Journal of Computational Mathematics
Vol.30, No.2, 2012, 197–222.

http://www.global-sci.org/jcm
doi:10.4208/jcm.1107-m3559

A FEASIBLE SEMISMOOTH GAUSS-NEWTON METHOD FOR SOLVING A CLASS OF SLCPS*

Changfeng Ma
School of Mathematics and Computer Science, Fujian Normal University, Fuzhou 350007, China
Email: macf@fjnu.edu.cn

Abstract

In this paper, we consider a class of the stochastic linear complementarity problems (SLCPs) with finitely many elements. A feasible semismooth damped Gauss-Newton algorithm for the SLCP is proposed. The global and locally quadratic convergence of the proposed algorithm are obtained under suitable conditions. Some numerical results are reported in this paper, which confirm the good theoretical properties of the proposed algorithm.

Mathematics subject classification: 90C33, 65K10.
Key words: Stochastic linear complementarity problems, Gauss-Newton algorithm, Convergence analysis, Numerical results.

1. Introduction

Assume that $(\Omega, \mathcal{F}, \mathcal{P})$ is a probability space with $\Omega \subseteq \Re^n$, where the probability distribution $\mathcal{P}$ is known. The stochastic linear complementarity problem (see [1–10]) is to find a vector $x \in \Re^n$ such that

$$x \geq 0, \quad M(\omega)x + q(\omega) \geq 0, \quad x^T\big[M(\omega)x + q(\omega)\big] = 0, \quad \text{a.e. } \omega \in \Omega, \tag{1.1}$$

where $\Omega \subset \Re^n$ is the underlying sample space and $\omega \in \Omega$ is a random vector with given probability distribution $\mathcal{P}$ and, for each ω, $M(\omega) \in \Re^{n\times n}$ and $q(\omega) \in \Re^n$.

Problem (1.1) is usually denoted by SLCP$(M(\omega), q(\omega))$ or SLCP, briefly. If Ω is a singleton, SLCP reduces to the intensively studied and standard linear complementarity problem (denoted by LCP); see [11–14].

In general there is no vector x satisfying (1.1) for all $\omega \in \Omega$. In order to obtain a reasonable solution of Problem (1.1), there have been several types of models being proposed. One of them is the expected value (EV) model [15] that formulates (1.1) as follows: Let $\bar{M} = E[M(\omega)]$ and $\bar{q} = E[q(\omega)]$ be mathematical expectations of $M(\omega)$ and $q(\omega)$, respectively. The EV model is to find an $x \in \Re^n$ such that

$$x \geq 0, \quad \bar{y} = \bar{M}x + \bar{q} \geq 0, \quad x^T\bar{y} = 0. \tag{1.2}$$

Another is the expected residual minimization (ERM) model (see [1,7]). The ERM model is to find an $x \in \Re^n_+$ that minimizes the expected total residual function

$$\min_{x\geq 0} f(x) = E\big[\|\tilde{\Phi}(x,\omega)\|^2\big] = \sum_{i=1}^n E\big\{[\varphi(x_i, M_i(\omega)x + q_i(\omega))]^2\big\}, \tag{1.3}$$

* Received October 20, 2010 / Revised version received May 18, 2011 / Accepted June 24, 2011 /
Published online February 24, 2012 /

where $M_i(\omega)\,(i=1,\cdots,n)$ is the i-th row of random matrix $M(\omega)$ and $\varphi:\Re^2\to\Re$ is an NCP-function that satisfies

$$\varphi(a,b)=0 \iff a\ge 0,\ b\ge 0,\ ab=0$$

and

$$\tilde{\Phi}(x,\omega)=\begin{pmatrix}\varphi(x_1,M_1(\omega)x+q_1(\omega))\\ \vdots\\ \varphi(x_n,M_n(\omega)x+q_n(\omega))\end{pmatrix}.$$

Recently, Zhou and Caccetta (see [16]) present a new model for a class of stochastic linear complementarity problems in which sample space Ω has only finitely many elements. Let $\Omega=\{\omega_1,\omega_2,\cdots,\omega_m\}$ and their model is to find an $x\in\Re^n$ such that

$$x\ge 0,\quad M(\omega_i)x+q(\omega_i)\ge 0,\quad x^T\big[M(\omega_i)x+q(\omega_i)\big]=0,\quad i=1,\cdots,m,\ m>1. \tag{1.4}$$

In their model it is assumed that $p_i=\mathcal{P}\{\omega_i\in\Omega\}>0$, $i=1,\cdots,m$, and let $\bar{M}$ and $\bar{q}$ be the expectation values of the random matrix $M(\omega)$ and random vector $q(\omega)$, i.e.,

$$\bar{M}=\sum_{i=1}^m p_iM(\omega_i),\quad \bar{q}=\sum_{i=1}^m p_iq(\omega_i). \tag{1.5}$$

They claim that problem (1.4) is equivalent to (1.6)-(1.7):

$$x\ge 0,\quad \bar{M}x+\bar{q}\ge 0,\quad x^T(\bar{M}x+\bar{q})=0, \tag{1.6}$$

$$M(\omega_i)x+q(\omega_i)\ge 0,\quad i=1,\cdots,m. \tag{1.7}$$

Furthermore, they define

$$\Phi_\alpha(x)=\begin{pmatrix}\varphi_\alpha(x_1,(\bar{M}x+\bar{q})_1)\\ \vdots\\ \varphi_\alpha(x_n,(\bar{M}x+\bar{q})_n)\end{pmatrix},$$

where, $\varphi_\alpha(a,b)=a+b-\sqrt{a^2+b^2}+\alpha[a]_+[b]_+$ with $\alpha>0$ and $[t]_+=\max\{0,t\}$. Then problem (1.4), if it has a solution, can be reformulated as the following minimization problem with nonnegative constraints

$$\begin{aligned}\min\quad & \theta(z)=\frac{1}{2}\|\tilde{H}(z)\|^2,\\ \text{s.t.}\quad & z\ge 0,\end{aligned} \tag{1.8}$$

where $z=(x,y)\in\Re^n\times\Re^{mn}$ and

$$\tilde{H}(z):=\tilde{H}(x,y)=\begin{pmatrix}\Phi_\alpha(x)\\ \tilde{M}(\omega)x+\tilde{q}(\omega)-y\end{pmatrix}.$$

Here

$$\tilde{M}(\omega)=\begin{pmatrix}M(\omega_1)\\ \vdots\\ M(\omega_m)\end{pmatrix}\in\Re^{mn\times n},\quad \tilde{q}(\omega)=\begin{pmatrix}q(\omega_1)\\ \vdots\\ q(\omega_m)\end{pmatrix}\in\Re^{mn}.$$

The authors of [16] propose a semismooth Newton method for solving the constrained minimization problem (1.8). They also examined the effectiveness of the algorithm by means of numerical experiments.

The method in [16] can be regarded as a direct application of the asymptotically feasible semismooth Newton method (AFSN) proposed in [17]. The AFSN yields a stationary point of (1.8) or generates an infinite sequence, the cluster point of which is a stationary point of (1.8). The search direction used in AFSN is the asymptotically Newton direction, i.e., the convex combination of the projected gradient direction and the projected Newton direction. Notice that, at the k-th iteration, if the Newton equation has no solution, then the Newton direction obtained by AFSN is not exact. Hence, the asymptotically Newton direction will turn to the projection gradient direction. In this case, the performance of AFSN may be dissatisfactory. Since AFSN always chooses the projected direction after certain iteration, the convergence rate of AFSN is poor for those test problems in [15]. Furthermore, we notice that, when m and n are large, problem (1.8) has $n(m+1)$ variables, which is difficult to solve. Based on these reasons, we are motivated to seek a new reformulation and algorithms for solving problem (1.8).

In the last decade, there have been strong interests in smoothing Newton-type methods for solving the classical nonlinear complementarity problems [18–23]. Motivated by Reference [16] and smoothing Newton-type method, in this paper, we present a smoothing Gauss-Newton method for solving the stochastic linear complementarity problem (1.4) based on the smoothing Fischer-Burmeister NCP-function. Under suitable assumptions, the proposed algorithm is proved to be convergent globally and superlinearly/quadratically.

The rest of this paper is organized as follows. In Section 2, we state some preliminary results. In Sections 3 and 4, we propose a feasible semismooth damped Gauss-Newton method for the stochastic linear complementarity problem (1.4) and prove the global and locally quadratic convergence of the proposed algorithm. Some numerical results are reported in Section 5.

The following notations will be used throughout this paper. All vectors are column vectors, the superscript ^{T} denotes transpose, $\Re^n$ (respectively, $\Re$) denotes the space of n-dimensional real column vectors (respectively, real numbers), $\Re^n_+$ and $\Re^n_{++}$ denote the nonnegative and positive orthants of $\Re^n$, $\Re_+$ (respectively, $\Re_{++}$) denotes the nonnegative (respectively, positive) orthant in $\Re$. For any $v \in \Re^n$, let $[v]_+ := \max\{v, 0\} = (\max\{v_1, 0\}, \cdots, \max\{v_n, 0\})^T$. We define $N := \{1, 2, \cdots, n\}$. For any vector $u \in \Re^n$, we denote the diagonal matrix whose ith diagonal element is u_i by $\mathrm{diag}\{u_i : i \in N\}$. For simplicity, we use (u, v) for the column vector $(u^T, v^T)^T$. The matrix I represents the identity matrix of arbitrary dimension. The symbol $\|\cdot\|$ stands for the 2-norm. For any α, $\beta \in R_{++}$, $\alpha = \mathcal{O}(\beta)$ (respectively, $\alpha = o(\beta)$) means α/β is uniformly bounded (respectively, tends to zero) as $\beta \to 0$.

2. Constrained Equations Reformulation

In this section, we reformulate problem (1.4) as a minimization problem with nonnegative constraints. Notice that (1.6) is a standard LCP and it can be reformulated as a system of semismooth equations by a smoothing NCP-function. Throughout this paper, we use the following NCP-function $\phi : \Re^2 \to \Re$ [16]:

$$\phi_\alpha(a, b) = (a + b) - \sqrt{a^2 + b^2} + \alpha a_+ b_+. \tag{2.1}$$

Then, it is easy to verify the following proposition.

Proposition 2.1. *Assume that ϕ_α is defined by* (2.1). *Then*

(1) $\phi_\alpha(a, b) = 0 \iff a \ge 0,\ b \ge 0,\ ab = 0.$

(2) $\phi_\alpha(\cdot)$ *is strongly semismooth on* $\Re^2$.

(3) $\phi_\alpha(\cdot)^2$ *is continuously differentiable on* $\Re^2$.

(4) *the generalized gradient* $\partial\phi_\alpha(a,b)$ *at a point* $(a,b)\in\Re^2$ *is equal to the set of all* (ξ_a,ξ_b) *such that*

$$(\xi_a,\xi_b)=\begin{cases}\Big(1-\frac{a}{\sqrt{a^2+b^2}},1-\frac{b}{\sqrt{a^2+b^2}}\Big)+\ \alpha(b_+\partial a_+,a_+\partial b_+), & \textit{if } (a,b)\neq(0,0),\\ (1-\zeta,1-\eta), & \textit{if } (a,b)=(0,0),\end{cases} \tag{2.2}$$

where (ζ,η) *is any vector satisfying* $\sqrt{\zeta^2+\eta^2}\le 1$ *and*

$$\partial c_+=\begin{cases}1, & \textit{if } c>0,\\ \{[0,1]\}, & \textit{if } c=0,\\ 0, & \textit{if } c<0.\end{cases}$$

(5) *Let* $\{a^k\},\{b^k\}\subset\Re$ *be any two sequences such that either* $a^k\to-\infty$, *or* $b^k\to-\infty$, *or* $a_+^k b_+^k\to+\infty$. *Then* $|\phi_\alpha(a^k,b^k)|\to+\infty$.

Let

$$\Phi_\alpha(\mu,x)=\begin{pmatrix}\phi_\alpha(x_1,\bar M_1x+\bar q_1)\\ \vdots\\ \phi_\alpha(x_n,\bar M_nx+\bar q_n)\end{pmatrix}, \tag{2.3}$$

where $\bar M_i$ and $\bar q_i$ $(i=1,\cdots,n)$ are the i-th row of matrix $\bar M$ and the i-th component of vector $\bar q$, respectively, $\bar M$ and $\bar q$ are defined in (1.5). Then, x solves (1.6) if and only if $\Phi_\alpha(x)=0$. Furthermore, let

$$M(\omega)=\begin{pmatrix}M(\omega_1)\\ M(\omega_2)\\ \vdots\\ M(\omega_m)\end{pmatrix},\quad q(\omega)=\begin{pmatrix}q(\omega_1)\\ q(\omega_2)\\ \vdots\\ q(\omega_m)\end{pmatrix}.$$

Thus, if (1.4) has a solution, then solving (1.4) is equivalent to solving the seimismooth system of equations with nonnegative constraints:

$$H(x)=0,\quad \text{with } x\ge 0, \tag{2.4}$$

where

$$H(x)=\begin{pmatrix}\Phi_\alpha(x)\\ G(x)\end{pmatrix}, \tag{2.5}$$

with

$$G(x)=[M(\omega)x+q(\omega)]_-. \tag{2.6}$$

Here, for any $x\in\Re^n$, the notation $[\cdot]_-$ implies that $[x]_-=\min(0,x)$ and the minimum is taken component-wise.

For function H defined in (2.5), we have

Proposition 2.2. *Assume that* H *is defined in* (2.4). *Then* $H(x)$ *is local Lipschitz and strongly semismooth.*

Proof. By the fact that $[\cdot]_-$ is strongly semismooth and Proposition 2.1 (2), it follows from Theorems 19 and 20 of [28] that the function $H(x)$ is strongly semismooth. □

Proposition 2.3. *Suppose that $LCP(\bar{M},\bar{q})$ is R-regular at a solution x^*. Then, all $V_\Phi \in \partial_C\Phi_\alpha(x^*) = \partial\Phi_{\alpha,1}(x^*) \times \partial\Phi_{\alpha,2}(x^*) \times \cdots \times \partial\Phi_{\alpha,n}(x^*)$ are nonsingular.*

For any $x \in \Re^n$, we have that

$$\partial_C H(x) = \left\{ \begin{pmatrix} V_\Phi \\ V_G \end{pmatrix} : V_\Phi \in \partial_C\Phi_\alpha(x),\ V_G \in \partial_C G(x) \right\}, \tag{2.7}$$

where $V_\Phi \in \Re^{n\times n}$, $V_G \in \Re^{mn\times n}$. Hence, by Proposition 2.3, we have:

Proposition 2.4. *Suppose that x^* is a solution of (2.4) and that the $LCP(\bar{M},\bar{q})$ is R-regular at the solution x^*. Then, for any*

$$V = \begin{pmatrix} V_\Phi \\ V_G \end{pmatrix} \in \partial_C H(x),$$

the matrix

$$V^T V = V_\Phi^T V_\Phi + V_G^T V_G$$

is nonsingular.

For any $x \in \Re^n$, an element of the C-subdifferential of $\partial_C\Phi_\alpha(x)$ or $\partial_C G(x)$ can be calculated as follwos.

Algorithm 2.1. (Procedure to calculate an element $\mathbf{V_\Phi} \in \partial_\mathbf{C}\mathbf{\Phi}_\alpha(\mathbf{x})$)

Step 0. *Let $x \in \Re^n$ be given, and let $V_{\Phi,i}$ denote the ith row of the matrix $V_\Phi \in \partial_C\Phi_\alpha(x)$.*

Step 1. *Set $S_1 = \{i : x_i = \bar{M}_i x + \bar{q}_i = 0\}$ and $S_2 = \{i : x_i > 0,\ \bar{M}_i x + \bar{q}_i > 0\}$.*

Step 2. *Set $c \in \Re^n$ such that $c_i = 1$ for $i \in S_1$ and $c_i = 0$ for $i \notin S_1$.*

Step 3. *For $i \in S_1$, set*

$$V_{\Phi,i} = \Big(1 - \frac{c_i}{\sqrt{c_i^2 + (\bar{M}_i c)^2}}\Big)e_i^T + \Big(1 - \frac{\bar{M}_i c}{\sqrt{c_i^2 + (\bar{M}_i c)^2}}\Big)\bar{M}_i^T.$$

Step 4. *For $i \in S_2$, set*

$$V_{\Phi,i} = \Big(1 - \frac{x_i}{\sqrt{x_i^2 + (\bar{M}_i x + \bar{q}_i)^2}} + \alpha(\bar{M}_i x + \bar{q}_i)\Big)e_i^T + \Big(1 - \frac{\bar{M}_i x + \bar{q}_i)^2}{\sqrt{c_i^2 + (\bar{M}_i x + \bar{q}_i)^2}} + \alpha x_i\Big)\bar{M}_i.$$

Step 5. *For $i \notin S_1 \cup S_2$, set*

$$V_{\Phi,i} = \Big(1 - \frac{x_i}{\sqrt{x_i^2 + (\bar{M}_i x + \bar{q}_i)^2}}\Big)e_i^T + \Big(1 - \frac{\bar{M}_i x + \bar{q}_i)^2}{\sqrt{c_i^2 + (\bar{M}_i x + \bar{q}_i)^2}}\Big)\bar{M}_i.$$

Algorithm 2.2. (**Procedure to calculate an element $\mathbf{V_G} \in \partial_\mathbf{C}\mathbf{G}(\mathbf{x})$**)

Step 0. *Let $x \in \Re^n$ be given, and let $V_{G,i}$ denote the ith row of the matrix $V_G \in \partial_C G(x)$, $i = 1, \cdots, mn$.*

Step 1. *Set $T_1 = \{i : M_i(\omega)x + q_i(\omega) < 0\}$ and $T_2 = \{i : M_i(\omega)x + q_i(\omega) \geq 0\}$.*

Step 2. *For $i \in T_1$, set $V_{G,i} = M_i(\omega)$.*

Step 3. *For $i \in T_2$, set $V_{G,i} = 0$.*

Now, we define the merit function of (2.4) by

$$\Psi(x) = \frac{1}{2}\|H(x)\|^2. \tag{2.8}$$

Let

$$\Psi_1(x) = \frac{1}{2}\|\Phi_\alpha(x)\|^2 \tag{2.9}$$

and

$$\Psi_2(x) = \frac{1}{2}\|G(x)\|^2. \tag{2.10}$$

It is not difficult to see that if (1.4) has a solution, so (2.4) has a solution which is equivalent to the following optimization problem has a global minimal solution with nonnegative constraints and the minimal value is zero.

$$\begin{aligned} \min \quad & \Psi(x) = \frac{1}{2}\|H(x)\|^2, \\ \text{s.t.} \quad & x \geq 0. \end{aligned} \tag{2.11}$$

One can easily see that x is a stationary point of (2.11) if only if

$$x \geq 0, \quad \nabla\Psi(x) \geq 0, \quad x^T\nabla\Psi(x) = 0. \tag{2.12}$$

In what follows, we will concern with finding a point x which satisfies (2.12).

From [12], we have the following lemma.

Lemma 2.1. *The function $\Psi(x)$ is continuously differentiable on $\Re^n$ with*

$$\nabla\Psi(x) := \nabla\Psi_1(x) + \nabla\Psi_2(x) = V_\Phi^T\Phi_\alpha(x) + V_G^T G(x).$$

and $\nabla\Psi(x)$ is strongly semismooth, i.e.,

$$\nabla\Psi(x) = \nabla\Psi(x^*) + U^T(x - x^*) + \mathcal{O}(\|x - x^*\|^2),$$

where $U \in \partial\nabla\Psi(x)$ is given by

$$U = V_\Phi^T V_\Phi + V_G^T V_G + \sum_{i=1}^{(m+1)n} H_i(x)T_i(x).$$

Here $T_i(x) \in \Re^{n\times n}$ is the generalized Hessian of $H_i(x)$.

Throughout this paper, we make the following assumption:

Assumption 2.1. *$T_i(x) \in \Re^{n\times n}$, the generalized Hessian of $H_i(x)$, is locally Lipschitz continuous.*

Now we examine the boudedness of the level set of problem (2.7). For a given $x^0 \in \Re^n_+$, the level set of $\Psi(x)$ is defined as follows:

$$\mathcal{L}_{\Psi(x)}(x^0) = \{x \in \Re^n_+ \mid \Psi(x) \le \Psi(x^0)\}. \tag{2.13}$$

We recall the definition of the stochastic R_0 matrix in [9]. A stochastic matrix $M(\omega)$ is called as a stochastic R_0 matrix if

$$x \ge 0,\ M(\omega)x \ge 0,\ x^T M(\omega)x = 0,\ a.e. \implies x = 0.$$

If the expected matrix of $M(\omega)$, $\bar{M} = E[M(\omega)]$, is an R_0 matrix, then $M(\cdot)$ is a stochastic R_0 matrix. However, the converse of this proposition is not true, that is, $M(\cdot)$ being a stochastic R_0 matrix does not imply that there is an $\omega \in \Omega$ such that $\bar{M} = E[M(\omega)]$ is an R_0 matrix (see [9] for more details about the stochastic R_0 matrix).

Next, we verify that the level set $\mathcal{L}_{\Psi(x)}(x^0)$ defined in (2.11) is bounded under the assumption that $M(\cdot)$ is a stochastic R_0 matrix.

Theorem 2.1. *The level set $\mathcal{L}_{\Psi(x)}(x^0)$ defined in* (2.11) *is bounded for any $q(\cdot)$ if and only if $M(\cdot)$ is a stochastic R_0 matrix.*

Proof. We first show the necessity. Let $q(\omega_i) \equiv 0\,(i = 1, \cdots, m)$ and $x^0 = 0$. Then

$$\Psi(x^0) = 0 \ \text{ and } \ 0 \in \mathcal{L}_{\Psi(x)}(x^0) = \{x \in \Re^n_+ \mid \Psi(x) \le \Psi(x^0)\}.$$

Suppose that $M(\cdot)$ is not a stochastic R_0 matrix. Then there exists an $\hat{x} \ge 0$ with $\hat{x} \ne 0$ such that

$$\hat{x} \ge 0, \quad M(\omega_i)\hat{x} \ge 0, \quad \hat{x}^T M(\omega_i)\hat{x} = 0, \quad i = 1, \cdots, m.$$

Thus, we obtain that

$$\hat{x}^T \bar{M}\hat{x} = 0 \qquad \text{and} \qquad \Psi_1(\hat{x}) = 0.$$

Furthermore, we have

$$(l\hat{x})^T \bar{M}(l\hat{x}) = 0 \ \text{ and } \ \Psi_1(l\hat{x}) = 0, \qquad \text{for all } l > 0.$$

Notice that $M(\omega)(l\hat{x}) + q(\omega) = M(\omega)(l\hat{x}) \ge 0$. Hence $\Psi_2(l\hat{x}) = 0$, which implies that $l\hat{x} \in \mathcal{L}_{\Psi(x)}(x^0)$. This contradicts with the boundedness of $\mathcal{L}_{\Psi(x)}(x^0)$.

Now we prove the sufficiency by contradiction. Suppose there is a sequence $\{x^k\} \subseteq \mathcal{L}_{\Psi(x)}(x^0)$ such that $\|x^k\| \to +\infty$ as $k \to +\infty$. Let

$$u^k = \frac{x^k}{\|x^k\|}.$$

Notice that the boundedness of the sequence $\{u^k\}$, without loss of generality, we can assume that

$$\lim_{k\to\infty} u^k = u^*.$$

Obviously $u^* \ge 0$ and $\|u^k\| = 1$. The following proof is divided into two steps (a) and (b).

(a) We claim $M(\omega_i)u^* \geq 0$ for all $i = 1, \cdots, m$ with reduction to absurdity. Suppose there exist $i_0 \in \{1, \cdots, m\}$ and $j_0 \in \{1, \cdots, n\}$ such that $M_{j_0}(\omega_{i_0})u^* < 0$, where $M_{j_0}(\cdot)$ denotes the j_0-th row. Since

$$\begin{aligned} M_{j_0}(\omega_i)u^* &= \lim_{k\to\infty} M_{j_0}(\omega_{i_0})u^k = \lim_{k\to\infty} M_{j_0}(\omega_{i_0})\frac{x^k}{\|x^k\|} \\ &= \lim_{k\to\infty} \frac{M_j(\omega_{i_0})x^k + q_j(\omega_{i_0})}{\|x^k\|}. \end{aligned}$$

Notice that $M_{j_0}(\omega_{i_0})u^* < 0$, so there exists a $k_0 > 0$ such that

$$\frac{M_{j_0}(\omega_{i_0})x^k + q_{j_0}(\omega_{i_0})}{\|x^k\|} < \frac{1}{2}M_{j_0}(\omega_{i_0})u^*, \quad \text{for all } k > k_0,$$

that is, we have for all $k > k_0$,

$$M_{j_0}(\omega_{i_0})x^k + q_{j_0}(\omega_{i_0}) < \frac{1}{2}[M_{j_0}(\omega_{i_0})u^*]\|x^k\|,$$

which implies that

$$M_{j_0}(\omega_{i_0})x^k + q_{j_0}(\omega_{i_0}) \to -\infty, \quad k \to +\infty.$$

Then, we have

$$\Psi_2(x^k, y^k) \to +\infty, \quad k \to +\infty. \tag{2.14}$$

This contradicts with the fact $\Psi_2(x^k, y^k) \leq \Psi(x^0)$. So there must be $M(\omega_i)u^* \geq 0$ for all $i = 1, \cdots, m$.

(b) We deduce that $(u^*)^T M(\omega_i)u^* = 0$ for all $i = 1, \cdots, m$. We prove this assertion with absurdity. Suppose that, combined (a), there exist $i_0 \in \{1, \cdots, m\}$ and $j_0 \in \{1, \cdots, n\}$ such that $u^*_{j_0} > 0$ and $M_{j_0}(\omega_{i_0})u^* > 0$. We have

$$M_{j_0}(\omega_i)u^* = \lim_{k\to\infty} \frac{M_{j_0}(\omega_{i_0})x^k}{\|x^k\|} = \lim_{k\to\infty} \frac{M_{j_0}(\omega_{i_0})x^k + q_{j_0}(\omega_{i_0})}{\|x^k\|}.$$

Thus, there exists a $k_1 > 0$ such that

$$\frac{M_{j_0}(\omega_{i_0})x^k + q_{j_0}(\omega_{i_0})}{\|x^k\|} > \frac{1}{2}M_{j_0}(\omega_i)u^*, \quad \text{for all } k \geq k_1,$$

that is,

$$M_{j_0}(\omega_{i_0})x^k + q_{j_0}(\omega_{i_0}) > \frac{1}{2}M_{j_0}(\omega_i)u^* \cdot \|x^k\|, \quad \text{for all } k \geq k_1,$$

which implies that

$$M_{j_0}(\omega_{i_0})x^k + q_{j_0}(\omega_{i_0}) \to +\infty \quad \text{as } k \to +\infty.$$

On the other hand, without loss of generality, assume that $M_j(\omega_i)x^k + q_j(\omega_i)$ are bounded below for all $i = 1, \cdots, m$ and $j = 1, \cdots, n$ (or else $\Psi_2(w^k) \to +\infty$). Notice that

$$\begin{aligned} \bar{M}_{j_0}x^k + \bar{q}_{j_0} =& \sum_{i=1}^{m} p_i[M_{j_0}(\omega_i)x^k + q_{j_0}(\omega_i)] \\ =& \sum_{i\neq i_0} p_i[M_{j_0}(\omega_i)x^k + q_{j_0}(\omega_i)] + p_{i_0}[M_{j_0}(\omega_{i_0})x^k + q_{j_0}(\omega_{i_0})] \\ \geq& C + p_{i_0}[M_{j_0}(\omega_{i_0})x^k + q_{j_0}(\omega_{i_0})], \end{aligned}$$

where C is some constant. Hence, we have

$$[\bar{M}_{j_0}x^k + \bar{q}_{j_0}]_+ \to +\infty, \quad \text{as } k \to +\infty. \tag{2.15}$$

Notice that,

$$\lim_{k\to\infty} \frac{x_{j_0}^k}{\|x^k\|} = \lim_{k\to\infty} u_{j_0}^k = u_{j_0}^* > 0. \tag{2.16}$$

Then, for $k > 0$ sufficiently large, we have

$$x_{j_0}^k > 0.$$

Thus

$$[x_{j_0}^k]_+[\bar{M}_{j_0}x^k + \bar{q}_{j_0}]_+ \to +\infty, \quad \text{as } k \to +\infty. \tag{2.17}$$

Therefore, by Proposition 2.1 (5), we obtain that

$$|\phi_\alpha(x_{j_0}^k, [\bar{M}_{j_0}x^k + \bar{q}_{j_0}])| \to +\infty, \quad \text{as } k \to +\infty,$$

which implies that

$$\Psi_1(x^k) \to +\infty, \quad \text{as } k \to +\infty.$$

This contradicts that $\Psi_1(\mu_k, x^k) \leq \Psi(z^0)$. Thus, it follows from (a) and (b) that for all $i = 1, \cdots, m$,

$$u^* \geq 0, \quad M(\omega_i)u^* \geq 0, \quad (u^*)^T M(\omega_i)u^* = 0.$$

Since $M(\cdot)$ is a stochastic R_0 matrix, we deduce that $u^* = 0$, which is inconsistent with $\|u^*\| = 1$. Hence, the level set $\mathcal{L}_{\Psi(x)}(x^0)$ must be bounded. So far, we complete the whole proof. □

Let $\mathcal{S}_{(x,y)}$ and $\mathcal{S}_x$ be the solution sets of (1.8) and (2.7). Let (2.7) be defined by the function $\phi_\alpha(\cdot,\cdot)$, and (1.8) be defined by the function $\phi_\alpha(\cdot,\cdot)$. Then we easily prove that $\mathcal{S}_{(x,y)}$ and $\mathcal{S}_x$ are empty or nonepmty simultaneously. In addition, $(x^*, y^*) \in \mathcal{S}_{(x,y)}$ with $y^* = [Mx^* + q]_+$ if and only if $x^* \in \mathcal{S}_x$ under the condition that both $\mathcal{S}_{(x,y)}$ and $\mathcal{S}_x$ are nonepmty.

By Theorem 2.1, we have the following corollary immediately.

Corollary 2.1. *The solution set $\mathcal{S}_x$ of (2.7) is nonempty and bounded for all $q(\cdot)$ if $M(\cdot)$ is a stochastic R_0 matrix.*

3. The Feasible Semismooth Gauss-Newton Algorithm

In this section, we will propose a feasible damped Gauss-Newton algorithm for solving (2.4). In order to reduce computation dimensions, we resort to the active set strategy. The active set is defined by $A_k = \{i : x_i^k > 0 \text{ or } \nabla\Psi(x^k)_i \leq 0\}$ for each x^k. Now we state the algorithm as follows.

Algorithm 3.1. (Feasible Damped Gauss-Newton Algorithm)

Step 0. *Choose parameters* $\eta, \rho, \sigma \in (0,1)$, $p \in [1,2]$. *Let* $x^0 \geq 0$. *Set* $k := 0$.

Step 1. *Choose* $V^k \in \partial_C H(x^k)$ *and compute* $\nabla\Psi(x^k) = (V^k)^T H(x^k)$. *Stop if* x^k *is a stationary point of* (2.11), *i.e.,* x^k *satisfies* (2.12). *Otherwise, go to Step* 2.

Step 2. *Let* $A_k = \{i : x_i^k > 0 \text{ or } \nabla\Psi(x^k)_i \leq 0\}$. *Solve the following linear equations to get Gauss-Newton direction* d^k :

$$W^k_{A_k,A_k} d = -\nabla\Psi(x^k)_{A_k}, \tag{3.1}$$

where

$$W^k_{A_k,A_k} = [(V^k)^T V^k]_{A_k,A_k} + \beta_k I.$$

Here, I *is an identity matrix in proper order and*

$$\beta_k = \begin{cases} 0, & \text{if } [(V^k)^T V^k]_{A_k,A_k} \text{ is nonsingular}, \\ \|\nabla\Psi(x^k)_{A_k}\|^p, & \text{otherwise}. \end{cases}$$

Set

$$(d_N^k)_i := \begin{cases} d_i^k, & i \in A_k, \\ 0, & i \notin A_k. \end{cases}$$

Step 3. *Let*

$$d_G^k = -\gamma_k \nabla\Psi(x^k),$$

where

$$\gamma_k = \min\Big\{1, -\frac{\eta\nabla\Psi(x^k)^T d_N^k}{\|\nabla\Psi(x^k)\|^2}\Big\}.$$

Step 4. *Let* m_k *be the smallest nonnegative integer* m *satisfying*

$$\Psi(x^k + \bar{d}^k(\rho^m)) \leq \Psi(x^k) + \sigma\nabla\Psi(x^k)^T \bar{d}_G^k(\rho^m), \tag{3.2}$$

where for any $\lambda \in (0,1]$,

$$\bar{d}^k(\lambda) = t_k^* \bar{d}_G^k(\lambda) + (1 - t_k^*)\bar{d}_N^k(\lambda),$$
$$\bar{d}_N^k(\lambda) = [x^k + \lambda d_N^k]_+ - x^k, \quad \bar{d}_G^k(\lambda) = [x^k + \lambda d_G^k]_+ - x^k,$$

and $t^*(\lambda)$ *is a minimal point of the following problem:*

$$\min_{0\leq t\leq 1} \Psi(x^k) + \nabla\Psi(x^k)^T \bar{d}^k + \frac{1}{2}(\bar{d}^k)^T (V^k)^T V^k \bar{d}^k, \tag{3.3}$$

where $\bar{d}^k = t\bar{d}_N^k(\lambda) + (1-t)\bar{d}_G^k(\lambda)$. *Let* $\lambda_k = \rho^{m_k}$, $z^{k+1} := x^k + \bar{d}^k(\lambda_k)$.

Step 5. *Let* $k := k+1$ *and go to Step* 1.

Remark 3.1. One is easy to find that (3.3) can be written as

$$\min_{t\in[0,1]} \frac{1}{2}at^2 + bt + c, \tag{3.4}$$

where

$$a = [\bar{d}_G^k(\lambda) - \bar{d}_N^k(\lambda)]^T (V^k)^T V^k [\bar{d}_G^k(\lambda) - \bar{d}_N^k(\lambda)], \tag{3.5}$$
$$b = [\nabla\Psi(x^k) + (V^k)^T V^k \bar{d}_N^k(\lambda)]^T [\bar{d}_G^k(\lambda) - \bar{d}_N^k(\lambda)], \tag{3.6}$$
$$c = \frac{1}{2}\|H(x^k) + V^k \bar{d}_N^k\|^2. \tag{3.7}$$

The minimum point of problem (3.4) is given by

$$t^*(\lambda) = \begin{cases} \max\{0, \min\{1, -\frac{b}{a}\}\}, & \text{if } a > 0, \\ 0, & \text{if } a = 0, b \geq 0, \\ 1, & \text{if } a = 0, b \leq 0. \end{cases} \tag{3.8}$$

The following lemma can be seen in [17]:

Lemma 3.1. *The following statements hold:*
(a) *For any $v \in \Re^n_+$, $([u]_+ - u)^T([u]_+ - v) \leq 0$ for all $u \leq \Re^n$.*
(b) *$\|[u]_+ - [v]_+\| \leq \|u - v\|$ for all $u, v \in \Re^n$.*
(c) *Given $x \in \Re^n$ and $d \in \Re^n$, the function g defined by*

$$g(\lambda) = \frac{\|[x + \lambda d]_+ - x\|}{\lambda}, \quad \lambda > 0$$

is nonincreasing.

By Lemma 3.1, we can prove the following result:

Lemma 3.2. *Suppose that the projection-gradient direction $\bar{d}_G^k(\lambda)$ is defined by Algorithm* 3.1. *Then we have*

$$\|\bar{d}_G^k(\lambda)\|^2 \leq \lambda (d_G^k)^T \bar{d}_G^k(\lambda). \tag{3.9}$$

Proof. By Lemma 3.1(a), taking $u := x^k + \lambda d_G^k$, $v := x^k$ we get

$$\begin{aligned} 0 \geq & ([u]_+ - u)^T([u]_+ - v) \\ = & ([x^k + \lambda d_G^k]_+ - [x^k + \lambda d_G^k])^T([x^k + \lambda d_G^k]_+ - x^k) \\ = & \{([x^k + \lambda d_G^k]_+ - x^k) - \lambda d_G^k\}^T([x^k + \lambda d_G^k]_+ - x^k) \\ = & (\bar{d}_G^k - \lambda d_G^k)^T \bar{d}_G^k, \end{aligned}$$

that is, (3.9) holds. □

Lemma 3.3. *Suppose that x^k is not a stationary point of* (2.11). *Then we have* $\|\bar{d}_G^k(1)\| > 0$.

Proof. We prove it by contradiction. Suppose that $\|\bar{d}_G^k(1)\| = 0$. Then

$$0 = [x^k + d_G^k]_+ - x^k = [x^k - \gamma_k \nabla\Psi(x^k)]_+ - x^k,$$

i.e.,

$$x^k = [x^k - \gamma_k \nabla\Psi(x^k)]_+,$$

which implies that x^k is a stationary point of (2.11). This is a contradiction, so we complete the proof. □

Next we show that Algorithm 3.1 is well-defined. We only need to prove that Step 4 is finitely terminated, i.e., $\bar{d}(\lambda)$ is a descent direction for all $\lambda > 0$ sufficiently small.

Theorem 3.1. *Suppose that x^k is not a stationary point of* (2.7)*. Then there exists a constant $\bar{\lambda} \in (0,1]$ such that $\bar{d}^k(\lambda)$ is a descent direction of $\Psi(x^k)$ at point x^k for any $\lambda \in (0,\bar{\lambda}]$ and*

$$\Psi(x^k + \bar{d}(\lambda)) \leq \Psi(x^k) + \sigma \nabla\Psi(x^k)^T \bar{d}_G^k(\lambda). \tag{3.10}$$

Proof. Notice that the smallest eigenvalue ς of the matrix W^k is positive, i.e., $\varsigma = \lambda_{\min}(W^k) > 0$. Then, from Step 3 of Algorithm 3.1 we have

$$\|d_N^k\| = \|[W^k]^{-1}\nabla\Psi(x^k)\| \leq \frac{1}{\varsigma}\|\nabla\Psi(x^k)\|.$$

Since x^k is not a stationary point of problem (2.11), it holds that $\nabla\Psi(x^k) \neq 0$. Thus we have

$$\begin{aligned}\gamma_k =& \min\left\{1, -\frac{\eta\nabla\Psi(x^k)^T d_N^k}{\|\nabla\Psi(x^k)\|^2}\right\}\\ =& \min\left\{1, \eta\frac{\nabla\Psi(x^k)_{A_k}^T (W_{A_k,A_k}^k)^{-1}\nabla\Psi(x^k)_{A_k}}{\|\nabla\Psi(x^k)\|^2}\right\}\\ >& 0.\end{aligned} \tag{3.11}$$

Hence, we have

$$\|d_G^k\| = \|-\gamma_k\nabla\Psi(x^k)\| \leq \|\nabla\Psi(x^k)\|.$$

Let $C_0 = \max\{1, 1/\varsigma\}\|\nabla\Psi(x^k)\|$. Then $\|d_N^k\| \leq C_0$ and $\|d_G^k\| \leq C_0$. By using nonexpansivity of the projection operator, we get

$$\|\bar{d}_N^k(\lambda)\| = \|\max(x^k + \lambda d_N^k, 0) - x^k\| \leq \lambda\|d_N^k\| \leq C_0\lambda,$$

$$\|\bar{d}_G^k(\lambda)\| = \|\max(x^k + \lambda d_G^k, 0) - x^k\| \leq \lambda\|d_G^k\| \leq C_0\lambda.$$

Thus, for all $\lambda \in [0,1]$,

$$\|\bar{d}^k(\lambda)\| = \|t_k^*\bar{d}_G^k + (1-t_k^*)\bar{d}_N^k\| \leq C_0\lambda.$$

So we have $\|\bar{d}_N^k(\lambda)\| = \mathcal{O}(\lambda)$, $\|\bar{d}_G^k(\lambda)\| = \mathcal{O}(\lambda)$ and $\|\bar{d}^k(\lambda)\| = \mathcal{O}(\lambda)$.

Next we show that $\nabla\Psi(x^k)^T\bar{d}^k(\lambda) < 0$. By (3.3) we have

$$\begin{aligned}&\nabla\Psi(x^k)^T\bar{d}^k(\lambda) + \frac{1}{2}\bar{d}^k(\lambda)^T(V^k)^T V^k\bar{d}^k(\lambda)\\ \leq&\nabla\Psi(x^k)^T\bar{d}_G^k(\lambda) + \frac{1}{2}\bar{d}_G^k(\lambda)^T(V^k)^T V^k\bar{d}_G^k(\lambda).\end{aligned}$$

Notice that there exists a constant $C_1 > 0$ such that $\|(V^k)^T V^k\| \leq \|W^k\| \leq C_1$ for all k. Hence, we have

$$\frac{1}{2}\bar{d}^k(\lambda)^T(V^k)^T V^k\bar{d}^k(\lambda) = \mathcal{O}(\lambda^2), \quad \frac{1}{2}\bar{d}_N^k(\lambda)^T(V^k)^T V^k\bar{d}_N^k(\lambda) = \mathcal{O}(\lambda^2).$$

Thus, one has

$$\nabla\Psi(x^k)^T\bar{d}^k(\lambda) \leq \nabla\Psi(x^k)^T\bar{d}_G^k(\lambda) + \mathcal{O}(\lambda^2) = \nabla\Psi(x^k)^T\bar{d}_G^k(\lambda) + o(\lambda).$$

By Lemmas 3.1(c), 3.2, 3.3 and Step 3 of Algorithm 3.1 we have

$$\begin{aligned}\nabla\Psi(x^k)^T\bar{d}_G^k(\lambda) =& -\frac{1}{\gamma_k}(d_G^k)^T\bar{d}_G^k(\lambda) \leq -\frac{1}{\gamma_k\lambda}\|\bar{d}_G^k(\lambda)\|^2\\ &\leq -\frac{\lambda\|\bar{d}_G^k(1)\|}{\gamma_k} \leq -\lambda\|\bar{d}_G^k(1)\| < 0.\end{aligned} \tag{3.12}$$

Therefore, we have $\nabla\Psi(x^k)^T\bar{d}^k(\lambda) < 0$ for $\lambda > 0$ sufficiently small, which implies $\bar{d}^k(\lambda)$ is a decent direction of $\Psi(x^k)$.

Now we come to prove that (3.13) holds. Notice that $\Psi(\cdot)$ is continuously differentiable. Then, by (3.12), we have for $\lambda > 0$ sufficiently small,

$$\begin{aligned}\Psi(x^k+\bar{d}^k(\lambda)) =&\Psi(x^k)+\nabla\Psi(x^k)^T\bar{d}^k(\lambda)+o(\lambda)\\ \le&\Psi(x^k)+\nabla\Psi(x^k)^T\bar{d}_G^k(\lambda)+o(\lambda)\\ =&\Psi(x^k)+\sigma\nabla\Psi(x^k)^T\bar{d}_G^k(\lambda)+(1-\sigma)\nabla\Psi(x^k)^T\bar{d}_G^k(\lambda)+o(\lambda)\\ \le&\Psi(x^k)+\sigma\nabla\Psi(x^k)^T\bar{d}_G^k(\lambda)-(1-\sigma)\lambda\|\bar{d}_G^k(1)\|^2+o(\lambda).\end{aligned} \tag{3.13}$$

Hence, there exists a constant $\bar{\lambda}\in(0,1]$ such that

$$\Psi(x^k+\bar{d}^k(\lambda)) \le \Psi(x^k)+\sigma\nabla\Psi(x^k)^T\bar{d}_G^k(\lambda), \quad \text{for all } \lambda\in(0,\bar{\lambda}],$$

which completes the proof. □

Corollary 3.1. *Algorithm* 3.1 *is well defined. Furthermore, let* $\{x^k\}$ *be a sequence generated by Algorithm* 3.1. *Then* $\{x^k\}\subseteq\Re^n_+$.

Proof. By the construction of Algorithm 3.1, we need only to prove Step 4 is well defined. From Theorem 3.1 we know that the line-search (3.2) is finitely terminated. So Algorithm 3.1 is well defined.

Furthermore, for any $\lambda_k\in(0,1]$ and $x^k\ge 0$, we have

$$\begin{aligned}x^{k+1} =&x^k+\bar{d}^k(\lambda_k)\\ =&x^k+t_k^*\bar{d}_G^k(\lambda)+(1-t_k^*)\bar{d}_N^k(\lambda)\\ =&t^*(\lambda_k)[x^k+\lambda_k d_G^k]_+ +[1-t^*(\lambda_k)][x^k+\lambda_k d_N^k]_+ \quad \in\Re^n_+,\end{aligned}$$

which completes the whole proof. □

4. Convergence Analysis

In this section, we prove the global and locally quadratic convergence of Algorithm 3.1. We first introduce the following lemma which is easily proved and hence, we omit it.

Lemma 4.1. $\hat{x}\ge 0$ *is a stationary point of* (2.11) *if and only if there exists a constant* $\lambda>0$ *such that* $\hat{x}=[\hat{x}-\lambda\nabla\Psi(\hat{x})]_+$.

We next prove the following lemma.

Lemma 4.2. *Suppose that* $\{x^k\}$ *generated by Algorithm* 3.1 *is an infinite sequence and* $\tilde{x}$ *is an accumulation point of it. Then if* $\tilde{x}$ *is not a stationary point of* (2.11), *there must be* $\liminf_{k\to\infty}\gamma_k>0$, *where* γ_k *is defined by Step* 3 *of Algorithm* 3.1.

Proof. By (3.11), let $\liminf\limits_{k\to\infty}\gamma_k=\tilde{\gamma}\ge 0$. Without loss of generality, we can assume that

$$\lim_{k\to\infty}x^k=\tilde{x},\quad \lim_{k\to\infty}\gamma_k=\tilde{\gamma},$$

and

$$\tilde{\beta} = \begin{cases} 0, \text{ if } [(\tilde{V})^T \tilde{V}]_{\tilde{A},\tilde{A}} \text{ is nonsingular,} \\ \|\nabla\Psi(\tilde{x})_{\tilde{A}}\|^p, \quad \text{otherwise,} \end{cases}$$

where $\tilde{V} \subset \partial_C H(\tilde{x})$, $\tilde{A} = \{i : \tilde{x}_i > 0 \text{ or } \nabla\Psi(\tilde{x})_i \le 0\}$. It is easy to find that for sufficiently large k,

$$\gamma_k = -\frac{\eta \nabla\Psi(x^k)^T d_N^k}{\|\nabla\Psi(x^k)\|^2}. \tag{4.1}$$

Notice that if $\tilde{x}$ is not a stationary point of (2.11), then $\nabla\Psi(\tilde{x})_{\tilde{A}} \ne 0$. Hence, by (4.1) we have

$$\begin{aligned}
\tilde{\gamma} = \lim_{k\to\infty} \gamma_k =& - \lim_{k\to\infty} \frac{\eta \nabla\Psi(x^k)^T d_N^k}{\|\nabla\Psi(x^k)\|^2} \\
=&\eta \lim_{k\to\infty} \frac{\nabla\Psi(x^k)_{A_k}^T (W_{A_k,A_k}^k)^{-1} \nabla\Psi(x^k)_{A_k}}{\|\nabla\Psi(x^k)\|^2} \\
=&\eta \frac{\nabla\Psi(\tilde{x})_{\tilde{A}}^T (\tilde{W}_{\tilde{A},\tilde{A}})^{-1} \nabla\Psi(\tilde{x})_{\tilde{A}}}{\|\nabla\Psi(\tilde{z})\|^2} \\
\ge&\eta \lambda_{\min}\big(\tilde{W}_{\tilde{A},\tilde{A}})^{-1}\big) \frac{\|\nabla\Psi(\tilde{x})_{\tilde{A}}\|^2}{\|\nabla\Psi(\tilde{x})\|^2} > 0,
\end{aligned}$$

which completes the proof. □

The next lemma indicates that both $\{\lambda_k\}$ and $\{\bar{d}_G^k(1)\}$ are bounded below if the cluster point of $\{x^k\}$ is not a stationary point of (2.11).

Lemma 4.3. *Suppose that $\{x^k\}$ generated by Algorithm* 3.1 *is an infinite sequence and $\tilde{x}$ is an accumulation point of it. If $\tilde{x}$ is not a stationary point of* (2.11)*, then*

(a) *there exists a constant $\xi > 0$ such that for all $k \ge 0$,*

$$\|\bar{d}_G^k(1)\| = \|[x^k - \gamma_k \nabla\Psi(x^k)]_+ - x^k\| \ge \xi; \tag{4.2}$$

(b) *there exists a constant $\lambda_0 > 0$ such that $\lambda_k \ge \lambda_0$ for all $k \ge 0$.*

Proof. (a) It follows from (3.11) that $\{\gamma_k\}$ is bounded. Notice that $\tilde{x}$ is an accumulation point of $\{x^k\}$. So, without loss of generality, we can assume that $\lim_{k\to\infty} x^k = \tilde{x}$ and $\lim_{k\to\infty} \gamma_k = \tilde{\gamma}_0$. Then, we have

$$\begin{aligned}
\lim_{k\to\infty} \|\bar{d}_G^k(1)\| = \lim_{k\to\infty}& \|[x^k - \gamma_k \nabla\Psi(x^k)]_+ - x^k\| \\
=&\|[\tilde{x} - \tilde{\gamma}_0 \nabla\Psi(\tilde{x})]_+ - \tilde{x}\|.
\end{aligned}$$

By Lemma 4.1, we get $\tilde{\gamma}_0 > 0$. In addition, it follows from Lemma 4.1 that $\|[\tilde{x} - \tilde{\gamma}_0 \nabla\Psi(\tilde{x})]_+ - \tilde{x}\| > 0$. Hence, there exists a positive integer $k_0 > 0$ such that for all $k \ge k_0$,

$$\|\bar{d}_G^k(1)\| \ge \frac{1}{2} \|[\tilde{x} - \tilde{\gamma}_0 \nabla\Psi(\tilde{x})]_+ - \tilde{x}\|.$$

Let

$$\xi = \min\big\{\|\bar{d}_G^0(1)\|, \|\bar{d}_G^1(1)\|, \cdots, \|\bar{d}_G^{k_0}(1)\|, \frac{1}{2}\|[\tilde{x} - \tilde{\gamma}_0 \nabla\Psi(\tilde{x})]_+ - \tilde{x}\|\big\}.$$

We obtain immediately that (4.2) holds.

(b) By (3.13) and (4.2), we have for all $\lambda_k \in (0,1]$ and $k \geq 0$

$$\begin{aligned}\Psi(x^k + \bar{d}^k(\lambda_k)) \leq & \Psi(x^k) + \sigma \nabla \Psi(x^k)^T \bar{d}^k_G(\lambda_k) - (1-\sigma)\lambda_k \|\bar{d}^k_G(1)\|^2 + o(\lambda_k) \\ \leq & \Psi(x^k) + \sigma \nabla \Psi(x^k)^T \bar{d}^k_G(\lambda_k) - (1-\sigma)\xi^2 \lambda_k + o(\lambda_k).\end{aligned}$$

Therefore, there exists $\lambda_1 \in (0,1]$ such that

$$-(1-\sigma)\xi^2\lambda_k + o(\lambda_k) < 0, \quad \text{for all } \lambda_k \in (0, \lambda_1].$$

Thus for all $\lambda_k \in (0, \lambda_1]$ and all $k \geq 0$,

$$\Psi(x^k + \bar{d}^k(\lambda_k)) \leq \Psi(x^k) + \sigma \nabla \Psi(x^k)^T \bar{d}^k_G(\lambda_k).$$

By the updating rule of λ_k, let $\lambda_0 = \rho\lambda_1 > 0$, we have $\lambda_k > \lambda_0$, for all $k \geq 0$. This completes the whole proof. □

Now we come to consider the global convergence of Algorithm 3.1. We have the following theorem.

Theorem 4.1. *Suppose that* $\{x^k\}$ *generated by Algorithm* 3.1 *is an infinite sequence and* x^* *is an accumulation point of* $\{x^k\}$*. Then* x^* *is a stationary point of* (2.11)*. Furthermore, if the matrix* $M(\cdot)$ *is a stochastic* R_0 *matrix, then such an accumulation point must exist.*

Proof. We prove the conclusion of the theorem by contradiction. Suppose that x^* is not a stationary point of (2.11). Since x^* is an accumulation point of $\{x^k\}$, there exists a subsequence $\{x^{k_i}\}$ of $\{x^k\}$ converges to x^*. By Lemma 4.3, there exist constants $\xi > 0$ and $\lambda_0 > 0$ such that $\|\bar{d}^{k_i}_G(1)\| \geq \xi$ and $\lambda_{k_i} \geq \lambda_0$ for all $k_i \geq 0$. It follows from (3.12) that

$$\nabla \Psi(x^{k_i})^T \bar{d}^{k_i}_G(\lambda_k) \leq -\lambda_k \|\bar{d}^{k_i}_G(1)\|^2 \leq -\xi^2\lambda_0.$$

Thus, by the line-search (3.2) we have

$$\Psi(x^{k_i+1}) \leq \Psi(x^{k_i}) - \sigma\xi^2\lambda_0.$$

Then,

$$\Psi(x^{k_i}) \leq \Psi(x^{k_i-1}) \leq \cdots \leq \Psi(x^{k_{i-1}+1}) \leq \Psi(x^{k_{i-1}}) - \sigma\xi^2\lambda_0. \tag{4.3}$$

From (4.3) we deduce that

$$\Psi(x^*) \leq \Psi(x^*) - \sigma\xi^2\lambda_0,$$

which implies that $\xi^2\lambda_0 \leq 0$. This is a contradiction, which proves the first part of the theorem. Furthermore, by Corollary 2.1, we obtain immediately the second conclusion of the theorem. The whole proof is completed. □

From now on, we come to prove locally quadratic convergence of Algorithm 3.1. Assume that x^* is a stationary point of (2.11). Define the function

$$r(x) = x - [x - \nabla\Psi(x)]_+. \tag{4.4}$$

We prove that $r(x^k)$ provides a local error bound for the sequence $\{x^k\}$ generated by Algorithm 3.1.

Lemma 4.4. *Suppose that the infinite sequence* $\{x^k\}$ *generated by Algorithm* 3.1 *converges to* x^*. *If* x^* *is a stationary point of* (2.11) *and every* $U^* \in \partial\nabla\Psi(x^*)$ *is positive definite, then for all* x^k *sufficiently close to* x^*, *we have*

$$\|x^k - x^*\| \le C\|r(x^k)\|, \tag{4.5}$$

where $C > 0$ *is a certain constant.*

Proof. By Lemma 2.1, we have for all x^k sufficiently close to x^*,

$$\nabla\Psi(x^k) - \nabla\Psi(x^*) = U^k(x^k - x^*) + \mathcal{O}(\|x^k - x^*\|^2), \tag{4.6}$$

where $U^k \in \partial\nabla\Psi(x^k)$. Notice that every $U^* \in \partial\nabla\Psi(x^*)$ is positive definite, so for all x^k sufficiently close to x^*, every $U^k \in \partial\nabla\Psi(x^k)$ is uniformly positive definite, i.e., there exists a constant $C_0 > 0$ such that

$$\frac{1}{C_0} \le \lambda_{min}((U^k)^{-1}) \le \lambda_{max}((U^k)^{-1}) \le C_0 \tag{4.7}$$

and

$$\frac{1}{C_0} \le \lambda_{min}(U^k) \le \lambda_{max}(U^k) \le C_0. \tag{4.8}$$

Thus, it follows from (4.6) that

$$\frac{1}{2C_0}\|x^k - x^*\| \le \|\nabla\Psi(x^k) - \nabla\Psi(x^*)\| \le 2C_0\|x^k - x^*\|.$$

Notice that

$$\begin{aligned}\nabla\Psi(x^*)^T[x^k - r(x^k) - x^*] =& \nabla\Psi(x^*)^T[x^k - r(x^k)]\\ =& \nabla\Psi(x^*)^T[x^k - \nabla\Psi(x^k)]_+ \ge 0.\end{aligned}$$

By Lemma 3.1, we have

$$\begin{aligned}&[\nabla\Psi(x^k) - r(x^k)]^T[x^* - x^k + r(x^k)]\\ =&([x^k - \nabla\Psi(x^k)]_+ - [x^k - \nabla\Psi(x^k)])^T(x^* - [x^k - \nabla\Psi(x^k)]_+) \ge 0.\end{aligned}$$

Adding the last two inequalities, we have

$$\nabla\Psi(x^*)^T[x^k - r(x^k) - x^*] + [\nabla\Psi(x^k) - r(x^k)]^T[x^* - x^k + r(x^k)] \ge 0,$$

that is,

$$[\nabla\Psi(x^k) - \nabla\Psi(x^*)]^T(x^* - x^k) + [\nabla\Psi(x^k) - \nabla\Psi(x^*)]^T r(x^k) - (x^* - x^k)^T r(x^k) - \|r(x^k)\|^2 \ge 0.$$

Thus,

$$\begin{aligned}&[\nabla\Psi(x^k) - \nabla\Psi(x^*)]^T(x^k - x^*)\\ \le&[\nabla\Psi(x^k) - \nabla\Psi(x^*)]^T r(x^k) + (x^k - x^*)^T r(x^k) - \|r(x^k)\|^2\\ \le&[\|\nabla\Psi(x^k) - \nabla\Psi(x^*)\| + \|x^k - x^*\|]\|r(x^k)\|\\ \le&(2C_0 + 1)\|x^k - x^*\|\|r(x^k)\|.\end{aligned}$$

Notice that

$$\begin{aligned}&\left[\nabla\Psi(x^k)-\nabla\Psi(x^*)\right]^T(x^k-x^*)\\=&(x^k-x^*)^T\left[U(x^k-x^*)+\mathcal{O}(\|x^k-x^*\|^2)\right]\\\geq&\frac{1}{2C_0}\|x^k-x^*\|^2.\end{aligned}$$

Therefore, we obtain that

$$\frac{1}{2C_0}\|x^k-x^*\|^2\leq(2C_0+1)\|x^k-x^*\|\|r(x^k)\|.$$

Let $C=2C_0(2C_0+1)$, we have immediately

$$\|x^k-x^*\|\leq C\|r(x^k)\|,$$

which completes the proof of the lemma. □

Lemma 4.5. *Assume that $\{x^k\}$ is an infinite sequence generated by Algorithm* 3.1 *and x^* is an accumulation point of it. Suppose that Assumption* 2.1 *is satisfied. If all $U^*\in\partial\nabla\Psi(x^*)$ are positive definite, then for every subsequence $\{x^{k_i}\}$ converging to x^*, we have*

$$\nabla\Psi(x^{k_i})_{A_{k_i}}=U^{k_i}_{A_{k_i},A_{k_i}}(x^{k_i}-x^*)+\mathcal{O}(\|x^{k_i}-x^*\|^2),\tag{4.9}$$

$$(d_N^{k_i})_{A_{k_i}}=-(x^{k_i}-x^*)_{A_{k_i}}+\mathcal{O}(\|x^{k_i}-x^*\|^2),\tag{4.10}$$

$$(\bar{d}_N^{k_i}(1))_{A_{k_i}}=-(x^{k_i}-x^*)_{A_{k_i}}+\mathcal{O}(\|x^{k_i}-x^*\|^2),\tag{4.11}$$

for all x^{k_i} sufficiently close to x^*, where $U^{k_i}\in\partial\nabla\Psi(x^{k_i})$ and $A_{k_i}=\{i:x_i^{k_i}>0$ or $\nabla\Psi(x^{k_i})_i\leq 0\}$.

Proof. By Lemma 2.1, we have

$$\nabla\Psi(x^{k+i})=\nabla\Psi(x^*)+U^{k_i}(x^{k_i}-x^*)+\mathcal{O}(\|x^{k_i}-x^*\|^2).$$

Let

$$I_1=\{i:\nabla\Psi(x^*)_i>0\},\quad I_2=\{i:x_i^*=\nabla\Psi(x^*)_i=0\}$$

and

$$I_3=\{i:x_i^*>0,\text{ and }\nabla\Psi(x^*)_i=0\}.$$

Thus,

$$\begin{aligned}\nabla\Psi(x^{k+i})_{I_2\cup I_3}=&\nabla\Psi(x^*)_{I_2\cup I_3}+U^{k_i}_{I_2\cup I_3}(x^{k_i}-x^*)+\mathcal{O}(\|x^{k_i}-x^*\|^2)\\=&\nabla\Psi(x^*)_{I_2\cup I_2}+U^{k_i}_{I_2\cup I_2,I_2\cup I_3}(x^{k_i}-x^*)_{I_2\cup I_3}+\mathcal{O}(\|x^{k_i}-x^*\|^2),\end{aligned}\tag{4.12}$$

where $U^{k_i}_{I_2\cup I_3}$ is the submatrix of U^{k_i} whose rows are indexed by $I_2\cup I_2$ and columns are the same as U^{k_i}. Notice that if x^{k_i} is sufficiently close to x^*, $\nabla\Psi(x^{k_i})_i>0$ for all $i\in I_1$ and $x_i^{k_i}>0$ for all $i\in I_3$. We recall $A_{k_i}=\{i:x_i^{k_i}>0$ or $\nabla\Psi(x^{k_i})_i\leq 0\}$ and denote

$$\bar{A}_{k_i}:=\{1,\cdots,n\}\backslash A_{k_i}=\left\{i:x_i^{k_i}=0\text{ and }\nabla\Psi(x^{k_i})_i>0\right\}.$$

Then we have $I_1 \subseteq \bar{A}_{k_i}$ and $I_2 \subseteq A_{k_i} \subseteq I_2 \cup I_3$ for all x^{k_i} sufficiently close to x^*. Here, it follows from (4.12) that

$$\begin{aligned}\nabla\Psi(x^{k+i})_{A_{k_i}} =&\nabla\Psi(x^*)_{A_{k_i}} + U^{k_i}_{A_{k_i},I_2\cup I_3}(x^{k_i}-x^*)_{I_2\cup I_3} + \mathcal{O}(\|x^{k_i}-x^*\|^2)\\ =&\nabla\Psi(x^*)_{A_{k_i}} + U^{k_i}_{A_{k_i},A_{k_i}}(x^{k_i}-x^*)_{A_{k_i}} + \mathcal{O}(\|x^{k_i}-x^*\|^2).\end{aligned}$$

Notice that $\nabla\Psi(x^*)_{A_{k_i}} = 0$, we have immediately,

$$\nabla\Psi(x^{k+i})_{A_{k_i}} = U^{k_i}_{A_{k_i},A_{k_i}}(x^{k_i}-x^*)_{A_{k_i}} + \mathcal{O}(\|x^{k_i}-x^*\|^2).$$

By local Lipschitz continuity of $T_i(x)$, we have

$$\begin{aligned}U^{k_i} =&(V^{k_i}_\Phi)^T V^{k_i}_\Phi + (V^{k_i}_G)^T V^{k_i}_G + \sum_{i=1}^{(m+1)n} H_i(x^{k_i})T_i(x^{k_i})\\ =&(V^{k_i}_\Phi)^T V^{k_i}_\Phi + (V^{k_i}_G)^T V^{k_i}_G + \mathcal{O}(\|x^{k_i}-x^*\|)\\ =&W^{k_i} + \mathcal{O}(\|x^{k_i}-x^*\|).\end{aligned}$$

Then, from (3.1) and (4.7)-(4.8) we can obtain that

$$(d^{k_i}_N)_{A_{k_i}} = -(x^{k_i}-x^*)_{A_{k_i}} + \mathcal{O}(\|x^{k_i}-x^*\|^2).$$

Furthermore, it follows from last equality that

$$(x^{k_i})_{A_{k_i}} + (d^{k_i}_N)_{A_{k_i}} = x^*_{A_{k_i}} + \mathcal{O}(\|x^{k_i}-x^*\|^2).$$

Thus,

$$\begin{aligned}(\bar{d}^{k_i}_N(1))_{A_{k_i}} =&[(x^{k_i}+d^{k_i}_N)_{A_{k_i}}]_+ - (x^{k_i})_{A_{k_i}}\\ =&\big([x^*_{A_{k_i}}]_+ - (x^{k_i})_{A_{k_i}}\big) + \big([x^*_{A_{k_i}} + \mathcal{O}(\|x^{k_i}-x^*\|^2)]_+ - [x^*_{A_{k_i}}]_+\big)\\ =&-(x^{k_i}-x^*)_{A_{k_i}} + \mathcal{O}(\|x^{k_i}-x^*\|^2),\end{aligned}$$

which completes the whole proof of the lemma. □

The following lemma shows that under the conditions of Lemma 4.5, the project Gauss-Newton direction is chosen as the search direction of Algorithm 3.1 and the search stepsize is one.

Lemma 4.6. *Suppose that the conditions of Lemma 4.5 hold. Then, we have*
(i) $\bar{d}^{k_i}(1) = -(x^{k_i}-x^*) + \mathcal{O}(\|x^{k_i}-x^*\|^2)$;
(ii) $x^{k_i+1} = x^{k_i} + \bar{d}^{k_i}(1)$.

Proof. (i) Notice that $[\bar{d}^{k_i}_G(1) - \bar{d}^{k_i}_N(1)]_{\bar{A}_{k_i}} = 0$, it follows from (3.5) that

$$\begin{aligned}b =&[\nabla\Psi(x^{k_i}) + W^{k_i}\bar{d}^{k_i}_N(1)]^T[\bar{d}^{k_i}_G(1) - \bar{d}^{k_i}_N(1)]\\ =&[\nabla\Psi(x^{k_i})_{A_{k_i}} + W^{k_i}_{A_{k_i},A_{k_i}}\bar{d}^{k_i}_N(1)_{A_{k_i}}]^T[\bar{d}^{k_i}_G(1) - \bar{d}^{k_i}_N(1)]_{A_{k_i}}.\end{aligned}$$

From Lemma 4.5 we can deduce that

$$\nabla\Psi(x^{k_i})_{A_{k_i}} + W^{k_i}_{A_{k_i},A_{k_i}}\bar{d}^{k_i}_N(1)_{A_{k_i}} = \mathcal{O}(\|x^{k_i}-x^*\|^2)$$

and

$$\begin{aligned}
&\|[\bar{d}_G^{k_i}(1)-\bar{d}_N^{k_i}(1)]_{A_{k_i}}\| \\
\leq&\|[d_G^{k_i}-d_N^{k_i}]_{A_{k_i}}\| \\
=&\|-\gamma_{k_i}\nabla\Psi(x^{k_i})_{A_{k_i}}-[-(x^{k_i}-x^*)_{A_{k_i}}+\mathcal{O}(\|x^{k_i}-x^*\|^2)]\| \\
\leq&(C_0+1)\|(x^{k_i}-x^*)_{A_{k_i}}\|+\mathcal{O}(\|x^{k_i}-x^*\|^2).
\end{aligned}$$

Hence, we have $b=\mathcal{O}(\|x^{k_i}-x^*\|^3)$. On the other hand, since

$$0<\gamma_{k_i}\leq-\frac{\eta\nabla\Psi(x^{k_i})_{A_{k_i}}^T(d_N^{k_i})_{A_{k_i}}}{\|\nabla\Psi(x^{k_i})\|^2}\leq-\frac{\eta\nabla\Psi(x^{k_i})_{A_{k_i}}^T(d_N^{k_i})_{A_{k_i}}}{\|\nabla\Psi(x^{k_i})_{A_{k_i}}\|^2},$$

we can deduce that

$$\begin{aligned}
[\bar{d}_G^{k_i}(1)-d_N^{k_i}]_{A_{k_i}}^T\nabla\Psi(x^{k_i})_{A_{k_i}}=&[\bar{d}_G^{k_i}(1)]^T\nabla\Psi(x^{k_i})_{A_{k_i}}-[d_N^{k_i}]_{A_{k_i}}^T\nabla\Psi(x^{k_i})_{A_{k_i}} \\
\geq&-\gamma_{k_i}\|\nabla\Psi(x^{k_i})_{A_{k_i}}\|^2-[d_N^{k_i}]_{A_{k_i}}^T\nabla\Psi(x^{k_i})_{A_{k_i}} \\
\geq&(\eta-1)[d_N^{k_i}]_{A_{k_i}}^T\nabla\Psi(x^{k_i})_{A_{k_i}} \\
=&(1-\eta)\nabla\Psi(x^{k_i})_{A_{k_i}}^T(W_{A_{k_i},A_{k_i}}^{k_i})^{-1}\nabla\Psi(x^{k_i})_{A_{k_i}} \\
\geq&\frac{1-\eta}{C_0}\|\nabla\Psi(x^{k_i})_{A_{k_i}}\|^2. \qquad (4.13)
\end{aligned}$$

Hence, by Schwartz inequality, we have

$$\begin{aligned}
\bar{d}_G^{k_i}(1)-d_N^{k_i}\geq&\frac{1-\eta}{C_0}\|\nabla\Psi(x^{k_i})_{A_{k_i}}\| \\
=&C_1\|x^{k_i}-x^*\|+\mathcal{O}(\|x^{k_i}-x^*\|^2),
\end{aligned}$$

where $C_0=(1-\eta)/C_0>0$. Notice that $[\bar{d}_N^{k_i}(1)-d_N^{k_i}]_{A_{k_i}}=\mathcal{O}(\|x^{k_i}-x^*\|^2)$. Thus,

$$\begin{aligned}
&\|[\bar{d}_G^{k_i}(1)-\bar{d}_N^{k_i}(1)]_{A_{k_i}}\|^2 \\
=&\|[\bar{d}_G^{k_i}(1)-d_N^{k_i}]_{A_{k_i}}\|^2-2[\bar{d}_G^{k_i}(1)-d_N^{k_i}]_{A_{k_i}}^T[\bar{d}_N^{k_i}(1)-d_N^{k_i}]_{A_{k_i}}+\|[\bar{d}_N^{k_i}(1)-d_N^{k_i}]_{A_{k_i}}\|^2 \\
\geq&C_1^2\|x^{k_i}-x^*\|^2+o(\|x^{k_i}-x^*\|^2). \qquad (4.14)
\end{aligned}$$

From (3.4) and (4.14), we get

$$\begin{aligned}
a=&[\bar{d}_G^{k_i}(1)-\bar{d}_N^{k_i}(1)]_{A_{k_i}}^TW_{A_{k_i},A_{k_i}}^{k_i}[\bar{d}_G^{k_i}(1)-\bar{d}_N^{k_i}(1)]_{A_{k_i}} \\
\geq&\frac{1}{2C_0}\|[\bar{d}_G^{k_i}(1)-\bar{d}_N^{k_i}(1)]_{A_{k_i}}\|^2 \\
\geq&\frac{C_1^2}{2C_0}\|x^{k_i}-x^*\|^2+o(\|x^{k_i}-x^*\|^2),
\end{aligned}$$

which implies,

$$-\frac{b}{a}=\mathcal{O}(\|x^{k_i}-x^*\|).$$

Hence, we have

$$t_{k_i}^*(1)=\mathcal{O}(\|x^{k_i}-x^*\|).$$

Notice that $\bar{d}^{k_i}(1) = t^*_{k_i}(1)\bar{d}^{k_i}_G(1) + [1 - t^*_{k_i}(1)]\bar{d}^{k_i}_N(1)$. Then by Lemma 4.5 we obtain that

$$\begin{aligned}\bar{d}^{k_i}(1) =& \bar{d}^{k_i}_N(1) + t^*_{k_i}(1)[\bar{d}^{k_i}_G(1) - \bar{d}^{k_i}_N(1)] \\ =& -(x^{k_i} - x^*) + \mathcal{O}(\|x^{k_i} - x^*\|^2),\end{aligned} \tag{4.15}$$

which proves (i).

(ii) Since $\bar{d}^{k_i}(1) = t^*_{k_i}(1)\bar{d}^{k_i}_G(1) + (1 - t^*_{k_i}(1))\bar{d}^{k_i}_N(1)$, we have

$$\begin{aligned}&\nabla\Psi(x^{k_i})^T[\bar{d}^{k_i}(1) - \bar{d}^{k_i}_G(1)] \\ =&[t^*_{k_i}(1) - 1]\nabla\Psi(x^{k_i})^T_{A_{k_i}}[\bar{d}^{k_i}_G(1) - \bar{d}^{k_i}_N(1)]_{A_{k_i}} \\ =&[t^*_{k_i}(1) - 1]\nabla\Psi(x^{k_i})^T_{A_{k_i}}[(\bar{d}^{k_i}_G(1) - d^{k_i}_N(1))_{A_{k_i}} + \mathcal{O}(\|x^{k_i} - x^*\|^2)] \\ =&[t^*_{k_i}(1) - 1]\nabla\Psi(x^{k_i})^T_{A_{k_i}}[\bar{d}^{k_i}_G(1) - d^{k_i}_N(1)]_{A_{k_i}} + o(\|x^{k_i} - x^*\|^2),\end{aligned} \tag{4.16}$$

By (4.9) we have

$$\|\nabla\Psi(x^{k_i})_{A_{k_i}}\|^2 \geq \frac{1}{C_0^2}\|x^{k_i} - x^*\|^2 + o(\|x^{k_i} - x^*\|^2). \tag{4.17}$$

From (4.13) and (4.17) we get

$$\begin{aligned}\nabla\Psi(x^{k_i})^T_{A_{k_i}}[\bar{d}^{k_i}_G(1) - d^{k_i}_N]_{A_{k_i}} \geq& \frac{1-\eta}{C_0}\|\nabla\Psi(x^{k_i})_{A_{k_i}}\|^2 \\ \geq& \frac{1-\eta}{C_0^3}\|x^{k_i} - x^*\|^2 + o(\|x^{k_i} - x^*\|^2).\end{aligned} \tag{4.18}$$

Using (4.16), (4.18) and $t^*_{k_i}(1) = \mathcal{O}(\|x^{k_i} - x^*\|)$, it is not difficult to see that

$$\nabla\Psi(x^{k_i})^T[\bar{d}^{k_i}(1) - \bar{d}^{k_i}_G(1)] \leq [t^*_{k_i}(1) - 1]\frac{1-\eta}{C_0^3}\|x^{k_i} - x^*\|^2 + o(\|x^{k_i} - x^*\|^2). \tag{4.19}$$

Further, we have

$$\begin{aligned}&-\nabla\Psi(x^{k_i})^T\bar{d}^{k_i}(1) = -\nabla\Psi(x^{k_i})^T[t^*_{k_i}(1)\bar{d}^{k_i}_G(1) + (1 - t^*_{k_i}(1))\bar{d}^{k_i}_N(1)] \\ =&-t^*_{k_i}(1)\nabla\Psi(x^{k_i})^T_{A_{k_i}}[\bar{d}^{k_i}_G(1)]_{A_{k_i}} + [t^*_{k_i}(1) - 1]\nabla\Psi(x^{k_i})^T_{A_{k_i}}[(d^{k_i}_N)_{A_{k_i}} + \mathcal{O}(x^{k_i} - x^*\|^2)] \\ =&[1 - t^*_{k_i}(1)]\nabla\Psi(x^{k_i})^T_{A_{k_i}}(W^{k_i}_{A_{k_i},A_{k_i}})^{-1}\nabla\Psi(x^{k_i})^T_{A_{k_i}} + o(\|x^{k_i} - x^*\|^2) \\ \geq&\frac{1 - t^*_{k_i}(1)}{C_0^3}\|x^{k_i} - x^*\|^2 + o(\|x^{k_i} - x^*\|^2).\end{aligned} \tag{4.20}$$

By (4.15) we know that

$$x^{k_i} + \bar{d}^{k_i}(1) - x^* = \mathcal{O}(\|x^{k_i} - x^*\|^2). \tag{4.21}$$

It follows from (4.19), (4.20) and (4.21) that

$$\begin{aligned}
&\Psi(x^{k_i}+\bar{d}^{k_i}(1))-\Psi(x^{k_i})\\
=&\nabla\Psi(x^{k_i})^T\bar{d}^{k_i}(1)+\mathcal{O}(\|x^{k_i}-x^*\|^2)\\
=&\sigma\nabla\Psi(x^{k_i})^T\bar{d}_G^{k_i}(1)+\sigma\nabla\Psi(x^{k_i})^T[\bar{d}^{k_i}(1)-\bar{d}_G^{k_i}(1)]\\
&\quad+\ (1-\sigma)\nabla\Psi(x^{k_i})^T\bar{d}^{k_i}(1)+\mathcal{O}(\|x^{k_i}-x^*\|^2)\\
\le&\sigma\nabla\Psi(x^{k_i})^T\bar{d}_G^{k_i}(1)+\sigma[t_{k_i}^*(1)-1]\frac{1-\eta}{C_0^3}\|x^{k_i}-x^*\|^2\\
&\quad-(1-\sigma)\frac{1-t_{k_i}^*(1)}{C_0^3}\|x^{k_i}-x^*\|^2+\mathcal{O}(\|x^{k_i}-x^*\|^2)\\
=&\sigma\nabla\Psi(x^{k_i})^T\bar{d}_G^{k_i}(1)-\frac{1-t_{k_i}^*(1)}{C_0^3}(1-\sigma\eta)\|x^{k_i}-x^*\|^2+\mathcal{O}(\|x^{k_i}-x^*\|^2)\\
\le&\sigma\nabla\Psi(x^{k_i})^T\bar{d}_G^{k_i}(1).
\end{aligned}$$

Therefore, we can deduce that

$$\Psi(x^{k_i}+\bar{d}^{k_i}(1))\le\Psi(x^{k_i})+\sigma\nabla\Psi(x^{k_i})^T\bar{d}_G^{k_i}(1)$$

for all x^{k_i} sufficiently close to x^*, which implies that

$$x^{k_i+1}=x^{k_i}+\bar{d}^{k_i}(1).$$

So far, we have proved (ii). □

Now, we come to prove that the sequence $\{x^k\}$ generated by Algorithm 3.1 is quadratically convergent under suitable conditions. We have the following theorem.

Theorem 4.2. *Suppose that the conditions of Lemma* 4.5 *hold. Then, the sequence* $\{x^k\}$ *generated by Algorithm* 3.1 *converges Q-quadratically to* x^*.

Proof. Let $\{x^{k_i}\}$ be any subsequence converging to x^*. Then, by Lemma 4.6, we have

$$\|x^{k_i+1}-x^*\|=\|x^{k_i}+\bar{d}^{k_i}(1)-x^*\|=\mathcal{O}(\|x^{k_i}-x^*\|^2)$$

for all x^{k_i} sufficiently close to x^*. This implies that

$$\|x^{k_i+1}-x^{k_i}\|\le\|x^{k_i+1}-x^*\|+\|x^{k_i}-x^*\|\to 0\quad\text{ as } k_i\to+\infty.$$

Since x^* is a stationary point of (2.11) if and only if $x^*=[x^*-\lambda\nabla\Psi(x^*)]_+$ for any $\lambda>0$, then by Lemma 4.4 we know that x^* is an isolated stationary point of (2.11). Thus, it follows from Proposition 8.3.10 of [12] that $\{x^k\}$ converges to x^*. Combined with Lemmas 4.5 and 4.6, we obtain

$$\|x^{k+1}-x^*\|=\|x^k+\bar{d}^k(1)-x^*\|=\mathcal{O}(\|x^k-x^*\|^2)$$

for all x^k sufficiently close to x^*, which completes the proof. □

5. Numerical Experiments

In this section, we will report some numerical results. The test problems are randomly generated. The procedure to generate a test problem of monotone SLCP (1.4) is from [7, 16]. We recall the procedure as follows. Suppose that

$$p_j = \mathcal{P}\{\omega_j \in \Omega\} = \frac{1}{m}, \quad j = 1, \cdots, m.$$

Procedure 5.1.

Step 1. *Generate a diagonal matrix D whose elements are determined as*

$$D_{jj} = \begin{cases} 1/\nu, & j = 1, \\ \nu^{\lambda_j}, & j = 2, \cdots, n-1, \\ \nu, & j = n, \end{cases}$$

where $\nu > 0$ and λ_j, $j = 2, \cdots, n-1$, are uniform variate in the interval $(-1, 1)$.

Step 2. *Generate a random matrix S. By using the singular decomposition of the random matrix S, we obtain a random orthogonal matrix $U \in \Re^{n\times n}$. Let $\bar{M} = UDU^T$.*

Step 3. *Generate m random matrices $B^i \in \Re^{n\times n}$, $i = 1, \cdots, m$, whose elements are in the interval $(0, 1)$. Let*

$$M^i = M(\omega_i) = \bar{M} + c_2(B^i - B^{m-i+1}), \quad i = 1, \cdots, m,$$

where $c_2 > 0$.

Step 4. *Generate a random vector $\bar{x} \in \Re^n$ such that $n_x(< n)$ elements are in the interval $(0, c_1)$, $c_1 > 0$, and all the other elements are zero. Let $\mathcal{J} = \{i : \bar{x}_i > 0\}$.*

Step 5. *For each $i = 1, \cdots, m$, let n_I be the number of elements in the index set $\mathcal{I}_i = \{j : \bar{x}_j = 0,\ [M(\omega_i)\bar{x} + q(\omega_i)]_j > 0\}$ and let n_K be the number of elements in the index set $\mathcal{K}_i = \{j : \bar{x}_j = 0,\ [M(\omega_i)\bar{x} + q(\omega_i)]_j = 0\}$. Let*

$$q^i_j = [q(\omega_i)]_j = \begin{cases} (-M^i\bar{x})_j, & j \in \mathcal{K}_i, \\ (-M^i\bar{x} + c_3u^i)_j, & j \in \mathcal{J}, \\ (-M^i\bar{x} + c_4u^i)_j, & j \in \mathcal{I}_i, \end{cases}$$

where $c_3, c_4 \geq 0$ and $u^i \in \Re^n$ is a random vector whose elements in the interval $(0, 1)$.

We can easily see that for the test problem generated by Procedure 5.1, if $c_3 = 0$ then $\bar{x}$ is a solution of (1.4) and also a global solution of (2.11) with $\Psi(\bar{x}) = 0$. If $c_3 > 0$, $\bar{x}$ is not necessarily a solution of (1.4) and the test problem may have no solutions.

In general, problem (1.4) has no solution. A measure of optimality and feasibility for (1.4) has been proposed in [7] and employed by [16]. We recall the definitions as follows. Let

$$\Gamma(x) := Op(x) + Fe(x), \tag{5.1}$$

where

$$Op(x) = \sum_{i=1}^{n} x^T[M(\omega_i)x + q(\omega_i)]_+, \quad x \in \Re^n_+, \tag{5.2}$$

Table 5.1: Comparison of Algorithm 3.1 and the AFSN for some SLCPs with $c_3 = 0$

Problem	Algor. 3.1				AFSN			
(n, n_x, c_2, c_3, le)	$Fe(x^1)$	$Op(x^1)$	Iter	CPU	$Fe(x^2)$	$Op(x^2)$	Iter	CPU
$(30, 10, 20, 0, e)$	1.11E-12	1.27E-11	4.0	0.0170	2.27E-08	2.98E-07	10.0	0.0252
$(30, 10, 20, 0, 10e)$	4.86E-12	5.53E-11	4.0	0.0172	2.90E-09	3.68E-08	13.0	0.0293
$(30, 10, 20, 0, 20e)$	8.13E-12	1.05E-10	4.0	0.0212	9.35E-09	1.34E-07	13.0	0.0306
$(30, 10, 20, 0, 30e)$	8.41E-12	9.07E-11	4.0	0.0174	4.45E-09	5.68E-08	14.0	0.0322
$(30, 10, 20, 0, 40e)$	2.13E-12	2.51E-11	4.0	0.0137	1.78E-09	2.42E-08	14.0	0.0368
$(30, 10, 20, 0, 50e)$	3.51E-12	4.01E-11	4.0	0.0197	8.66E-09	1.10E-07	14.0	0.0340
$(90, 30, 20, 0, e)$	1.67E-12	3.36E-10	4.0	0.2347	2.93E-09	7.15E-08	11.0	0.2879
$(90, 30, 20, 0, 10e)$	9.92E-13	2.23E-11	4.0	0.2034	9.94E-09	2.11E-07	13.0	0.2989
$(90, 30, 20, 0, 20e)$	8.17E-13	1.75E-11	4.0	0.2038	1.08E-08	2.34E-07	13.5	0.3100
$(90, 30, 20, 0, 30e)$	1.56E-12	3.78E-11	4.0	0.2258	8.77E-10	2.12E-08	14.0	0.3407
$(90, 30, 20, 0, 40e)$	1.24E-12	2.69E-11	4.0	0.2013	9.06E-09	2.02E-07	14.0	0.3394
$(90, 30, 20, 0, 50e)$	1.49E-12	3.21E-11	4.0	0.2077	1.31E-08	3.07E-07	14.0	0.3621
$(150, 50, 15, 0, e)$	1.56E-12	4.41E-11	4.0	0.6977	3.46E-08	1.04E-06	11.0	0.8211
$(150, 50, 15, 0, 10e)$	1.41E-12	3.94E-11	4.0	0.9187	1.37E-11	4.11E-10	13.0	1.0693
$(150, 50, 15, 0, 20e)$	2.08E-12	6.27E-11	4.0	0.7406	2.89E-08	8.82E-07	13.0	0.9952
$(150, 50, 15, 0, 30e)$	2.25E-12	6.48E-11	4.0	0.7299	2.96E-09	9.62E-08	14.0	1.0280
$(150, 50, 15, 0, 40e)$	1.33E-12	3.89E-11	4.0	0.7647	6.12E-09	1.88E-07	14.0	1.1009
$(150, 50, 15, 0, 50e)$	1.21E-12	3.16E-11	4.0	0.7129	9.61E-09	3.09E-07	14.0	1.0239

Table 5.2: Comparison of Algorithm 3.1 and the AFSN for some SLCPs with $c_3 > 0$

Problem	Algor. 3.1				AFSN			
(n, n_x, c_2, c_3, le)	$Fe(x^1)$	$Op(x^1)$	Iter	CPU	$Fe(x^2)$	$Op(x^2)$	Iter	CPU
$(30, 10, 20, 10, e)$	1.22E-02	5.45E+02	8.0	0.0619	2.42E+02	2.54E+03	100	0.4978
$(30, 10, 20, 10, 10e)$	1.01E-02	4.92E+02	8.0	0.0556	2.32E+02	2.20E+03	100	0.5303
$(30, 10, 20, 10, 20e)$	1.25E-02	5.28E+02	8.0	0.0342	2.06E+02	2.22E+03	100	0.5292
$(30, 10, 20, 10, 30e)$	8.90E-03	5.12E+02	8.0	0.0365	2.28E+02	2.25E+03	100	0.5109
$(30, 10, 20, 10, 40e)$	1.19E-02	5.57E+02	8.0	0.0413	2.41E+02	2.69E+03	100	0.5171
$(30, 10, 20, 10, 50e)$	1.21E-02	5.25E+02	8.0	0.0459	2.13E+02	2.20E+03	100	0.5753
$(90, 30, 20, 10, e)$	1.24E-02	1.62E+03	8.0	0.5206	6.71E+02	1.17E+04	100	4.4764
$(90, 30, 20, 10, 10e)$	1.06E-02	1.50E+03	8.5	0.4133	6.96E+02	1.10E+04	100	4.6689
$(90, 30, 20, 10, 20e)$	1.07E-02	1.57E+03	9.0	0.4099	7.17E+02	1.19E+04	100	4.6406
$(90, 30, 20, 10, 30e)$	1.15E-02	1.57E+03	8.0	0.3093	6.96E+02	1.14E+04	100	4.5007
$(90, 30, 20, 10, 40e)$	1.14E-02	1.60E+03	8.0	0.3633	6.93E+02	1.20E+04	100	4.7189
$(90, 30, 20, 10, 50e)$	1.20E-02	1.52E+03	9.0	0.4110	6.02E+02	1.00E+04	100	4.5162
$(150, 50, 20, 10, e)$	1.07E-02	2.67E+03	8.5	1.2304	1.19E+03	2.55E+04	100	12.489
$(150, 50, 20, 10, 10e)$	1.14E-02	2.57E+03	9.5	1.5679	1.21E+03	2.50E+04	100	12.379
$(150, 50, 20, 10, 20e)$	1.16E-02	2.57E+03	9.0	1.7362	1.12E+03	2.33E+04	100	15.638
$(150, 50, 20, 10, 30e)$	1.04E-02	2.62E+03	8.0	1.3889	1.21E+03	2.54E+04	100	12.921
$(150, 50, 20, 10, 40e)$	1.09E-02	2.59E+03	9.0	1.5000	1.19E+03	2.50E+04	100	13.349
$(150, 50, 20, 10, 50e)$	1.16E-02	2.54E+03	9.0	1.7380	1.16E+03	2.15E+04	100	13.326

and

$$Fe(x) = \sum_{i=1}^{n} \| \min\{0, M(\omega_i)x + q(\omega_i)\}\|, \qquad x \in \Re^n_+. \tag{5.3}$$

Here, the function $Op(x)$ is a measure of optimality and $Fe(x)$ is a measure of feasibility.

We implemented Algorithm 3.1 by our programm codes in Matlab 7.9. All these problems were done at a PC with Intel(R) Core(TM)2 Duo CPU P7350 @ 2.0GHz and RAM of 2GB.

Parameters c_1, c_2, c_3, c_4 and μ are needed to generate the problem. A vector $\bar{x} \in \Re^n_+$ is randomly generated. The generated expected matrix $\bar{M}$ is positive definite.

We make a comparison between the Algorithm 3.1 and the AFSN method in [16] on some problems with $c_3 = 0$ and $c_3 > 0$. For each problem, the starting points are taken as $x^0 = le$, where $l = 1,\ 10,\ 20,\ 30,\ 40,\ 50$, $e = (1, 1, \cdots, 1)$. The problems are generated with

$$c_1 = 20, \quad c_4 = 15, \quad \nu = 10.$$

We set $m = 100$ and $n = 30, 60$ and 150. The other parameters used in Algorithm 3.1 are as follows:

$$\eta = 0.9, \quad \rho = 0.5, \quad \sigma = 10^{-2}, \quad \alpha = 10^{-10}.$$

The iteration is terminated if

$$\max_{i \le i \le n} \left\{ |x_i \Psi(x^k)_i| \right\} < 10^{-6} \ \text{ and } \ \max_{i \le i \le n} \left\{ |([\nabla \Psi(x^k)]_-)_i| \right\} < 10^{-6},$$

or

$$k_{max} = 100.$$

For each starting point, we generate ten problems and present the average values of the function $Fe(x)$ and $Op(x)$. In addition, the CPU time and the number of iterations are also the average values of these ten problems.

In Tables 1 and 2, x^1 and x^2 are the computed solutions of Algorithm 3.1 and the AFSN, respectively. Iter denotes the number of iterations. CPU records the CPU time in second for solving each problem.

The results reported in Tables 1 and 2 indicate that Algorithm 3.1 may yield a solution in a smaller number of iterations and less CPU time than the AFSN. In addition, we can also see that most of $Fe(x^1)$ and $Op(x^1)$ are smaller than $Fe(x^2)$ and $Op(x^2)$, respectively. Notice that the fact that $H(x, y) = 0$ has a solution is necessary for quadratic convergence of the AFSN [16]. Whereas it follows from Theorem 4.2 that the Q-quadratic convergence of Algorithm 3.1 does not depend on that Eq. (2.4) has a solution, from which we can conclude that the conditions for the locally Q-quadratic convergence of Algorithm 3.1 is weaker than that of the AFSN in [16].

Notice that problem (1.4) has no solution for $c_3 > 0$. Then, the Q-quadratic convergence of the AFSN does not hold when $c_3 > 0$. In fact, we can find from Table 2 that Algorithm 3.1 can yield a solution in no more than 10 iterations, whereas AFSN always stops at maximum number of iterations $k_{max} = 100$. Comparing with AFSN, we can conclude that Algorithm 3.1 yields a reasonable solution with higher safety and within smaller number of iterations based on these numerical results in Tables 1 and 2.

Acknowledgments. This project is supported by National Natural Science Foundation of China (11071041) and Fujian Natural Science Foundation (2009J01002).

References

[1] X. Chen, M. Fukushima, Expected residual minimization method for stochastic linear complementarity problems, *Math. Oper. Res.*, **30** (2005), 1022-1038.

[2] Ling Chen, Liqun Qi, Guanglu Zhou, Louis Caccettac The SC^1 property of an expected residual function arising from stochastic complementarity problems, *Oper. Res. Lett.*, **36** (2008), 456-460.

[3] Steven A Gabriel, Jifang Zhuang, Ruud Egging, Solving stochastic complementarity problems in energy market modeling using scenario reduction, *Eur. J Oper. Res.*, **197** (2009), 1028-1040.

[4] M. Wang, M.M. Ali, Stochastic Nonlinear Complementarity Problems: Stochastic Programming Reformulation and Penalty Based Approximation Method, *J Optimiz. Theory App.*, **144** (2010), 597-614.

[5] M.J. Luo, G.H. Lin, Convergence Results of the ERM Method for Nonlinear Stochastic Variational Inequality Problems, *J Optimiz. Theory App.*, **142** (2009), 569-581.

[6] M.J. Luo, G.H. Lin, Expected Residual Minimization Method for Stochastic Variational Inequality Problems, *J Optimiz Theory App.*, **140** (2009), 103-116.

[7] Xiaojun Chen, Chao Zhang, Masao Fukushima, Robust solution of monotone stochastic linear complementarity problems, *Math. Program., Ser. B*, **117** (2009), 51-80.

[8] C. Zhang and X. Chen, Stochastic nonlinear complementarity problem and applications to traffic equilibrium under uncertainty, *J Optimiz. Theory App.*, **137** (2008), 277-295.

[9] H. Fang, X. Chen, M. Fukushima, Stochastic R_0 matrix linear complementarity problem, *SIAM, J. Optimiz.*, **18** (2007), 482-506.

[10] H. Liu, Y. Huang, X. Li, New reformulation and feasible semismooth Newton method for a class of stochastic linear complementarity problems, *Appl. Math. Comput.*, **217** (2011), 9723-9740

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

[12] F. Facchinei, J.-S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems, I and II, Springer-Verlag, New York, 2003.

[13] M.C. Ferris, J.-S. Pang, Engineering and economic applications of complementarity problems, *SIAM Rev.*, **39** (1997), 669-713.

[14] R.W. Cottle, J.-S. Pang, R.E Stone, *The Linear Complementarity Problem*, Academic Press, New York, 1992.

[15] G. Gürkan, A Y Özge, S M Robinson, Sample-path solution of stochastic variational inequalities, *Math. Program.*, **84** (1999), 313-333.

[16] G.L. Zhou, L. Caccetta, Feasible Semismooth Newton Method for a Class of Stochastic Linear Complementarity Problems, *J Optimiz. Theory App.*, **139** (2008), 379-392.

[17] D.-F. Sun, R.S. Womersly, H.-D Qi, A feaible semismooth asymptically Newton method for mixed complementarity probloms, *Math. Program.*, **94** (2002), 167-187.

[18] B. Chen and N. Xiu, A global linear and local quadratic non-interior continuation method for nonlinear complementarity problems based on Chen-Mangasarian smoothing functions, *SIAM J Optimiz.*, **9** (1999), 605-623.

[19] X. Chen, L. Qi, and D. Sun, Global and superlinear convergence of the smoothing Newton method and its application to general box-constrained variational inequalities, *Math. Comput*, **67** (1998), 519-540.

[20] B. Chen and P.T. Harker, Smoothing approximations to nonlinear complementarity problems, *SIAM J Optimiz.*, **7** (1997), 403- 420.

[21] H. Qi, A regularized smoothing Newton method for box constrained variational inequality problems with P0-functions, *SIAM J Optimiz.*, **10** (200), 315-330.

[22] L. Qi and D. Sun, Improving the convergence of non-interior point algorithm for nonlinear complementarity problems, *Math. Comput.*, **69** (2000), 283-304.

[23] L. Zhang, J. Han, J. and Z. Huang, Superlinear/Quadratic one-step smoothing Newton method for P_0-NCP, *Acta Math. Sin.*, **26** (2005), 117-128.

[24] C. Kanzow, Global convergence properities of some interative methods for linear complementarity problems, *SIAM J Optimiz.*, **6** (1996), 326-341.

[25] L. Qi and J. Sun, A nonsmooth version of Newton's method, *Math. Program.*, **58** (1993), 353-367.

[26] L. Qi, Convegence analysis of some algorithms for solving nonsmooth equations, *Math. Oper. Res.*, **18** (1993), 227-244.

[27] J.S. Pang and L. Qi, Nonsmooth equations: Motivation and algorithms, *SIAM J Optimiz.*, **3** (1993), 443-465.

[28] A. Fischer, Solution of monotone complementarity problems with locally Lipschitzian functions, *Math. Program.*, **76** (1997), 513-532.