

## A Feasible method for solving the KKT system of variational inequalities\*

Xiaojiao TONG and Shixiao XIAO

(Received March 12, 2004; Revised February 4, 2005)

**Abstract.** In this paper the KKT system of a general variational inequality problem (denoted by  $VIP(X,F)$ ) is reformulated as a constrained optimization problem. A sufficient condition, which ensures a stationary point of the optimization problem being a solution of the KKT system of  $VIP(X,F)$ , is analyzed. A projection-type method for solving the KKT system of  $VIP(X,F)$  with closed convex set  $X$  is presented. The new algorithm has nice properties such as retaining feasibility, easy computation if the region  $X$  is a box or a ball, and strongly global and local convergence. Numerical examples show that the new algorithm is promising.

*Key words:* Variational inequality problem, KKT system, projection method, global and local convergence.

### 1. Introduction

Consider the variational inequality problem ( $VIP(X,F)$  for abbreviation), which is to find a vector  $x^* \in X$  such that

$$F(x^*)^T(x-x^*) \geq 0 \quad \text{for all } x \in X, \quad (1.1)$$

where  $X := \{x \in R^n \mid h(x) = 0, g(x) \leq 0\} \subset R^n$  is a nonempty set;  $F: U \rightarrow R^n$  is once and  $h(x): U \rightarrow R^p$  and  $g: U \rightarrow R^m$  are twice continuously differentiable defined on the open set  $U$  containing  $X$ . The following system is called the KKT system of  $VIP(X,F)$  (1.1)

$$\begin{aligned} F(x) + \nabla h(x)y + \nabla g(x)z &= 0, \\ h(x) &= 0, \\ g(x) \leq 0, z \geq 0, g(x)^T z &= 0. \end{aligned} \quad (1.2)$$

A triple  $(x, y, z) \in R^{n+p+m}$  satisfying (1.2) is called a KKT point of

---

2000 Mathematics Subject Classification : 90C30, 90C33, 65H10.

\*This work is partially supported by the National Science Foundation of China (No. 60474070) and the Hunan Provincial Natural Science Foundation of China (No. 04JJ3031, 03JJY6002), Hunan Provincial Education Foundation of China (04C1330).

VIP(X,F).

There is a strong relationship between the KKT conditions (1.2) and the solution of VIP(X,F) (see [4], [9] and [15]). Namely, if  $x^* \in X$  is a solution of VIP(X,F) and a constraint qualification holds, then multiplier vectors  $y^* \in R^p$ ,  $z^* \in R^m$  exist such that  $(x^*, y^*, z^*) \in R^{n+p+m}$  is a KKT point of VIP(X,F). Conversely, if  $h_i$  for  $i = 1, \dots, p$  are affine and  $g_i$  for  $i = 1, \dots, m$  are convex, and if  $(x^*, y^*, z^*)$  solves (1.2), then  $x^*$  solves VIP(X,F). This conclusion causes that most of the methods for solving VIP(X,F) have focused on the KKT system of VIP(X,F) (1.2).

It is well-known that, by using the Fischer-Burmeister function

$$\phi(a, b) := \sqrt{a^2 + b^2} - (a + b),$$

the system (1.2) can be reformulated equivalently as a system of nonsmooth equations

$$\begin{aligned} F(x) + \nabla h(x)y + \nabla g(x)z &= 0, \\ h(x) &= 0, \\ \Phi(-g(x), z) &= 0, \end{aligned} \tag{1.3}$$

where  $\Phi(-g(x), z) = (\phi(-g_1(x), z_1), \dots, \phi(-g_m(x), z_m))^T \in R^m$ .

Denote  $w := (x, y, z)$  and

$$\begin{aligned} L(w) &:= F(x) + \nabla h(x)y + \nabla g(x)z; \\ \Phi(w) &:= \Phi(-g(x), z); \\ H(w) &:= \begin{pmatrix} L(w) \\ h(x) \\ \Phi(w) \end{pmatrix}. \end{aligned}$$

Then we rewrite system (1.3) as

$$H(w) = 0. \tag{1.4}$$

By using the natural merit function, i.e.,

$$\Psi(w) := \frac{1}{2} \|H(w)\|^2,$$

a popular method for solving (1.4) is equivalent to finding a solution of the following unconstrained optimization problem

$$\min \Psi(w). \tag{1.5}$$

This approach has been studied by many authors (see [12], [10], [6], [4] and references therein).

It is well-known that for optimization problems, most of the methods can only find a stationary point, therefore it is required to provide some conditions which ensure a stationary point of a reformulated optimization problem being a solution of the system (1.4). We call the condition *Stationary Condition*. Some papers have discussed this problem (see [1], [4]–[7], [13], [14]). For general VIP(X,F) (1.1), Facchinei et al. [4] gave the stationary condition ensuring that the stationary point of the unconstrained optimization problem (1.5) is a solution of the system (1.4) (see Theorem 4.3 in [4]). Following the work of [4], Facchinei et al. [5] found that, even for a strongly monotone VIP(X,F) with a convex set defined by nonlinear inequalities, one cannot ensure that the method converges to the unique solution of the VIP(X,F). Based on this observation, [5] presented a simply constrained optimization reformulation for the KKT system of VIP(X,F) where the equivalent problem is

$$\min \Psi(w), \quad \text{subject to } z \geq 0. \quad (1.6)$$

Then a suitable algorithm is given, and the stationary condition is studied in [5].

We note that both of the reformulation (1.5) and (1.6) require that  $F(x)$  is well defined on the whole space  $R^n$ . This requirement for  $F$  is not satisfied in many VIP(X,F) arising from economic equilibrium problems (see [3] and [8]). Moreover, even if  $F$  is defined on the whole space  $R^n$ , some important properties of  $F$  (for example, the monotone character) holds on  $X$  and may not hold outside  $X$ . These observations lead us to focus on feasible methods, i.e., we take into account a more generally constrained optimization reformulation of the KKT system of VIP(X,F) as follows:

$$\begin{aligned} \min \Psi(w) &= \frac{1}{2} \|H(w)\|^2 \\ \text{s.t. } h(x) &= 0, \\ g(x) &\leq 0, \\ z &\geq 0. \end{aligned} \quad (1.7)$$

Based on the reformulation (1.7), two questions arise:

(1) what conditions ensure that a stationary point of (1.7) is a solution of the KKT system (1.3)?

(2) problem (1.7) is only a  $C^1$  optimization problem, how to design a suitable algorithm for solving it?

In the next section, we shall first study the stationary condition of (1.7), a sufficient condition is given and the first question above is answered. In Section 3, for the case that  $X$  is a closed convex set, we present a projection-type algorithm and show its convergence. Section 4 provides some preliminary numerical examples to test the new algorithm. Some conclusions are drawn in the last section.

Some words about notations: If  $F: R^n \supset U \rightarrow R^n$  is a differentiable function, the Jacobian of  $F$  at a point  $x \in R^n$  is denoted by  $F'(x)$ , whereas  $\nabla F(x)$  is the transposed Jacobian. Throughout this paper  $\|\cdot\|$  denotes the Euclidean norm.  $P_\Omega(w)$  denotes the projection of vector  $w$  in the convex set  $\Omega$ . A superscript  $k$  such as  $w^k$  refers to a specific vector and usually denotes an iteration index. If  $v \in R^m$  and  $J \subseteq \{1, 2, \dots, m\}$  is a subindex set,  $v_J$  denotes the subvector with elements in  $i \in J$ .

Let  $\varphi: R^n \supset D \rightarrow R^n$  be a locally Lipschitzian vector function,  $D_\varphi$  denotes the set of points where  $\varphi$  is differentiable. Then the B-subdifferential of  $\varphi$  at  $x \in D$  is defined to be

$$\partial_B \varphi(x) = \left\{ \lim_{\substack{x^k \rightarrow x \\ x^k \in D_\varphi}} \nabla \varphi(x^k)^T \right\}, \quad (1.8)$$

whereas Clarke's generalized Jacobian of  $\varphi$  at  $x$  [2] is defined to be

$$\partial \varphi(x) = \text{conv } \partial_B \varphi(x). \quad (1.9)$$

The following concepts can be found in [15] and [16].

$\varphi$  is called *semismooth* at  $x$  if  $\varphi$  is directionally differentiable at  $x$  and for all  $V \in \partial \varphi(x+h)$  and  $h \rightarrow 0$ ,

$$\varphi(x+h) - \varphi(x) = Vh + o(\|h\|); \quad (1.10)$$

$\varphi$  is called *strongly semismooth* at  $x$  if for all  $V \in \partial \varphi(x+h)$  and  $h \rightarrow 0$ ,

$$\varphi(x+h) - \varphi(x) = Vh + O(\|h\|^2); \quad (1.11)$$

A vector  $x^*$  is called *BD-regular* for  $\varphi(x)$  if all elements  $V \in \partial_B \varphi(x^*)$  are nonsingular.

## 2. Stationarity Conditions

In this section, we shall study conditions ensuring that a stationary point of (1.7) is a solution of (1.3) or (1.4).

First we have the following properties for the nonsmooth function  $H(w)$  and the merit function  $\Psi(w)$  (see Proposition 3.1 and Proposition 4.1 in [4]).

**Lemma 2.1** *Suppose that  $F$  is  $C^1$ -function and  $h$  and  $g$  are  $C^2$ -functions. Then*

(i) *each element  $V \in \partial H(w)$  can be represented as*

$$V = \begin{pmatrix} \nabla_x L(w) & \nabla h(x) & \nabla g(x) \\ \nabla h(x)^T & 0 & 0 \\ -D_a(w)\nabla g(x)^T & 0 & D_b(w) \end{pmatrix}, \quad (2.1)$$

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

$$a_i(w) = \begin{cases} \frac{-g_i(x)}{\sqrt{(-g_i(x))^2 + z_i^2}} - 1, & \text{if } (-g_i(x), z_i) \neq 0, \\ a_i(w) = \xi_i - 1, & \text{otherwise;} \end{cases}$$

$$b_i(w) = \begin{cases} \frac{z_i}{\sqrt{(-g_i(x))^2 + z_i^2}} - 1, & \text{if } (-g_i(x), z_i) \neq 0, \\ \eta_i - 1, & \text{otherwise,} \end{cases}$$

where  $(\xi_i, \eta_i)$  satisfies  $\|(\xi_i, \eta_i)\| \leq 1$ .

(ii)  $\Psi(w)$  is continuously differentiable and for  $V \in \partial H(w)$ , it holds

$$\nabla \Psi(w) = V^T H(w). \quad (2.2)$$

We review that a point  $(w, \nu, \mu)$  with  $w \in \Omega := \{w \in \mathbb{R}^{n+p+m} \mid x \in X, z \geq 0\}$  is a stationary point of the problem (1.7) if  $(w, \nu, \mu)$  satisfies

$$\begin{aligned} \nabla_x \Psi(w) + \nabla h(x)\nu + \nabla g(x)\mu &= 0, \\ \nabla_y \Psi(w) &= 0, \\ h(x) &= 0, \\ z_i > 0, &\implies (\nabla_z \Psi(w))_i = 0, \quad i \in \mathcal{I}, \\ z_i = 0, &\implies (\nabla_z \Psi(w))_i \geq 0, \quad i \in \mathcal{I}, \\ g_i(x) \leq 0, \mu_i \geq 0, &g_i(x)\mu_i = 0, \quad i \in \mathcal{I}, \end{aligned} \quad (2.3)$$

where  $\mathcal{I} = \{1, \dots, m\}$ . From (2.1) and (2.2), (2.3) can be rewritten as

$$\begin{aligned}
& \nabla_x L(w)L(w) + \nabla h(x)h(x) - \nabla g(x)D_a(w)\Phi(w) \\
& \quad + \nabla h(x)\nu + \nabla g(x)\mu = 0, \\
& \quad \nabla h(x)^T L(w) = 0, \\
& \quad h(x) = 0, \tag{2.4} \\
& z_i > 0, \implies (\nabla g(x)^T L(w) + D_b(w)\Phi(w))_i = 0, \quad i \in \mathcal{I}, \\
& z_i = 0, \implies (\nabla g(x)^T L(w) + D_b(w)\Phi(w))_i \geq 0, \quad i \in \mathcal{I}, \\
& \quad g_i(x) \leq 0, \quad \mu_i \geq 0, \quad g_i(x)\mu_i = 0, \quad i \in \mathcal{I}.
\end{aligned}$$

Denote

$$\begin{aligned}
I_z^0 &= \{i \in I \mid z_i = 0\}, & I_z^+ &= \{i \in I \mid z_i > 0\}, \\
I_\mu^0 &= \{i \in I \mid \mu_i = 0\}, & I_\mu^+ &= \{i \in I \mid \mu_i > 0\}. \tag{2.5}
\end{aligned}$$

Then we have the following main theorem.

**Theorem 2.1** *Let  $(w^*, \nu^*, \mu^*)$  be a stationary point of (1.7). Assume that*

- (a)  $\nabla_x L(w^*)$  is positive semidefinite on  $R^n$ ;
- (b)  $\nabla_x L(w^*)$  is positive definite on the cone

$$\begin{aligned}
\mathcal{C}(w^*, \mu^*) &= \{v \in R^n \mid \nabla h(x^*)^T v = 0, \nabla g_{I_z^+}(x^*)^T v = 0, \\
& \quad \nabla g_{I_z^0}(x^*)^T v \geq 0, (\nabla g_{I_z^0}(x^*)^T v)^T \mu_{I_z^0}^* = 0\}.
\end{aligned}$$

Then  $w^*$  is a solution of the KKT system (1.3).

*Proof.* We drop the symbol  $*$  for simplicity in the process of proof.

Multiply the first expression of (2.4) by  $L(w)^T$ . Then from the definition of (2.5), (2.4) follows

$$\begin{aligned}
& L(w)^T \nabla_x L(w)L(w) - L(w)^T \nabla g(x)D_a(w)\Phi(w) \\
& \quad + L(w)^T \nabla g(x)\mu = 0, \\
& \quad \nabla h(x)^T L(w) = 0, \\
& \quad h(x) = 0, \tag{2.6} \\
& (\nabla g(x)^T L(w) + D_b(w)\Phi(w))_{I_z^+} = 0, \\
& (\nabla g(x)^T L(w) + D_b(w)\Phi(w))_{I_z^0} \geq 0, \\
& \quad g(x) \leq 0, \quad \mu \geq 0, \quad g(x)^T \mu = 0.
\end{aligned}$$

From the definition of  $\Phi(w) = \Phi(-g(x), z)$  and  $g(x) \leq 0$  we obtain

$$\phi(-g_i(x), z_i) = 0, \quad \text{for } i \in I_z^0. \quad (2.7)$$

So (2.6) can be expressed as

$$\begin{aligned} L(w)^T \nabla_x L(w) L(w) - L(w)^T \nabla g(x) D_a(w) \Phi(w) \\ + L(w)^T \nabla g(x) \mu = 0, \\ \nabla h(x)^T L(w) = 0, \\ h(x) = 0, \\ (\nabla g(x)^T L(w) + D_b(w) \Phi(w))_{I_z^+} = 0, \\ (\nabla g(x)^T L(w))_{I_z^0} \geq 0, \\ g(x) \leq 0, \quad \mu \geq 0, \quad g(x)^T \mu = 0. \end{aligned} \quad (2.8)$$

Now we pay attention to the first expression of (2.8) and analyze some terms in it.

$$\begin{aligned} -L(w)^T \nabla g(x) D_a(w) \Phi(w) &= -(\nabla g(x)^T L(w))^T (D_a(w) \Phi(w)) \\ &= -(\nabla g(x)^T L(w))_{I_z^0}^T (D_a(w) \Phi(w))_{I_z^0} \\ &\quad - (\nabla g(x)^T L(w))_{I_z^+}^T (D_a(w) \Phi(w))_{I_z^+} \\ &= -(\nabla g(x)^T L(w))_{I_z^+}^T (D_a(w) \Phi(w))_{I_z^+} \\ &= (D_b(w) \Phi(w))_{I_z^+}^T (D_a(w) \Phi(w))_{I_z^+} \quad (2.9) \\ &= \Phi_{I_z^+}(w)^T (D_b(w) D_a(w))_{I_z^+} \Phi_{I_z^+}(w) \\ &= \Phi(w)^T (D_b(w) D_a(w)) \Phi(w), \end{aligned}$$

where the third and last equality are due to (2.7), the fourth equality comes from the fourth expression of (2.8).

Using a similar analysis, we can obtain

$$\begin{aligned} L(w)^T \nabla g(x) \mu &= (\nabla g(x)^T L(w))_{I_z^0}^T \mu_{I_z^0} + (\nabla g(x)^T L(w))_{I_z^+}^T \mu_{I_z^+} \\ &= (\nabla g(x)^T L(w))_{I_z^0}^T \mu_{I_z^0} - (D_b(w) \Phi(w))_{I_z^+}^T \mu_{I_z^+}. \end{aligned} \quad (2.10)$$

On the other hand,  $I_z^+$  can be written as

$$I_z^+ = (I_z^+ \cap I_\mu^0) \cup (I_z^+ \cap I_\mu^+).$$

We consider two cases:

- (i)  $i \in I_z^+ \cap I_\mu^0$ . It is obviously that  $(D_b(w)\Phi(w))_i \mu_i = 0$ ;  
(ii)  $i \in I_z^+ \cap I_\mu^+$ . From the last term of (2.8) we have  $g_i(x) = 0$ , which results  $\phi(-g_i(x), z_i) = 0$ . This implies  $(D_b(w)\Phi(w))_i \mu_i = 0$ .

Therefore we have  $(D_b(w)\Phi(w))_{I_z^+}^T \mu_{I_z^+} = 0$ . (2.10) becomes

$$L(w)^T \nabla g(x) \mu = (\nabla g(x)^T L(w))_{I_z^0}^T \mu_{I_z^0}, \quad (2.11)$$

which together with (2.9) and the first expression of (2.8) yields

$$L(w)^T \nabla_x L(w) L(w) + \Phi(w)^T (D_b(w) D_a(w)) \Phi(w) + (\nabla g(x)^T L(w))_{I_z^0}^T \mu_{I_z^0} = 0. \quad (2.12)$$

From the condition (a) of theorem, the definition of  $D_a(w)$ ,  $D_b(w)$ ,  $\Phi(w)$  and the last two expressions of (2.8), we know that each term above is nonnegative. Then it follows

$$\begin{aligned} L(w)^T \nabla_x L(w) L(w) &= 0, \\ \Phi(w)^T (D_b(w) D_a(w)) \Phi(w) &= 0, \\ (\nabla g(x)^T L(w))_{I_z^0}^T \mu_{I_z^0} &= 0. \end{aligned} \quad (2.13)$$

Moreover, from the definition of  $a_i(w)$  and  $b_i(w)$  we can derive that

$$a_i(w) = 0 \text{ or } b_i(w) = 0 \implies \phi(-g_i(x), z_i) = 0, \quad (2.14)$$

which together with the second expression of (2.13) yields

$$\Phi(w) = 0. \quad (2.15)$$

Note that for each  $J \subseteq \{1, \dots, m\}$ ,  $(\nabla g(x)^T L(w))_J = \nabla g_J(x)^T L(w)$ . From (2.13) and (2.15), (2.8) is simplified as

$$\begin{aligned} L(w)^T \nabla_x L(w) L(w) &= 0, \\ (\nabla g_{I_z^0}(x)^T L(w))_{I_z^0}^T \mu_{I_z^0} &= 0, \\ \nabla h(x)^T L(w) &= 0, \\ h(x) &= 0, \\ \nabla g_{I_z^+}(x)^T L(w) &= 0, \\ \nabla g_{I_z^0}(x)^T L(w) &\geq 0, \\ g(x) \leq 0, \mu \geq 0, g(x)^T \mu &= 0. \end{aligned} \quad (2.16)$$

(2.16) shows that  $L(w) \in \mathcal{C}(w, \mu)$ . From the assumption (b) of the theorem, we have

$$L(w) = 0. \tag{2.17}$$

In addition,  $h(x) = 0$  is included in (2.16), which combining with (2.15) and (2.17) implies

$$H(w) = 0.$$

This result shows that  $w$  is a solution of the KKT system (1.3). □

**Remark 2.1** *Theorem 2.1 is similar to Theorem 4.3 [4] and Theorem 3.1 [5] where the former gave some conditions ensuring that a stationary point of the unconstrained optimization (i.e., problem (1.5)) is the solution of (1.4), and the latter gave conditions ensuring that a stationary point of simple constrained optimization (problem (1.6)) is the solution of (1.4). Note that both problems (1.6) and (1.7) require  $z \geq 0$ , and there exists a relationship between the cone  $\mathcal{C}(x^*)$  defined in Theorem 3.1 [5] and the cone  $\mathcal{C}(w^*, \mu^*)$  defined in Theorem 2.1 as  $\mathcal{C}(w^*, \mu^*) \subset \mathcal{C}(x^*)$ . Hence, the assumed condition in Theorem 2.1 is weaker than one in Theorem 3.1 [5].*

Next we consider the special VIP(X,F) where  $X$  is a box or ball. From Theorem 2.1 it is not difficult to get the following corollary. Note that for this case, solving the KKT system and VIP(X,F) is equivalent.

**Corollary 2.1** *Let  $X$  be  $X = \{x \in R^n \mid b \geq x \geq a\}$  with  $a, b \in R^n$ ,  $b > a$  or  $X = \{x \in R^n \mid \|x\| \leq r\}$  with  $r > 0$ , and  $(w^*, \mu^*)$  be a stationary point of (1.7). Assume that*

- (a)  $F(x)$  is monotone,
- (b)  $\nabla_x F(x)$  is positive definite on the cone  $\mathcal{C}(w^*, \mu^*)$ .

*Then  $x^*$  is a solution of VIP(X,F) .*

### 3. Projected Algorithm

In this section we present a suitable algorithm for solving the system (1.7) with a closed convex set  $X$ . Obviously, if  $X$  is a closed convex set, so is the set  $\Omega$  defined by

$$\Omega = \{w \in R^{n+p+m} \mid h(x) = 0, g(x) \leq 0, z \geq 0\}.$$

The new method proposed in this section is motivated by the recent

works of [18] and [21]. A key technique of these algorithms in [18, 21] is the the choice of the search direction. By using an optimal combination of the projected gradient direction and the projected trust region direction, we obtain a feasible and descent direction. It is proved that by using the combination direction, the algorithm has nice globally and locally convergent property.

Next we give the outline of the algorithm. Let  $w^k$  be the current point. We omit the superscript  $k$  and denote the current point as  $w$  for simplicity. Let  $\Delta > 0$ ,  $\Delta_{\max} \geq \Delta$ ,  $\gamma > 0$  be some given constants, which are correspondent to the point  $w$ . The search direction is obtained by three steps as follows:

**(1) Compute a projected gradient direction  $\bar{d}_G(\Delta)$**

Compute the projected gradient direction by

$$\begin{aligned} d_G(\Delta) &:= -\frac{\Delta}{\Delta_{\max}}\gamma\nabla\Phi(w), \\ \bar{d}_G(\Delta) &:= P_{\Omega}[w + d_G(\Delta)] - w. \end{aligned} \quad (3.1)$$

$\bar{d}_G(\Delta)$  has some nice properties such as feasibility and descent. However, it can not ensure locally superlinear convergence. Therefore we shall combine the projected trust region direction to speed up the convergent rate.

**(2) Compute a projected trust region direction  $\bar{d}_T(\Delta)$**

Choose  $V \in \partial H(w)$ . Let  $d_T(\Delta)$  be a solution of the trust region subproblem

$$\begin{aligned} \min \quad & \frac{1}{2}\|H(w) + Vd\|^2 \triangleq q(d) \\ \text{s.t.} \quad & \|d\| \leq \Delta. \end{aligned} \quad (3.2)$$

Then the projected trust region direction is computed by

$$\bar{d}_T(\Delta) := P_{\Omega}[w + d_T(\Delta)] - w. \quad (3.3)$$

**(3) Compute an optimal search direction  $\bar{d}(\Delta)$**

Consider that the projected trust region direction  $\bar{d}_T(\Delta)$  may not be a descent direction of the merit function  $\Psi(w)$  when a point  $w$  is far from a solution, whereas  $\bar{d}_G(\Delta)$  is a descent direction of  $\Psi(w)$ . Hereby, we shall consider an optimal combination of the two directions.

Define the search direction by

$$\bar{d}(\Delta) := t^*(\Delta)\bar{d}_G(\Delta) + (1-t^*(\Delta))\bar{d}_T(\Delta), \quad (3.4)$$

where  $t^*(\Delta) \in (0, 1)$  is an optimal solution of the following problem

$$\min_{t \in [0,1]} \frac{1}{2} \|H(w) + V[t\bar{d}_G(\Delta) + (1-t)\bar{d}_T(\Delta)]\|^2 \triangleq q_\Delta(t). \quad (3.5)$$

Let  $t(\Delta)$  be a solution of  $\nabla q_\Delta(t) = 0$ . Then for any  $\Delta > 0$ , we have

$$t(\Delta) = \begin{cases} -\frac{[H(w) + V\bar{d}_T(\Delta)]^T V[\bar{d}_G(\Delta) - \bar{d}_T(\Delta)]}{\|V[\bar{d}_G(\Delta) - \bar{d}_T(\Delta)]\|^2}, & \text{if } V\bar{d}_G(\Delta) \neq V\bar{d}_T(\Delta), \\ \text{any number in } (-\infty, +\infty), & \text{if } V\bar{d}_G(\Delta) = V\bar{d}_T(\Delta). \end{cases} \quad (3.6)$$

It is not difficult to deduce the following result for the solution of (3.5).

**Lemma 3.1** *Suppose that  $w \in \Omega$ . Then for any  $\Delta > 0$ , the optimal solution  $t^*(\Delta)$  of (3.5) is*

$$t^*(\Delta) = \max\{0, \min\{1, t(\Delta)\}\}, \quad (3.7)$$

where  $t(\Delta)$  is defined by (3.6).

We now state the new algorithm as follows.

**Algorithm 3.1** (Projection Algorithm)

**Step 0.** Choose  $w^0 \in \Omega$ , give constants  $\alpha_1, \alpha_2, \rho_1, \rho_2, \sigma \in (0, 1)$ ,  $\eta, \Delta_{\min}, \Delta_{\max}$  and  $\Delta_0$  be such that  $0 < \alpha_1 < 1 < \alpha_2$ ,  $0 < \rho_1 < \rho_2 < 1$ ,  $\eta \in (0, 1)$ ,  $\Delta_0 > 0$ ,  $\Delta_{\max} > \Delta_{\min} > 0$ , set  $k := 0$ .

**Step 1.** If  $w^k$  is a stationary point of problem (1.7) stop.

Otherwise let  $\Delta^k := \min\{\Delta_{\max}, \max\{\Delta_{\min}, \Delta_k\}\}$ ,  $\hat{\Delta} := \Delta^k$  and choose  $V_k \in \partial H(w^k)$ .

**Step 2.** Solve the trust region subproblem (3.2) by a suitable algorithm.

Denote the solution by  $d_T^k(\hat{\Delta})$ .

**Step 3.** Let

$$\gamma_k := \min\left\{1, \frac{\Delta_{\max}}{\|\nabla \Psi(w^k)\|}, \eta \frac{\|H(w^k)\|}{\|\nabla \Psi(w^k)\|}, \eta \frac{\Psi(w^k)}{\|\nabla \Psi(w^k)\|^2}\right\}, \quad (3.8)$$

compute  $\bar{d}_G^k(\hat{\Delta}), \bar{d}_T^k(\hat{\Delta}), t_k^*(\hat{\Delta})$  by (3.1), (3.3) and (3.7). Let

$$\bar{d}^k(\hat{\Delta}) := t_k^*(\hat{\Delta})\bar{d}_G^k(\hat{\Delta}) + (1 - t_k^*(\hat{\Delta}))\bar{d}_T^k(\hat{\Delta}). \quad (3.9)$$

**Step 4.** Compute

$$\hat{r}_k := \frac{\Psi(w^k + \bar{d}^k(\hat{\Delta})) - \Psi(w^k)}{(1/2)\|H(w^k) + V_k \bar{d}^k(\hat{\Delta})\|^2 - \Psi(w^k)}. \quad (3.10)$$

If

$$\begin{aligned} \Psi(w^k) - \frac{1}{2}\|H(w^k) + V_k \bar{d}^k(\hat{\Delta})\|^2 &\geq -\sigma \nabla \Psi(w^k)^T \bar{d}_G^k(\hat{\Delta}) \\ \hat{r}_k &\geq \rho_1, \end{aligned} \quad (3.11)$$

let  $s^k = \bar{d}^k(\hat{\Delta})$ ,  $w^{k+1} := w^k + s^k$ ,  $\delta_k = \hat{\Delta}$ . Set

$$\Delta_{k+1} := \begin{cases} \hat{\Delta}, & \text{if } \rho_1 \leq \hat{r}_k < \rho_2, \\ \alpha_2 \hat{\Delta}, & \text{if } \hat{r}_k \geq \rho_2; \end{cases}$$

Let  $k := k + 1$  return to step 1.

Otherwise set  $\hat{\Delta} := \alpha_1 \hat{\Delta}$  and repeat to do step 2.

**Remark 3.1**  $\Delta^k$  expresses the initial trust region radius at each iteration, and  $\delta_k$  is the radius of the corresponding trial step  $s^k$ . This also implies that for all  $k$ ,  $\Delta^k \geq \Delta_{\min}$ . The trust region subproblem (3.2) can be solved approximately by many existing methods such as the truncated conjugate gradient method (see [20] and [24]).

**Remark 3.2** Algorithm 3.1 is designed based on the case where the projection on  $X$  can be carried out easily. We know that, except for some special cases such as box and ball constraints, the projected calculation onto a general closed convex set  $X$  is expensive and is still an open question. This difficulty also restricts the application of Algorithm 3.1 for some VIP(X,F) problems.

Algorithm 3.1 is an extension of the algorithm proposed in [18] in which the constraint is a box. Therefore we can obtain the following convergence results. Here we omit the detailed proofs and refer to [18].

**Theorem 3.1** *Let  $X$  be a closed convex set and  $\{w^k\}$  be generated by Algorithm 3.1. Then any accumulation point  $w^*$  of  $\{w^k\}$  is a stationary point of (1.7). Moreover, if  $w^*$  satisfies the conditions of Theorem 2.1,  $w^*$  is a solution of the KKT system (1.3).*

*Proof.* The first conclusion is proved as Theorem 4.1 in [18]. The second conclusion follows from Theorem 2.1 directly.  $\square$

Next theorem shows the locally superlinear convergence of Algorithm 3.1. In reference [18], it is proved that for  $w^k$  sufficiently close to a solution  $w^*$ , the trust region direction solved by (3.2) is reduced to the Newton direction (see Lemma 5.4 in [18]). This conclusion implies that the trust region constraint becomes inactive near the solution. Furthermore, Lemma 5.7 in [18] proved that  $t_k^*(\Delta^k) = o(1)$ . Hence, from (3.4), the search direction used in this paper is reduced asymptotically to the projected Newton direction, then we have the following superlinear convergence (See Theorem 5.2 in [18] for the process of proof).

**Theorem 3.2** *Suppose that  $w^*$  is an accumulation point of the sequence  $\{w^k\}$  generated by Algorithm 3.1 and a BD-regular solution of the KKT system (1.3). Then the whole sequence  $\{w^k\}$  converges superlinearly. Moreover, if  $\nabla F$ ,  $\nabla^2 h_i$  ( $i = 1, \dots, p$ ) and  $\nabla^2 g_i$  ( $i = 1, \dots, m$ ) are locally Lipschitzian, the convergent rate is quadratic.*

#### 4. Preliminary Numerical Experiments

In this section, we present some preliminary numerical examples for Algorithm 3.1, and the problems are extracted from some references as follows.

**Example 1** The problem comes from Ralph and Wright problem 3 in [19], which is a convex optimization problem. The corresponding VIP problem (1.1) is

$$F(x) = \begin{pmatrix} 2x_1 + x_2 + 1 \\ x_1 + 4x_2 + 1 \end{pmatrix} \tag{4.1}$$

$$X := \{x \in \mathbb{R}^2 \mid x_1 \geq 0, x_2 \geq 0, (x_1 - 2)^2 + (x_2 - 1)^2 \leq 5\}$$

**Example 2** This example is from Taji, Fukushima and Ibaraki in [22] except that the polyhedron constraint is replaced by a ball constraint in the feasible region, i.e.,

$$F(x) = \begin{pmatrix} 0.726 & -0.949 & 0.266 & -1.193 & -0.504 \\ 1.645 & 0.678 & 0.333 & -0.217 & -1.443 \\ -1.016 & -0.225 & 0.769 & 0.934 & 1.007 \\ 1.063 & 0.587 & -1.144 & 0.550 & -0.548 \\ -0.256 & 1.453 & -1.073 & 0.509 & 1.026 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}$$

$$+ 10 \begin{pmatrix} \arctan(x_1 - 2) \\ \arctan(x_2 - 2) \\ \arctan(x_3 - 2) \\ \arctan(x_4 - 2) \\ \arctan(x_5 - 2) \end{pmatrix} + \begin{pmatrix} 5.308 \\ 0.008 \\ -0.938 \\ 1.024 \\ -1.312 \end{pmatrix} \quad (4.2)$$

and

$$X := \left\{ x \in R^5 \mid \sum_{i=1}^5 (x_i - 2)^2 \leq 5 \times 2^2, x_i \geq 0, i = 1, \dots, 5 \right\}. \quad (4.3)$$

**Example 3** The problem is the minimization of a sum of Euclidean norms:

$$\min_{x \in R^n} \sum_{i=1}^m \|b_i - A_i^T x\|, \quad (4.4)$$

where  $b_1, b_2, \dots, b_m \in R^d$  are column vectors in the Euclidean  $d$ -space,  $A_1, A_2, \dots, A_m \in R^{n \times d}$  are  $n$ -by- $d$  matrices. Problem (4.4) is a convex programming problem, but its objective function is not differentiable at any point  $x$  with  $b_i - A_i^T x = 0$  for some  $i$ . We have the following conclusion for problem (4.4) (see Lemma 5.1 and 5.2 in [17]).

**Lemma 4.1**  $x^*$  is a solution of (4.4) if and only if  $u^* = (x^*, y^*)$  satisfies

$$F(u^*)^T (u - u^*) \geq 0 \quad \text{for all } u \in X, \quad (4.5)$$

where  $y^* = (y_1^*, \dots, y_m^*) \in R^{md}$  with  $y_i^* \in R^d$  ( $i = 1, \dots, m$ ),  $X = R^n \times B \times \dots \times B \subset R^{n+md}$  with  $B = \{s \in R^d \mid \|s\| \leq 1\}$ ,

$$F(u) = \begin{pmatrix} Ay \\ A^T x - b \end{pmatrix},$$

$$A = [A_1, A_2, \dots, A_m], \quad b^T = [b_1^T, b_2^T, \dots, b_m^T].$$

In (4.4), the dimension of the reformulated problem (4.5) is  $n + md$ , and  $X$  can be written as

$$X := \left\{ u = (x, y) \in R^{n+md} \mid g_i(u) = \|y_i\|^2 - 1 \leq 0, \right. \\ \left. y_i \in R^d, i = 1, \dots, m \right\}. \quad (4.6)$$

(4.4) can be reformulated as the second order cone program

$$\begin{aligned} \min_{x \in R^n, y \in R^m} \quad & \sum_{i=1}^m y_i \\ \text{s.t.} \quad & \|b_i - A_i^T x\| \leq y_i \quad (i = 1, \dots, m) \end{aligned}$$

with dimension  $m + n$ . Then the dimension of the reformulated problem is significantly reduced compared with problem (4.5). Here the motivation for choosing reformulation (4.5) is to convert the problem (4.4) into VIP(X,F) (1.1) so that Algorithm 3.1 can be tested.

In our numerical example (4.4), for  $i = 1, \dots, m$ ,  $A_i \in R^{n \times d}$  is a matrix with elements

$$(A_i)_{kl} = \begin{cases} 1, & \text{if } k = l \\ 0, & \text{otherwise,} \end{cases}$$

and each component of  $b_i$  is selected randomly from a uniform distribution of  $(-5, 5)$ .

The parameters used in Algorithm 3.1 are set

$$\begin{aligned} \alpha_1 = 0.5, \quad \alpha_2 = 2, \quad \rho_1 = 0.0001, \quad \rho_2 = 0.75, \quad \eta = 0.9, \quad \sigma = 0.5, \\ \Delta_0 = 5, \quad \Delta_{\min} = 0.0001, \quad \Delta_{\max} = 10 \end{aligned}$$

We use  $\min\{\Psi(w^k), \|\nabla\Psi(w^k)\|\} \leq 1.0e^{-10}$  as the stopping rule. The trust region subproblem in Algorithm 3.1 is solved by the truncated preconditioned conjugate gradient method proposed in [20] and [24]. All examples were done on a **P-III** with MATLAB code.

For Example 1 and Example 2, the starting points  $x^0$  are chosen randomly in  $(0, 1)$ , and  $z^0$  is set  $z^0 = e$ , where  $e$  denotes the vector of all ones. We test these two examples for 10 times with random  $x^0$ , and the results listed in Table 4.1 are the average value of 10 times. We report results of Example 3 in Table 4.2, in which  $\mathbf{m}$  denotes the number of terms in sum;  $\mathbf{d}$  denotes the dimension of  $b_i$ ;  $\mathbf{n}$  indicates the dimension of  $x$ . The starting point of Example 3 is chosen as  $u^0 = (x^0, y^0) = 0$ ,  $z^0 = 0.5e$ .

In Table 4.1 and Table 4.2, **Iter** denotes the number of iterations, which is also equal to the number of Jacobian calculations of  $F(u)$ ; **NF** indicates the number of function calculations, the number of solving the trust region subproblem then is  $NF - 1$ ;  $\Psi(w^f)$  denotes the value of  $\Psi(w)$  at the final iterate;  $\|\nabla\Psi(w^f)\|$  denotes the value of  $\|\nabla\Psi(w)\|$  at the final iterate;

Table 4.1. (Computing Results of Example 1 and Example 2)

Example 1.	Iter	NF	$\Psi(w^f)$	$\ \nabla\Psi(w^f)\ $	$t^*(\Delta)_{ave}$
1	3.5	4.5	2.1843e-25	2.0970e-12	0.1701
2	4.2	5.2	1.1204e-11	7.9197e-06	0.0017

Table 4.2. (Computing Results of Example 3)

m	d	n	Iter	NF	$\Psi(w^f)$	$\ \nabla\Psi(w^f)\ $	$t^*(\Delta)_{ave}$
5	50	100	7	8	8.8409e-15	3.6800e-06	0.0403
	100	100	9	10	1.2513e-13	1.6713e-05	0
	200	100	20	30	2.5651e-13	3.9582e-05	0.3343
	100	1000	9	10	9.3376e-12	1.0853e-04	5.9008e-05
10	10	100	6	7	2.9888e-13	1.1463e-05	0.2441
	50	100	9	11	5.0705e-13	2.0286e-05	0.1962
	50	500	9	11	1.0827e-13	1.4576e-05	0.0394
	100	1000	16	28	1.1550e-11	1.8243e-04	0.4568
20	50	100	10	11	6.1714e-12	9.9883e-05	0.1369
	50	1000	10	11	1.1920e-12	5.0466e-05	0.0655
	100	100	11	12	7.0359e-13	4.5551e-05	0.0635
	100	1000	11	12	2.3270e-12	8.7587e-05	0.0781

$t^*(\Delta)_{ave}$  denotes the average of all  $t_k^*(\Delta^k)$ .

We give some explanations to  $t^*(\Delta)_{ave}$ . From the locally convergent analysis we know that under some assumptions, the value of  $t_k^*(\Delta^k)$  specifies the superlinear property of Algorithm 3.1, i.e.,  $t_k^*(\Delta^k) = o(1)$  means that the search direction  $\bar{d}_T(\Delta^k)$  is used eventually and the superlinear conclusion is obtained. Note that  $t^*(\Delta)_{ave}$  has a different meaning with  $t_k^*(\Delta^k)$ . The former, i.e.  $t^*(\Delta)_{ave}$ , expresses the average value of  $t_k^*(\Delta^k)$  during the total iterates, whereas the latter is the weight coefficient of the search direction (see (3.9)) at the  $k$ th iterate. The value  $t^*(\Delta)_{ave}$  just shows which direction (the projected gradient direction or the projected trust region direction) is used mostly in all iterates. In other words, if  $t^*(\Delta)_{ave}$  is close to zero, the projected trust region direction is used most of the time in the total iterates, otherwise, if it is close to one, then the projected gradient direction is used most of the time.

Table 4.1 and Table 4.2 show that our algorithm is promising. The computing effect is represented by the iterative number of examples. We also note that for all examples, the average value  $t^*(\Delta)_{ave}$  is small, which

implies that the trust region direction was used very often. This feature also shows that the proposed method is computationally better than the projected gradient method.

We note that Example 1 and Example 2 are strongly monotone variational inequality problems in the whole space, so the conditions assumed in Theorem 2.1 are satisfied. Though we can not prove that Example 3 satisfies the assumptions of sufficient conditions in Theorem 2.1, Algorithm 3.1 is still flexible for the problem.

## 5. Conclusion

In this paper we consider how to solve the KKT system of a general VIP(X,F). We reformulate the KKT system as a constrained optimization problem. The motivation of the new reformulation is based on some real cases where  $F(x)$  is not defined well on the whole space, or some important properties of  $F(x)$  hold just on  $X$ . The so-called stationary condition, which ensures a stationary point of the reformulated optimization problem being a solution of the KKT system, is discussed. A projection-type method for solving the reformulation problem with a closed convex set is also presented. The new algorithm has nice global and local convergence properties. Numerical examples show that the new algorithm is promising. Since how to compute efficiently the projection of a general closed convex set is an open question, the new algorithm is suitable for some special VIP(X,F) where the projection on  $X$  is easy to be computed. On the other hand, from the numerical example 3 we note that the conditions proposed in Theorem 2.1 may be reduced in the further research. Moreover, to require the iterates to stay in the whole set  $X$  is quite questionable. In order to reduce the difficulties in solving the reformulated constrained problem, we may just consider the constraints which are necessary. These are our further research topics.

**Acknowledgement** The authors would like to thank two anonymous referees for their detailed comments which considerably improved the presentation of the paper.

## References

- [1] Andreani R., Friedlander A. and Martinez J.M., *Solution of finite-dimensional variational inequalities using smooth optimization with simple bounds*, JOTA, **94** (1997), 635–657.

- [2] Clarke F.H., *Optimization and Nonsmooth Analysis*, John Wiley, New York, 1983.
- [3] Dirkse S.P. and Ferris M.C., *MCPLIB: A collection of nonlinear mixed complementarity problems*, *Optim. Method and Software*, **5** (1995), 319–345.
- [4] Facchinei F., Fischer A. and Kanzow C., *Regularity properties of a semismooth reformulation of variational inequalities*, *SIAM J. Optim.*, **8** (1998), 850–869.
- [5] Facchinei F., Fischer A. and Kanzow C., Peng J-M., *A simply constrained optimization reformulation of KKT systems arising from variational inequalities*, *Appl. Math. Optim.*, **40** (1999), 19–37.
- [6] Facchinei F. and Kanzow C., *On unconstrained and constrained stationary points of the implicit Lagrangian*, *JOTA*, **92** (1997), 99–115.
- [7] Ferris M.C., Kanzow C. and Munson T.S., *Feasible descent algorithms for mixed complementarity problems*, *Math. Prog.*, **86** (1999), 475–497.
- [8] Ferris M.C. and Pang J.S., *Engineering and economic applications of complementarity problems*, *SIAM Rev.*, **39** (1997), 669–713.
- [9] Harker P.T. and Pang J.S., *Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications*, *Math. Prog.*, **48** (1990), 161–220.
- [10] Jiang H., *Global convergence analysis of the generalized Newton and Gauss-Newton methods of the Fischer-Burmeister equation for complementarity problem*, *Math. Oper. Res.* **24** (1999), 529–542.
- [11] Jiang H., Fukushima M., Qi L. and Sun D., *A trust region method for solving generalized complementarity problems*, *SIAM J. Optim.*, **8** (1998), 140–157.
- [12] Jiang H. and Qi L., *A new nonsmooth equations approach to nonlinear complementarity problems*, *SIAM J. Control Optim.*, **35** (1997), 178–193.
- [13] Luca T.D., Facchinei F. and Kanzow C., *A semismooth equation approach to the solution of nonlinear complementarity problems*, *Math. Prog.*, **75** (1996), 407–439.
- [14] Moré J., *Global methods for nonlinear complementarity problems*, *Mathematics of Operations research*, **21** (1996), 589–614.
- [15] Pang J.S. and Qi L., *Nonsmooth equations: Motivation and algorithms*, *SIAM J. Optim.*, **3** (1993), 443–465.
- [16] Qi L., *Convergence analysis of some algorithms for solving nonsmooth equations*, *Math. Oper. Res.*, **18** (1993), 227–245.
- [17] Qi L. and Zhou G., *A smoothing Newton method for ball constrained variational inequalities with applications*, *Applied Mathematics Report AMR00/28*, School of Mathematics, The University of New South Wales, November, 2000.
- [18] Qi L., Tong X. and Li D., *An active-set projected trust region algorithm for box constrained nonsmooth equations*, *J. Optim. Theory and Applications*, **120** (2004), 601–625.
- [19] Ralph D. and Wright S.J., *Superlinear Convergence of an Interior-Point Method for Monotone Variational Inequalities*, in *Complementarity and Variational Problem, State of the Art*, M.C. Ferris and J.S. Pang, eds., SIAM Philadelphia, PA, pp. 345–385, 1997.

- [20] Steihaug T., *The conjugate gradient method and trust region in large scale optimization*, SIAM J. Numer. Anal., **20** (1983), 626–637.
- [21] Sun D., Womersley R.S. and Qi H., *A feasible semismooth asymptotically Newton method for mixed complementarity problems*, Math. Prog., **94** (2002), 167–187.
- [22] Taji K., Fukushima M. and Ibaraki T., *A globally convergent Newton method for solving strongly monotone variational inequalities*, Math. Prog., **58** (1993), 369–383.
- [23] Ulbrich M., *Nonmonotone trust-region method for bound-constrained semismooth equations with applications to nonlinear mixed complementarity problem*, SIAM J. Optim., **11** (2001), 889–917.
- [24] Yuan Y., *On the truncated conjugate gradient method*, Math. Prog., **87** (2000), 561–573.

X. Tong  
Institute of Mathematics,  
Changsha University of Science and Technology,  
Changsha 410077  
E-mail: tongxj@csust.edu.cn

S. Xiao  
Institute of Mathematics,  
Changsha University of Science and Technology,  
Changsha 410077