Chin. Ann. Math.
29B(3), 2008, 273–290
DOI: 10.1007/s11401-007-0082-6

Chinese Annals of Mathematics, Series B

An Affine Scaling Interior Trust Region Method via Optimal Path for Solving Monotone Variational Inequality Problem with Linear Constraints***

Yunjuan WANG*　　Detong ZHU**

Abstract Based on a differentiable merit function proposed by Taji et al. in "Math. Prog. Stud., **58**, 1993, 369–383", the authors propose an affine scaling interior trust region strategy via optimal path to modify Newton method for the strictly monotone variational inequality problem subject to linear equality and inequality constraints. By using the eigensystem decomposition and affine scaling mapping, the authors form an affine scaling optimal curvilinear path very easily in order to approximately solve the trust region subproblem. Theoretical analysis is given which shows that the proposed algorithm is globally convergent and has a local quadratic convergence rate under some reasonable conditions.

Keywords Trust region, Affine scaling, Interior point, Optimal path, Variational inequality problem
2000 MR Subject Classification 90C33, 49M99

1 Introduction

In this paper, we analyze the following variational inequality problem subject to both linear equality and linear inequality constraints (VIP):

$$\text{“Find } x_* \in \mathcal{S} \stackrel{\text{def}}{=} \{x \mid A_1 x = b_1, A_2 x \geq b_2\} \text{ such that } \langle F(x_*), x - x_* \rangle \geq 0 \text{ for all } x \in \mathcal{S}\text{”},$$

where $A_1^T = [a_1, \cdots, a_l] \in \mathbb{R}^{n\times l}$ and $A_2^T = [a_{l+1}, \cdots, a_m] \in \mathbb{R}^{n\times(m-l)}$ $(m > l)$ are two matrices, $b \stackrel{\text{def}}{=} \binom{b_1}{b_2} = (b^1, \cdots, b^l, b^{l+1}, \cdots, b^m)^T \in \mathbb{R}^m$ is a vector. We denote the strict interior feasible (or 'strictly feasible') set $\text{int}(\mathcal{S}) \stackrel{\text{def}}{=} \{x \mid A_1 x = b_1, A_2 x > b_2\}$ for the inequality constraints. Assume that F is a continuous mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$, and $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathbb{R}^n$.

The problem (VIP) is widely used to study various equilibrium models arising in economic, operations research, transportation and regional sciences. To solve these problems, many iterative methods have been developed, such as projection method, the nonlinear Jacobian method,

Manuscript received March 5, 2007. Revised May 8, 2007. Published online April 23, 2008.
*Mathematics and Science College, Shanghai Normal University, Shanghai 200234, China; Arts and Science School, Shanghai Dianji University, Shanghai 200240, China.
E-mail: yunjuanwang@163.com
**Business College, Shanghai Normal University, Shanghai 200234, China. E-mail: dtzhu@shnu.edu.cn
***Project supported by the National Natural Science Foundation of China (No. 10471094), the Doctoral Programmer Foundation of the Ministry of Education of China (No. 0527003), the Shanghai Leading Academic Discipline Project (No. T0401), and the Science Foundation Grant of Shanghai Municipal Education Committee (Nos. 05DZ11, 06A110).

the successive overrelaxation method, and generalized gradient method. These methods usually converge to a solution of (VIP) under certain conditions on the mapping F and the rates of convergence are generally linear.

It is well-known that Newton method for solving nonlinear equations and the unconstrained minimization problems converges locally and quadratically. When applied to (VIP), Newton method generates a sequence of iterates $\{x_k\}$, where x_0 is chosen in $\mathcal{S}$ and x_{k+1} is determined to be a solution of the variational inequality problem obtained by linearizing F at the current iterate x_k, i.e., $x_{k+1} \in \mathcal{S}$ and

$$\text{“Find } x \in \mathcal{S} \text{ such that } \langle F(x_k) + \nabla F(x_k)^T (x - x_k), y - x\rangle \geq 0 \text{ for all } y \in \mathcal{S}\text{”.}$$

It has been shown that, under the assumptions that x_* is a regular solution of (VIP) and $\nabla F(x)$ is Lipschitz continuous around x_*, the sequence converges quadratically to x_* if the starting point x_0 is sufficiently close to x_*.

Recently, Marcotte and Dussault in 1989, Taji, Fujushima and Ibaraki in 1993 presented a globally convergent Newton method for (VIP) by incorporating line search strategies. Mcrcotte and Dussault's method uses the gap function

$$\phi(x) = \max\{\langle F(x), x - y\rangle \,|\, y \in \mathcal{S}\}$$

as a merit function. The function ϕ is generally nondifferentiable and achieves its minimum at a solution of (VIP) on $\mathcal{S}$. The set $\mathcal{S}$ is assumed to be compact in order that the function ϕ is well-defined. It is shown that, when F is monotone, the method is globally convergent when line searches are exact and, under the joint assumptions of strong monotonicity and strict complementarity, the rate of convergence is quadratical. Taji's method employs a differentiable merit function proposed by Fukushima, whose minimizer on $\mathcal{S}$ coincides with the solution of (VIP). The method allows inexact line searches and does not rely on the compact assumption of the set $\mathcal{S}$. When F is strongly monotone, the method is globally convergent and, under additional assumptions that the set $\mathcal{S}$ is polyhedral convex, $\nabla F(x)$ is locally Lipschitz continuous, the strictly complementarity condition holds at the unique solution x_* of (VIP), and the rate of convergence is quadratic.

In this paper, we propose an affine scaling trust region strategy via optimal curvilinear path to modify Newton method with interior point backtracking technique for solving (VIP). The method makes use of a differentiable merit function proposed by Fukushima, which also has the property that its minimum on $\mathcal{S}$ coincides with the solution of problem convergent to a solution of problem (VIP). As for how to solve its minimum on $\mathcal{S}$, we build an affine scaling interior trust region subproblem based on the idea of Coleman and Li's method, typically called the double affine-scaling interior-point Newton method, for solving the optimization problem with linear inequality constraints in [3]. When F is strictly monotone rather than strongly monotone, the proposed algorithm is well-defined and converges globally to the unique solution of (VIP). Under the same assumptions as made in [8], it is shown that for k sufficiently large, no trust region subproblem is involved and the line search step is full; therefore, the algorithm reduces to the basic Newton method and hence the rate of convergence is also quadratic.

The paper is organized as follows. In Section 2, we review some preliminary results concerning the projection operators and monotone mappings and the merit function that are useful in the subsequent sections. In Section 3, we describe the affine scaling interior trust region method via optimal curvilinear path to modify Newton method for solving monotone variational inequality problem subject to both linear equality and linear inequality constraints. In Section 4, we prove that the proposed algorithm is well defined and weak global convergence of the algorithm is established. Local convergence rate is discussed in Section 5.

2 Preliminaries

In this section, we summarize some basic concepts of monotone mapping F and Fukushima's differentiable merit function and their properties used in subsequent sections.

A mapping $F : \mathbb{R}^n \to \mathbb{R}^n$ is said to be monotone on $\mathcal{S}$ if

$$\langle F(x) - F(x'), x - x' \rangle \geq 0 \quad \text{for all } x, x' \in \mathcal{S} \tag{2.1}$$

and strictly monotone on $\mathcal{S}$ if strict inequality holds whenever $x \neq x'$. If F is continuously differentiable and its Jacobian matrix $\nabla F(x)$ is positive definite for all $x \in \mathcal{S}$, then F is strictly monotone on $\mathcal{S}$. Note that $\nabla F(x)$ may not be symmetric. A mapping F is said to be strongly (or uniformly) monotone with modulus $\mu > 0$ on $\mathcal{S}$ if

$$\langle F(x) - F(x'), x - x' \rangle \geq \mu \|x - x'\|^2 \quad \text{for all } x, x' \in \mathcal{S}. \tag{2.2}$$

Throughout the presentation in this paper, $\|\cdot\|$ denotes the 2-norm. When F is continuously differentiable, a necessary and sufficient condition for (2.2) is

$$\langle d, \nabla F(x) d \rangle \geq \mu \|d\|^2 \quad \text{for all } x \in \mathcal{S} \text{ and } d \in \mathbb{R}^n. \tag{2.3}$$

It is clear that strongly monotone implies strictly monotone.

Let G be any given $n \times n$ symmetric positive matrix. The G-norm projection of a point $x \in \mathbb{R}^n$ onto a set $\mathcal{S}$, denoted by $\mathrm{proj}_{\mathcal{S},G}(x)$, is defined as the unique solution to the following constrained optimization problem

$$\min \|y - x\|_G \quad \text{s.t.} \quad y \in \mathcal{S},$$

where $\|x\|_G = \langle x, Gx \rangle^{\frac{1}{2}}$ denotes the G-norm of a vector x in $\mathbb{R}^n$. The projection operator $\mathrm{proj}_{\mathcal{S},G}(\cdot)$ is nonexpansive (see [1]), i.e.,

$$\|\mathrm{proj}_{\mathcal{S},G}(x) - \mathrm{proj}_{\mathcal{S},G}(x')\|_G \leq \|x - x'\|_G \quad \text{for all } x, x' \in \mathbb{R}^n. \tag{2.4}$$

Suppose that an $n \times n$ symmetric positive matrix G is given and $x \in \mathbb{R}^n$ is an given point. Since the minimization problem

$$\min \langle F(x), y - x \rangle + \frac{1}{2} \langle y - x, G(y - x) \rangle \quad \text{s.t.} \quad y \in \mathcal{S} \tag{2.5}$$

is essentially equivalent to the problem

$$\min \|y - (x - G^{-1} F(x))\|_G^2 \quad \text{s.t.} \quad y \in \mathcal{S}, \tag{2.6}$$

$H(x)\overset{\text{def}}{=}\text{proj}_{\mathcal{S},G}(x-G^{-1}F(x))$, the unique optimal solution of (2.6), is also the unique optimal solution of problem (2.5). It follows from (2.4) that $H:\mathbb{R}^n\to\mathcal{S}$ is continuous whenever F is continuous. The mapping H yields a fixed point characterization of the solution of (VIP).

Proposition 2.1 (see [8]) *Let G be an $n\times n$ symmetric positive matrix and let $H(x)$ be the unique optimal solution of minimization problem* (2.6) *for each given $x\in\mathbb{R}^n$. Then x solves* (VIP) *if and only if x is a fixed point of the mapping H, i.e., $x=H(x)$.*

For any given $x\in\mathcal{S}$, the linearized variational inequality problem of (VIP) at x is

$$\text{“Find } z\in\mathcal{S} \text{ such that } \langle F(x)+\nabla F(x)^T(z-x), y-z\rangle\geq 0 \text{ for all } y\in\mathcal{S}\text{”}. \tag{2.7}$$

When F is continuously differentiable and $\nabla F(x)$ is positive definite, a unique solution, denoted by $z(x)$, exists. The mapping $z:\mathcal{S}\to\mathcal{S}$ has the following property.

Proposition 2.2 (see [8]) *If F is continuously differentiable and strongly monotone on $\mathcal{S}$, then the mapping $z:\mathcal{S}\to\mathcal{S}$ is continuous on $\mathcal{S}$. Furthermore, x is the solution of* (VIP) *if and only if x satisfies $x=z(x)$.*

In fact, from the proof of this proposition (see [8]) it can be concluded that when F is continuously differentiable and its Jacobian matrix $\nabla F(x)$ is positive definite on $\mathcal{S}$, the second part of the proposition is also true.

For given mapping $F:\mathbb{R}^n\to\mathbb{R}^n$ and given $n\times n$ positive definite symmetric matrix G, we define the function $f:\mathbb{R}^n\to\mathbb{R}$ as

$$f(x)\overset{\text{def}}{=}-\langle F(x), H(x)-x\rangle-\frac{1}{2}\langle H(x)-x, G(H(x)-x)\rangle, \tag{2.8}$$

where $H(x)$ is the unique solution of minimization problem (2.6). It has been shown that, for any nonempty closed convex set $\mathcal{S}$, the function f has the following property.

Proposition 2.3 (see [5]) *If the mapping $F:\mathbb{R}^n\to\mathbb{R}^n$ is continuous, then the function $f:\mathbb{R}^n\to\mathbb{R}$ is also continuous. Furthermore, if F is continuously differentiable, then f is also continuously differentiable and its gradient is given by*

$$\nabla f(x)=F(x)-[\nabla F(x)-G](H(x)-x). \tag{2.9}$$

By using the function f, an equivalent optimization problem can be formulated for any variational inequality problem.

Proposition 2.4 (see [5]) *$f(x)\geq 0$ for all $x\in\mathcal{S}$ and $f(x_*)=0$ if and only if x_* solves* (VIP). *Hence x_* solves* (VIP) *if and only if x_* solves the following optimization problem* (2.10) *and $f(x_*)=0$:*

$$\min f(x) \quad \text{s.t.} \quad x\in\mathcal{S}. \tag{2.10}$$

Though the function f generally is not convex, it has the desirable property that, if $\nabla F(x)$ is positive definite for all $x\in\mathcal{S}$, any stationary point of problem (2.10) is also a global optimal solution of minimization problem (2.10).

Proposition 2.5 (see [5]) *Assume that the mapping $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable and its Jacobian matrix $\nabla F(x)$ is positive definite for all $x \in \mathcal{S}$. If x is a stationary point of problem* (2.10)*, i.e., $\langle \nabla f(x), y - x \rangle \geq 0$ for all $y \in \mathcal{S}$, then x is a global optimal solution of minimization problem* (2.10) *and hence solves* (VIP)*.*

Proposition 2.5 indicates that the function f can be used as a merit function for a descent method to solve a kind of strictly monotone variational inequality problems.

Next, let us consider the problem of computing a local minimizer of minimization problem (2.10).

Coleman and Li [3] presented a trust region affine scaling interior point algorithm, called double trust region interior point method (TRAM), for solving the minimization problem subject only to linear inequality constraints. Based on the two forces of nonlinearity and feasibility adjusted by the trust region radius, the global and local convergence properties of the TRAM algorithm were established in [3]. Recently, combining trust region strategy with line search technique, Zhu [10] proposed a new affine scaling trust region algorithm with nonmonotonic interior point backtracking technique for problem (2.10). The global convergence and fast locally convergent rate of the proposed algorithm are established under some reasonable smooth conditions. Based on the idea of the affine scaling interior trust region method with interior point backtracking technique in [3, 10], we consider an interior affine scaling algorithm via optimal path with nonmonotonic interior backtracking technique for minimization problem (2.10). The affine scaling optimal path involves choosing a scaling matrix and a quadratic model. By examining the first-order necessary conditions for minimization problem (2.10), we motivate our choice of affine scaling matrix.

The first-order necessary optimality conditions for minimization problem (2.10) are well established. A feasibility $x_* \in \mathcal{S}$ is said to be stationary point for minimization problem (2.10) called the first order necessary conditions, if there exist two vectors $\lambda_* \in \mathbb{R}^l$, $0 \leq \mu_* \in \mathbb{R}^{m-l}$ such that

$$\nabla f(x_*) - A_1^T \lambda_* - A_2^T \mu_* = 0 \quad \text{and} \quad \mathrm{diag}\{A_2 x_* - b_2\}\mu_* = 0. \tag{2.11}$$

Strict complementarity is said to hold at x_* if $|\lambda_*^j| > 0$, $j = 1, \cdots, l$ and at least one of the two inequalities $a_{l+j}^T x_* - b^{l+j} > 0$ and $|\mu_*^j| > 0$ ($j = 1, \cdots, m-l$) holds, that is, $|\lambda_*^j| > 0$, $j = 1, \cdots, l$ and $|a_{l+j}^T x_* - b^{l+j}| + |\mu_*^j| > 0$, $j = 1, \cdots, m-l$, where λ_*^j , b^{l+j} and μ_*^j are the jth components of the vectors λ_*, b_2 and μ_*, respectively. The affine scaling optimal path can arise naturally from the Newton step for the first order necessary conditions of the problem (2.10). Ignoring primal and dual feasibility of the inequality constraints, the first order necessary conditions of minimization problem (2.10) can be expressed as an $(m+n) \times (m+n)$ system of nonlinear equations

$$\begin{aligned} &\nabla f(x) - A_1^T \lambda - A_2^T \mu = 0, \\ &A_1 x = b_1, \\ &\mathrm{diag}\{A_2 x - b_2\}\mu = 0. \end{aligned} \tag{2.12}$$

For the kth iteration, to globalize, we generate a modified Newton step by replacing diag$\{\mu_k\}$ by $C_k \stackrel{\text{def}}{=} \text{diag}\{|\mu_k|\}$ suggested by Coleman and Li [3] which was a descent direction for $f(x)$ far away from a solution, that is,

$$\begin{pmatrix} \nabla^2 f(x_k) & -A_1^T & -A_2^T \\ A_1 & 0 & 0 \\ C_k A_2 & 0 & D_k \end{pmatrix} \begin{bmatrix} \Delta p_k^N \\ \Delta \lambda_k^N \\ \Delta \mu_k^N \end{bmatrix} = - \begin{bmatrix} \nabla f_k - A_1^T \lambda_k - A_2^T \mu_k \\ A_1 x_k - b_1 \\ D_k \mu_k \end{bmatrix}, \tag{2.13}$$

where $D_k \stackrel{\text{def}}{=} \text{diag}\{A_2 x_k - b_2\}$. It can be shown that the modified Newton step sufficiently approximates the exact Newton step, asymptotically, to achieve fast convergence. Using the augmented quadratic as the objective function of the model, we see that a trust region consistent with the modified Newton step Δp_k^N in the null subspace of A_1 is

$$\begin{aligned} \min \quad & \nabla f_k^T d + \frac{1}{2} d^T B_k d + \frac{1}{2} d^T A_2^T D_k^{-\frac{1}{2}} C_k D_k^{-\frac{1}{2}} A_2 d, \\ \text{s.t.} \quad & A_1 d = 0, \\ & \|(d; D_k^{-\frac{1}{2}} A_2 d)\| \le \Delta_k, \end{aligned} \tag{2.14}$$

where $d = x - x_k$, B_k is either the Hessian of f at x_k or its approximation, Δ_k is the trust region radius. Set the transformation $\widehat{d} = D_k^{-\frac{1}{2}} A_2 d$, the trust region subproblem (2.14) is equivalent to the following problem in original variable space,

$$\begin{aligned} & \min_{(d;\widehat{d}) \in \mathbb{R}^{n+m}} \quad && \psi_k(d;\widehat{d}) \stackrel{\text{def}}{=} \nabla f_k^T d + \frac{1}{2} d^T B_k d + \frac{1}{2} \widehat{d}^T C_k \widehat{d}, \\ (\overline{S}_k) \qquad & \text{s.t.} && A_1 d = 0, \ D_k^{\frac{1}{2}} \widehat{d} = A_2 d, \\ & && \|(d;\widehat{d})\| \le \Delta_k. \end{aligned}$$

Let $(d_k; \widehat{d}_k)$ denote a solution to the subproblem $(\overline{S}_k)$. It is easy to see that $(d_k; \widehat{d}_k)$ satisfies the necessary and sufficient conditions concerning $(d_k; \widehat{d}_k)$, $\nu_k \ge 0$ and $\lambda_{k+1},\ \mu_{k+1}$ that

$$\left(\begin{bmatrix} B_k & 0 \\ 0 & C_k \end{bmatrix} + \nu_k I \right) \begin{bmatrix} d_k \\ \widehat{d}_k \end{bmatrix} = - \begin{bmatrix} \nabla f_k \\ 0 \end{bmatrix} + \begin{bmatrix} A_1^T \\ 0 \end{bmatrix} \lambda_{k+1} + \begin{bmatrix} A_2^T \\ -D_k^{\frac{1}{2}} \end{bmatrix} \mu_{k+1}; \tag{2.15}$$

$$\nu_k \left(\Delta_k - \left\| \begin{bmatrix} d_k \\ \widehat{d}_k \end{bmatrix} \right\| \right) = 0, \quad A_1 d_k = 0. \tag{2.16}$$

Here the least squares Lagrangian multipliers λ_{k+1} and μ_{k+1} are defined as follows:

$$\begin{bmatrix} A_1^T & A_2^T \\ 0 & -D_k^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \lambda_{k+1} \\ \mu_{k+1} \end{bmatrix} \stackrel{\text{L.S.}}{=} \begin{bmatrix} \nabla f_k \\ 0 \end{bmatrix}. \tag{2.17}$$

Denote the projected gradient direction as

$$g_k \stackrel{\text{def}}{=} g(x_k) \stackrel{\text{def}}{=} -(\nabla f_k - A_1^T \lambda_{k+1} - A_2^T \mu_{k+1}), \tag{2.18}$$

and at the same time, the augmented gradient direction of the objective function of the subproblem $(\overline{S}_k)$ as

$$\widetilde{g}_k \stackrel{\text{def}}{=} \begin{bmatrix} \nabla f_k \\ 0 \end{bmatrix} - \begin{bmatrix} A_1^T \\ 0 \end{bmatrix} \lambda_{k+1} - \begin{bmatrix} A_2^T \\ -D_k^{\frac{1}{2}} \end{bmatrix} \mu_{k+1} = \begin{bmatrix} -g_k \\ D_k^{\frac{1}{2}} \mu_{k+1} \end{bmatrix}.$$

Let P_k denote the orthogonal projection onto the null space of $\begin{bmatrix} A_1 & 0 \\ A_2 & -D_k^{\frac{1}{2}} \end{bmatrix}$. Then

$$\nabla f_k^T g_k = -\left\| P_k \begin{bmatrix} \nabla f_k \\ 0 \end{bmatrix} \right\|^2 = -(\|\nabla f_k - A_1^T \lambda_{k+1} - A_2^T \mu_{k+1}\|^2 + \|D_k^{\frac{1}{2}} \mu_{k+1}\|^2) = -\|\widetilde{g}_k\|^2. \quad (2.19)$$

It is obvious that from (2.19), a sufficient decrease of $\psi_k(d; \widehat{d})$ measured against the decrease from the damped minimizer $-\nabla f_k^T g_k$ leads to satisfaction of complementarity:

$$\lim_{k\to\infty} \|\nabla f_k - A_1^T \lambda_{k+1} - A_2^T \mu_{k+1}\| = 0 \quad \text{and} \quad \lim_{k\to\infty} \|D_k^{\frac{1}{2}} \mu_{k+1}\| = 0. \quad (2.20)$$

So, if $|\nabla f_k^T g_k|^{\frac{1}{2}} = 0$, stop with the solution x_k of minimization problem (2.10).

3 Algorithm

In this section, we present an affine scaling interior trust region approach via optimal path for solving monotone variational inequality problem. Throughout this section we assume that the set $\mathcal{S}$ in nonempty and that the mapping $F : \mathbb{R}^n \to \mathbb{R}^n$ is positive definite for $x \in \mathcal{S}$.

For any given $x_k \in \mathcal{S}$, consider the following linearized variational inequality problem:

(LVIP(x_k)) "Find $z \in \mathcal{S}$ such that $\langle F(x_k) + \nabla F(x_k)^T (z - x_k), x - z\rangle \geq 0$ for all $x \in \mathcal{S}$".

The assumptions on F ensure that the linearized problem (LVIP(x_k)) can be rewritten as a linear complementarity problem and can be solved in a finite number of steps by Lemke's complementary pivoting method in [4].

In the neighborhood of a local minimizer, the Newton step defined by (2.13) for (2.12) is a solution to the trust region subproblem ($\overline{S}_k$) when the trust region constraint is inactive. In order to avoid solving the trust region subproblem frequently, we pay no attention to the trust region constraint and go to search for the trial step at each iteration along some curvilinear path. By using the idea of the optimal path of general trust region subproblem in [2], we now form an approximate affine scaling interior optimal path of trust region subproblem ($\overline{S}_k$).

For convenience, we denote $M_k \stackrel{\text{def}}{=} \begin{bmatrix} B_k & 0 \\ 0 & C_k \end{bmatrix}$. In order to define the path in a closed form, we shall use the eigensystem decomposition of M_k. We begin with the factorization of the sysmetric matrix B_k. Due to its symmetry, we can factorize the matrix B_k into the form $B_k = U_k \Lambda_k U_k^T$, where U_k is an identity orthonormal matrix and Λ_k a diagonal matrix. From the definition of M_k, we can expand the matrix U_k and Λ_k such that

$$M_k = \begin{bmatrix} U_k & \\ & I_l \end{bmatrix} \times \begin{bmatrix} \Lambda_k & \\ & C_k \end{bmatrix} \times \begin{bmatrix} U_k^T & \\ & I_l \end{bmatrix},$$

so its eigenvalues $\phi_1, \phi_2, \cdots, \phi_{n+m-l}$ are all real numbers, among which n eigenvalues are the diagonal elements of the matrix Λ_k, and the others are the diagonal elements of the matrix C_k, where I_l is the unit matrix on $\mathbb{R}^l$. Let $u^1, u^2, \cdots, u^{n+m-l}$ be orthonormal eigenvectors associated with the eigenvalues $\phi_1, \phi_2, \cdots, \phi_{n+m-l}$. By the definition of $\widetilde{g}_k$ and (2.19), we have that $\widetilde{g}_k$ belongs to the null subspace of the matrix $\begin{bmatrix} A_1 & 0 \\ A_2 & -D_k^{\frac{1}{2}} \end{bmatrix}$. In this paper, we assume that the matrix $\begin{bmatrix} A_1 & 0 \\ A_2 & -D_k^{\frac{1}{2}} \end{bmatrix}$ is full row rank. Then the number of the basis of the null subspace of

$\begin{bmatrix} A_1 & 0 \\ A_2 & -D_k^{\frac{1}{2}} \end{bmatrix}$ is $n-l$. Therefore, for orthonormal vectors $u^1, u^2, \cdots, u^{n+m-l}$ in the original space, there only are $n-l$ vectors belonging to $\begin{bmatrix} A_1 & 0 \\ A_2 & -D_k^{\frac{1}{2}} \end{bmatrix}$. Without loss of generality, denote these vectors as $\widetilde{u}^1, \widetilde{u}^2, \cdots, \widetilde{u}^{n-l}$. Accordingly, let $\widetilde{\phi}_1, \widetilde{\phi}_2, \cdots, \widetilde{\phi}_{n-l}$ be corresponding orthonormal eigenvectors and assume that $\widetilde{\phi}_1 \le \widetilde{\phi}_2 \le \cdots \le \widetilde{\phi}_{n-l}$. We partition the set $\{1, 2, \cdots, n-l\}$ into $\widetilde{\mathcal{I}}^+$, $\widetilde{\mathcal{I}}^-$ and $\widetilde{\mathcal{N}}$ according to $\widetilde{\phi}_i > 0$, $\widetilde{\phi}_i < 0$ and $\widetilde{\phi}_i = 0$. We now give the affine scaling interior optimal path.

Based on the eigensystems of M_k, the optimal path $\Gamma(\tau)$ can be expressed as

$$\Gamma(\tau) = \Gamma_1(t_1(\tau)) + \Gamma_2(t_2(\tau)), \tag{3.1}$$

where

$$\begin{aligned} \Gamma_1(t_1(\tau)) &= -\Big[\sum_{i \in \widetilde{\mathcal{I}}} \frac{t_1(\tau)}{\widetilde{\phi}_i t_1(\tau) + 1} g_k^i \widetilde{u}^i + t_1(\tau) \sum_{i \in \widetilde{\mathcal{N}}} g_k^i \widetilde{u}^i \Big], \\ \Gamma_2(t_2(\tau)) &= t_2(\tau) \widetilde{u}^1, \end{aligned}$$

and

$$t_1(\tau) = \begin{cases} \tau, & \text{if } \tau < \dfrac{1}{T}, \\ \dfrac{1}{T}, & \text{if } \tau \ge \dfrac{1}{T}, \end{cases} \quad \text{and} \quad t_2(\tau) = \begin{cases} 0, & \text{if } \tau < \dfrac{1}{T}, \\ \tau - \dfrac{1}{T}, & \text{if } \tau \ge \dfrac{1}{T}, \end{cases}$$

where $\widetilde{\mathcal{I}} = \{i \,|\, \widetilde{\phi}_i \ne 0, i = 1, 2, \cdots, n-l\}$, $\widetilde{\mathcal{N}} = \{i \,|\, \widetilde{\phi}_i = 0, i = 1, 2, \cdots, n-l\}$, $g_k^i = g_k^T \widetilde{u}^i$, $i = 1, 2, \cdots, n-l$, $g_k = \sum_{i=1}^{n-l} g_k^i \widetilde{u}^i$, $T = \max\{0, -\widetilde{\phi}_1\}$ and $\frac{1}{T}$ is defined as $+\infty$ if $T = 0$. It should be noted that $\Gamma_2(t_2(\tau))$ is defined only when M_k is indefinite and $g_k^i = 0$ for all $i \in \{1, 2, \cdots, n-l\}$ with $\widetilde{\phi}_i = \widetilde{\phi}_1 < 0$, which is referred to hard case (see [2]) for unconstrained optimization, and for other cases, $\Gamma(\tau)$ is defined only for $0 \le \tau < \frac{1}{T}$, that is, $\Gamma(\tau) = \Gamma_1(t_1(\tau))$.

Algorithm

Step 1 Choose parameters $\beta \in (0, \frac{1}{2})$, $\omega \in (0, 1)$, $0 < \eta_1 < \eta_2 < 1$, $0 < \gamma_1 < \gamma_2 < 1 < \gamma_3$, $\theta_0 \in (0, 1)$. Choose an initial matrix B_0 approximate to the Hessian at x_0. Select an initial trust region radius $\Delta_0 > 0$ and a maximal trust region radius $\Delta_{\max} > 0$. Give a starting strict feasible interior point $x_0 \in \mathrm{int}(\mathcal{S})$. Set $k = 0$ and go to the following step.

Step 2 If $f(x_k) = 0$, stop with the approximate solution x_k.

Step 3 Find the solution $z(x_k) \in \mathcal{S}$ of problem LVIP(x_k) and let $d_k = z(x_k) - x_k$.

Step 4 If $f(x_k + d_k) = f(z(x_k)) \le \kappa f(x_k)$, then $x_{k+1} = z(x_k)$, $k \leftarrow k+1$ and go to Step 2.

Step 5 Let $\Delta_k = \min\{\|d_k\|, \sigma_k\}$, and $\sigma_0 = \|d_0\|$.

Step 6 Solve a step $(p_k; \widehat{p}_k)$ based on the following problem (S_k) via affine scaling optimal path:

$$\begin{array}{lll} & \displaystyle\min_{(p;\widehat{p}) \in \mathbb{R}^{n+m}} & \psi_k(p; \widehat{p}) \stackrel{\text{def}}{=} \nabla f_k^T p + \frac{1}{2} p^T B_k p + \frac{1}{2} \widehat{p}^T C_k \widehat{p}, \\ (S_k) & \text{s.t.} & A_1 p = 0, \quad D_k^{\frac{1}{2}} \widehat{p} = A_2 p, \\ & & \|(p; \widehat{p})\| \le \Delta_k, \quad (p; \widehat{p}) \in \Gamma_k(\tau). \end{array}$$

Step 7 Choose $\alpha_k = 1, w, w^2, \cdots$, such that the following inequality are satisfied:

$$f(x_k + \alpha_k p_k) \le f(x_k) + \beta \alpha_k \nabla f_k^T p_k, \tag{3.2}$$

$$x_k + \alpha_k p_k \in \mathcal{S}. \tag{3.3}$$

Step 8 Set

$$\delta_k = \begin{cases} \alpha_k p_k, & \text{if } x_k + \alpha_k p_k \in \text{int}(\mathcal{S}), \\ \theta_k \alpha_k p_k, & \text{otherwise}, \end{cases} \tag{3.4}$$

where $\theta_k \in (\theta_0, 1]$, $0 < \theta_0 < 1$ and $\theta_k - 1 = o(\|p_k\|)$, and correspondingly set $\widehat{\delta}_k = D_k^{-\frac{1}{2}} A_2 \delta_k$. Set

$$x_{k+1} = x_k + \delta_k. \tag{3.5}$$

Step 9 Compute

$$\begin{aligned} \text{Pred}(\delta_k) &= -\Big[\nabla f_k^T \delta_k + \frac{1}{2}\delta_k^T B_k \delta_k\Big], \\ \text{Ared}(\delta_k) &= f(x_k) - f(x_k + \delta_k), \\ \rho_k &= \frac{\text{Ared}(\delta_k)}{\text{Pred}(\delta_k)}. \end{aligned}$$

Step 10 Update trust region size Δ_{k+1} from Δ_k,

$$\Delta_{k+1} = \begin{cases} [\gamma_1 \Delta_k,\ \gamma_2 \Delta_k], & \text{if } \rho_k \le \eta_1, \\ (\gamma_2 \Delta_k,\ \Delta_k], & \text{if } \eta_1 < \rho_k < \eta_2, \\ (\Delta_k,\ \min\{\gamma_3 \Delta_k, \Delta_{\max}\}], & \text{if } \rho_k \ge \eta_2. \end{cases} \tag{3.6}$$

Step 11 Let $k \leftarrow k+1$, $\sigma_k = \Delta_k$, and go to step 2.

4 Global Convergence

The following assumptions are commonly used in the convergence analysis of most methods for solving the variational inequality problem with both linear equality and linear inequality constraints. Since x_k solves (VIP) if algorithm stops at Step 2, it is assumed, without loss of generality, that the algorithm generates an infinite sequence $\{x_k\}$. Given $x_0 \in \text{int}(\mathcal{S}) \subseteq \mathbb{R}^n$, the algorithm generates a sequence $\{x_k\} \subset \mathcal{S} \subset \mathbb{R}^n$. In our analysis, we denote the level set of f by

$$\mathcal{L}(x_0) = \{x \in \mathbb{R}^n \mid f(x) \le f(x_0),\ A_1 x = b_1,\ A_2 x \ge b_2\}.$$

Assumption A1 *The sequence $\{x_k\}$ generated by the proposed algorithm is contained in a compact set $\mathcal{L}(x_0)$ on $\mathbb{R}^n$.*

Assumption A2 *There exist positive scalars κ_f and κ_g such that $\|\nabla f(x_k)\| \le \kappa_f$ and $\|g(x_k)\| \le \kappa_g$ for all k, where $g(x_k)$ is given in* (2.18). *Further, we assume that there exists a positive scalar κ_B such that $\|B_k\| \le \kappa_B$ for all k.*

Assumptions A1–A2 imply that there exist $\kappa_D,\ \kappa_M > 0$ such that $\|D_k^{-1}\| \le \kappa_D$ and $\|M_k\| \le \kappa_M$ for all k .

Assumption A3 $\begin{bmatrix} A_1 & 0 \\ A_2 & -D(x)^{\frac{1}{2}} \end{bmatrix}$ *is assumed to have full row rank for all* $x \in \mathcal{L}(x_0)$.

Similar to the proof of Lemma 2.3 in [9] and Lemmas 3.1 and 3.2 in [10], we can also obtain that at the kth iteration the trial step p_k is a sufficiently descent direction and the predicted reduction satisfies a sufficient descent condition. In this paper we only summarize the properties of optimal path as the following lemma, whose proofs are similar to those in [9, 10].

Lemma 4.1 *Let the step* $(p_k; \widehat{p}_k)$ *in trust region subproblem be obtained from the affine scaling interior optimal path. Then the norm function of the path is monotonically increasing for* $\tau \in (0, +\infty)$, *and there exists* $\tau_k \in (0, +\infty)$ *such that the point* $\begin{bmatrix} p_k \\ \widehat{p}_k \end{bmatrix} = \Gamma(\tau_k)$ *on the path with* $\|\Gamma(\tau_k)\| = \Delta_k$ *satisfies the following systems that there exist* λ_{k+1}, μ_{k+1} *such that*

$$\left(\begin{bmatrix} B_k & 0 \\ 0 & C_k \end{bmatrix} + \nu_k I\right)\begin{bmatrix} p_k \\ \widehat{p}_k \end{bmatrix} = -\begin{bmatrix} \nabla f_k \\ 0 \end{bmatrix} + \begin{bmatrix} A_1^T \\ 0 \end{bmatrix}\lambda_{k+1} + \begin{bmatrix} A_2^T \\ -D_k^{\frac{1}{2}} \end{bmatrix}\mu_{k+1}, \tag{4.1}$$

$$\nu_k\left(\Delta_k - \left\|\begin{bmatrix} p_k \\ \widehat{p}_k \end{bmatrix}\right\|\right) = 0, \quad A_1 p_k = 0; \tag{4.2}$$

and $\nu_k \ge 0$ *is given as follows:*

$$\nu_k = \frac{1}{\tau_k}, \quad \text{as } \tau_k < \frac{1}{T_k}, \tag{4.3}$$

$$\nu_k = \frac{1}{T_k}, \quad t_2(\tau_k) = \tau_k - \frac{1}{T_k}, \quad \text{as } \tau_k \ge \frac{1}{T_k}, \tag{4.4}$$

where $T_k = \max\{0, -\widetilde{\phi}_1^k\}$. *Furthermore, there exist* $\omega_1 > 0$ *and* $\omega_2 > 0$ *such that the step* p_k *satisfies the following the first and the second order sufficient descent conditions*

$$-\nabla f_k^T p_k \ge \omega_1 |\nabla f_k^T g_k|^{\frac{1}{2}} \min\Big\{\Delta_k, \frac{|\nabla f_k^T g_k|^{\frac{1}{2}}}{\|M_k\|}\Big\}, \tag{4.5}$$

$$-\psi_k(p_k; \widehat{p}_k) \ge \omega_2 |\nabla f_k^T g_k|^{\frac{1}{2}} \min\Big\{\Delta_k, \frac{|\nabla f_k^T g_k|^{\frac{1}{2}}}{\|M_k\|}\Big\} \tag{4.6}$$

for all $\nabla f_k,\ g_k,\ \|M_k\|$, *and* Δ_k, *where* $\psi_k(p_k; \widehat{p}_k)$ *is given in the subproblem* (S_k). *In fact, here* $\omega_1 \in (0, \frac{1}{4}],\ \omega_2 \in (0, \frac{1}{2}]$ *and*

$$|\nabla f_k^T g_k|^{\frac{1}{2}} \overset{\text{def}}{=} \sqrt{\|g_k\|^2 + \|D_k^{-\frac{1}{2}} A_2 g_k\|^2}.$$

Lemma 4.2 *Let the step* δ_k *be obtained from Step* 8. *Then there exists* $\omega_3 > 0$ *such that*

$$\text{Pred}(\delta_k) \ge \omega_3 \alpha_k \theta_k |\nabla f_k^T g_k|^{\frac{1}{2}} \min\Big\{\Delta_k, \frac{|\nabla f_k^T g_k|^{\frac{1}{2}}}{\|M_k\|}\Big\}. \tag{4.7}$$

Proof By the definition of $\text{Pred}(\delta_k)$ and C_k being positive semi-definite, we have

$$\text{Pred}(\delta_k) = -\psi_k(\delta_k; \widehat{\delta}_k) + \frac{1}{2}\widehat{\delta}_k^T C_k \widehat{\delta}_k \ge -\alpha_k\theta_k g_k^T p_k - \frac{1}{2}\theta_k^2\alpha_k^2[p_k^T B_k p_k + \widehat{p}_k^T C_k \widehat{p}_k], \tag{4.8}$$

where $\widehat{\delta}_k = D_k^{-\frac{1}{2}} A_2 \delta_k$. Note that $0 < \alpha_k \theta_k \le 1$. If $p_k^T B_k p_k + \widehat{p}_k^T C_k \widehat{p}_k \ge 0$, then from (4.6) and (4.8) we have

$$\begin{aligned}\operatorname{Pred}(\delta_k) &\ge -\alpha_k \theta_k g_k^T p_k - \frac{1}{2}\theta_k \alpha_k [p_k^T B_k p_k + \widehat{p}_k^T C_k \widehat{p}_k] \\ &= -\alpha_k \theta_k \psi_k(p_k; \widehat{p}_k) \\ &\ge \omega_2 \alpha_k \theta_k |\nabla f_k^T g_k|^{\frac{1}{2}} \min\Big\{\Delta_k, \frac{|\nabla f_k^T g_k|^{\frac{1}{2}}}{\|M_k\|}\Big\}. \end{aligned} \tag{4.9}$$

On the other hand, if $p_k^T B_k p_k + \widehat{p}_k^T C_k \widehat{p}_k < 0$, then from (4.5) and (4.8) we have

$$\operatorname{Pred}(\delta_k) \ge -\alpha_k \theta_k g_k^T p_k \ge \omega_1 \alpha_k \theta_k |\nabla f_k^T g_k|^{\frac{1}{2}} \min\Big\{\Delta_k, \frac{|\nabla f_k^T g_k|^{\frac{1}{2}}}{\|M_k\|}\Big\}. \tag{4.10}$$

From (4.9) and (4.10), we can easily conclude that

$$\operatorname{Pred}(\delta_k) \ge \omega_3 \alpha_k \theta_k |\nabla f_k^T g_k|^{\frac{1}{2}} \min\Big\{\Delta_k, \frac{|\nabla f_k^T g_k|^{\frac{1}{2}}}{\|M_k\|}\Big\},$$

where $\omega_3 = \min\{\omega_1, \omega_2\}$.

Lemma 4.3 *At the kth iteration, if $f(x_k) \ne 0$, then x_{k+1} is obtained either in Step 4 or by repeating Steps 6, 7 and 8 a finite number of times.*

Proof It is clear that we only need to prove that x_{k+1} can be computed by repeating Step 6, 7 and 8 a finite number of times if x_{k+1} is not obtained at Step 4.

Since $f(x_k) \ne 0$, x_k is not a solution and hence not a stationary point of the minimization problem (2.10) by Propositions 2.4 and 2.5. Suppose that $(p_k; \widehat{p}_k)$ is the solution of subproblem (S_k). In the following, we show that α_k which is obtained by Step 7 is not decreased for sufficiently large k and hence bounded away from zero. Thus, $\{\alpha_k\}$ cannot converge to zero.

Since $|\nabla f_k^T g_k|^{\frac{1}{2}} \ne 0$, by continuity there exist $\delta > 0$ and $\epsilon > 0$ such that $|\nabla f(x)^T g(x)|^{\frac{1}{2}} \ge \epsilon$ for all x with $\|x_k - x\| \le \delta$. Clearly, in a finite number of backtracking reductions, α_k will satisfy

$$\alpha_k \le t_k \overset{\text{def}}{=} \min\Big\{ -\frac{a_{l+i}^T x_k - b^{l+i}}{a_{l+i}^T p_k} \,\Big|\, -\frac{a_{l+i}^T x_k - b^{l+i}}{a_{l+i}^T p_k} > 0,\ i = 1, \cdots, m-l\Big\}, \tag{4.11}$$

with $t_k \overset{\text{def}}{=} +\infty$ if $\frac{a_{l+i}^T x_k - b^{l+i}}{a_{l+i}^T p_k} \ge 0$ for all i. Using the mean value theorem, we have that with $0 \le \vartheta_k \le 1$, the equality satisfies

$$\begin{aligned} f(x_k + \alpha_k p_k) = f(x_k) &+ \beta \alpha_k \nabla f(x_k)^T p_k + \alpha_k \{(1-\beta)\nabla f(x_k)^T p_k \\ &+ [\nabla f(x_k + \vartheta_k \alpha_k p_k) - \nabla f(x_k)]^T p_k\}. \end{aligned} \tag{4.12}$$

Under the assumptions on F, it is not difficult to show that $\nabla f(x)$ is locally Lipschitz continuous, and there exist sufficiently small α_k and $\delta' > 0$ when $\|\vartheta_k \alpha_k p_k\| \le \delta'$ such that

$$|[\nabla f(x_k + \vartheta_k \alpha_k p_k) - \nabla f(x_k)]^T p_k| \le (1-\beta)\omega_1 \epsilon \min\Big\{1, \frac{\epsilon}{2\kappa_M \Delta_{\max}}\Big\}\|p_k\|.$$

From (4.5), we get

$$\nabla f(x_k)^T p_k \le -\omega_1 \epsilon \min\Big\{\Delta_k, \frac{\epsilon}{\kappa_M}\Big\} \le -\omega_1 \epsilon \min\Big\{1, \frac{\epsilon}{\kappa_M \Delta_{\max}}\Big\}\Delta_k.$$

Hence, after a finite number of reductions, the last term in brackets in the right-handed side of (4.12) will become negative and the corresponding α_k will be acceptable, that is, we have that in a finite number of backtracking steps, α_k must satisfy (3.2). Hence, the conclusion of the lemma holds.

The above proof implies that after a finite number of reduction of α_k, by choosing correspondingly θ_k, the conditions (3.2), (3.3) and (3.4) must be satisfied and $x_{k+1} = x_k + \delta_k$ is well defined.

Lemma 4.4 *Assume that Assumptions* A1–A3 *hold. Assume further that* $\liminf_{k\to\infty} \Delta_k = 0$. *Then*

$$\lim_{k\to\infty} \theta_k = 1 \quad \textit{and} \quad \liminf_{k\to\infty} |\nabla f_k^T g_k|^{\frac{1}{2}} = 0.$$

Furthermore, if $\liminf_{k\to\infty} |\nabla f_k^T g_k|^{\frac{1}{2}} > 0$, *then*

$$\lim_{k\to\infty} \alpha_k = 1.$$

Proof By the assumptions, there exists an infinite subset $\mathcal{K}_0 \stackrel{\text{def}}{=} \{k_i\}$ of $\mathcal{K} \stackrel{\text{def}}{=} \{1, 2, \cdots\}$ such that

$$\lim_{i\to\infty} \Delta_{k_i} = 0.$$

The mechanism of the trust region radius update ensures that

$$\lim_{k\to\infty} \Delta_k = 0, \tag{4.13}$$

which means that $p_k \to 0$. Therefore, by the condition on the strictly feasible step size $\theta_k \in (\theta_0, 1]$, $0 < \theta_0 < 1$ and $\theta_k - 1 = o(\|p_k\|)$, $\lim_{k\to\infty} \theta_k = 1$ comes from $\lim_{k\to\infty} p_k = 0$.

If the conclusion of the lemma is not true, then there exists $\varepsilon > 0$ such that

$$|\nabla f_k^T g_k|^{\frac{1}{2}} \geq \varepsilon$$

for all k. Since $\nabla f(\cdot)$ is Lipschitz continuous on the neighborhood $\mathcal{N}$ of any accumulation point of $\{x_k\}$, there exists a Lipschitz constant $L > 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad \text{for all } x, y \in \mathcal{N}.$$

Further, we can obtain

$$\begin{aligned} |f(x_k+\delta_k) - f(x_k) - \psi_k(\delta_k; \widehat{\delta}_k)| &= \Big| f(x_k+\delta_k) - f(x_k) - \nabla f_k^T \delta_k - \frac{1}{2}\delta_k^T B_k \delta_k - \frac{1}{2}\widehat{\delta}_k^T C_k \widehat{\delta}_k \Big| \\ &= \Big| \int_0^1 \langle \nabla f(x_k + t\delta_k) - \nabla f_k, \delta_k \rangle \mathrm{d}t - \frac{1}{2}\delta_k^T B_k \delta_k - \frac{1}{2}\widehat{\delta}_k^T C_k \widehat{\delta}_k \Big| \\ &\leq L \int_0^1 t\|\delta_k\|^2 \mathrm{d}t + \frac{1}{2}\kappa_M \left\| \begin{bmatrix} \delta_k \\ \widehat{\delta}_k \end{bmatrix} \right\|^2 \\ &= \frac{1}{2}(L + \kappa_M) \left\| \begin{bmatrix} \delta_k \\ \widehat{\delta}_k \end{bmatrix} \right\|^2 \leq \frac{1}{2}(L + \kappa_M)\alpha_k^2 \theta_k^2 \Delta_k^2. \end{aligned} \tag{4.14}$$

From Lemma 4.2, we have

$$\begin{aligned}\operatorname{Pred}(\delta_k) &\geq \theta_k\alpha_k\omega_3|\nabla f_k^T g_k|^{\frac{1}{2}}\min\Big\{\Delta_k, \frac{|\nabla f_k^T g_k|^{\frac{1}{2}}}{\|M_k\|}\Big\}\\ &\geq \alpha_k\theta_k\omega_3\varepsilon\min\Big\{\Delta_k, \frac{\varepsilon}{\|M_k\|}\Big\}\\ &= \omega_3\alpha_k\theta_k\varepsilon\Delta_k \quad \Big(\text{when } \Delta_k \leq \frac{\varepsilon}{\kappa_M}\Big).\end{aligned} \tag{4.15}$$

From (4.14), (4.15) and $\Delta_k \to 0$, we can get, from $\alpha_k \leq 1$,

$$|\rho_k - 1| = \Big|\frac{\operatorname{Ared}(\delta_k) - \operatorname{Pred}(\delta_k)}{\operatorname{Pred}(\delta_k)}\Big| \leq \frac{\frac{1}{2}(L+\kappa_M)\alpha_k^2\theta_k^2\Delta_k^2}{\alpha_k\theta_k\omega_3\varepsilon\Delta_k} = \frac{(L+\kappa_M)\alpha_k\theta_k\Delta_k}{\omega_3\varepsilon} \to 0.$$

Thus $\rho_k \to 1$, which contradicts (4.13). So, the first conclusion of the lemma is true.

Now, we prove the second conclusion of the lemma. If (4.13) holds, taking norm in (4.1), we can obtain

$$\begin{aligned}\nu_k\Delta_k &= \nu_k\|(p_k;\widehat{p}_k)\|\\ &\geq (\|\nabla f(x_k) - A_1^T\lambda_{k+1} - A_2^T\mu_{k+1}\|^2 + \|D_k^{\frac{1}{2}}\lambda_k\|^2)^{\frac{1}{2}} - \|M_k\|\|(p_k;\widehat{p}_k)\|\\ &= |\nabla f_k^T g_k|^{\frac{1}{2}} - \|M_k\|\|(p_k;\widehat{p}_k)\|.\end{aligned}$$

Noting $\|(p_k;\widehat{p}_k)\| \leq \Delta_k$, we get

$$\nu_k \geq \frac{|\nabla f_k^T g_k|^{\frac{1}{2}}}{\Delta_k} - \|M_k\|.$$

Since $\left[\begin{smallmatrix} A_1 & 0 \\ A_2 & -D_k^{\frac{1}{2}}\end{smallmatrix}\right]$ has full row rank in the compact set $\mathcal{L}(x_0)$, and $\{\lambda_{k+1}\},\ \{\mu_{k+1}\},\ \{\|M_k\|\}$ are bounded, $\Delta_k \to 0$ implies that

$$\lim_{k\to\infty}\nu_k = +\infty. \tag{4.16}$$

If α_k given in Step 7 is the stepsize to the boundary of inequality constraints along p_k, then

$$\alpha_k \stackrel{\text{def}}{=} \min\Big\{-\frac{a_{l+i}^T x_k - b^{l+i}}{a_{l+i}^T p_k}\ \Big|\ -\frac{a_{l+i}^T x_k - b^{l+i}}{a_{l+i}^T p_k} > 0,\ i = 1,\cdots,m-l\Big\}, \tag{4.17}$$

with $\alpha_k \stackrel{\text{def}}{=} +\infty$ if $\frac{a_{l+i}^T x_k - b^{l+i}}{a_{l+i}^T p_k} \geq 0$ for all i. Assume that α_k given in (3.3) is the stepsize to the boundary of inequality constraints along p_k. From (4.1) and $\widehat{p}_k = D_k^{-\frac{1}{2}}A_2 p_k$, there exist μ_{k+1} and ν_k such that

$$a_{l+i}^T p_k = (a_{l+i}^T x_k - b^{l+i})^{\frac{1}{2}}\widehat{p}_k^i = -\frac{(a_{l+i}^T x_k - b^{l+i})\mu_{k+1}^i}{\nu_k + |\mu_{k+1}^i|},$$

where $\widehat{p}_k^i$ and μ_{k+1}^i are the ith components of the vectors $\widehat{p}_k$ and μ_{k+1}, respectively. Hence, there exists $j \in \{1,\cdots,m-l\}$ such that

$$\alpha_k = -\frac{a_{l+j}^T x_k - b^{l+j}}{a_{l+j}^T p_k} \geq \frac{\nu_k + |\mu_{k+1}^j|}{|\mu_{k+1}^j|} \geq \frac{\nu_k + |\mu_{k+1}^j|}{\|\mu_{k+1}\|_\infty}. \tag{4.18}$$

This inequality, together with (4.16), means that as α_k given in Step 7 is the step size to the boundary of inequality constraints along p_k,

$$\lim_{k\to\infty} \alpha_k = +\infty, \tag{4.19}$$

where α_k is given in the step size to the boundary of inequality constraints along p_k.

Next, we prove that if

$$\Delta_k \le \min\Big\{\frac{\varepsilon}{\kappa_M}, \frac{\omega_1\varepsilon(1-\beta)}{L}\Big\}, \tag{4.20}$$

then $\alpha_k = 1$ must satisfy the accepted condition (3.2) in Step 7, that is,

$$f(x_k + p_k) \le f(x_k) + \beta \nabla f_k^T p_k. \tag{4.21}$$

If the above formula is not true, by the gradient $\nabla f(x)$ being Lipschitz continuous with Lipschitz constant L, we have

$$\begin{aligned} 0 &< f(x_k + p_k) - f(x_k) - \beta\nabla f_k^T p_k \\ &= \nabla f(x_k + \xi_k p_k)^T p_k - \nabla f(x_k)^T p_k + (1-\beta)\nabla f_k^T p_k \\ &\le (1-\beta)\nabla f_k^T p_k + L\Delta_k^2, \end{aligned}$$

where $\xi_k \in (0,1)$, which implies that

$$0 < (1-\beta)\nabla f_k^T p_k + L\Delta_k^2. \tag{4.22}$$

By (4.5), we can obtain

$$-\omega_1\varepsilon(1-\beta)\min\Big\{\Delta_k, \frac{\varepsilon}{\kappa_M}\Big\} + L\Delta_k^2 > 0. \tag{4.23}$$

From (4.20), we have

$$[-\omega_1\varepsilon(1-\beta) + L\Delta_k]\Delta_k > 0.$$

This means that by $\Delta_k > 0$, $w_1\varepsilon(1-\beta) < L\Delta_k$, which contradicts (4.20).

From the above, we can see that if (4.21) holds, then the step size will be determined by (3.2). So, $\lim\limits_{k\to\infty} \alpha_k = 1$, that is, the second conclusion of the lemma is also true.

Theorem 4.1 *Suppose that the mapping F is continuously differentiable and its Jacobian matrix $\nabla F(x)$ is positive definite for all $x \in \mathcal{S}$ and that the set $\mathcal{S}$ is nonempty. Further, assume that Assumptions A1–A3 hold and the strict complementarity of the minimization problem* (2.10) *holds at every limit point of $\{x_k\}$. Then, for any starting point $x_0 \in \mathrm{int}(\mathcal{S})$, the sequence $\{x_k\}$ converges to the unique solution of* (VIP) *wherever $\{x_k\}$ is bounded.*

Proof If $f(x_k + d_k) \le \kappa f(x_k)$ holds infinitely often, then from $\kappa \in (0,1)$ we can get $\lim\limits_{k\to\infty} f(x_k) = 0$. Since f is continuous by Proposition 2.3, $f(\overline{x}) = 0$ for any accumulation point $\overline{x}$ of $\{x_k\}$ and hence $\overline{x}$ is a solution of (VIP). Since (VIP) has at most a solution, it follows that $\overline{x}$ is the unique solution of (VIP) and the entire sequence $\{x_k\}$ has a unique accumulation point $\overline{x}$ and necessarily converges to $\overline{x}$.

Now we consider the case when $f(x_k + d_k) \le \kappa f(x_k)$ holds for only finitely many k. In this case, the sequence $\{x_k\}$ is generated by Steps 6, 7 and 8 and satisfies (3.2), (3.3) and (3.4) for k sufficiently large. Let $\{x_k\}_{k\in\mathcal{K}}$ be any convergent subsequence of $\{x_k\}$ and let $\overline{x}$ be its limit point.

If $\inf_{k\in\mathcal{K}} \|d_k\| = 0$, then there exists an infinite subset $\mathcal{K}_1$ of $\mathcal{K}$ such that $\lim_{k\in\mathcal{K}_1} \|d_k\| = 0$. Since $d_k = z(x_k) - x_k$ and $\lim_{k\in\mathcal{K}_1} x_k = \overline{x}$, if necessary, taking a subset of $\mathcal{K}_1$ which without loss of generality we still denote by $\mathcal{K}_1$, we can obtain $\lim_{k\in\mathcal{K}_1} z(x_k) = \overline{x}$. From the continuity of f by Proposition 2.3 and the fact that $f(z(x_k)) > \kappa f(x_k)$ $(0 < \kappa \le 1)$ by the algorithm, we have $f(\overline{x}) \ge \kappa f(\overline{x})$. Therefore, as $\kappa \in (0,1)$ and $f(\overline{x}) \ge 0$, $f(\overline{x}) = 0$ and $\overline{x}$ solves problem (VIP) by Proposition 2.4.

If $\inf_{k\in\mathcal{K}} \|d_k\| > 0$, then we should still consider the case of σ_k given in Step 5. If $f(\overline{x}) \ne 0$ for any accumulation point $\overline{x}$ of $\{x_k\}$, then by continuity there exists $\varepsilon > 0$ such that $|\nabla f_k^T g_k|^{\frac{1}{2}} \ge \varepsilon$ for all k.

(a) If $\inf_{k\in\mathcal{K}} \sigma_k = 0$, then

$$\inf_{k\in\mathcal{K}} \Delta_k = 0.$$

Thus there exists an infinite subset $\mathcal{K}_2$ of $\mathcal{K}$ such that

$$\lim_{k\in\mathcal{K}_2} \Delta_k = 0.$$

Lemma 4.4 means that if $\liminf_{k\to\infty} \Delta_k = 0$, then

$$\liminf_{k\to\infty} |\nabla f_k^T g_k|^{\frac{1}{2}} = 0,$$

which contradicts $|\nabla f_k^T g_k|^{\frac{1}{2}} \ge \varepsilon$ for all k. So the case $\inf_{k\in\mathcal{K}} \sigma_k = 0$ does not hold.

Now we consider the other case that $\inf_{k\in\mathcal{K}} \sigma_k > 0$.

(b) If $\inf_{k\in\mathcal{K}} \sigma_k > 0$, then

$$\inf_{k\in\mathcal{K}} \Delta_k > 0.$$

According to the acceptance rule in Step 7, we have

$$f(x_k) - f(x_k + \alpha_k p_k) \ge -\alpha_k \beta \nabla f_k^T p_k. \tag{4.24}$$

From the first-order expansions of $f(x_k + \alpha_k p_k)$ and $f(x_k + \theta_k \alpha_k p_k)$ at the point x_k, we have

$$\begin{aligned} f(x_k + \theta_k\alpha_k p_k) - f(x_k + \alpha_k p_k) &= [f(x_k + \theta_k\alpha_k p_k) - f_k] - [f(x_k + \alpha_k p_k) - f_k] \\ &= -(1-\theta_k)\alpha_k \nabla f_k^T p_k + o(\|\alpha_k p_k\|). \end{aligned}$$

By the definition of $(1-\theta_k)$, (4.24) and (4.5), we have

$$\begin{aligned} f(x_k) - f(x_k + \theta_k\alpha_k p_k) &= f(x_k) - f(x_k + \alpha_k p_k) - [f(x_k + \theta_k\alpha_k p_k) - f(x_k + \alpha_k p_k)] \\ &\ge -\alpha_k\beta\nabla f_k^T p_k - (1-\theta_k)\alpha_k \nabla f_k^T p_k + o(\|\alpha_k p_k\|) \\ &\ge \alpha_k \omega_1 \varepsilon \min\left\{\Delta_k, \frac{\varepsilon}{\|M_k\|}\right\} + o(\|\alpha_k p_k\|). \end{aligned} \tag{4.25}$$

Since $\lim\limits_{\substack{k\to\infty\\ k\in\mathcal{K}}} x_k = \overline{x}$ and $f(x_k)$ is monotonically decreasing and is bounded, we have

$$\lim_{\substack{k\to\infty\\ k\in\mathcal{K}}} [f(x_{k+1}) - f(x_k)] = 0.$$

So from (4.25), we can obtain that $\inf\limits_{k\in\mathcal{K}} \Delta_k > 0$ means

$$\liminf_{\substack{k\to\infty\\ k\in\mathcal{K}}} \alpha_k = 0. \tag{4.26}$$

Assume $a_{l+j}^T \overline{x} > b^{l+j}$ for all j. Recalling (4.15), we conclude that for k sufficiently large and for all j,

$$a_{l+j}^T x_k - b^{l+j} \geq \frac{1}{2}(a_{l+j}^T \overline{x} - b^{l+j}) > 0.$$

Hence

$$\alpha_k \overset{\text{def}}{=} \min\Big\{ -\frac{a_{l+i}^T x_k - b^{l+i}}{a_{l+i}^T p_k} \;\Big|\; -\frac{a_{l+i}^T x_k - b^{l+i}}{a_{l+i}^T p_k} > 0,\ i = 1, \cdots, m-l \Big\} > 0.$$

Let the corresponding Lagrange multipliers be $\overline{\mu}$ and $\overline{\lambda}$ at $\overline{x}$, and hence, without loss of generality, assume $a_{l+j}^T \overline{x} - b^{l+j} = 0$ for some j. Since the strict complementarity of the problem (2.10) holds, we have that from $(\mu_k)_j \to \overline{\mu}_j$, for k sufficiently large,

$$|(\mu_k)_j| \geq \frac{1}{2}|\overline{\mu}_j| > 0. \tag{4.27}$$

Recalling (4.16) again, we conclude that

$$\alpha_k = -\frac{a_{l+j}^T x_k - b^{l+j}}{a_{l+j}^T p_k} \geq \frac{\nu_k + |\mu_{k+1}^j|}{|\mu_{k+1}^j|} \geq \frac{\nu_k + |\mu_{k+1}^j|}{\|\mu_{k+1}\|_\infty} \geq \frac{\frac{1}{2}|\overline{\mu}_j|}{\|\mu_{k+1}\|_\infty} > 0.$$

Hence

$$\lim_{k\to\infty} \alpha_k \neq 0.$$

From the above discussion, we can obtain that if α_k is determined by (3.3), then

$$\lim_{k\to\infty} \alpha_k \neq 0. \tag{4.28}$$

So, $\lim\limits_{k\to\infty} \alpha_k = 0$ holds only in (3.2). The acceptance rule (3.2) means that, for large enough k,

$$f\Big(x_k + \frac{\alpha_k}{\omega} p_k\Big) - f(x_k) > \beta \frac{\alpha_k}{\omega} \nabla f_k^T p_k. \tag{4.29}$$

Since

$$f\Big(x_k + \frac{\alpha_k}{\omega} p_k\Big) - f(x_k) = \frac{\alpha_k}{\omega} \nabla f_k^T p_k + o\Big(\frac{\alpha_k}{\omega}\|p_k\|\Big),$$

we have

$$(1-\beta)\frac{\alpha_k}{\omega} \nabla f_k^T p_k + o\Big(\frac{\alpha_k}{\omega}\|p_k\|\Big) \geq 0. \tag{4.30}$$

Dividing (4.30) by $\frac{\alpha_k}{\omega}\|p_k\|$ and noting that $1-\beta > 0$ and $\nabla f_k^T p_k \leq 0$, we obtain

$$\lim_{k\to\infty} \frac{\nabla f_k^T p_k}{\|p_k\|} = 0. \tag{4.31}$$

From

$$\nabla f_k^T p_k \leq -\omega_1 |\nabla f_k^T g_k|^{\frac{1}{2}} \min\left\{\Delta_k, \frac{|\nabla f_k^T g_k|^{\frac{1}{2}}}{\kappa_M}\right\} \leq -\omega_1 \varepsilon \min\left\{\Delta_k, \frac{\varepsilon}{\kappa_M}\right\} \tag{4.32}$$

we have that (4.30) means

$$\lim_{k\to\infty} \frac{\Delta_k}{\|p_k\|} = 0, \tag{4.33}$$

which contradicts $\|(p_k; \widehat{p}_k)\| \leq \Delta_k$. From (a) and (b), we have that $|\nabla f_k^T g_k|^{\frac{1}{2}} \geq \varepsilon$ is not true, which implies that $\overline{x}$ is a solution to (VIP).

From the uniqueness of the solution to (VIP), we can conclude that the entire sequence $\{x_k\}$ has a unique accumulation point $\overline{x}$ and converges to $\overline{x}$.

5 Properties of the Local Convergent Rate

In this section, we will show that, under the assumption that F is strongly monotone on $x \in \mathcal{S}$, the algorithm is locally quadratically convergent.

Proposition 5.1 (see [8]) *Let x_* be a solution to* (VIP). *If F is strongly monotone with modulus μ on $x \in \mathcal{S}$, then f of* (2.8) *satisfies the inequality*

$$f(x) \geq \Big(\mu - \frac{1}{2}\|G\|\Big)\|x - x_*\|^2 \quad \text{for all } x \in \mathcal{S}. \tag{5.1}$$

In particular, if the matrix G is chosen to be sufficiently small so that $\|G\| < 2\mu$, then

$$\lim_{\substack{x\in\mathcal{S} \\ \|x\|\to\infty}} f(x) = +\infty.$$

It is obvious that the decreasing of $\{f(x_k)\}$ and Proposition 5.1 imply that, when F is strongly monotone on $x \in \mathcal{S}$ and when the matrix G is sufficiently small, the sequence $\{x_k\}$ generated by the proposed algorithm is bounded. To obtain the second order convergence result, we need the following strict complementarity condition in [8], which is a generalization of the strict complementarity condition for inequality constraints and corresponding Lagrange multipliers that appear in the Karush-Kuhn-Tucher conditions in nonlinear programming.

Definition 5.1 (see [8]) *Suppose that $x \in \mathcal{S}$ is polyhedral and that* (VIP) *has a unique solution x_*. Let T^* denote the minimal face of $x \in \mathcal{S}$ containing x_*. Then the strict complementarity holds at x_* if $x \in \mathcal{S}$ and that $\langle F(x_*), x - x_*\rangle = 0$ imply $x \in T^*$.*

Similar to the conclusion presented in [6], now we give the following rate of convergence result whose proof is standard.

Theorem 5.1 (Quadratic Convergence) *Suppose that the set $x \in \mathcal{S}$ is polyhedral convex, the mapping F is strongly monotone with modulus μ on $x \in \mathcal{S}$ and $\nabla F(\cdot)$ is Lipschitz continuous on a neighborhood $\mathcal{N}$ of the unique solution x_* of* (VIP). *If the matrix G is sufficiently small such that $\|G\| < 2\mu$ and the strict complementarity condition holds at x_*, then the sequence $\{x_k\}$ generated by algorithm converges quadratically to x_*.*

Theorem 5.1 means that the local convergence rate for the proposed algorithm depends on the quality of the the matrix G. We have studied the convergence properties of affine scaling interior trust region strategy via optimal path to modify Newton method for the strictly monotone variational inequality problem subject to linear equality and inequality constraints. One of the future research topics is to carry out the numerical experience of the proposed algorithm.

References

[1] Bertsekas, D. P. and Tsitsiklis, J. N., Parallel and Distributed Computation: Numerical Methods, Prentice-Hall, Englewood Cliffs, New York, 1989.

[2] Bulteau, J. P. and Vial, J. Ph., Curvilinear path and trust region in unconstrained optimization, a convergence analysis, *Math. Prog. Stud.*, **30**, 1987, 82–101.

[3] Coleman, T. and Li, Y., A trust region and affine scaling interior point method for nonconvex minimization with linear inequality constraints, *Math. Prog. Stud., Ser. A,* **88**, 2000, 1–31.

[4] Eaves, B., Computing stationary points, *Math. Prog. Stud.*, **7**, 1978, 1–14.

[5] Fukushima, M., Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems, *Math. Prog. stud.*, **53**, 1992, 99–110.

[6] Liang, X., Xu, C. and Qian, J., A trust region-type method for solving monotone variational inequality, *J. Comp. Math.*, **18**, 2000, 13–24.

[7] Moré, J. J., Recent developments in algorithms and software for trust region methods, in Mathematical Programming, The State of the Art, M. G. A. Bachem and E. B. Dorte (eds.), Springer-Verlag, Berlin, 1983.

[8] Taji, K., Fukushima, M. and Ibaraki, T., A globally convergent Newton method for solving strongly monotone variational inequalities, *Math. Prog. Stud.*, **58**, 1993, 369–383.

[9] Zhu, D., Curvilinear paths and trust region methods with nonmonotonic back tracking technique for unconstrained optimization, *J. Comp. Math.*, **19**, 2001, 241–258.

[10] Zhu, D., A new affine scaling interior point algorithm for nonlinear optimization subject to linear equality and inequality constraints, *J. Comp. Appl. Math.*, **161**, 2003, 1–25.