A Computable Generalized Hessian of the D-Gap Function and Newton-Type Methods for Variational Inequality Problems

Defeng Sun [1]
School of Mathematics
The University of New South Wales
Sydney, New South Wales 2052, Australia
E-mail: sun@solution.maths.unsw.edu.au

Masao Fukushima [2]
Department of Applied Mathematics and Physics
Graduate School of Engineering
Kyoto University
Kyoto 606-01, Japan
E-mail: fuku@kuamp.kyoto-u.ac.jp

and

Liqun Qi [1]
School of Mathematics
The University of New South Wales
Sydney, New South Wales 2052, Australia
E-mail: L.Qi@unsw.edu.au

Revised: May 9, 1996

Abstract. It is known that the variational inequality problem (VIP) can be converted to a differentiable unconstrained optimization problem via a merit function first considered by Peng and later studied further by Yamashita, Taji and Fukushima. This merit function, called the D-gap function, though is differentiable, is not twice differentiable and its generalized Hessian with existing definitions is very difficult to compute and may not exist in some cases. This paper introduces a computable generalized Hessian (CGH) for the D-gap function in the case that the closed convex set for the VIP is defined by several twice continuously differentiable convex functions. Local superlinear convergence for the generalized Newton method to minimize the D-gap function with the CGH is established. A globally and superlinearly convergent trust region algorithm for the VIP, which utilizes the D-gap function and its CGH, is presented.

Key Words. Variational inequality problems, merit function, generalized Hessian, trust region algorithm, superlinear convergence.

[1]The research of this author is supported by the Australian Research Council.

[2]The research of this author is supported in part by the Science Research Grant-in-Aid from the Ministry of Education, Science and Culture, Japan under grant 06650443. This paper was completed while he was visiting The University of New South Wales; the visit was funded by the Australian Research Council.

1. Introduction

Let $F : \Re^n \to \Re^n$ be a continuously differentiable mapping and S be a nonempty closed convex set in $\Re^n$. The variational inequality problem (VIP) is to find a vector $x \in S$ such that

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

where $\langle \cdot, \cdot \rangle$ denotes the inner product in $\Re^n$. In the special case where $S = \Re^n_+$, the VIP reduces to the complementarity problem. A comprehensive survey of the VIP is given in [13].

In the last several years, much effort has been made to derive merit functions for the VIP, thereby reformulating the VIP as an equivalent optimization problem with certain desirable properties. Recent developments of such approaches are summarized in [10]. Early merit functions such as the regularized gap function [9] are intended to reformulate the VIP as a constrained differentiable optimization problem. Recently, Peng [24] showed that the difference of two regularized gap functions constitutes an unconstrained differentiable optimization problem equivalent to the VIP. Later, Yamashita, Taji and Fukushima [31] extended the idea of Peng [24] and investigated some important properties related to this merit function. Specifically, the latter authors considered the function $g_{\alpha\beta} : \Re^n \to \Re$ defined by

$$g_{\alpha\beta}(x) = f_\alpha(x) - f_\beta(x), \tag{1.1}$$

where α and β are arbitrary positive parameters such that $\alpha < \beta$ and f_α is the regularized gap function

$$f_\alpha(x) = \max_{y \in S} \left\{ \langle F(x), x - y \rangle - \frac{\alpha}{2} \|x - y\|^2 \right\}. \tag{1.2}$$

(The function f_β is defined similarly with α replaced by β.) In the special case $\beta = 1/\alpha$ and $\alpha < 1$ in (1.1), the function $g_{\alpha\beta}$ reduces to the merit function studied by Peng [24]. We call the function $g_{\alpha\beta}$ the D-gap function, where D stands for the word "difference". Note that, in [31], the quadratic term in the definition (1.1) of $g_{\alpha\beta}$ is replaced by a more general function. In the present paper, however, we restrict ourselves to the quadratic case, because it makes the analysis significantly simpler.

As shown in [24, 31], the function $g_{\alpha\beta}$ has a number of interesting properties. Among other things, $g_{\alpha\beta}(x)$ is nonnegative for all $x \in \Re^n$, and $g_{\alpha\beta}(x) = 0$ if and only if x is a solution of the VIP. Thus we may say that the VIP is equivalent to the unconstrained minimization problem

$$\text{minimize}_{x \in \Re^n} \quad g_{\alpha\beta}(x), \tag{1.3}$$

whenever the VIP has a solution. Moreover, it is easy to see that the function $g_{\alpha\beta}$ is continuously differentiable whenever so is F, and its gradient is given by

$$\nabla g_{\alpha\beta}(x) = \nabla f_\alpha(x) - \nabla f_\beta(x), \tag{1.4}$$

in which

$$\nabla f_\alpha(x) = F(x) + (\nabla F(x) - \alpha I)(x - y_\alpha(x)), \tag{1.5}$$

where $y_\alpha(x)$ is the unique maximizer of the right-hand side of the definition (1.2) of f_α. $\nabla f_\beta(x)$ and $y_\beta(x)$ are similarly defined.

Although the function $g_{\alpha\beta}$ is in general nonconvex, its stationary point becomes a global minimum, provided that the mapping involved in the VIP has positive definite Jacobian [31]. Therefore it is quite natural to attempt to solve the minimization problem (1.3) by a rapidly convergent iterative algorithm. One thing one should keep in mind is, however, that the function $g_{\alpha\beta}$ is once continuously differentiable but not twice differentiable even if F is twice continuously differentiable. More specifically, (1.4) reveals that $\nabla g_{\alpha\beta}(x)$ is represented in terms of $y_\alpha(x)$ and $y_\beta(x)$. Since $y_\alpha(x)$ may alternatively be written as

$$y_\alpha(x) = \Pi_S(x - \alpha^{-1}F(x)), \tag{1.6}$$

where Π_S denotes the projection operator on the set S, the function $y_\alpha(\cdot)$ is in general nondifferentiable. Nevertheless, it is often verified that the projection operator enjoys the property called semismoothness, under appropriate assumptions on the set S. Therefore we may still expect to have a rapidly convergent algorithm for minimizing $g_{\alpha\beta}$ by utilizing the idea from the recently developed theory for superlinear convergence of generalized Newton methods that relies on the semismoothness of the gradient mapping [5, 15, 21, 22, 26, 27, 28].

In the remainder of the paper, we suppose that the parameters α and β are fixed in the definition (1.1) of $g_{\alpha\beta}$. Thus, to simplify the notation, we shall write g for $g_{\alpha\beta}$. Moreover, we shall often denote the gradient mapping ∇g as G.

A key ingredient for a rapidly convergent algorithm for minimizing g is to calculate some generalized Hessian of g at an iterative point. To do this, we focus our attention to some special convex set S. We assume that

$$S = \{y \in \Re^n \mid h_i(y) \le 0, i = 1, \cdots, m\}, \tag{1.7}$$

where each h_i is twice continuously differentiable and convex. However, even if S is defined by (1.7), it is still very difficult to calculate a generalized Hessian of g at a given point x. The difficulty arises from three aspects:

i) generalized Jacobians of Π_S at $x - \alpha^{-1}F(x)$ and $x - \beta^{-1}F(x)$ are needed but not easy to compute.

ii) from existing definitions of generalized Jacobians of a vector function, it is unavoidable to compute the second-derivative of F or the generalized Jacobian of ∇F. This is not practical in computation even if $\nabla^2 F$ exists. If F is only continuously differentiable but ∇F is not locally Lipschitz continuous, then the generalized Jacobian of ∇F with existing definitions does not exist.

iii) the difference of generalized Hessians of f_α and f_β is not necessarily a generalized Hessian of g with existing definitions.

There are two existing definitions of generalized Jacobians of a locally Lipschitz continuous function $H : \Re^n \to \Re^n$. By Rademacher's theorem, H is differentiable almost everywhere and the B-differential of H at x is given by

$$\partial_B H(x) = \{V \in \Re^{n\times n} \mid V = \lim_{x^k \to x} \nabla H(x^k)^T,\ x^k \in \Omega_H\},$$

where $\Omega_H = \{x \in \Re^n \mid H \text{ is differentiable at } x\}$ [26]. For any $x \in \Re^n$, $\partial_B H(x)$ is a nonempty compact set consisting of $n \times n$ matrices. On the other hand, the Clarke Jacobian of H [2] at x is defined by

$$\partial H(x) = \text{conv } \partial_B H(x).$$

The above three difficulties occur for both the B-differential and the Clarke Jacobian.

If ∇F is not locally Lipschitz continuous, then G may also not be locally Lipschitz continuous. In this case, neither ∂G nor $\partial_B G$ exists. For example, let $S = \Re^1$, $\alpha = \frac{1}{2}$, $\beta = 1$ and $F(x) =$

$\int_0^x p(t)dt$, where

$$p(t) = \begin{cases} \sqrt{|t|}\sin\frac{1}{t} + |t| + 1 & \text{if } t \neq 0 \\ 1 & \text{otherwise.} \end{cases}$$

Then $\nabla F(x) = p(x)$, $g(x) = \frac{1}{2}F(x)^2$ and $G(x) = \nabla F(x)F(x) = p(x)F(x)$. Note that ∇F is everywhere continuous but fails to be Lipschitz continuous around $x = 0$. Since $F(0) = 0$ and $\nabla F(0) = 1$, we have, for all $|x|$ sufficiently small,

$$F(|x|) = F(0) + \nabla F(0)|x| + o(|x|) \geq \frac{1}{2}|x|.$$

By taking $x^k = \frac{1}{2k\pi+\frac{\pi}{2}}$ and $y^k = \frac{1}{2k\pi+\frac{3}{2}\pi}$, where $k = 0, 1, 2, \cdots$, we have

$$\begin{aligned} \frac{|G(x^k) - G(y^k)|}{|x^k - y^k|} &= \frac{|\nabla F(x^k)(F(x^k) - F(y^k)) + (\nabla F(x^k) - \nabla F(y^k))F(y^k)|}{|x^k - y^k|} \\ &\geq \frac{|(\nabla F(x^k) - \nabla F(y^k))F(y^k)|}{|x^k - y^k|} - \frac{|\nabla F(x^k)(F(x^k) - F(y^k))|}{|x^k - y^k|} \\ &\geq \frac{(\sqrt{x^k} + \sqrt{y^k})|F(y^k)|}{|x^k - y^k|} - |F(y^k)| - O(1) \\ &\geq \frac{1}{2}\frac{(\sqrt{x^k} + \sqrt{y^k})y^k}{|x^k - y^k|} - O(1) \\ &= \frac{1}{2}\frac{\left(\sqrt{\frac{1}{2k\pi+\frac{\pi}{2}}} + \sqrt{\frac{1}{2k\pi+\frac{3\pi}{2}}}\right)\frac{1}{2k\pi+\frac{3\pi}{2}}}{\frac{1}{2k\pi+\frac{\pi}{2}} - \frac{1}{2k\pi+\frac{3\pi}{2}}} - O(1) \\ &= O(\sqrt{k}) - O(1) \end{aligned}$$

for k sufficiently large. So, G is not Lipschitz continuous around $x = 0$. In this case, neither $\partial G(0)$ nor $\partial_B G(0)$ is defined at all.

The above example indicates that the conventional generalized Hessians of g may not be defined by ∂G or $\partial_B G$, when F is only continuously differentiable. In this paper, we aim to define a generalized Hessian of g which covers such unfavorable cases and is also computable in practice.

In Section 2, we propose a computable generalized Jacobian (CGJ) at a given point for the projection operator Π_S with S defined by (1.7). We denote it by $\partial_C \Pi_S(\cdot)$, where C stands for the word "computable". Based on $\partial_C \Pi_S(\cdot)$, in Section 3 we discuss a way to define a computable generalized Hessian (CGH) of f_α at x by using only the first-order information of F at a given point. We denote it by $\tilde{H}_C f_\alpha(x)$ and define a CGH for g at x by

$$H_C g(x) = \tilde{H}_C f_\alpha(x) - \tilde{H}_C f_\beta(x).$$

The CGH of g may not coincide with the Hessian of g even if the latter exists, but it plays a role of the Hessian matrix of g in our generalized Newton method. Using the CGH of g, we establish local superlinear convergence of a generalized Newton method for minimizing g in Section 4. We propose a trust region algorithm in Section 5 and prove its global convergence in Section 6. In Section 7, based upon the theory developed in Section 4, we establish local superlinear convergence of the trust region algorithm.

It is noted that globally convergent Newton methods for the VIP have already been proposed by several authors [17, 30]. Those algorithms, however, require the solution of a linearized variational inequality problem at each iteration. In contrast with them, the algorithm proposed in this paper solves at each iteration a quadratic minimization problem with a bound constraint on the variables. For the inequality constrained VIP, people also consider merit functions for its Karush-Kuhn-Tucker (KKT) system [1, 6, 7].

2. A computable generalized Jacobian for the projection operator

In this section, we show how to compute a CGJ of $\Pi_S(\cdot)$, where S is defined by (1.7). Let $\bar{y} = \Pi_S(x)$. Then it is the unique solution of the following nonlinear programming problem in y :

$$\begin{aligned} \min \quad & \frac{1}{2}\|y-x\|^2 \\ \text{s.t.} \quad & h_i(y) \le 0, \quad i = 1, \cdots, m. \end{aligned} \tag{2.1}$$

Let $\mathcal{M}(x)$ denote the (possibly empty) set of multipliers $\lambda \in \Re^m$ that satisfy KKT optimality conditions for (2.1) at $\bar{y}$:

$$\begin{gathered} \bar{y} - x + \textstyle\sum_{i=1}^m \lambda_i \nabla h_i(\bar{y}) = 0, \\ \lambda_i \ge 0, \quad h_i(\bar{y}) \le 0, \quad \lambda_i h_i(\bar{y}) = 0, \quad i = 1, \cdots, m. \end{gathered} \tag{2.2}$$

For any $y \in \Re^n$, we will denote the active set by

$$I(y) = \{i \mid h_i(y) = 0\}.$$

In order to ensure the nonemptiness of $\mathcal{M}(x)$, we need some constraint qualifications. Here we will use the so-called constant rank constraint qualification (CRCQ), which was used by Janin [14] for studying the stability of nonlinear programming and recently used by Pang and Ralph [23] to investigate conditions for piecewise smoothness of $\Pi_S(\cdot)$. The CRCQ is said to hold at $\bar{y} = \Pi_S(x)$, if there exists a neighborhood $N(\bar{y})$ of $\bar{y}$ such that for every set $J \subseteq I(\bar{y})$, the family of gradient vectors

$$\{\nabla h_i(y) \mid i \in J\}$$

has the same rank (which depends on J) for all vectors $y \in N(\bar{y})$. The CRCQ is weaker than the linear independence constraint qualification (LICQ), i.e., the family of vectors

$$\{\nabla h_i(\bar{y}) \mid i \in I(\bar{y})\}$$

are linearly independent, and will hold automatically on the whole space $\Re^n$ if S is a convex polyhedral set. It is known that if the CRCQ holds at $\bar{y}$, then $\mathcal{M}(x)$ is nonempty [14]. For a nonnegative vector $d \in \Re^m$, we let supp(d), called the support of d, be the subset of $\{1, \cdots, m\}$

consisting of the indices i for which $d_i > 0$. Let $\mathcal{B}(x)$ be a family of subsets of $\{1, \cdots, m\}$ defined as follows: $J \in \mathcal{B}(x)$ if and only if $\mathrm{supp}(\lambda) \subseteq J \subseteq I(\bar{y})$ for some $\lambda \in \mathcal{M}(x)$ and the vectors

$$\{\nabla h_i(\bar{y}) \mid i \in J\}$$

are linearly independent. Here we allow the empty index set to be a member of $\mathcal{B}(x)$. So if $I(\bar{y}) = \emptyset$, $\mathcal{B}(x) = \{\emptyset\}$. When $I(\bar{y}) \neq \emptyset$, $\emptyset \in \mathcal{B}(x)$ if and only if $0 \in \mathcal{M}(x)$. In particular, when $I(\bar{y}) \neq \emptyset$, $\mathcal{B}(x) = \{\emptyset\}$ if and only if all $\nabla h_i(\bar{y})$, $i \in I(\bar{y})$, are zero vectors. From the CRCQ, the latter implies that $\mathcal{B}(z) = \{\emptyset\}$ for each z in a neighborhood of x.

The following lemma is proved by Pang and Ralph [23] for the case $x \notin S$. By considering the above observations, the results of this lemma also hold for $x \in S$.

Lemma 2.1 *[23] If the CRCQ holds at $\bar{y} = \Pi_S(x)$, then there exists a neighborhood $N(x)$ of x such that for all $z \in N(x)$,*
(i) the CRCQ holds at $\Pi_S(z)$;
(ii) $\mathcal{B}(z) \subseteq \mathcal{B}(x)$.

Now suppose that the CRCQ holds at $\bar{y} = \Pi_S(x)$. By definition, for each $J \in \mathcal{B}(x)$, there exists $\lambda \in \mathcal{M}(x)$ such that

$$\mathrm{supp}(\lambda) \subseteq J \subseteq I(\bar{y}). \tag{2.3}$$

Consider the following system of nonlinear equations:

$$H(y, \mu, z; J) \equiv \begin{pmatrix} y - z + \sum_{i=1}^m \mu_i \nabla h_i(y) \\ \\ h_J(y) \\ \\ \mu_{\bar{J}} \end{pmatrix} = 0, \tag{2.4}$$

where $(y, \mu) \in \Re^n \times \Re^m$ are variables, $z \in \Re^n$ are parameters and $\bar{J}$ is the complement of J in $\{1, \cdots, m\}$, i.e., $\bar{J} = \{1, \cdots, m\} \backslash J$. For any partition $J \cup \bar{J} = \{1, \cdots, m\}$, we write $\mu = (\mu_J, \mu_{\bar{J}})$ and $h(y) = (h_J(y), h_{\bar{J}}(y))$.

For the vectors $\bar{y} = \Pi_S(x)$, $\bar{\mu} = \lambda$ and $\bar{z} = x$, it follows from the KKT conditions (2.2) and the inclusions (2.3) that

$$H(\bar{y}, \bar{\mu}, \bar{z}; J) = 0.$$

The partial derivative of $H(\cdot, \cdot, \cdot; J)$ with respect to (y, μ) is given by

$$A(y, \mu) \equiv \nabla_{y,\mu} H(y, \mu, z; J)^T = \begin{pmatrix} I + \sum_{i=1}^m \mu_i \nabla^2 h_i(y) & \nabla h_J(y) & \nabla h_{\bar{J}}(y) \\ \\ \nabla h_J(y)^T & 0 & 0 \\ \\ 0 & 0 & I \end{pmatrix}.$$

It is easy to check that, by the CRCQ, $A(\bar{y}, \bar{\mu})$ is nonsingular when h_i is twice continuously differentiable and convex. So the implicit function theorem [18] ensures that there exist open neighborhoods $N(\bar{z}; J)$ of $\bar{z}(= x)$ and $N(\bar{y}, \bar{\mu}; J)$ of $(\bar{y}, \bar{\mu})$ such that $H(y, \mu, z; J) = 0$ has a unique solution $(y(z; J), \mu(z; J)) \in \mathrm{cl}\ N(\bar{y}, \bar{\mu}; J)$ whenever $z \in \mathrm{cl}\ N(\bar{z}; J)$. Moreover $(y(z; J), \mu(z; J))$ is continuously differentiable in z. After easy computations, we have

$$\nabla y(z; J) = C^{-1} - C^{-1} D \left(D^T C^{-1} D \right)^{-1} D^T C^{-1}, \tag{2.5}$$

where

$$C \equiv C(z;J) \equiv I + \sum_{i=1}^{m} \mu_i(z;J)\nabla^2 h_i(y(z;J)), \quad D \equiv D(z;J) \equiv \nabla h_J(y(z;J)). \tag{2.6}$$

Notice that

$$y(z;\emptyset) \equiv z$$

and

$$\nabla y(z;\emptyset) \equiv I.$$

Lemma 2.2 *The matrix $\nabla y(x;J)$ is symmetric positive semidefinite and $\|\nabla y(x;J)\| \le 1$.*

Proof. Let C and D, respectively, denote $C(z;J)$ and $D(z;J)$ evaluated at $z = x$. It is clear from (2.5) and (2.6) that $\nabla y(x;J)$ is symmetric and C is symmetric positive definite. Let

$$B = C^{-\frac{1}{2}}D\left((C^{-\frac{1}{2}}D)^T(C^{-\frac{1}{2}}D)\right)^{-1}(C^{-\frac{1}{2}}D)^T.$$

It is easy to check that $B^T = B$, $B^2 = B$, $\|B\| \le 1$ and $\|I - B\| \le 1$. So for any $d \in \Re^n$, we have

$$\begin{aligned}\langle d, \nabla y(x;J)d\rangle &= \langle d, C^{-1}d\rangle - \langle C^{-\frac{1}{2}}d, BC^{-\frac{1}{2}}d\rangle \\ &\ge \langle d, C^{-1}d\rangle - \langle C^{-\frac{1}{2}}d, C^{-\frac{1}{2}}d\rangle \\ &= 0,\end{aligned}$$

which means that $\nabla y(x;J)$ is positive semidefinite and

$$\|\nabla y(x;J)\| = \|C^{-\frac{1}{2}}(I-B)C^{-\frac{1}{2}}\| \le \|C^{-\frac{1}{2}}\|^2\|I-B\| \le 1.$$

This completes the proof. $\square$

From Lemma 2.1, if the CRCQ holds at $\Pi_S(x)$, then there exists a neighborhood $N(x)$ of x such that the CRCQ holds at $\Pi(z)$ and $\mathcal{B}(z) \subseteq \mathcal{B}(x)$ whenever $z \in N(x)$. So by the definition of $y(\bar{z};J)$, $J \in \mathcal{B}(x)$, we have

$$\Pi_S(x) = y(\bar{z};J), \quad J \in \mathcal{B}(x). \tag{2.7}$$

Based on these observations, we define the CGJ of $\Pi_S(x)$ as follows:

$$\partial_C \Pi_S(x) = \{\nabla y(x;J) \mid J \in \mathcal{B}(x)\}. \tag{2.8}$$

Note that, when S is a polyhedral set, we have $\nabla^2 h_i(y) = 0$ for all $y \in \Re^n$. Han and Sun [12] used the set $\partial_C \Pi_S(x)$ to construct Newton and quasi-Newton methods for solving variational inequalities with a polyhedral set.

From the definition of $\partial_C \Pi_S(x)$, to find one element $P \in \partial_C \Pi_S(x)$ is equivalent to find an index set $J \in \mathcal{B}(x)$. This is often not difficult after we have the value of $\Pi_S(x)$. For instance, if the LICQ holds at $\Pi_S(x)$, we can choose $J = I(\bar{y})$. In fact, any $\lambda \in \mathcal{M}(x)$ with minimal support gives an index set $J = \text{supp}(\lambda) \in \mathcal{B}(x)$ no matter whether or not the LICQ holds. A multiplier $\lambda \in \mathcal{M}(x)$ with minimal $\text{supp}(\lambda)$ can be obtained easily if we have some element of $\mathcal{M}(x)$. But such an element of $\mathcal{M}(x)$ is often a by-product of computing $\Pi_S(x)$. In particular, if S is a polyhedral set and the LICQ holds at $\Pi_S(x)$, the work to find a $P \in \partial_C \Pi_S(x)$ is approximately equal to the work to calculate inverses of some matrices (see (2.5) and (2.8)).

The proof of the following lemma is stimulated by the arguments in [23].

Lemma 2.3 *Suppose that the CRCQ holds at* $\bar{y} = \Pi_S(x)$. *Let* $(y(z;J), \mu(z;J))$ *be the solution of (2.4) for given* z *and* J *satisfying (2.3). Then there exists a neighborhood* $U(x)$ *of* x *such that for each* $z \in U(x)$,
(i) $\Pi_S(z) = y(z;J), \quad J \in \mathcal{B}(z)$;
(ii) $\partial_C \Pi_S(z) = \{\nabla y(z;J) \mid J \in \mathcal{B}(z)\}$.

Remark. Note that (i) and (ii) in Lemma 2.3 does not follow immediately from (2.7) and (2.8), because $y(z;J)$ is related to the equation (2.4) with J determined from the point x rather than z (see (2.3)).

Proof of Lemma 2.3. From Lemma 2.1, there exists a neighborhood $N(x)$ of x such that for any $z \in N(x)$, the CRCQ holds at $\Pi_S(z)$ and $\mathcal{B}(z) \subseteq \mathcal{B}(x)$.

Let

$$U(x) \subseteq \bigcap_{J \in \mathcal{B}(x)} N(\bar{z};J) \cap N(x)$$

be an open neighborhood of $x (= \bar{z})$ such that for any $z \in U(x)$, any $\lambda^z \in \mathcal{M}(z)$ and $J \in \mathcal{B}(z)$ satisfying $\mathrm{supp}(\lambda^z) \subseteq J \subseteq I(\Pi_S(z))$,

$$(\Pi_S(z), \lambda^z) \in N(\bar{y}, \bar{\mu}; J).$$

Such $U(x)$ can be chosen because there are only finitely many J's and, as $z \to x$, we have $J \in \mathcal{B}(z) (\subseteq \mathcal{B}(x))$,

$$\Pi_S(z) \to \Pi_S(x) = \bar{y} = y(\bar{z};J)$$

and

$$\begin{aligned} \lambda^z = (\lambda^z_J, \lambda^z_{\bar{J}}) &= (\Psi(z)(z - \Pi_S(z)), 0) \\ &\to (\Psi(x)(x - \Pi_S(x)), 0) \\ &= (\lambda_J, 0) = \lambda = \bar{\mu} = \mu(\bar{z};J) \end{aligned}$$

with $\lambda \in \mathcal{M}(x)$ satisfying $\mathrm{supp}(\lambda) \subseteq J \subseteq I(\bar{y})$, where $\Psi(v)$ is the matrix defined for any $v \in N(x;J)$ by

$$\Psi(v) = \left(\nabla h_J(\Pi_S(v))^T \nabla h_J(\Pi_S(v))\right)^{-1} \nabla h_J(\Pi_S(v))^T.$$

For $J \in \mathcal{B}(x)$, let

$$U(x;J) = \{z \mid z \in U(x),\ J \in \mathcal{B}(z)\}.$$

Then

$$U(x) = \bigcup_{J \in \mathcal{B}(x)} U(x;J).$$

For any $z \in U(x;J)$,

$$H(\Pi_S(z), \lambda^z, z; J) = 0$$

and

$$(\Pi_S(z), \lambda^z) \in N(\bar{y}, \bar{\mu}; J).$$

Thus we obtain

$$(\Pi_S(z), \lambda^z) = (y(z;J), \mu(z;J)),$$

since the solution of $H(y, \mu, v; J) = 0$ is unique in cl $N(\bar{y}, \bar{\mu}; J)$ for each $v \in N(\bar{z};J)$. So (i) follows.

Next we prove (ii). From the definition of $\partial_C \Pi_S(z)$, for each $\lambda^z \in \mathcal{M}(z)$ and $J \in \mathcal{B}(z)$ satisfying $\mathrm{supp}(\lambda^z) \subseteq J \subseteq I(\Pi_S(z))$, there exist two open neighborhoods $N(z;J)$ ($\subseteq N(\bar{z};J)$) of z and $N(\Pi_S(z),\lambda^z;J)$ of $(\Pi_S(z),\lambda^z)$ such that $H(y,\mu,v;J)=0$ has a unique continuously differentiable solution $(y^z(v;J),\mu^z(v;J)) \in \mathrm{cl}\ N(\Pi_S(z),\lambda^z;J)$ whenever $v \in \mathrm{cl}\ N(z;J)$. So

$$\partial_C \Pi_S(z) = \{\nabla y^z(z;J) \mid J \in \mathcal{B}(z)\}.$$

Since

$$(y^z(z;J),\mu^z(z;J)) = (\Pi_S(z),\lambda^z) \in N(\bar{y},\bar{\mu};J)$$

and $N(\bar{y},\bar{\mu};J)$ is an open set, we can assume $N(z;J)$ sufficiently small so that for any $v \in \mathrm{cl}\ N(z;J)$

$$(y^z(v;J),\mu^z(v;J)) \in N(\bar{y},\bar{\mu};J).$$

Then from the uniqueness of the solution of $H(y,\mu,v;J)=0$ in $\mathrm{cl}\ N(\bar{y},\bar{\mu};J)$ for each $v \in \mathrm{cl}\ N(\bar{z};J)$, it follows that for any $v \in N(z;J) \subseteq N(\bar{z};J)$

$$(y^z(v;J),\mu^z(v;J)) = (y(v;J),\mu(v;J)).$$

So we have

$$\nabla y^z(z;J) = \nabla y(z;J), \quad J \in \mathcal{B}(z),$$

and hence,

$$\partial_C \Pi_S(z) = \{\nabla y(z;J) \mid J \in \mathcal{B}(z)\}.$$

This completes the proof. □

From Lemmas 2.1 and 2.3, if the CRCQ holds at $\Pi_S(x)$, there exists a neighborhood $N(x)$ of x such that for any $z \in N(x)$,

$$\Pi_S(z) \in \{y(z;J) \mid J \in \mathcal{B}(x)\}.$$

So we have

$$\partial_B \Pi_S(x) \subseteq \{\nabla y(x;J) \mid J \in \mathcal{B}(x)\} = \partial_C \Pi_S(x).$$

If the LICQ holds at $\Pi_S(x)$, then we have

$$\partial_B \Pi_S(x) = \partial_C \Pi_S(x)$$

according to Theorem 3.2 and Corollary 3.2.2 of [20]. In [19, 20] the generalized Jacobian has been discussed for more general parametric VIP under the LICQ assumption. When the LICQ does not hold but the CRCQ holds, the above equality does not hold in general. A counterexample is given in [12].

3. A computable generalized Hessian for the D-gap function

Now let us define the CGH of f_α at x as

$$\tilde{H}_C f_\alpha(x) = \{V \in \Re^{n\times n} \mid \quad \begin{array}{l} V = \nabla F(x)^T + (\nabla F(x) - \alpha I)(I - P_\alpha(I - \alpha^{-1}\nabla F(x))^T), \\ P_\alpha \in \partial_C \Pi_S(x - \alpha^{-1}F(x))\} \end{array}$$

and define $\tilde{H}_C f_\beta(x)$ similarly. Using these sets, we define the CGH of g at x as

$$H_C g(x) = \tilde{H}_C f_\alpha(x) - \tilde{H}_C f_\beta(x).$$

By rearrangements, we have

$$H_C g(x) = \{V \in \Re^{n\times n} \mid \quad \begin{array}{l} V = (\beta-\alpha)I - V_\beta + V_\alpha, \\ V_\beta \in \beta^{-1}(\beta I - \nabla F(x))\partial_C \Pi_S(x - \beta^{-1}F(x))(\beta I - \nabla F(x))^T, \\ V_\alpha \in \alpha^{-1}(\alpha I - \nabla F(x))\partial_C \Pi_S(x - \alpha^{-1}F(x))(\alpha I - \nabla F(x))^T\}. \end{array}$$

The reason that we use $H_C g(x)$ here is that an element $V \in H_C g(x)$ is often easier to compute than $V \in \partial_B G(x)$ but it still leads to superlinear (and quadratic) convergence of generalized Newton methods. The above set $H_C g(x)$ is, in general, not equal to the generalized Jacobian $\partial_B G(x)$. In fact, when $S = \Re^n$ and $F \in C^2$, we have

$$H_C g(x) = \{(\alpha^{-1} - \beta^{-1})\nabla F(x)\nabla F(x)^T\} \tag{3.1}$$

and

$$\partial_B G(x) = \{\nabla G(x)\} = \{(\alpha^{-1} - \beta^{-1})\nabla F(x)\nabla F(x)^T + (\alpha^{-1} - \beta^{-1})\nabla^2 F(x)F(x)\}. \tag{3.2}$$

It is generally difficult to establish a relation between $\partial_B G$ and $H_C g$ since the second-order derivative is not used in $H_C g$. But, if F is affine, i.e., $F(x) = Mx + c$ with $M \in \Re^{n\times n}$ and $c \in \Re^n$, we have

Lemma 3.1 *If $F(x) = Mx + c$ and the CRCQ holds at $y_\alpha(x)$ and $y_\beta(x)$, then*

$$\partial_B G(x) \subseteq H_C g(x).$$

Proof. From Lemmas 2.1 and 2.3, there exists a neighborhood $N(x)$ of x such that for any $z \in N(x)$,

$$y_\alpha(z) = \Pi_S(z - \alpha^{-1}F(z)) \in \{y(z - \alpha^{-1}F(z); L) \mid L \in \mathcal{B}(x - \alpha^{-1}F(x))\}$$

and

$$y_\beta(z) = \Pi_S(z - \beta^{-1}F(z)) \in \{y(z - \beta^{-1}F(z); J) \mid J \in \mathcal{B}(x - \beta^{-1}F(x))\}.$$

So for each $z \in N(x)$,

$$G(z) \in \{G^{LJ}(z) \mid L \in \mathcal{B}(x - \alpha^{-1}F(x)),\ J \in \mathcal{B}(x - \beta^{-1}F(x))\},$$

where

$$\begin{aligned} G^{LJ}(z) \ &\equiv M^T(y(z - \beta^{-1}F(z); J) - y(z - \alpha^{-1}F(z); L)) + \beta(z - y(z - \beta^{-1}F(z); J)) \\ &\quad -\alpha(z - y(z - \alpha^{-1}F(z); L)). \end{aligned}$$

Then,

$$\begin{aligned} \partial_B G(x) \ &\subseteq \{\nabla G^{LJ}(x) \mid L \in \mathcal{B}(x - \alpha^{-1}F(x)),\ J \in \mathcal{B}(x - \beta^{-1}F(x))\} \\ &= H_C g(x). \end{aligned}$$

This completes the proof. □

In the above lemma, the equality does not hold in general. For example, let $F(x) = x$, $S = \Re^1_+$ and $0 < \alpha < 1 < \beta$. Then

$$\begin{aligned} &\partial_B G(0) = \{2 - \alpha - \beta^{-1}, \beta + \alpha^{-1} - 2\} \\ &\subset \{2 - \alpha - \beta^{-1}, \beta + \alpha^{-1} - 2, \beta - \alpha, \alpha^{-1} - \beta^{-1}\} = H_C g(0). \end{aligned}$$

Notice that for solving nonlinear equations, the classical Gauss-Newton method uses (3.1) instead of (3.2) (with $\alpha^{-1} - \beta^{-1} = 1$) to avoid computing $\nabla^2 F(x)$. Now the question is: Can we still obtain local superlinear convergence even if one only uses $V \in H_C g(x)$ instead of $V \in \partial_B G(x)$? The answer is "yes". In the next section, we will show that it retains local superlinear (and quadratic) convergence.

In order to prove superlinear convergence, we need that all members of $H_C g(x^*)$ are positive definite for a solution x^* of the VIP.

Theorem 3.1 *Let $x \in \Re^n$ be given. Suppose that the CRCQ holds at $y_\alpha(x)$ and $y_\beta(x)$, and that $\nabla F(x)$ is positive definite. If $\lambda_{\min}(\nabla F(x) + \nabla F(x)^T) > \alpha + \beta^{-1}\|\nabla F(x)\|^2$, where $\lambda_{\min}(A)$ denotes the smallest eigenvalue of a symmetric matrix A, then all $V \in H_C g(x)$ are positive definite.*

Proof. For any $V \in H_C g(x)$, there exist two positive semidefinite matrices $P_\beta \in \partial_C \Pi_S(x - \beta^{-1}F(x))$ and $P_\alpha \in \partial_C \Pi_S(x - \alpha^{-1}F(x))$ such that

$$V = (\beta - \alpha)I - V_\beta + V_\alpha,$$

where

$$V_\beta = \beta^{-1}(\beta I - \nabla F(x))P_\beta(\beta I - \nabla F(x))^T$$

and

$$V_\alpha = \alpha^{-1}(\alpha I - \nabla F(x))P_\alpha(\alpha I - \nabla F(x))^T.$$

Since, from Lemma 2.2, P_α is symmetric positive semidefinite and $\|P_\beta\| \le 1$, we have for any $0 \ne d \in \Re^n$,

$$\begin{aligned}
\langle d, Vd\rangle &= (\beta - \alpha)\langle d, d\rangle - \langle d, V_\beta d\rangle + \langle d, V_\alpha d\rangle \\
&\ge (\beta - \alpha)\langle d, d\rangle - \beta^{-1}\|(\beta I - \nabla F(x))^T d\|^2 \\
&= \langle d, (\nabla F(x) + \nabla F(x)^T)d\rangle - \alpha\langle d, d\rangle - \beta^{-1}\langle d, \nabla F(x)\nabla F(x)^T d\rangle. \qquad (3.3)
\end{aligned}$$

Then from the assumption and (3.3), it follows that V is positive definite. □

Note that, for each x, the condition $\lambda_{\min}(\nabla F(x) + \nabla F(x)^T) > \alpha + \beta^{-1}\|\nabla F(x)\|^2$ is satisfied if we choose β sufficiently large and α sufficiently small.

Remark. If $x = x^*$, then we have $x^* = y_\gamma(x^*)$ for all $\gamma > 0$. So, in this case, the CRCQ at x^* means that it holds at $y_\gamma(x^*)$ for each $\gamma > 0$.

When F is affine, i.e., $F(x) = Mx + c$ with $M \in \Re^{n\times n}$ and $c \in \Re^n$, we have the following result:

Corollary 3.1 *Suppose that $F(x) = Mx + c$ and M is positive definite. If the CRCQ holds everywhere, in particular if S is polyhedral, and $\lambda_{\min}(M + M^T) > \alpha + \beta^{-1}\|M\|^2$, then matrices $V \in H_C g(x)$, $x \in \Re^n$, are uniformly positive definite, and hence g is strongly convex.*

Proof. By (3.3) in the proof of Theorem 3.1, $V \in H_C g(x)$ are uniformly positive definite. The strong convexity of g follows from Lemma 3.1 and the positive definiteness of $V \in H_C g(x)$. □

4. Local superlinear convergence of a generalized Newton method

In this section, we will consider the local convergence properties of the following generalized Newton method:

$$x^{k+1} = x^k - V_k^{-1} G(x^k), \quad k = 0, 1, \cdots, \tag{4.1}$$

where $V_k \in H_C g(x^k)$.

To prove the superlinear convergence of (4.1), we need the following lemma.

Lemma 4.1 *Let x^* be a solution of the VIP. Suppose that all h_i are twice continuously differentiable and convex. If the CRCQ holds at x^*, then for any $V \in H_C g(x)$, we have*

$$G(x) - G(x^*) - V(x - x^*) = o(\|x - x^*\|). \tag{4.2}$$

Furthermore if ∇F and all $\nabla^2 h_i$ are Lipschitz continuous at x^*, then

$$G(x) - G(x^*) - V(x - x^*) = O(\|x - x^*\|^2). \tag{4.3}$$

Proof. Recall that $x^* = \Pi_S(x^* - \alpha^{-1}F(x^*)) = \Pi_S(x^* - \beta^{-1}F(x^*))$. By the assumption that the CRCQ holds at x^*, it follows from Lemma 2.1 that there exists a neighborhood $N(x^*)$ of x^* such that, for each $x \in N(x^*)$,

$$\mathcal{B}(x - \beta^{-1}F(x)) \subseteq \mathcal{B}(x^* - \beta^{-1}F(x^*))$$

and

$$\mathcal{B}(x - \alpha^{-1}F(x)) \subseteq \mathcal{B}(x^* - \alpha^{-1}F(x^*)).$$

For any $V \in H_C g(x)$, there exist $P_\beta \in \partial_C \Pi_S(x - \beta^{-1}F(x))$ and $P_\alpha \in \partial_C \Pi_S(x - \alpha^{-1}F(x))$ such that

$$V = (\beta - \alpha)I - V_\beta + V_\alpha,$$

where

$$V_\beta = \beta^{-1}(\beta I - \nabla F(x)) P_\beta (\beta I - \nabla F(x))^T$$

and

$$V_\alpha = \alpha^{-1}(\alpha I - \nabla F(x)) P_\alpha (\alpha I - \nabla F(x))^T.$$

So we can write

$$G(x) - G(x^*) - V(x - x^*) = T_1 + T_2 + T_3,$$

where

$$\begin{aligned} T_1 &= \nabla F(x)(y_\beta(x) - y_\alpha(x)) - \nabla F(x^*)(y_\beta(x^*) - y_\alpha(x^*)) \\ &\quad -\nabla F(x)\left(P_\beta(I - \beta^{-1}\nabla F(x))^T - P_\alpha(I - \alpha^{-1}\nabla F(x))^T\right)(x - x^*), \end{aligned}$$

$$T_2 = -\beta\left(y_\beta(x) - y_\beta(x^*) - P_\beta(I - \beta^{-1}\nabla F(x))^T(x - x^*)\right)$$

and

$$T_3 = \alpha\left(y_\alpha(x) - y_\alpha(x^*) - P_\alpha(I - \beta^{-1}\nabla F(x))^T(x - x^*)\right).$$

From Lemma 2.3, we can assume that $N(x^*)$ is sufficiently small so that, for each $x \in N(x^*)$, there exists $J \in \mathcal{B}(x^* - \beta^{-1}F(x^*))$ (J may be the empty index set) such that

$$y_\beta(x) = y(x - \beta^{-1}F(x); J)$$

and

$$P_\beta = \nabla y(x - \beta^{-1}F(x); J).$$

Therefore, we have

$$\begin{aligned}
& y_\beta(x) - y_\beta(x^*) - P_\beta(I - \beta^{-1}\nabla F(x))^T(x - x^*) \\
&= y(x - \beta^{-1}F(x); J) - y(x^* - \beta^{-1}F(x^*); J) \\
&\quad -\nabla y(x - \beta^{-1}F(x); J)(I - \beta^{-1}\nabla F(x))^T(x - x^*) \\
&= y(x - \beta^{-1}F(x); J) - y(x^* - \beta^{-1}F(x^*); J) \\
&\quad -\nabla y(x - \beta^{-1}F(x); J)(x - \beta^{-1}F(x) - (x^* - \beta^{-1}F(x^*))) \\
&\quad -\beta^{-1}\nabla y(x - \beta^{-1}F(x); J)(F(x) - F(x^*) - \nabla F(x)^T(x - x^*)) \\
&= o(\|x - \beta^{-1}F(x) - (x^* - \beta^{-1}F(x^*))\|) + o(\|x - x^*\|) \\
&= o(\|x - x^*\|). \qquad (4.4)
\end{aligned}$$

Since there are only finitely many J's, it follows from (4.4) that $T_2 = o(\|x - x^*\|)$. Similarly we have $T_3 = o(\|x - x^*\|)$. To prove (4.2), it remains to show $T_1 = o(\|x - x^*\|)$. Since $y_\beta(x^*) - y_\alpha(x^*) = x^* - x^* = 0$, we can write $T_1 = -\nabla F(x)(\beta^{-1}T_2 + \alpha^{-1}T_3)$. Hence $T_1 = o(\|x - x^*\|)$ follows from $T_2 = o(\|x - x^*\|)$ and $T_3 = o(\|x - x^*\|)$. Thus we obtain (4.2). Finally, when ∇F and all $\nabla^2 h_i$ are Lipschitz continuous, we can easily modify the above arguments to get (4.3). □

Theorem 4.1 *Let x^* be a solution of the VIP. Suppose that all h_i are twice continuously differentiable and convex. If the CRCQ holds at x^* and all $V \in H_C g(x^*)$ are positive definite, then there exists a neighborhood $N(x^*)$ of x^* such that when the initial point x^0 is chosen in $N(x^*)$, the sequence generated by (4.1) is well defined and converges to x^* Q-superlinearly. Furthermore, if ∇F and all $\nabla^2 h_i$ are Lipschitz continuous, then the convergence rate is Q-quadratic.*

Proof. From Lemma 2.3 and the definition of $H_C g(\cdot)$, it is easy to see that $H_C g(x^*)$ is compact, and $H_C g(\cdot)$ is upper-semicontinuous at x^*, i.e., for any $\varepsilon > 0$, there exists a positive number δ such that for all $x \in \{y \in \Re^n | \|y - x^*\| \le \delta\}$ we have

$$H_C g(x) \subseteq H_C g(x^*) + \varepsilon B,$$

where B is the unit ball of $\Re^n$. Then from the given assumptions, there exists a neighborhood $N(x^*)$ such that for any $x \in N(x^*)$, all $V \in H_C g(x)$ are uniformly positive definite. So for

$k = 0$, (4.1) is well defined. From Lemma 4.1 and (4.1), we have

$$\begin{aligned}\|x^{k+1} - x^*\| &= \|x^k - x^* - V_k^{-1}G(x^k)\| \\ &= \|V_k^{-1}(G(x^k) - G(x^*) - V_k(x^k - x^*))\| \\ &= o(\|x^k - x^*\|).\end{aligned}$$

This proves the Q-superlinear convergence of $\{x^k\}$. If ∇F and all $\nabla^2 h_i$ are Lipschitz continuous, by modifying the above arguments, we have the Q-quadratic convergence of $\{x^k\}$. □

Remark. Notice that for the superlinear convergence of the generalized Newton method (4.1), we only require F to be continuously differentiable. It is easy to see that the function F in the example in Section 1 satisfies the superlinear convergence conditions of Theorem 4.1.

The above method (4.1) only has a local convergence property. There are many ways to globalize such a method. In the next section we will provide a trust region algorithm to globalize it.

5. A trust region algorithm

The k-th iteration of the trust region algorithm for solving the unconstrained minimization problem

$$\text{minimize}_{x\in\Re^n} \quad g(x)$$

is stated as follows: Given $x^k \in \Re^n$ and $\Delta_k > 0$, solve the minimization problem

$$\begin{array}{ll}\text{minimize} & g_k(d) \equiv g(x^k) + \langle G(x^k), d\rangle + \frac{1}{2}\langle d, V_k d\rangle \\ \text{subject to} & \|d\| \le \Delta_k,\end{array} \tag{5.1}$$

where V_k is an element of $H_C g(x^k)$ or some approximation to it. Let d^k denote an optimal solution of subproblem (5.1). If $d^k = 0$, then we terminate the iteration. Otherwise, compute the ratio

$$\rho_k = \frac{g(x^k) - g(x^k + d^k)}{g(x^k) - g_k(d^k)} \tag{5.2}$$

and determine x^{k+1} and Δ_{k+1}, respectively, by

$$x^{k+1} = \begin{cases} x^k + d^k & \text{if } \rho_k > \eta_1, \\ x^k & \text{if } \rho_k \le \eta_1, \end{cases} \tag{5.3}$$

$$\Delta_{k+1} = \begin{cases} \gamma_1\Delta_k & \text{if } \rho_k \le \eta_1, \\ \Delta_k & \text{if } \eta_1 < \rho_k \le \eta_2, \\ \gamma_2\Delta_k & \text{if } \rho_k > \eta_2, \end{cases} \tag{5.4}$$

where $\eta_1, \eta_2, \gamma_1, \gamma_2$ are predetermined constants such that $0 < \eta_1 < \eta_2 < 1$ and $0 < \gamma_1 < 1 < \gamma_2$.

Note that if V_k is positive definite, and the constraint $\|d\| \le \Delta_k$ is inactive at the solution d^k of (5.1), and $\rho_k > \eta_1$, then the iteration $x^{k+1} = x^k + d^k$ reduces to the Newton iteration (4.1).

6. Global convergence of the trust region algorithm

The global convergence of trust region algorithms for unconstrained differentiable optimization problems has been studied quite extensively [25, 29, 8]. For example, Powell [25] showed that the trust region algorithm as described in the previous section generates a sequence $\{x^k\}$ such that

$$\liminf_{k\to\infty} \|G(x^k)\| = 0, \tag{6.1}$$

if the following conditions are satisfied:

(a) $g(x)$ is bounded below;

(b) $G(x)$ is uniformly continuous;

(c) $\{V_k\}$ satisfies either

$$\|V_k\| \le c_1 + c_2 \sum_{i=1}^{k} \Delta_i \quad \text{for all } k \tag{6.2}$$

or

$$\|V_k\| \le c_1 + c_2 k \quad \text{for all } k, \tag{6.3}$$

where c_1 and c_2 are some positive constants.

In the present case where g is the D-gap function defined by (1.1), condition (a) is automatically satisfied, since $g(x) \ge 0$ for all $x \in \Re^n$. For condition (b), we have the following lemma. This lemma has essentially been proved in [11], but we give the proof here for completeness.

Lemma 6.1 *Suppose that $\nabla F : \Re^n \to \Re^{n\times n}$ is uniformly continuous and bounded. Then the gradient mapping $G = \nabla g$ is uniformly continuous.*

Proof. First we show that F is also uniformly continuous. From the mean-value theorem [18], we have

$$\begin{aligned} F(x) - F(y) &= \int_0^1 \nabla F(x + t(y-x))^T (y-x) dt \\ &= \int_0^1 [\nabla F(x + t(y-x)) - \nabla F(x)]^T (y-x) dt + \nabla F(x)^T (y-x). \end{aligned}$$

Then it is not difficult to deduce the uniform continuity of F from the uniform continuity and the boundedness of ∇F. Now since the gradient mapping $G = \nabla g$ is given by

$$G(x) = \nabla F(x)(y_\beta(x) - y_\alpha(x)) + \beta(x - y_\beta(x)) - \alpha(x - y_\alpha(x)),$$

(see (1.4) and (1.5)), and since y_α and y_β are both composite functions of F and the projection operator Π_S (see (1.6)), the uniform continuity of G follows from the uniform continuity of F and ∇F and the boundedness of ∇F. □

From the above results, the following global convergence theorem is readily established.

Theorem 6.1 *Suppose that $\nabla F : \Re^n \to \Re^{n\times n}$ is uniformly continuous and bounded. Then the sequence $\{x^k\}$ generated by the trust region algorithm satisfies (6.1), provided that the sequence of matrices $\{V_k\}$ satisfies either of the conditions (6.2) and (6.3).*

Proof. As mentioned just before Lemma 6.1, the function g is bounded below. By Lemma 6.1, G is uniformly continuous. Consequently, it follows from the result established by Powell [25] that (6.1) is satisfied. □

It has been shown [31] that if F is strongly monotone and if either F is Lipschitz continuous or S is compact, then the function g has bounded level sets. So when these conditions are satisfied, the descent property of the trust region algorithm ensures the boundedness of the generated sequence $\{x^k\}$. Moreover, if $\{\Delta_k\}$ is bounded, which is usually the case in the trust region algorithm, the uniform continuity condition on G, i.e., condition (b), may be replaced by the weaker condition that G is uniformly continuous on a bounded set containing $\{x^k\}$ and $\{x^k + d^k\}$. The latter condition is particularly satisfied under the present standing assumption of continuous differentiability of F. Thus, in this case, we need not require ∇F to be uniformly continuous and bounded. To summarize, we obtain the next theorem.

Theorem 6.2 *Suppose that F is strongly monotone and that either F is Lipschitz continuous or S is compact. Suppose also that $\{V_k\}$ satisfies either (6.2) or (6.3) and that $\{\Delta_k\}$ is bounded. Then the sequence $\{x^k\}$ generated by the trust region algorithm contains a subsequence whose limit point x^* is the unique solution of the VIP.*

Proof. As mentioned in the paragraph preceding the theorem, the boundedness of the generated sequence $\{x^k\}$ is guaranteed under the given hypotheses and the same conclusion as that of Theorem 6.1 remains true. Hence, there is a subsequence $\{x^k\}_{k\in K}$ such that

$$\lim_{\substack{k\to\infty \\ k\in K}} G(x^k) = 0.$$

Therefore, by the continuity of G, we may deduce that there exists a subsequence whose limit point x^* satisfies the stationarity condition $G(x^*) = 0$. Moreover, the strong monotonicity of F ensures not only the existence of a unique solution of the VIP but also the fact that any stationary point of g solves the VIP [31]. This completes the proof. □

When the generated sequence $\{x^k\}$ is not bounded, we cannot say much about its asymptotic behavior. In fact, the above-mentioned result by Powell [25] only says that there exists a subsequence $\{x^k\}_{k\in K}$ which is a stationary sequence for g in the sense that

$$\lim_{\substack{k\to\infty \\ k\in K}} G(x^k) = 0.$$

In general, this does not imply that $\{x^k\}_{k\in K}$ is a minimizing sequence for g in the sense that

$$\lim_{\substack{k\to\infty \\ k\in K}} g(x^k) = \inf_{x\in\Re^n} g(x),$$

unless some additional conditions are met [11].

7. Superlinear convergence of the trust region algorithm

In this section, we establish local superlinear convergence of the algorithm. In order to do this, we need the following lemma.

Lemma 7.1 *Let x^* be a solution of the VIP. Suppose that all h_i are twice continuously differentiable and convex. If the CRCQ holds at x^*, then for any $V \in H_C g(x)$, we have*

$$g(x) - g(x^*) - \langle G(x^*), x - x^* \rangle - \frac{1}{2}\langle x - x^*, V(x - x^*)\rangle = o(\|x - x^*\|^2). \tag{7.1}$$

Proof. Let $N(x^*)$ be a neighborhood of x^* such that for any $x \in N(x^*)$,

$$\mathcal{B}(x - \alpha^{-1}F(x)) \subseteq \mathcal{B}(x^* - \alpha^{-1}F(x^*)),$$

$$\mathcal{B}(x - \beta^{-1}F(x)) \subseteq \mathcal{B}(x^* - \beta^{-1}F(x^*)),$$

and

$$y_\alpha(x) = y(x - \alpha^{-1}F(x); L),$$
$$y_\beta(x) = y(x - \beta^{-1}F(x); J)$$

for any $L \in \mathcal{B}(x - \alpha^{-1}F(x))$ and $J \in \mathcal{B}(x - \beta^{-1}F(x))$, respectively. The existence of $N(x^*)$ follows from Lemmas 2.1 and 2.3. Let us define

$$S^L = \{y \in \Re^n \mid h_i(y) \le 0,\ i \in L\}$$

and

$$f_\gamma^L(z) = \max_{y \in S^L}\left\{\langle F(z), z - y\rangle - \frac{\gamma}{2}\|z - y\|^2\right\},$$

where $\gamma = \alpha$ or β. It is known [9] that f_γ^L is continuously differentiable and

$$\nabla f_\gamma^L(z) = F(z) + \nabla F(z)(z - y_\gamma^L(z)) - \gamma(z - y_\gamma^L(z)),$$

where $y_\gamma^L(z) = \Pi_{S^L}(z - \gamma^{-1}F(z))$.

Let $g^{LJ} : \Re^n \to \Re$ be defined by

$$g^{LJ}(z) = f_\alpha^L(z) - f_\beta^J(z).$$

Since all h_i are twice continuously differentiable and convex, it follows from Lemma 3 of [23] that for any z and any $L \in \mathcal{B}(z)$ we have

$$\Pi_{S^L}(z) = \Pi_S(z).$$

Then, for any $L \in \mathcal{B}(x - \alpha^{-1}F(x))$ and $J \in \mathcal{B}(x - \beta^{-1}F(x))$, we have

$$\begin{aligned}
g^{LJ}(x) &= f_\alpha^L(x) - f_\beta^J(z) \\
&= \left[\langle F(x), (x - y_\alpha^L(x))\rangle - \frac{\alpha}{2}\|x - y_\alpha^L(x)\|^2\right] - \left[\langle F(x), (x - y_\beta^J(x))\rangle - \frac{\beta}{2}\|x - y_\beta^J(x)\|^2\right] \\
&= \left[\langle F(x), (x - y_\alpha(x))\rangle - \frac{\alpha}{2}\|x - y_\alpha(x)\|^2\right] - \left[\langle F(x), (x - y_\beta(x))\rangle - \frac{\beta}{2}\|x - y_\beta(x)\|^2\right] \\
&= f_\alpha(x) - f_\beta(x)
\end{aligned}$$

and

$$\begin{aligned}
\nabla g^{LJ}(x) &= \nabla f_\alpha^L(x) - \nabla f_\beta^J(x) \\
&= \Big[F(x) + \nabla F(x)(x - y_\alpha^L(x)) - \alpha(x - y_\alpha^L(x))\Big] \\
&\quad - \Big[F(x) + \nabla F(x)(x - y_\beta^J(x)) - \beta(x - y_\beta^J(x))\Big] \\
&= \nabla F(x)(y_\beta^J(x) - y_\alpha^L(x)) + \beta(x - y_\beta^J(x)) - \alpha(x - y_\alpha^L(x)) \\
&= \nabla F(x)(y_\beta(x) - y_\alpha(x)) + \beta(x - y_\beta(x)) - \alpha(x - y_\alpha(x)) \\
&= \nabla g(x) = G(x).
\end{aligned}$$

For any $V \in H_C g(x)$, there exist $L \in \mathcal{B}(x - \alpha^{-1}F(x))$ and $J \in \mathcal{B}(x - \beta^{-1}F(x))$ such that

$$V = (\beta - \alpha)I - V_\beta + V_\alpha, \tag{7.2}$$

where

$$V_\beta = \beta^{-1}(\beta I - \nabla F(x))\nabla y(x - \beta^{-1}F(x); J)(\beta I - \nabla F(x))^T$$

and

$$V_\alpha = \alpha^{-1}(\alpha I - \nabla F(x))\nabla y(x - \alpha^{-1}F(x); L)(\alpha I - \nabla F(x))^T.$$

Let $x_t = x + t(x - x^*),\ t \in [0, 1]$ and consider

$$\begin{aligned}
\delta^{LJ}(x) &\equiv g^{LJ}(x) - g^{LJ}(x^*) - \langle \nabla g^{LJ}(x^*), x - x^*\rangle - \frac{1}{2}\langle x - x^*, V(x - x^*)\rangle \\
&= \int_0^1 \langle \nabla g^{LJ}(x_t) - \nabla g^{LJ}(x^*), x - x^*\rangle dt - \frac{1}{2}\langle x - x^*, V(x - x^*)\rangle \\
&= T_4 + T_5 + T_6 - \frac{1}{2}\langle x - x^*, V(x - x^*)\rangle,
\end{aligned}$$

where, by the definitions of ∇g^{LJ},

$$T_4 = \int_0^1 (\beta - \alpha)\langle x_t - x^*, x - x^*\rangle dt = \frac{1}{2}(\beta - \alpha)\langle x - x^*, x - x^*\rangle,$$

$$\begin{aligned}
T_5 &= \int_0^1 \langle \nabla F(x_t) y(x_t - \beta^{-1}F(x_t); J) - \nabla F(x^*) y(x^* - \beta^{-1}F(x^*); J), x - x^*\rangle dt \\
&\quad - \int_0^1 \langle \nabla F(x_t) y(x_t - \alpha^{-1}F(x_t); L) - \nabla F(x^*) y(x^* - \alpha^{-1}F(x^*); L), x - x^*\rangle dt
\end{aligned}$$

and

$$\begin{aligned}
T_6 &= -\beta \int_0^1 \langle y(x_t - \beta^{-1}F(x_t); J) - y(x^* - \beta^{-1}F(x^*); J), x - x^*\rangle dt \\
&\quad + \alpha \int_0^1 \langle y(x_t - \alpha^{-1}F(x_t); L) - y(x^* - \alpha^{-1}F(x^*); L), x - x^*\rangle dt.
\end{aligned}$$

By noting

$$y(x^* - \alpha^{-1}F(x^*); L) = y(x^* - \beta^{-1}F(x^*); J) = x^*,$$

we have

$$\begin{aligned}
T_5 &= \int_0^1 \langle \nabla F(x^*)\nabla y(x^* - \beta^{-1}F(x^*); J)(I - \beta^{-1}\nabla F(x^*))^T(x_t - x^*), x - x^*\rangle dt \\
&\quad - \int_0^1 \langle \nabla F(x^*)\nabla y(x^* - \alpha^{-1}F(x^*); L)(I - \alpha^{-1}\nabla F(x^*))^T(x_t - x^*), x - x^*\rangle dt \\
&\quad + \int_0^1 o(\|x_t - x^*\|\|x - x^*\|)dt \\
&= \frac{1}{2}\langle \nabla F(x^*)\nabla y(x^* - \beta^{-1}F(x^*); J)(I - \beta^{-1}\nabla F(x^*))^T(x - x^*), x - x^*\rangle \\
&\quad -\frac{1}{2}\langle \nabla F(x^*)\nabla y(x^* - \alpha^{-1}F(x^*); L)(I - \alpha^{-1}\nabla F(x^*))^T(x - x^*), x - x^*\rangle \\
&\quad + o(\|x - x^*\|^2)
\end{aligned}$$

and

$$\begin{aligned}
T_6 &= -\beta\int_0^1 \langle \nabla y(x^* - \beta^{-1}F(x^*); J)(I - \beta^{-1}\nabla F(x^*))^T(x_t - x^*), x - x^*\rangle dt \\
&\quad + \alpha\int_0^1 \langle \nabla y(x^* - \alpha^{-1}F(x^*); L)(I - \alpha^{-1}\nabla F(x^*))^T(x_t - x^*), x - x^*\rangle dt \\
&\quad + \int_0^1 o(\|x_t - x^*\|\|x - x^*\|)dt \\
&= -\frac{\beta}{2}\langle \nabla y(x^* - \beta^{-1}F(x^*); J)(I - \beta^{-1}\nabla F(x^*))^T(x - x^*), x - x^*\rangle \\
&\quad +\frac{\alpha}{2}\langle \nabla y(x^* - \alpha^{-1}F(x^*); L)(I - \alpha^{-1}\nabla F(x^*))^T(x - x^*), x - x^*\rangle \\
&\quad + o(\|x - x^*\|^2).
\end{aligned}$$

Now let

$$V^* = (\beta - \alpha)I - V^*_\beta + V^*_\alpha,$$

where

$$V^*_\beta = \beta^{-1}(\beta I - \nabla F(x^*))\nabla y(x^* - \beta^{-1}F(x^*); J)(\beta I - \nabla F(x^*))^T$$

and

$$V^*_\alpha = \alpha^{-1}(\alpha I - \nabla F(x^*))\nabla y(x^* - \alpha^{-1}F(x^*); L)(\alpha I - \nabla F(x^*))^T.$$

Then we have

$$T_4 + T_5 + T_6 = \frac{1}{2}\langle x - x^*, V^*(x - x^*)\rangle + o(\|x - x^*\|^2).$$

So it follows that

$$\delta^{LJ}(x) = \frac{1}{2}\langle x - x^*, (V^* - V)(x - x^*)\rangle + o(\|x - x^*\|^2).$$

In view of (7.2), we get $\|V - V^*\| \to 0$ as $x \to x^*$. Therefore,

$$\delta^{LJ}(x) = o(\|x - x^*\|^2).$$

Since there are only finitely many L's and J's, we have

$$\begin{aligned} & g(x) - g(x^*) - \langle G(x^*), x - x^* \rangle - \frac{1}{2}\langle x - x^*, V(x - x^*)\rangle \\ &= g^{LJ}(x) - g^{LJ}(x^*) - \langle \nabla g^{LJ}(x^*), x - x^* \rangle - \frac{1}{2}\langle x - x^*, V(x - x^*)\rangle \\ &= \delta^{LJ}(x) \\ &= o(\|x - x^*\|^2), \end{aligned}$$

which proves (7.1). □

Now we can establish a superlinear convergence result of the trust region algorithm.

Theorem 7.1 *Let x^* be a solution of the VIP. Suppose that the CRCQ holds at x^* and all $V \in H_C g(x^*)$ are positive definite. Suppose also that the sequence $\{x^k\}$ generated by the trust region algorithm converges to x^* and $V_k \in H_C g(x^k)$ for all k. If the bound constraint in subproblem (5.1) is inactive for all sufficiently large k, then the sequence $\{x^k\}$ converges to x^* Q-superlinearly. Furthermore, if ∇F and all $\nabla^2 h_i$ are Lipschitz continuous, the convergence is Q-quadratic.*

Proof. From the proof of Theorem 4.1, there exists a neighborhood $N(x^*)$ of x^* such that when $x \in N(x^*)$, all $W \in H_C g(x)$ are uniformly positive definite. Because the bound constraint is inactive for all sufficiently large k, we have

$$d^k = -V_k^{-1} G(x^k)$$

when k is sufficiently large. From Theorem 4.1,

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

which means that

$$\|d^k\| = \|x^k - x^*\| + o(\|x^k - x^*\|).$$

So from Lemma 7.1, for all $W_k \in H_C g(x^k + d^k)$ we have

$$\begin{aligned} g(x^k + d^k) &= g(x^*) + \langle G(x^*), x^k + d^k - x^* \rangle \\ &\quad + \frac{1}{2}\langle x^k + d^k - x^*, W_k(x^k + d^k - x^*)\rangle + o(\|x^k + d^k - x^*\|^2) \\ &= o(\|x^k - x^*\|^2) = o(\|d^k\|^2) \end{aligned}$$

and

$$\begin{aligned} g(x^k) &= g(x^*) + \langle G(x^*), x^k - x^* \rangle + \frac{1}{2}\langle x^k - x^*, V_k(x^k - x^*)\rangle + o(\|x^k - x^*\|^2) \\ &= \frac{1}{2}\langle d^k, V_k d^k \rangle + o(\|d^k\|^2) \\ &= -\frac{1}{2}\langle G(x^k), d^k \rangle + o(\|d^k\|^2). \end{aligned}$$

Therefore, for any $\eta' \in (\eta_1, 1)$, we have

$$\begin{aligned} g(x^k + d^k) - g(x^k) - \frac{\eta'}{2}\langle G(x^k), d^k\rangle &= \frac{1-\eta'}{2}\langle G(x^k), d^k\rangle + o(\|d^k\|^2) \\ &= -\frac{1-\eta'}{2}\langle d^k, V_k^{-1} d^k\rangle + o(\|d^k\|^2) \\ &< 0, \end{aligned}$$

when k is sufficiently large. From this inequality, we can deduce that, for all k sufficiently large,

$$\rho_k = \frac{g(x^k) - g(x^k + d^k)}{g(x^k) - g_k(d^k)} > \eta_1,$$

which means

$$x^{k+1} = x^k + d^k.$$

So the superlinear (and quadratic) convergence of $\{x^k\}$ follows from Theorem 4.1. □.

Remark 7.1. Theorem 7.1 assumes that the bound constraint in subproblem (5.1) becomes inactive eventually. However, G is not continuously differentiable in general, and it does not seem possible to show that this assumption always holds, as in the smooth case. A possible remedy for this shortcoming is to use a hybrid technique: In the k-th iteration of the trust region algorithm, if V_k is positive definite, then let $d^k := -V_k^{-1} G(x^k)$ and compute ρ_k as in (5.2). If

$$\rho_k > \eta_1 \tag{7.3}$$

and

$$-\langle G(x^k), d^k\rangle > \rho\|G(x^k)\|^{2+\varepsilon} \tag{7.4}$$

for given constants $\rho, \varepsilon \in (0, \infty)$, let $x^{k+1} = x^k + d^k$ and $\Delta_{k+1} = \Delta_k$. Otherwise, solve (5.1) for d^k and continue the trust region algorithm as described in Section 5. Note that condition (7.4) was used in [3] for solving nonlinear complementarity problems, and will hold if x^k is close enough to a solution x^* and all $V \in H_C g(x^*)$ are positive definite. So superlinear convergence for such a hybrid method can be obtained. Also we may expect (6.1) to hold, because conditions (7.3) and (7.4) still guarantee a sufficient decrease in the objective value at every iteration such that $x^{k+1} \neq x^k$. Such a hybrid technique may be avoided if we adopt a line search strategy to globalize the Newton method, because we can always check the unit step size (Newton step) first. In this case, however, it may be necessary to modify matrix V_k in such a way that it becomes positive definite, thereby yielding a descent direction for the function g.

Remark 7.2. Very recently, Kanzow and Fukushima [16] refined the properties of the D-gap function for box constrained variational inequalities and presented a Gauss-Newton type algorithm with a line search strategy for minimizing the D-gap function. They tested the whole set of problems in the MCPLIB test problems with MATLAB version, see [4]. The numerical results reported in [16] show that most problems in [4] can be solved successfully. In [7], good numerical results for a similar iterative method with another merit function are also reported for variational inequalities with box constraints. The advantage of the approach adopted here may exist in that we handle the problem directly without increasing the dimension of the problem, while the approach presented in [7] is based on the KKT system involving Lagrange multipliers associated with the constraints of the problem.

Acknowledgement We are thankful to Jong-Shi Pang for drawing our attention to references [19, 20], and to two referees for their helpful suggestions.

References

[1] R. Andreani, A. Friedlander and J.S. Martínez, On the solution of finite-dimensional variational inequalities using smooth optimization with simple bounds, Preprint, Department of Applied Mathematics, University of Campinas, Campinas, Brazil (September 1995).

[2] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.

[3] T. De Luca, F. Facchinei and C. Kanzow, A semismooth equation approach to the solution of nonlinear complementarity problems, *Mathematical Programming*, to appear.

[4] S.P. Dirkse, M.C. Ferris, P.V. Preckel and T. Rutherfud, The GAMS callable program library for variational and complementarity solvers, Technical Report 94-07, Computer Sciences Department, University of Wisconsin, Madison, WI, 1994.

[5] F. Facchinei, Minimization of SC^1 functions and the Maratos effect, *Operations Research Letters*, 17 (1995), 131-137.

[6] F. Facchinei, A. Fischer and C. Kanzow, A semismooth Newton method for variational inequalities: Theoretical results and preliminary numerical experience, Preprint 102, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany, December 1995.

[7] F. Facchinei, A. Fischer and C. Kanzow, A semismooth Newton method for variational inequalities: The case of box constraints, Preprint 103, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany, December 1995.

[8] R. Fletcher, *Practical Methods of Optimization*, Wiley, New York, 1987.

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

[10] M. Fukushima, Merit functions for variational inequality and complementarity problems, *Nonlinear Optimization and Applications*, G. Di Pillo and F. Giannessi eds., (Plenum Publishing Corporation, New York, 1996), to appear.

[11] M. Fukushima and J.S. Pang, Minimizing and stationary sequences of merit functions for complementarity problems and variational inequalities, TR-IS-95026, Nara Institute of Science and Technology, Nara, Japan (September 1995).

[12] J. Han and D. Sun, Newton and quasi-Newton methods for normal maps with polyhedral sets, Manuscript, Institute of Applied Mathematics, Academia Sinica, Beijing, China (Revised version: November 1995).

[13] 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), 161–220.

[14] R. Janin, Directional derivative of the marginal function in nonlinear programming, *Mathematical Programming Study*, 21 (1984), 110-126.

[15] H. Jiang and L. Qi, Globally and superlinearly convergent trust region algorithm for convex SC^1 minimization problems and its application to stochastic programs, *Journal of Optimization Theory and Applications*, 90 (1996), 653-673.

[16] C. Kanzow and M. Fukushima, Theoretical and numerical investigation of the D-gap function for box constrained variational inequalities, Preprint 106, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany, March 1996.

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

[18] J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[19] J.V. Outrata, On optimization problems with variational inequality constraints, *SIAM Journal on Optimization*, 4 (1994), 340-357.

[20] J.V. Outrata and J. Zowe, A numerical approach to optimization problems with variational inequality constraints, *Mathematical Programming*, 68 (1995), 105-130.

[21] J.S. Pang and L. Qi, Nonsmooth equations: Motivation and algorithms, *SIAM Journal on Optimization*, 3 (1993), 443–465.

[22] J.S. Pang and L. Qi, A globally convergent Newton method for convex SC^1 minimization problems, *Journal of Optimization Theory and Applications*, 85 (1995), 633–648.

[23] J.S. Pang and D. Ralph, Piecewise smoothness, local invertibility, and parametric analysis of normal maps, *Mathematics of Operations Research*, 21 (1996).

[24] J.M. Peng, Equivalence of variational inequality problems to unconstrained optimization, Technical Report, State Key Laboratory of Scientific and Engineering Computing, Academia Sinica, Beijing, China (April 1995).

[25] M.J.D. Powell, On the global convergence of trust region algorithms for unconstrained minimization, *Mathematical Programming*, 29 (1984), 297–303.

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

[27] L. Qi, Superlinearly convergent approximate Newton methods for LC^1 optimization problems, *Mathematical Programming*, 64 (1994), 277–294.

[28] L. Qi and J. Sun, A nonsmooth version of Newton's method, *Mathematical Programming*, 58 (1993), 353–367.

[29] G.A. Shultz, R.B. Schnabel and R.H. Byrd, A family of trust-region-based algorithms for unconstrained minimization with strong global convergence properties, *SIAM Journal on Numerical Analysis*, 22 (1985), 47–67.

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

[31] N. Yamashita, K. Taji and M. Fukushima, Unconstrained optimization reformulations of variational inequality problems, *Journal of Optimization Theory and Applications*, to appear.