1}) \leq \gamma \Psi(w^k)$$

for infinitely many k which would imply $\Psi^* = 0$ (recall that $\gamma \in (0,1)$). Therefore, we can assume without loss of generality that all steps are safe steps.

Let w^* denote an accumulation point of $\{w^k\}$, and let $\{w^k\}_{K_1}$ be a subsequence converging to w^*. Since the index set $\mathcal{J}$ is finite, we can find a subsequence $\{w^k\}_{K_2}, K_2 \subseteq K_1$, such that the index set J_k remains unchanged, i.e., $J_k = J$ for all $k \in K_2$ and for a fixed set $J \subseteq \mathcal{J}$.

In view of the decreasing property of $\{\Psi(w^k)\}$ and since $\Psi^* > 0$, we have

$$\delta_k = \min\{\delta, c\sqrt{\|\Phi(w^k)\|}\} \geq \min\{\delta, c\sqrt{\|\Phi(w^*)\|}\} =: \delta_* > 0.$$

Since the sequence $\{g_J^k\}_{K_2}$ converges to $g_J^* := [\nabla\Psi(w^*)]_J$, the sequence of subvectors $\{\tilde{d}_J^k\}_{K_2}$ converges to $\tilde{d}_J^* := -\min\{w_J^*, g_J^*\}$. From the upper semicontinuity of the generalized Jacobian (see [1, Proposition 2.6.2 (c)]), it also follows that the sequence $\{H_k\}_{K_2}$ remains bounded. So, again, we can take a subsequence $\{H_k\}_{K_3}, K_3 \subseteq K_2$, such that $\{H_k\}_{K_3} \to H_*$ for a matrix H_* which must belong to $\partial\Phi(w^*)$ since the generalized Jacobian is a closed mapping (see [1, Proposition 2.6.2 (b)]). Let $\bar{J} := \mathcal{I} \cup \mathcal{P} \cup (\mathcal{J} \setminus J)$ and let $d_{\bar{J}}^*$ be the unique solution of the nonsingular linear system

$$[(H_{\cdot\bar{J}}^*)^T H_{\cdot\bar{J}}^* + \rho(\Psi(w^*))I]\, d_{\bar{J}} = -[\nabla\Psi(w^*)]_{\bar{J}}. \tag{9}$$

Then it follows immediately from Lemma 4.1 that $d_{\bar{J}}^k \to d_{\bar{J}}^*$. Therefore, we have

$$\{\tilde{d}^k\}_{K_3} = \left\{ \begin{pmatrix} \tilde{d}_J^k \\ d_{\bar{J}}^k \end{pmatrix} \right\}_{K_3} \to \begin{pmatrix} \tilde{d}_J^* \\ d_{\bar{J}}^* \end{pmatrix} =: \tilde{d}^*.$$

Since $\{\Psi(w^k)\}$ converges to Ψ^*, we have

$$\lim_{k \in K_3} (\Psi(w^{k+1}) - \Psi(w^k)) = 0$$

and therefore by our line search rule

$$\lim_{k \in K_3} \tau_k t_k^2 \Psi(w^k) = 0. \tag{10}$$

We now show that $\{\tau_k\}_{K_3}$ is bounded from below by some positive number. From the boundedness of $\{\tilde{d}^k\}_{K_3}$, say $\|\tilde{d}^k\| \leq \kappa_1$ for a constant $\kappa_1 > 0$ and all $k \in K_3$, we obtain for all $j \in \mathcal{J} \setminus J$, all $k \in K_3$ and all $\tau \geq 0$:

$$w_j^k + \tau d_j^k \geq \delta_k - \tau\|\tilde{d}^k\| \geq \delta_* - \tau\kappa_1,$$

where the first inequality is an immediate consequence of the definition of the index set J. Hence $\bar{\tau}_k$ in Step (S.4) satisfies the inequality

$$\bar{\tau}_k \geq \delta_*/\kappa_1,$$

so that

$$\tau_k \geq \min\{1, \delta_*/\kappa_1\} > 0.$$

This together with $\Psi^* > 0$ and (10) implies

$$\{t_k\}_{K_3} \to 0.$$

Let ℓ_k be the unique exponent such that $t_k = \beta^{\ell_k}$ in Step (S.5) of Algorithm 3.1. Then $\{\ell_k\}_{K_3} \to \infty$ and

$$\frac{\Psi(w^k + \tau_k \beta^{\ell_k - 1} \tilde{d}^k) - \Psi(w^k)}{\tau_k \beta^{\ell_k - 1}} > -\sigma \beta^{\ell_k - 1} \Psi(w^k)$$

for all $k \in K_3$. Taking the limit $k \to_{K_3} \infty$, we thus obtain

$$\nabla \Psi(w^*)^T \tilde{d}^* \geq 0.$$

Since $\tilde{d}^* = (\tilde{d}^*_J, d^*_{\bar{J}})$, where $d^*_{\bar{J}}$ is the solution of the linear system (9) and $\tilde{d}^*_J = -\min\{z^*_J, g^*_J\}$, we obtain from Lemma 3.3 that w^* is a stationary point of (2). $\square$

Theorem 4.2 provides subsequential convergence to stationary points of (2) only. Fortunately, Proposition 2.6 gives a relatively mild condition for such a stationary point to be a KKT point of VIP(X, F) and hence, quite often, to be a solution of VIP(X, F) itself, see the corresponding discussion in Section 1.

5 Local Convergence

In this section, we want to show that Algorithm 3.1 is locally superlinearly/quadratically convergent under Robinson's [32] strong regularity condition. In particular, we do not assume strict complementarity in order to verify local fast convergence.

The proof of this local convergence result is given in a step-by-step way and based on a sequence of lemmas. Basically, our first aim is to prove that the entire sequence $\{w^k\}$ converges to a strongly regular KKT point w^* whenever this KKT point is an accumulation point of the sequence $\{w^k\}$, see Lemma 5.5 below.

We then show that eventually only fast steps will be taken in Step (S.5) of Algorithm 3.1. This result, in turn, is based on the fact that, for all w^k sufficiently close to w^*, we can prove that $\tau_k = 1$ and that J_k identifies the set of active constraints J_*. Using these facts, it is quite easy to summarize all lemmas into our main local convergence result, Theorem 5.10 below.

We begin our local analysis with a relatively simple consequence of strong regularity.

Lemma 5.1 *Assume that $w^* = (x^*, y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ is a strongly regular KKT point of VIP(X, F). Then there is a constant $c_3 > 0$ such that the matrices $(H^k_{.\bar{J}_k})^T H^k_{.\bar{J}_k}$ are nonsingular and*

$$\| \left((H^k_{.\bar{J}_k})^T H^k_{.\bar{J}_k} \right)^{-1} \| \leq c_3$$

for all $w^k = (x^k, y^k, z^k) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ in a sufficiently small ball around w^*.

Proof. First note that the matrices $H^k_{.\bar{J}_k}$ have full column rank for all w^k sufficiently close to w^* since the matrices H_k themselves are nonsingular by Proposition 2.4 (a). Therefore, the matrices $(H^k_{.\bar{J}_k})^T H^k_{.\bar{J}_k}$ are nonsingular.

Hence, if the claim is not true, there is a sequence $\{w^k\} \subseteq \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ converging to w^* and a corresponding sequence $\{H_k\}$ with $H_k \in \partial\Phi(w^k)$ such that

$$\| \left((H^k_{.\bar{J}_k})^T H^k_{.\bar{J}_k} \right)^{-1} \| \to \infty \tag{11}$$

From the upper semicontinuity of the generalized Jacobian, it follows that $\{H_k\}$ remains bounded. By passing to a subsequence, we may assume that $\{H_k\}$ converges to a matrix H_* which must belong to $\partial\Phi(w^*)$ since the generalized Jacobian is a closed mapping. Since $\mathcal{J}$ is finite, we can find a subsequence $\{H_k\}_K$ such that $J_k = J$ for all $k \in K$ and a fixed set $J \subseteq \mathcal{J}$. Denote $H_* = (H^*_{.J}, H^*_{.\bar{J}})$. Then we have $H^k_{.\bar{J}} \to H^*_{.\bar{J}}$ and therefore

$$\|(H^k_{.\bar{J}})^T H^k_{.\bar{J}} - (H^*_{.\bar{J}})^T H^*_{.\bar{J}}\| \to 0.$$

By (11), $(H^*_{.\bar{J}})^T H^*_{.\bar{J}}$ must be singular. On the other hand, however, reasoning as in the first paragraph of this proof, we see that the matrix $(H^*_{.\bar{J}})^T H^*_{.\bar{J}}$ is nonsingular. This contradiction completes the proof. □

An application of Lemma 5.1 is given in the following

Lemma 5.2 *Assume that $w^* = (x^*, y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ is a strongly regular KKT point of VIP(X, F). Then there is a constant $c_4 > 0$ such that*

$$\|\tilde{d}^k\| \le c_4 \|\Phi(w^k)\|$$

for all $w^k = (x^k, y^k, z^k) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ sufficiently close to w^* with $z^k \ge 0$, where $\tilde{d}^k$ denotes the vector computed in Step (S.4) of Algorithm 3.1.

Proof. In view of the upper semicontinuity of the generalized Jacobian, there exists a constant $\kappa_2 > 0$ such that

$$\|\tilde{d}^k_{J_k}\| \le \|g^k_{J_k}\| \le \|g^k\| = \|H_k^T \Phi(w^k)\| \le \|H_k\| \, \|\Phi(w^k)\| \le \kappa_2 \|\Phi(w^k)\| \tag{12}$$

for all w^k in a sufficiently small ball around w^* (note that we also applied Proposition 2.5 and the definition of $\tilde{d}^k_{J_k}$ in the above chain of inequalities). Since

$$\left[(H^k_{.\bar{J}_k})^T H^k_{.\bar{J}_k} + \rho(\Psi(w^k)) I \right] d^k_{\bar{J}_k} + \left[\nabla\Psi(w^k) \right]_{\bar{J}_k} = 0$$

by (3), we further obtain from Proposition 2.5:

$$\begin{array}{rcl} 0 & = & (d^k_{\bar{J}_k})^T \left[H_k^T \Phi(w^k) \right]_{\bar{J}_k} + (d^k_{\bar{J}_k})^T \left[(H^k_{.\bar{J}_k})^T H^k_{.\bar{J}_k} + \rho(\Psi(w^k)) I \right] d^k_{\bar{J}_k} \\ & \ge & \|H^k_{.\bar{J}_k} d^k_{\bar{J}_k}\|^2 + (d^k_{\bar{J}_k})^T (H^k_{.\bar{J}_k})^T \Phi(w^k) \\ & \ge & \|H^k_{.\bar{J}_k} d^k_{\bar{J}_k}\|^2 - \|H^k_{.\bar{J}_k} d^k_{\bar{J}_k}\| \, \|\Phi(w^k)\|. \end{array}$$

Hence

$$\|H^k_{\cdot\bar{J}_k} d^k_{\bar{J}_k}\| \le \|\Phi(w^k)\|$$

for all w^k sufficiently close to w^*. Using Lemma 5.1, we obtain

$$d^k_{\bar{J}_k} = \left((H^k_{\cdot\bar{J}_k})^T H^k_{\cdot\bar{J}_k}\right)^{-1}\left((H^k_{\cdot\bar{J}_k})^T H^k_{\cdot\bar{J}_k} d^k_{\bar{J}_k}\right)$$

and therefore

$$\begin{aligned} \|d^k_{\bar{J}_k}\| &\le \|\left((H^k_{\cdot\bar{J}_k})^T H^k_{\cdot\bar{J}_k}\right)^{-1}\| \, \|H^k_{\cdot\bar{J}_k}\| \, \|H^k_{\cdot\bar{J}_k} d^k_{\bar{J}_k}\| \\ &\le c_3\|H_k\| \, \|H^k_{\cdot\bar{J}_k} d^k_{\bar{J}_k}\| \\ &\le c_3\kappa_2\|\Phi(w^k)\| \end{aligned} \tag{13}$$

for all w^k close enough to w^*. From (12) and (13), we obtain

$$\|\tilde{d}^k\| \le \|\tilde{d}^k_{J_k}\| + \|d^k_{\bar{J}_k}\| \le c_4\|\Phi(w^k)\|$$

with $c_4 := \kappa_2(1 + c_3)$. □

Before stating the next result, we recall that

$$J_* = \{j \in \mathcal{J} \,|\, w^*_j = 0\}$$

denotes the set of active constraints for the reformulation (2) of the KKT system. The following result gives a relation between J_* and its approximation J_k. Later, in Lemma 5.7, we will prove a stronger relationship which, however, is also based on a stronger assumption.

Lemma 5.3 *Let $w^* = (x^*, y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ be a KKT point of VIP(X, F). Then $J_k \subseteq J_*$ for all $w^k = (x^k, y^k, z^k) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ sufficiently close to w^* with $z^k \ge 0$.*

Proof. Define

$$\nu := \min\{w^*_j \,|\, j \notin J_*\} > 0,$$

and let w^k be sufficiently close to w^* such that

$$\|w^k - w^*\| \le \nu/4$$

and

$$c\sqrt{\|\Phi(w^k)\|} \le \nu/4.$$

Choose $j \in J_k$. Then

$$|w^k_j - w^*_j| \le \|w^k - w^*\| \le \nu/4$$

and

$$w^k_j \le \delta_k \le c\sqrt{\|\Phi(w^k)\|} \le \nu/4.$$

Hence

$$w^*_j \le \nu/4 + w^k_j \le \nu/2$$

so that $j \in J_*$. □

In order to prove convergence of the entire sequence $\{w^k\}$, we will apply the following result by Moré and Sorensen [23].

Proposition 5.4 *Assume that $w^* \in \mathbb{R}^t$ is an isolated accumulation point of a sequence $\{w^k\} \subseteq \mathbb{R}^t$ such that, for every subsequence $\{w^k\}_K$ converging to w^*, there is an infinite subset $\tilde{K} \subseteq K$ such that $\{\|w^{k+1} - w^k\|\}_{\tilde{K}} \to 0$. Then the whole sequence $\{w^k\}$ converges to w^*.*

We stress that Proposition 5.4 is slightly different from the original version given in [23]. The difference is that Moré and Sorensen assume that $\{\|w^{k+1} - w^k\|\}_K \to 0$ on the whole subset K, whereas we assume that this limit holds only on an infinite subsubset $\tilde{K} \subseteq K$. It is easy to see, however, that the proof given by Moré and Sorensen in [23] actually shows that the slightly stronger result given in Proposition 5.4 holds.

The following sequential convergence result is an application of Proposition 5.4.

Lemma 5.5 *Assume that $w^* = (x^*, y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ is a strongly regular KKT point of VIP(X, F), and assume that w^* is an accumulation point of a sequence $\{w^k\}$ generated by Algorithm 3.1. Then the entire sequence $\{w^k\}$ converges to w^*.*

Proof. Since w^* is a strongly regular KKT point, w^* is a locally isolated KKT point of VIP(X, F). Since $\{\Psi(w^k)\}$ decreases monotonically and $\{w^k\} \to w^*$ on a subsequence, we have $\{\Psi(w^k)\} \to \Psi(w^*) = 0$ on the whole sequence. Hence every accumulation point of $\{w^k\}$ is a global minimum of Ψ and therefore a KKT point of VIP(X, F). The isolatedness of the KKT point w^* therefore implies that w^* is necessarily an isolated accumulation point of $\{w^k\}$.

Now let $\{w^k\}_K$ be a subsequence converging to w^*. Then, by continuity, we have

$$\{\|\Phi(w^k)\|\}_K \to \|\Phi(w^*)\| = 0. \tag{14}$$

We now consider two cases.
Case 1: There is an infinite subset $\tilde{K} \subseteq K$ such that Algorithm 3.1 takes safe steps for all $k \in \tilde{K}$. Then we obtain from (14) and Lemma 5.2 that

$$\{\|\tilde{d}^k\|\}_{\tilde{K}} \to 0.$$

This implies

$$\{\|w^{k+1} - w^k\|\}_{\tilde{K}} \to 0$$

since $\tau_k t_k \leq 1$.
Case 2: There are only finitely many $k \in K$ for which Algorithm 3.1 takes a safe step. Then we can assume without loss of generality that all steps $k \in K$ are fast steps. Then we still have

$$\|d^k_{\bar{J}_k}\| \leq \|\tilde{d}^k\| \leq c_4 \|\Phi(w^k)\|$$

for all $k \in K$ by the very definition of $d^k_{\bar{J}_k}, \tilde{d}^k$ and Lemma 5.2, so that at least

$$\{\|d^k_{\bar{J}_k}\|\}_K \to 0.$$

If $j \in J_k$, we have $j \in J_*$ by Lemma 5.3 so that also

$$|d^k_j| = |w^k_j| \to w^*_j = 0,$$

where the limit is taken on the subset K. Hence we also have

$$\{\|d^k\|\}_K \to 0$$

which again implies

$$\{\|w^{k+1} - w^k\|\}_K \to 0.$$

From these two cases, we obtain

$$\{\|w^{k+1} - w^k\|\}_{\tilde{K}} \to 0,$$

where $\tilde{K}$ denotes the infinite subset of K defined in Case 1. Hence it follows from Proposition 5.4 that the entire sequence $\{w^k\}$ converges to w^*. □

We next show that we eventually have $\tau_k = 1$ in Step (S.4) of Algorithm 3.1.

Lemma 5.6 *Assume that $w^* = (x^*, y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ is a strongly regular KKT point of VIP(X, F), and let $\{w^k\}$ denote a sequence generated by Algorithm 3.1 which converges to w^*. Then $\tau_k = 1$ for all k sufficiently large.*

Proof. Using Lemma 5.2 and the definition of the vector $\tilde{d}^k$, we have

$$\|d^k_{\bar{J}_k}\| \leq \|\tilde{d}^k\| \leq c_4\|\Phi(w^k)\|$$

for all k sufficiently large. Hence we obtain for $j \in \mathcal{J} \setminus J_k$:

$$\begin{aligned} w_j^k + d_j^k &\geq \delta_k + d_j^k \\ &\geq \delta_k - c_4\|\Phi(w^k)\| \\ &= \min\{\delta, c\sqrt{\|\Phi(w^k)\|}\} - c_4\|\Phi(w^k)\|. \end{aligned}$$

Since $\{w^k\}$ converges to w^*, we have $\{\|\Phi(w^k)\|\} \to 0$ and therefore

$$\begin{aligned} w_j^k + d_j^k &\geq c\sqrt{\|\Phi(w^k)\|} - c_4\|\Phi(w^k)\| \\ &= \left(c/\sqrt{\|\Phi(w^k)\|} - c_4\right)\|\Phi(w^k)\| \\ &\geq 0 \end{aligned}$$

for all k sufficiently large. Thus $\bar{\tau}_k \geq 1$ and hence $\tau_k = 1$ for all w^k sufficiently close to w^*. □

The following lemma shows that the approximation J_k is eventually equal to the set of active constraints J_* under the strong regularity condition.

Lemma 5.7 *Assume that $w^* = (x^*, y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ is a strongly regular KKT point of VIP(X, F), and let $\{w^k\}$ denote a sequence generated by Algorithm 3.1 which converges to w^*. Then $J_k = J_*$ for all k sufficiently large.*

Proof. From Lemma 5.3, we already know that $J_k \subseteq J_*$ for all k sufficiently large. Conversely, assume that $j \in J_*$. Then $w_j^* = 0$, and we obtain from Proposition 2.4 (b) and all $w^k = (x^k, y^k, z^k)$ sufficiently close to w^*:

$$|w_j^k| = |w_j^k - w_j^*| \leq \|w^k - w^*\| \leq \|\Phi(w^k)\|/c_2.$$

Since $\{\|\Phi(w^k)\|\} \to 0$ and $\delta_k = O(\sqrt{\|\Phi(w^k)\|})$, we have

$$\|\Phi(w^k)\|/c_2 \leq \delta_k$$

for all k sufficiently large. Combining the last two inequalities, we obtain

$$w_j^k = |w_j^k| \leq \|\Phi(w^k)\|/c_2 \leq \delta_k,$$

so that $j \in J_k$. Hence $J_k = J_*$ for all k large enough. □

The following result has already been shown by Facchinei and Soares [8] under the additional assumption that the mapping G is semismooth. Note that we do not need this semismoothness assumption here, and that our proof is considerably simpler than the one given in [8].

Proposition 5.8 *Let $G : \mathbb{R}^t \to \mathbb{R}^t$ be locally Lipschitz continuous, $w^* \in \mathbb{R}^t$ with $G(w^*) = 0$ such that all elements in $\partial G(w^*)$ are nonsingular, and assume that there are two sequences $\{w^k\} \subseteq \mathbb{R}^t$ and $\{d^k\} \subseteq \mathbb{R}^t$ with $\{w^k\} \to w^*$ and $\|w^k + d^k - w^*\| = o(\|w^k - w^*\|)$. Then $\|G(w^k + d^k)\| = o(\|G(w^k)\|)$.*

Proof. Let $L > 0$ denote the local Lipschitz constant of G around w^*. In view of our nonsingularity assumption, we can apply Clarke's [1] inverse function theorem and conclude that, in a sufficiently small neighbourhood around $G(w^*)$, the inverse function G^{-1} exists and is also locally Lipschitz continuous. Hence, for all k sufficiently large, we have

$$\begin{aligned}
\|G(w^k + d^k)\| &= \|G(w^k + d^k) - G(w^*)\| \\
&\leq L\|w^k + d^k - w^*\| \\
&= o(\|w^k - w^*\|) \\
&= o(\|G^{-1}(G(w^k)) - G^{-1}(G(w^*))\|) \\
&\leq o(\|G(w^k) - G(w^*)\|) \\
&= o(\|G(w^k)\|).
\end{aligned}$$

This is precisely the statement of Proposition 5.8. □

Proposition 5.8 is the main ingredient in order to show that, eventually, we will only take fast steps in Algorithm 3.1.

The next lemma basically shows that the main assumption used in Proposition 5.8 is satisfied by the fast search direction d^k taken in Algorithm 3.1.

Lemma 5.9 *Assume that $w^* = (x^*, y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ is a strongly regular KKT point of VIP(X, F), and let $\{w^k\}$ denote any sequence (not necessarily generated by Algorithm 3.1) converging to w^*. Then*

$$\|w^k + d^k - w^*\| = o(\|w^k - w^*\|),$$

where d^k denotes the fast search direction computed in Step (S.4) of Algorithm 3.1. Moreover, if F is an LC^1 function, h, g are LC^2 functions and $\rho(\Psi(w^k)) = O(\sqrt{\Psi(w^k)})$, then

$$\|w^k + d^k - w^*\| = O(\|w^k - w^*\|^2).$$

Proof. First note that $J_k = J_*$ for all k sufficiently large by Lemma 5.7. In the following, we always assume that k is large enough so that this equality holds.

Then, for $j \in J_*$, we have

$$|w_j^k + d_j^k - w_j^*| = |w_j^k + d_j^k| = |w_j^k - w_j^k| = 0 = o(\|w^k - w^*\|). \tag{15}$$

Now consider indices which do not belong to J_*. From the linear system (3), we obtain

$$\begin{aligned}
& \left[(H^k_{.\bar{J}_k})^T H^k_{.\bar{J}_k} + \rho(\Psi(w^k))I\right](w^k_{\bar{J}_k} + d^k_{\bar{J}_k} - w^*_{\bar{J}_k}) \\
= \; & \left[(H^k_{.\bar{J}_k})^T H^k_{.\bar{J}_k} + \rho(\Psi(w^k))I\right] d^k_{\bar{J}_k} + \left[(H^k_{.\bar{J}_k})^T H^k_{.\bar{J}_k} + \rho(\Psi(w^k))I\right](w^k_{\bar{J}_k} - w^*_{\bar{J}_k}) \\
= \; & -g^k_{\bar{J}_k} + (H^k_{.\bar{J}_k})^T H^k_{.\bar{J}_k}(w^k_{\bar{J}_k} - w^*_{\bar{J}_k}) + \rho(\Psi(w^k))(w^k_{\bar{J}_k} - w^*_{\bar{J}_k}) \\
= \; & -\left[H_k^T \Phi(w^k)\right]_{\bar{J}_*} + (H^k_{.\bar{J}_*})^T H^k_{.\bar{J}_*}(w^k_{\bar{J}_*} - w^*_{\bar{J}_*}) + \rho(\Psi(w^k))(w^k_{\bar{J}_*} - w^*_{\bar{J}_*}) \\
= \; & -(H^k_{.\bar{J}_*})^T(\Phi(w^k) - \Phi(w^*)) + (H^k_{.\bar{J}_*})^T H^k_{.\bar{J}_*}(w^k_{\bar{J}_*} - w^*_{\bar{J}_*}) + \rho(\Psi(w^k))(w^k_{\bar{J}_*} - w^*_{\bar{J}_*}) \\
= \; & -(H^k_{.\bar{J}_*})^T\left(\Phi(w^k) - \Phi(w^*) - H^k_{.\bar{J}_*}(w^k_{\bar{J}_*} - w^*_{\bar{J}_*})\right) + \rho(\Psi(w^k))(w^k_{\bar{J}_*} - w^*_{\bar{J}_*}).
\end{aligned}$$

In view of Lemma 5.6, there exists an integer k_0 such that $\tau_k = 1$ for all $k \geq k_0$. We therefore obtain for all $j \in J_*$ and all $k \geq k_0$:

$$w_j^{k+1} = w_j^k + \tau_k d_j^k = w_j^k - w_j^k = 0.$$

By definition of J_*, we also have $w_j^* = 0$. Hence it follows that

$$H^k_{.\bar{J}_*}(w^k_{\bar{J}_*} - w^*_{\bar{J}_*}) = H_k(w^k - w^*)$$

for all $k > k_0$. Since $\rho(\Psi(w^k)) \to 0$, we obtain from Lemma 5.1 that

$$\|\left((H^k_{.\bar{J}_k})^T H^k_{.\bar{J}_k} + \rho(\Psi(w^k))I\right)^{-1}\| \leq 2c_3$$

for all k sufficiently large. The last displayed formulas together yield

$$\begin{aligned}
\|w^k_{\bar{J}_k} + d^k_{\bar{J}_k} - w^*_{\bar{J}_k}\| & \leq 2c_3\|H^k_{.\bar{J}_*}\|\,\|\Phi(w^k) - \Phi(w^*) - H_k(w^k - w^*)\| + \\
& \quad\; \rho(\Psi(w^k))\|w^k - w^*\| \\
& = o(\|w^k - w^*\|),
\end{aligned}$$

where the equality follows from Proposition 2.2.

The second part can be shown similarly by noting that

$$\|\Phi(w^k)\| = \|\Phi(w^k) - \Phi(w^*)\| = O(\|w^k - w^*\|)$$

in view of the local Lipschitz continuity of Φ, so that $\rho(\Psi(w^k)) = O(\sqrt{\Psi(w^k)}) = O(\|\Phi(w^k)\|)$ implies $\rho(\Psi(w^k)) = O(\|w^k - w^*\|)$. $\square$

We are now in the position to state the main local convergence result for Algorithm 3.1.

Theorem 5.10 *Assume that $w^* = (x^*, y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ is a strongly regular KKT point of VIP(X, F), and suppose that w^* is an accumulation point of a sequence $\{w^k\}$ generated by Algorithm 3.1. Then the following statements hold:*

(a) The whole sequence $\{w^k\}$ converges to w^*.

(b) Eventually, the algorithm takes only fast steps.

(c) The rate of convergence is Q-superlinear.

(d) The rate of convergence is Q-quadratic if, in addition, the assumptions of the second part of Lemma 5.9 are satisfied.

Proof. Statement (a) follows immediately from Lemma 5.5. Hence Lemma 5.9 together with Proposition 5.8 being applied to $G = \Phi$ shows that

$$\|\Phi(w^k + d^k)\|/\|\Phi(w^k)\| \to 0.$$

Since $\tau_k = 1$ for all k sufficiently large by Lemma 5.6, this shows that the descent test (4) in Algorithm 3.1 is eventually satisfied so that only fast steps will be taken for all k sufficiently large, i.e., statement (b) holds.

The remaining two statements on the local rate of convergence are therefore immediate consequences of Lemma 5.9. $\square$

6 Preliminary Numerical Results

We implemented Algorithm 3.1 in MATLAB and tested it on a SUN Sparc 20 station. As test problems, we took a couple of variational inequality problems from the literature. A brief description of them is given in the following.

Example 6.1 Here $n = 5$ and $F : \mathbb{R}^5 \to \mathbb{R}^5$ is a nonlinear function whose precise definition can be found in Table 1 of [34]. The feasible set for this example is given by

$$X := \{x \in \mathbb{R}^5 | \sum_{i=1}^{5} x_i \geq 10, x \geq 0\}.$$

Example 6.2 Also in this example the dimension is $n = 5$, and F is again a nonlinear function. The details can be found in Table 6 of [34]. The feasible set in this example is polyhedral and given by

$$X := \{x \in \mathbb{R}^5 \,|\, Ax \leq b, x \geq 0\}$$

for a constant matrix $A \in \mathbb{R}^{4\times 5}$ and a constant vector $b \in \mathbb{R}^4$ whose entries are also specified in Table 6 of [34].

Example 6.3 Here $n = 4$ and F is a nonlinear function which is defined precisely in Table 1 of [22]. The feasible set for this example is given by

$$X := \{x \in \mathbb{R}^4 \,|\, \sum_{i=1}^{4} x_i = 1, x \geq 0\}.$$

Example 6.4 Also this example is taken from [22]; the dimension is $n = 5$ and F is a nonlinear function described in Table 4 of that paper. The feasible set for this example is defined by

$$X := \{x \in \mathbb{R}^5 \,|\, \sum_{i=1}^{5} x_i = 1, x \geq 0\}.$$

Example 6.5 This is the first example used in the paper by Fukushima [13]. F is a nonlinear function of dimension $n = 3$ and the feasible set is, in contrast to all previous examples, nonlinear; more precisely, it is given by

$$X := \{x \in \mathbb{R}^3 \,|\, 1 - x_1^2 - 0.4x_2^2 - 0.6x_3^2 \geq 0\}.$$

Example 6.6 This example is also taken from [13]. The dimension of this problem is $n = 5$, the function F is linear, and the feasible set is described by

$$X := \{x \in \mathbb{R}^5 \,|\, x_1 + x_2 + x_3 = 210, x_4 + x_5 = 120, x \geq 0\}.$$

Example 6.7 Our last example is a variational inequality reformulation of the convex optimization problem 35 from the test problem collection [16] by Hock and Schittkowski. Its dimension is $n = 3$, and the feasible set is given by

$$X := \{x \in \mathbb{R}^3 \,|\, 3 - x_1 - x_2 - 2x_3 \geq 0, x \geq 0\}.$$

We run our algorithm on these examples using the termination criterion

$$\Psi(w^k) \leq 10^{-12}$$

and the parameter setting

$$\sigma = 10^{-4}, \;\; \beta = 0.5, \;\; \gamma = 0.9, \;\; c = 1, \;\; \delta = 1.$$

The forcing function used in our implementation is

$$\rho(\Psi(w)) = \min\{10^{-6}, \sqrt{\Psi(w)}\},$$

so that the regularization is quite small even in the first few iterations.

The element $H \in \partial\Phi(w)$ was chosen as described in Proposition 2.1 with

$$a_j(w) = -1 \quad \text{and} \quad b_j(w) = 0$$

if $(g_j(x), z_j) = (0, 0)$ (more precisely, if $\sqrt{g_j(x)^2 + z_j^2} \le 10^{-8}$). It is easy to see that this gives an element from the generalized Jacobian $\partial\Phi(w)$.

Finally, we replaced the monotone Armijo-condition (5) by a nonmonotone line search, see Grippo et al. [14]. This means that we replaced the line search rule (5) by the condition

$$\Psi(w^k + \tau_k t_k \tilde{d}^k) \le \mathcal{R}_k - \sigma \tau_k t_k^2 \Psi(w^k),$$

where the reference value $\mathcal{R}_k$ is given by

$$\mathcal{R}_k := \max_{j=k-l_k,\ldots,k} \Psi(w^j)$$

and l_k is a nonnegative integer which is updated at each iteration by the following rules: If the safe search direction $\tilde{d}^k$ satisfies the angle condition

$$-\nabla\Psi(w^k)^T \tilde{d}^k \ge 10^{-6} \|\nabla\Psi(w^k)\| \, \|\tilde{d}^k\|,$$

then

$$l_k := \min\{l_k + 1, 10\},$$

otherwise

$$l_k := 0;$$

note that the latter case corresponds to the standard Armijo-rule (5).

Our numerical results are summarized in Table 6.1, where we present the following data:

Example:	number of test example
x^0:	x-part of the starting vector $w^0 = (x^0, y^0, z^0)$
it:	number of iterations needed until termination
Φ-eval.:	number of Φ-evaluations needed until termination
$\Psi(w^f)$:	value of $\Psi(\cdot)$ at the final iterate w^f
safe/fast:	number of safe and fast steps taken during the iteration
$J_k = J_*$:	iteration from which on the active set was identified correctly.

Apart from the different choices of the starting vector x^0, we set all components of the initial Lagrange-multipliers y^0 and z^0 to one. If available, we took for x^0 the starting point(s) from the literature.

The results in Table 6.1 indicate that Algorithm 3.1 is quite promising. In general, the number of iterations and function evaluations is very small. Moreover, the second to last column in Table 6.1 clearly shows that the vast majority of steps taken are fast steps.

The last column shows that the set of active constraints is identified correctly at least two or three iterations before termination. We also observed that, usually, $\tau_k = 1$ during these iterations, and sometimes this value has been accepted much earlier.

Table 6.1: Numerical results for Algorithm 3.1.

Example	x^0	it	Φ-eval.	$\Psi(w^f)$	safe/fast	$J_k = J_*$
6.1	(25,0,0,0,0)	11	15	1.4e-18	2/9	$k \geq 9$
	(10,0,10,0,10)	7	11	6.7e-26	1/6	$k \geq 5$
	(10,0,0,0,0)	10	14	9.7e-15	2/8	$k \geq 8$
	(0,2.5,2.5,2.5,2.5)	9	11	5.5e-15	1/8	$k \geq 7$
6.2	(0,0,100,0,0)	36	49	2.3e-16	7/29	$k \geq 32$
	(100,0,0,0,0)	81	96	1.0e-18	10/71	$k \geq 76$
	(1,2,3,4,5)	18	23	1.4e-17	4/14	$k \geq 13$
	(0,0,0,0,0)	16	21	1.8e-18	3/13	$k \geq 11$
6.3	(1,0,0,0)	7	8	3.7e-15	0/7	$k \geq 1$
	(0,1,0,0)	6	7	1.3e-13	0/6	$k \geq 1$
	(0,0,0,0)	7	9	1.3e-13	1/6	$k \geq 3$
	(1,1,1,1)	5	6	2.5e-13	0/5	$k \geq 1$
6.4	(1,0,0,0,0)	9	10	2.5e-15	0/9	$k \geq 1$
	(0,1,0,0,0)	9	10	2.8e-15	0/9	$k \geq 1$
	(0,0,0,0,0)	9	13	2.0e-13	2/7	$k \geq 3$
	(1,1,1,1,1)	9	10	3.5e-15	0/9	$k \geq 1$
6.5	(1,1,0)	17	21	6.0e-13	1/16	$k \geq 1$
	(4,3,2)	89	100	9.6e-13	10/79	$k \geq 1$
	(1,1,1)	39	45	7.2e-13	4/35	$k \geq 1$
	(1,2,3)	198	301	8.1e-13	95/103	$k \geq 1$
6.6	(70,70,70,60,60)	30	40	1.3e-13	8/22	$k \geq 21$
	(0,0,0,0,0)	18	23	5.9e-13	3/15	$k \geq 9$
	(1,1,1,1,1)	18	23	8.7e-13	3/15	$k \geq 9$
	(1,2,3,4,5)	15	18	5.8e-13	2/13	$k \geq 6$
6.7	(0.5,0.5,0.5)	8	12	1.0e-14	2/6	$k \geq 6$
	(0,0,0)	5	7	5.1e-13	1/4	$k \geq 3$
	(4,3,2)	8	11	5.7e-14	2/6	$k \geq 6$
	(1,2,3)	8	11	5.7e-14	2/6	$k \geq 6$

The only test example for which our algorithm had some problems is Example 6.5. For this example we observed that the descent test (4) was satisfied for the fast steps quite often without improving the function value $\Psi(w^k)$ considerably. It is also interesting to note that Example 6.5 is the only one which has a nonlinear feasible set. In fact, it is our feeling that solving a nonlinearly constrained variational inequality problem is a considerably more difficult task than solving a variational inequality on polyhedral sets. However, as far as we know, there is not much numerical experience for solving nonlinearly constrained variational inequalities in the literature.

7 Final Remarks

In this paper, we introduced a new constrained Newton-type method for the solution of the general variational inequality problem which has a number of desirable properties. The fact that this method has to solve only one linear system of equations at each iteration makes it also applicable to large-scale problems which often arise from the discretization of continuous problems.

Since complementarity and mixed complementarity problems can also be viewed as variational inequality problems, it would be possible to apply Algorithm 3.1 to these classes of problems. However, due to the special structure of these problems, it should be possible to avoid the introduction of any Lagrange multipliers.

In fact, we believe that it is relatively easy to adapt our new method in order to solve (mixed) complementarity problems in a similar way as this was done in [17, 18] for the QP-based method from [7]. In fact, such an algorithm seems to be quite promising since the mapping F involved in complementarity problems is often not defined everywhere. A suitable modification of our new method for complementarity problems would avoid this problem since it would guarantee that all iterates would stay feasible.

Acknowledgement: The authors would like to thank Liqun Qi, Paul Tseng and the two anonyous referees for some useful comments on the original version of this manuscript.

References

[1] F.H. Clarke: *Optimization and Nonsmooth Analysis.* John Wiley & Sons, New York, 1983. (Reprinted by SIAM, Philadelphia, PA, 1990.)

[2] J.E. Dennis, Jr., and R.B. Schnabel: *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Englewood Cliffs, 1983 (reprinted by SIAM, Philadelphia, 1996).

[3] F. Facchinei, A. Fischer and C. Kanzow: *Inexact Newton methods for semismooth equations with applications to variational inequality problems.* In: G. Di Pillo and F. Giannessi (eds.): *Nonlinear Optimization and Applications.* Plenum Press, New York, 1996, pp. 125–149.

[4] F. Facchinei, A. Fischer and C. Kanzow: *Regularity properties of a semismooth reformulation of variational inequalities.* SIAM Journal on Optimization, to appear.

[5] F. Facchinei, A. Fischer and C. Kanzow: *A semismooth Newton method for variational inequalities: The case of box constraints.* In: M.C. Ferris and J.-S. Pang (eds.): *Complementarity and Variational Problems: State of the Art.* SIAM, Philadelphia, 1997, pp. 76–90.

[6] F. Facchinei, A. Fischer and C. Kanzow: *On the accurate identification of active constraints.* SIAM Journal on Optimization, to appear.

[7] F. Facchinei, A. Fischer, C. Kanzow and J.-M. Peng: *A simply constrained optimization reformulation of KKT systems arising from variational inequalities.* Preprint 112, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany, October 1996 (revised October 1997).

[8] F. Facchinei and J. Soares: *A new merit function for nonlinear complementarity problems and a related algorithm.* SIAM Journal on Optimization 7, 1997, pp. 225–247.

[9] M.C. Ferris and J.-S. Pang: *Engineering and economic applications of complementarity problems.* SIAM Review, to appear.

[10] A. Fischer: *A special Newton-type optimization method.* Optimization 24, 1992, pp. 269–284.

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

[12] A. Friedlander, J.M. Martínez and S.A. Santos: *A new strategy for solving variational inequalities in bounded polytopes.* Numerical Functional Analysis and Optimization 16, 1995, pp. 653–668.

[13] M. Fukushima: *A relaxed projection method for variational inequalities.* Mathematical Programming 35, 1986, pp. 58–70.

[14] L. Grippo, F. Lampariello and S. Lucidi: *A nonmonotone line search technique for Newton's method.* SIAM Journal on Numerical Analysis 23, 1986, pp. 707–716.

[15] 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.

[16] W. Hock and K. Schittkowski: *Test Examples for Nonlinear Programming Codes.* Lecture Notes in Economics and Mathematical Systems 187, Springer–Verlag, Berlin, Germany, 1981.

[17] C. Kanzow: *Semismooth Newton-type Methods for the Solution of Nonlinear Complementarity Problems.* Habilitation Thesis, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany, April 1997.

[18] C. Kanzow: *An inexact QP-based method for nonlinear complementarity problems.* Preprint 120, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany, February 1997 (revised September 1997).

[19] C. Kanzow and H. Jiang: *A continuation method for (strongly) monotone variational inequalities.* Mathematical Programming, to appear.

[20] J. Liu: *Strong stability in variational inequalities.* SIAM Journal on Control and Optimization 33, 1995, pp. 725–749.

[21] P. Marcotte: *A sequential linear programming algorithm for solving monotone variational inequalities.* SIAM Journal on Control and Optimization 27, 1989, pp. 1260–1278.

[22] P. Marcotte and J.-P. Dussault: *A note on a globally convergent Newton method for solving monotone variational inequalities.* Operations Research Letters 6, 1987, pp. 35–42.

[23] J.J. Moré and D.C. Sorensen: *Computing a trust region step.* SIAM Journal on Scientific and Statistical Computing 4, 1983, pp. 553–572.

[24] A. Nagurney: *Network Economics — A Variational Inequality Approach.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 1993.

[25] J.-S. Pang: *Newton's method for B-differentiable equations.* Mathematics of Operations Research 15, 1990, pp. 311–341.

[26] J.-S. Pang: *A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems.* Mathematical Programming 51, 1991, pp. 101–131.

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

[28] J.-M. Peng: *A global method for monotone variational inequality problems with inequality constraints.* Journal of Optimization Theory and Applications, to appear.

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

[30] L. Qi and H. Jiang: *Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton and quasi-Newton methods for solving these equations.* Mathematics of Operations Research 22, 1997, pp. 301–325.

[31] L. Qi and J. Sun: *A nonsmooth version of Newton's method.* Mathematical Programming 58, 1993, pp. 353–368.

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

[33] H. Sellami and S.M. Robinson: *Implementation of a continuation method for normal maps.* Mathematical Programming 76, 1997, pp. 563–578.

[34] K. Taji, M. Fukushima and T. Ibaraki: *A globally convergent Newton method for solving strongly monotone variational inequality problems.* Mathematical Programming 58, 1993, pp. 369–383.

[35] B. Xiao and P.T. Harker: *A nonsmooth Newton method for variational inequalities, I: Theory.* Mathematical Programming 65, 1994, pp. 151–194.

[36] B. Xiao and P.T. Harker: *A nonsmooth Newton method for variational inequalities, II: Numerical results.* Mathematical Programming 65, 1994, pp. 195–216.