STOCHASTIC R_0 MATRIX LINEAR COMPLEMENTARITY PROBLEMS*

HAITAO FANG†, XIAOJUN CHEN‡, AND MASAO FUKUSHIMA§

Abstract. We consider the expected residual minimization formulation of the stochastic R_0 matrix linear complementarity problem. We show that the involved matrix being a stochastic R_0 matrix is a necessary and sufficient condition for the solution set of the expected residual minimization problem to be nonempty and bounded. Moreover, local and global error bounds are given for the stochastic R_0 matrix linear complementarity problem. A stochastic approximation method with acceleration by averaging is applied to solve the expected residual minimization problem. Numerical examples and applications of traffic equilibrium and system control are given.

Key words. stochastic linear complementarity problem, R_0 matrix, expected residual minimization

AMS subject classifications. 90C33, 90C15

1. Introduction. Let $(\Omega, \mathcal{F}, P)$ be a probability space, where Ω is a subset of $\mathbb{R}^m$, and $\mathcal{F}$ is a σ-algebra generated by $\{\Omega \cap U : U \text{ is an open set in } \mathbb{R}^m\}$. We consider the stochastic linear complementarity problem (SLCP):

$$x \geq 0,\ M(\omega)x + q(\omega) \geq 0,\ \ x^T(M(\omega)x + q(\omega)) = 0,$$

where $M(\omega) \in \mathbb{R}^{n\times n}$ and $q(\omega) \in \mathbb{R}^n$ for $\omega \in \Omega$. We denote this problem by SLCP$(M(\omega), q(\omega))$ for short. Throughout this paper, we assume that $M(\omega)$ and $q(\omega)$ are measurable functions of ω with the following property:

$$E\{\|M(\omega)^T M(\omega)\|\} < \infty \quad \text{and} \quad E\{\|q(\omega)\|^2\} < \infty,$$

where E stands for the expectation. If Ω only contains a single realization, then the SLCP reduces to the standard LCP. For the standard LCP, much effort has been made in developing theoretical analysis for the existence of a solution, numerical methods for finding a solution, and applications in engineering and economics [5,7,9]. On the other hand, in many practical applications, some data in the LCP cannot be known with certainty. The SLCP is aimed at a practical treatment of the LCP under uncertainty. However, only a little attention has been paid on the SLCP in the literature.

In general, there is no x satisfying the SLCP$(M(\omega), q(\omega))$ for almost all $\omega \in \Omega$. A deterministic formulation for the SLCP provides a decision vector which is optimal in a certain sense. Different deterministic formulations may yield different solutions that are optimal in different senses.

Gürkan,Özge and Robinson [12] considered the sample-path approach for stochastic variational inequalities and provided convergence theory and applications for the

* This work is partly supported by the Information Research Center for Development of Knowledge Society Infrastructure, Graduate School of Informatics, Kyoto University, Japan.

† Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100080, China (htfang@iss.ac.cn). The work of this author is also supported by National Natural Science Foundation of China.

‡ Department of Mathematical Sciences, Hirosaki University, Hirosaki 036-8561, Japan (chen@cc.hirosaki-u.ac.jp). The work of this author is also supported by a Grant-in-Aid from Japan Society for the Promotion of Science.

§ Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan (fuku@i.kyoto-u.ac.jp). The work of this author is also supported by a Grant-in-Aid from Japan Society for the Promotion of Science.

approach. When applied to the SLCP$(M(\omega), q(\omega))$, the approach is the same as the *Expected Value* (EV) method, which uses the expected function of the random function $M(\omega)x + q(\omega)$ and solves the deterministic problem

$$x \geq 0, \quad E\{M(\omega)x + q(\omega)\} \geq 0, \; x^T E\{M(\omega)x + q(\omega)\} = 0.$$

Using a simulation-based algorithm in [12], we can find a solution of this problem.

Recently, Chen and Fukushima [3] proposed a new deterministic formulation called the *Expected Residual Minimization* (ERM) method, which is to find a vector $x \in \mathbb{R}^n_+$ that minimizes the expected residual of the SLCP$(M(\omega), q(\omega))$, i.e.,

$$\min_{x \in \mathbb{R}^n_+} E\{\|\Phi(x, \omega)\|^2\}, \tag{1.1}$$

where $\Phi : \mathbb{R}^n \times \Omega \to \mathbb{R}^n$ is defined by

$$\Phi(x, \omega) = \begin{pmatrix} \phi([M(\omega)x]_1 + q_1(\omega), x_1) \\ \vdots \\ \phi([M(\omega)x]_n + q_n(\omega), x_n) \end{pmatrix},$$

and $[x]_i$ denotes the ith component of the vector x. Here $\phi : \; \mathbb{R}^2 \to \mathbb{R}$ is an NCP function which has the property

$$\phi(a, b) = 0 \iff a \geq 0, \, b \geq 0, \, ab = 0.$$

Various NCP functions have been studied for solving complementarity problems [7]. In this paper, we will concentrate on the "min" function

$$\phi(a, b) = \min(a, b).$$

Similar results can be obtained for other NCP functions, such as the Fischer-Burmeister (FB) function [10], which have the same growth behavior as the "min" function.

Let ERM$(M(\cdot), q(\cdot))$ denote problem (1.1) and define

$$G(x) = \int_\Omega \|\Phi(x, \omega)\|^2 dF(\omega), \tag{1.2}$$

where $F(\omega)$ is the distribution function of ω. Then ERM$(M(\cdot), q(\cdot))$ is rewritten as

$$\text{minimize } G(x) \qquad \text{subject to } x \geq 0. \tag{1.3}$$

Recall that an $n \times n$ matrix A is called an R_0 matrix if

$$x \geq 0, \; Ax \geq 0, \; x^T Ax = 0 \quad \Longrightarrow \quad x = 0.$$

It is known [5, Theorem 3.9.23] that the solution set of the standard LCP(A, b)

$$x \geq 0, \; Ax + b \geq 0, \; x^T(Ax + b) = 0$$

is bounded for every $b \in \mathbb{R}^n$, if and only if A is an R_0 matrix. In addition, when A is a P_0 matrix, the LCP(A, b) has a nonempty solution set if and only if A is an R_0 matrix [5, Theorem 3.9.22]. Example 1 in [3] shows that the solution set of LCP$(M(\bar{\omega}), q(\bar{\omega}))$ being nonempty and bounded for some $\bar{\omega} \in \Omega$ does not imply that the ERM$(M(\cdot), q(\cdot))$ has a solution. The following results on the existence of a solution of ERM$(M(\cdot), q(\cdot))$ are given in [3].

(i) If $M(\cdot)$ is continuous in ω and there is an $\bar{\omega} \in \Omega$ such that $M(\bar{\omega})$ is an R_0 matrix, then the solution set of ERM$(M(\cdot), q(\cdot))$ is nonempty and bounded.
(ii) When $M(\omega) \equiv M$, the solution set of ERM$(M(\cdot), q(\cdot))$ is nonempty and bounded for any $q(\cdot)$ if and only if M is an R_0 matrix.

In this paper, we substantially extend and refine the results established in [3]. In particular, we introduce the concept of a stochastic R_0 matrix and show that $M(\cdot)$ being a stochastic R_0 matrix is a necessary and sufficient condition for the solution set of ERM$(M(\cdot), q(\cdot))$ to be nonempty and bounded. Moreover, we will extend the local and global error bound results for the R_0 matrix LCP given by Mangasarian and Ren [16] to the stochastic R_0 matrix LCP in the ERM formulation.

Throughout the paper, the norm $\|\cdot\|$ denotes the Euclidean norm and $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x \geq 0\}$. For a given vector $x \in \mathbb{R}^n$, we denote $I(x) = \{i : x_i = 0\}$ and $J(x) = \{i : x_i \neq 0\}$. For vectors $x, y \in \mathbb{R}^n$, $\min(x, y)$ denotes the vector with components $\min(x_i, y_i)$, $i = 1, \cdots, n$.

The remainder of the paper is organized as follows: In Section 2, the definition and some properties of a stochastic R_0 matrix are given. In Section 3, we show a necessary and sufficient condition for the existence of a minimizer of the ERM problem with an arbitrary $q(\cdot)$ is that $M(\cdot)$ is a stochastic R_0 matrix. In Section 4, the differentiability of G is considered. Some optimality conditions and error bounds of the ERM problem are given in Section 5. In Section 6, we use a stochastic approximation method [2, 14] with acceleration by averaging [18] to solve the general ERM problem, and use a Newton-type method to solve the ERM problem with $M(\omega) \equiv M$. Furthermore, applications to traffic equilibrium and system control are provided. Preliminary numerical results show that the ERM formulation has various advantages.

2. Stochastic R_0 matrix.

A stochastic R_0 matrix is formally defined as follows.

DEFINITION 2.1. *$M(\cdot)$ is called a stochastic R_0 matrix if*

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

If Ω only contains a single realization, then the definition of a stochastic R_0 matrix reduces to that of an R_0 matrix.

Let G be defined by (1.2). We call $x^* \in \mathbb{R}^n_+$ a *local solution* of the ERM$(M(\cdot),\ q(\cdot))$, if there is $\gamma > 0$ such that $G(x) \geq G(x^*)$ for all $x \in \mathbb{R}^n_+ \cap B(x^*, \gamma) := \{x : \|x - x^*\| \leq \gamma\}$, and call x^* a *global solution* of ERM$(M(\cdot),\ q(\cdot))$, if $G(x) \geq G(x^*)$ for all $x \in \mathbb{R}^n_+$.

THEOREM 2.2. *The following statements are equivalent.*

(i) *$M(\cdot)$ is a stochastic R_0 matrix.*
(ii) *For any $x \geq 0\ (x \neq 0)$, at least one of the following two conditions is satisfied:*
  (a) *$P\{\omega : [M(\omega)x]_i \neq 0\ \} > 0$ for some $i \in J(x)$;*
  (b) *$P\{\omega : [M(\omega)x]_i < 0\} > 0$ for some $i \in I(x)$.*
(iii) *ERM$(M(\cdot), q(\cdot))$ with $q(\omega) \equiv 0$ has zero as its unique global solution.*

Proof. The proof is given in the order (i) $\Rightarrow$ (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i).

(i) $\Rightarrow$ (iii): It is easy to see that zero is a global solution of ERM$(M(\cdot), q(\cdot))$ with $q(\omega) \equiv 0$, since $G(x) \geq 0$ for all $x \in \mathbb{R}^n_+$ and $G(0) = 0$. Now we show the uniqueness of the solution. Let $\bar{x} \in \mathbb{R}^n_+$ be an arbitrary vector such that $G(\bar{x}) = 0$. By the definition of G, we have

$$\Phi(\bar{x}, \omega) = \min(M(\omega)\bar{x}, \bar{x}) = 0, \quad \text{a.e.},$$

which implies

$$\bar{x} \geq 0, \quad M(\omega)\bar{x} \geq 0, \quad \bar{x}^T M(\omega)\bar{x} = 0, \quad \text{a.e.}$$

By the definition of a stochastic R_0 matrix, we deduce $\bar{x} = 0$.

(iii) $\Rightarrow$ (ii): Suppose (ii) does not hold, that is, there exists a nonzero $x^0 \geq 0$ such that

$$P\{\omega : \ [M(\omega)x^0]_i = 0\} = 1 \text{ for all } i \in J(x^0),$$
$$P\{\omega : \ [M(\omega)x^0]_i \geq 0\} = 1 \text{ for all } i \in I(x^0).$$

Then it follows from $q(\omega) \equiv 0$ that $G(x^0) = 0$. Moreover, it is easy to see that for any $\lambda > 0$, λx^0 is a solution of ERM$(M(\cdot), 0)$, i.e., zero is not the unique solution of ERM$(M(\cdot), 0)$.

(ii) $\Rightarrow$ (i): Assume that there exists $x \neq 0$ such that $x \geq 0$, $M(\omega)x \geq 0$ and $x^T M(\omega)x = 0$, a.e. Then, since $x^T M(\omega)x = 0$, we have for almost all ω, $[M(\omega)x]_i = 0$ for all $i \in J(x)$ and $[M(\omega)x]_i \geq 0$ for all $i \in I(x)$. This contradicts (ii). ∎

For $\nu > 0$, let us denote $B_\Omega(\bar{\omega}, \nu) := \{\omega : \|\omega - \bar{\omega}\| < \nu\}$ and

$$\text{supp}\Omega := \{\bar{\omega} \in \Omega : \int_{B_\Omega(\bar{\omega},\nu)\cap\Omega} dF(\omega) > 0 \text{ for any } \nu > 0\}.$$

Here suppΩ is called the support set of Ω. When Ω consists of countable discrete points, i.e., $\Omega = \{\omega_1, \cdots, \omega_i, \cdots\}$ and $P(\omega_i) = p_i > 0$ for all i, we have supp$\Omega = \Omega$. In the case that there is a density function ρ such that $dF(\omega) = \rho(\omega)d\omega$, we have supp$\Omega = \bar{S}$, where $\bar{S}$ is the closure of set $S = \{\omega \in \Omega : \ \rho(\omega) > 0\}$.

Corollary 2.3. *Suppose that $M(\omega)$ is a continuous function of ω. Then $M(\cdot)$ is a stochastic R_0 matrix if and only if for any $x \geq 0\,(x \neq 0)$, at least one of the following two conditions is satisfied:*

(a) *there exists $\bar{\omega} \in$ suppΩ such that $[M(\bar{\omega})x]_i \neq 0$ for some $i \in J(x)$;*

(b) *there exists $\bar{\omega} \in$ suppΩ such that $[M(\bar{\omega})x]_i < 0$ for some $i \in I(x)$.*

Proof. By the continuity of $M(\omega)$ and the definition of suppΩ, conditions (a) and (b) in this corollary imply (a) and (b) in Theorem 2.2 (ii), respectively. ∎

Corollary 2.4. *Suppose that $M(\omega)$ is a continuous function of ω and $M(\bar{\omega})$ is an R_0 matrix for some $\bar{\omega} \in$ suppΩ. Then $M(\cdot)$ is a stochastic R_0 matrix.*

The following example shows that the condition that $M(\cdot)$ is a stochastic R_0 matrix is weaker than the condition that $M(\omega)$ is continuous in ω and there is an $\bar{\omega} \in$ suppΩ such that $M(\bar{\omega})$ is an R_0 matrix.

Example 2.1. Let

$$M(\omega) = \begin{pmatrix} -2\omega & \omega - |\omega| & 0 \\ 0 & \omega + |\omega| & -2\omega \\ 0 & 0 & 0 \end{pmatrix},$$

where $\omega \in \Omega = [-0.5, 0.5]$ and ω is uniformly distributed on Ω. Clearly, for $\omega < 0$, $M(\omega) = \begin{pmatrix} -2\omega & 2\omega & 0 \\ 0 & 0 & -2\omega \\ 0 & 0 & 0 \end{pmatrix}$. Then $x = (1, 1, 0)^T$ satisfies $M(\omega)x = 0$. On the other hand, for $\omega > 0$, $M(\omega) = \begin{pmatrix} -2\omega & 0 & 0 \\ 0 & 2\omega & -2\omega \\ 0 & 0 & 0 \end{pmatrix}$. Then $x = (0, 1, 1)^T$ satisfies $M(\omega)x = 0$. In this example, there is no $\omega \in \Omega$ such that $M(\omega)$ is an R_0 matrix. However $M(\cdot)$ is a stochastic R_0 matrix as verified by Theorem 2.2 (ii). For any $x \geq 0$ with $x \neq 0$, if $x_1 \neq 0$, then for any $\omega > 0$, $[M(\omega)x]_1 = -2\omega x_1 < 0$. If $x_1 = 0$

but $x_2 \neq 0$, then for any $\omega < 0$, $[M(\omega)x]_1 = 2\omega x_2 < 0$. If only $x_3 \neq 0$, then for any $\omega > 0$, $[M(\omega)x]_2 = -2\omega x_3 < 0$.

The following proposition shows a relation between $M(\cdot)$ and $\bar{M} := E\{M(\omega)\}$.

PROPOSITION 2.5. *If $\bar{M}$ is an R_0 matrix, then $M(\cdot)$ is a stochastic R_0 matrix.*

Proof. If $M(\cdot)$ were not a stochastic R_0 matrix, then by Theorem 2.2 (ii), there exists $x \geq 0$ such that $x \neq 0$ and, for almost all ω, $[M(\omega)x]_i = 0$ for $i \in J(x)$ and $[M(\omega)x]_i \geq 0$ for $i \in I(x)$. Therefore, $[\bar{M}x]_i = 0$ for $i \in J(x)$ and $[\bar{M}x]_i \geq 0$ for $i \in I(x)$. This is impossible, since $\bar{M}$ is an R_0 matrix. □

This proposition implies that for any given $\tilde{M}$, if $\tilde{M}$ is an R_0 matrix, then $M(\cdot) = \tilde{M} + M_0(\cdot)$ with $E\{M_0(\omega)\} = 0$ is a stochastic R_0 matrix. The converse of this proposition is not true. The next proposition gives a way to construct a stochastic R_0 matrix $M(\cdot)$ from a given $\tilde{M}$ which is not necessarily an R_0 matrix. Let

$$\Xi(M) := \{x : x \geq 0, x \neq 0,\ [Mx]_i = 0, i \in J(x) \text{ and } [Mx]_i \geq 0, i \in I(x)\}. \tag{2.1}$$

Obviously, if $\Xi(\tilde{M}) = \emptyset$, then $\tilde{M}$ is an R_0 matrix, and hence, by Proposition 2.5, $M(\cdot) = \tilde{M} + M_0(\cdot)$ with $E\{M_0(\omega)\} = 0$ is a stochastic R_0 matrix.

PROPOSITION 2.6. *Let $\tilde{M}$ and $M_0(\cdot)$ be such that $\Xi(\tilde{M}) \neq \emptyset$ and $E\{M_0(\omega)\} = 0$. Suppose that for any $x \in \Xi(\tilde{M})$, at least one of the following two conditions is satisfied:*

(1) *For some $i \in J(x)$, $E\{([M_0(\omega)x]_i)^2\} > 0$;*

(2) *For some $i \in I(x)$, $P\{\omega : [M_0(\omega)x]_i < -b\} > 0$ for any $b > 0$.*

Then $M(\cdot) = \tilde{M} + M_0(\cdot)$ is a stochastic R_0 matrix.

Proof. For $x \in \Xi(\tilde{M})$, these two conditions imply that the conditions in Theorem 2.2 (ii) hold for $M(\cdot)$. For $x \notin \Xi(\tilde{M})$, the same conditions also hold trivially. So $M(\cdot)$ is a stochastic R_0 matrix. □

This proposition suggests a way to obtain a stochastic R_0 matrix $M(\cdot)$ from an arbitrary matrix $\tilde{M}$. Specifically, we can construct a simple stochastic matrix $M_0(\cdot)$ such that $\tilde{M} + M_0(\cdot)$ is a stochastic R_0 matrix, as illustrated in the following example.

EXAMPLE 2.2. We consider the following matrix [3]:

$$\tilde{M} = \begin{pmatrix} 0 & 0 & 1 & -2 & -3 \\ 0 & 0 & 1 & -6 & -3 \\ -1 & -1 & 0 & 0 & 0 \\ 2 & 6 & 0 & 0 & 0 \\ 3 & 3 & 0 & 0 & 0 \end{pmatrix},$$

which arises from a linear programming problem in [13]. Clearly, $\tilde{M}$ is not an R_0 matrix, and $\Xi(\tilde{M}) = \{x : x = (0, 0, \lambda, \alpha, \beta)^T, \lambda > 0, \alpha, \beta \geq 0, \lambda - 6\alpha - 3\beta \geq 0\}$. Let ω_0 be a random variable whose distribution is $\mathcal{N}(0, 1)$. Let

$$M_0(\omega_0) = \begin{pmatrix} 0 & 0 & 0.5\omega_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -0.5\omega_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Then for any $b > 0$, $P\{\omega_0 : [M_0(\omega_0)x]_1 < -b\} > 0$ holds for any $x \in \Xi(\tilde{M})$. Hence, by Proposition 2.6, $\tilde{M} + M_0(\cdot)$ is a stochastic R_0 matrix.

The following proposition shows that the sum of a stochastic R_0 matrix $M(\cdot)$ and a matrix $M_1(\cdot)$ with $E\{M_1(\omega_1)\} = 0$ yields a stochastic R_0 matrix.

PROPOSITION 2.7. *Let $\omega = (\omega_0, \omega_1)$ and $\hat{M}(\omega) = M(\omega_0) + M_1(\omega_1)$, where $M(\cdot)$ is a stochastic R_0 matrix, $E\{M_1(\omega_1)\} = 0$, and $M(\omega_0)$ is independent of $M_1(\omega_1)$. Then $\hat{M}(\cdot)$ is a stochastic R_0 matrix.*

Proof. If $\tilde{M} := E\{M(\omega_0)\}$ is an R_0 matrix, then from $E\{M_1(\omega_1)\} = 0$ and Proposition 2.5, $M(\cdot) + M_1(\cdot)$ is a stochastic R_0 matrix. Otherwise, let $M_0(\omega_0) = M(\omega_0) - \tilde{M}$ and choose any $x \in \Xi(\tilde{M})$. Suppose that the first condition of Proposition 2.6 holds for $M_0(\omega_0)$. Since $M(\omega_0)$ is independent of $M_1(\omega_1)$, we have

$$E\{([(M_0(\omega_0) + M_1(\omega_1))x]_i)^2\} = E\{([M_0(\omega_0)x]_i)^2\} + E\{([M_1(\omega_1)x]_i)^2\} > 0$$

for some $i \in J(x)$. Now, suppose that the second condition of Proposition 2.6 holds for $M_0(\omega_0)$, i.e., $P\{\omega_0 : [M_0(\omega_0)x]_i < -b\} > 0$ for some $i \in I(x)$. Note that

$$\begin{aligned} &P\{\omega : [(M_0(\omega_0) + M_1(\omega_1))x]_i < -b\} \\ &\geq P\{(\omega_0, \omega_1) : [M_0(\omega_0)x]_i < -b \text{ and } [M_1(\omega_1)x]_i \leq 0\} \\ &= P\{\omega_0 : [M_0(\omega_0)x]_i < -b\} P\{\omega_1 : [M_1(\omega_1)x]_i \leq 0\}. \end{aligned}$$

Since $E\{[M_1(\omega_1)x]_i\} = 0$, we have $P\{\omega_1 : [M_1(\omega_1)x]_i \leq 0\} > 0$. Thus, we have

$$P\{\omega : [(M_0(\omega_0) + M_1(\omega_1))x]_i < -b\} > 0,$$

i.e., the second condition of Proposition 2.6 also holds for $M_0(\omega_0 + M_1(\omega_1)$. Since

$$\hat{M}(\omega) = M(\omega_0) + M_1(\omega_1) = \tilde{M} + M_0(\omega_0) + M_1(\omega_1),$$

Proposition 2.6 ensures that $\hat{M}(\cdot)$ is a stochastic R_0 matrix. ☐

3. Boundedness of solution set.

In this section, the boundedness of the solution set of the ERM problem (1.3) is studied.

THEOREM 3.1. *Let $q(\cdot)$ be arbitrary. Then $G(x) \to \infty$ as $\|x\| \to \infty$ with $x \in \mathbb{R}^n_+$ if and only if $M(\cdot)$ is a stochastic R_0 matrix.*

Proof. First we prove the "if" part. For simplicity, we denote $|x| = (|x_1|, \cdots, |x_n|)^T$ and $\text{sign}(x) = (\text{sign}(x_1), \cdots, \text{sign}\ (x_n))^T$ for a vector x, where

$$\text{sign}(x_i) = \begin{cases} 1, & x_i > 0, \\ 0, & x_i = 0, \\ -1, & x_i < 0. \end{cases}$$

Note that for any $a, b \in \mathbb{R}$, we have

$$\begin{aligned} 2\min(a, b) &= a + b - \text{sign}(a - b)(a - b) \\ &= (1 - \text{sign}(a - b))a + (1 + \text{sign}(a - b))b \end{aligned}$$

and

$$\begin{aligned} 4(\min(a, b))^2 &= a(1 - \text{sign}(a - b))^2 a + b(1 + \text{sign}(a - b))^2 b + 2b(1 - \text{sign}^2(a - b))a \\ &= 2a(1 - \text{sign}(a - b))a + 2b(1 + \text{sign}(a - b))b. \end{aligned}$$

For any $x \in \mathbb{R}^n$ and $\omega \in \Omega$, we define the diagonal matrix

$$D(x, \omega) = \text{diag}(\text{sign}(M(\omega)x + q(\omega) - x)).$$

Then we have

$$\|\Phi(x,\omega)\|^2 = \frac{1}{2}[(M(\omega)x+q(\omega))^T(I-D(x,\omega))(M(\omega)x+q(\omega)) \\ + x^T(I+D(x,\omega))x]. \tag{3.1}$$

Consider an arbitrary $x \geq 0$ with $\|x\| = 1$. Suppose condition (a) in Theorem 2.2 (ii) holds. Choose $i \in J(x)$ such that $P\{\omega : [M(\omega)x]_i \neq 0\} > 0$. Then there exists a sufficiently large $K > 0$ such that $P\{\omega : [M(\omega)x]_i \neq 0,\ |q_i(\omega)| \leq K\} > 0$.

First consider the case where $P\{\omega : [M(\omega)x]_i < x_i,\ |q_i(\omega)| \leq K\} > 0$. Let

$$\Omega_1 := \{\omega : [M(\omega)x]_i < (1-\delta)x_i,\ |q_i(\omega)| \leq K\},$$

where $\delta > 0$. Then we have $P\{\Omega_1\} > 0$ whenever δ is sufficiently small. Moreover, for any sufficiently large $\lambda > 0$, $\mathrm{sign}(\lambda[M(\omega)x]_i + q_i(\omega) - \lambda x_i) = -1$ for any $\omega \in \Omega_1$. Therefore, by (1.2) and (3.1), we have

$$G(\lambda x) \geq \int_{\Omega_1} (\lambda[M(\omega)x]_i + q_i(\omega))^2 dF(\omega) \to \infty \text{ as } \lambda \to \infty. \tag{3.2}$$

Next, consider the case where $P\{\omega : [M(\omega)x]_i > x_i, |q_i(\omega)| \leq K\} > 0$. Let

$$\Omega_2 := \{\omega : [M(\omega)x]_i > (1+\delta)x_i, |q_i(\omega)| \leq K\}.$$

Then we have $P(\Omega_2) > 0$ for a sufficiently small $\delta > 0$. Moreover, for any sufficiently large $\lambda > 0$, $\mathrm{sign}(\lambda[M(\omega)x]_i + q_i(\omega) - \lambda x_i) = 1$ for $\omega \in \Omega_2$. Hence we have

$$G(\lambda x) \geq \int_{\Omega_2} (\lambda x_i)^2 dF(\omega) \to \infty \text{ as } \lambda \to \infty. \tag{3.3}$$

Finally consider the case where $P\{\omega : [M(\omega)x]_i = x_i, |q_i(\omega)| \leq K\} > 0$. Let

$$\Omega_3 := \{\omega : [M(\omega)x]_i = x_i, |q_i(\omega)| \leq K\}.$$

Then we have

$$G(\lambda x) \geq \int_{\Omega_3} \{(\lambda x_i + q_i(\omega))^2 1_{\{q_i(\omega)<0\}} + (\lambda x_i)^2 1_{\{q_i(\omega)\geq 0\}}\} dF(\omega) \to \infty \text{ as } \lambda \to \infty. \tag{3.4}$$

Combining (3.2)–(3.4), we see that $G(\lambda x) \to \infty$ as $\lambda \to \infty$.

Now, suppose condition (b) in Theorem 2.2 (ii) holds. Choose $i \in I(x)$ such that $P\{\omega : [M(\omega)x]_i < 0\} > 0$. Let

$$\Omega_4 := \{\omega :\ [M(\omega)x]_i < -\delta,\ |q_i(\omega)| < K\}.$$

Then we have $P\{\Omega_4\} > 0$ for any sufficiently small $\delta > 0$ and sufficiently large $K > 0$. Moreover, for any $\lambda > 0$ large enough, $\lambda[M(\omega)x]_i + q_i(\omega) < 0$ for $\omega \in \Omega_4$. Thus we have

$$(1 - \mathrm{sign}(\lambda[M(\omega)x]_i + q_i(\omega)))(\lambda[M(\omega)x]_i + q_i(\omega))^2 = 2(\lambda[M(\omega)x]_i + q_i(\omega))^2,$$

which yields

$$G(\lambda x) \geq \int_{\Omega_4} (\lambda[M(\omega)x]_i + q_i(\omega))^2 dF(\omega) \to \infty \text{ as } \lambda \to \infty.$$

Since x is an arbitrary nonzero vector such that $x \geq 0$, we deduce from the above arguments that $G(x) \to \infty$ as $\|x\| \to \infty$ with $x \geq 0$, provided the statement (ii) in Theorem 2.2 holds.

Let us turn to proving the "only if" part. Suppose that $M(\cdot)$ is not a stochastic R_0 matrix, i.e., there exists $x \geq 0$ with $x \neq 0$ such that $[M(\omega)x]_i = 0$ for all $i \in J(x)$ and $[M(\omega)x]_i \geq 0$ for all $i \in I(x)$, a.e. For any $\lambda > 0$, from (1.2) and (3.1), we have

$$\begin{aligned} G(\lambda x) = \frac{1}{2}\sum_{i=1}^{n} E\{(1 - &\ \mathrm{sign}(\lambda[M(\omega)x]_i + q_i(\omega) - \lambda x_i))(\lambda[M(\omega)x]_i + q_i(\omega))^2 \\ &+ (1 + \mathrm{sign}(\lambda[M(\omega)x]_i + q_i(\omega) - \lambda x_i))(\lambda x_i)^2\}. \end{aligned} \tag{3.5}$$

The ith term of the right-hand side of (3.5) with $x_i \neq 0$ equals

$$\begin{aligned} &E\{(1 - \mathrm{sign}(q_i(\omega) - \lambda x_i))q_i(\omega)^2 + (1 + \mathrm{sign}(q_i(\omega) - \lambda x_i))(\lambda x_i)^2\} \\ &\qquad = 2E\{q_i(\omega)^2 1_{\{q_i(\omega\} \leq \lambda x_i\}} + (\lambda x_i)^2 1_{\{q_i(\omega) > \lambda x_i\}}\} \leq 2E\{q_i(\omega)^2\}, \end{aligned}$$

while the ith term of the right-hand side of (3.5) with $x_i = 0$ equals

$$\begin{aligned} &E\{(1 - \mathrm{sign}(\lambda[M(\omega)x]_i + q_i(\omega)))(\lambda[M(\omega)x]_i + q_i(\omega))^2\} \\ &\qquad = 2E\{(\lambda[M(\omega)x]_i + q_i(\omega))^2 1_{\{\lambda[M(\omega)x]_i < -q_i(\omega)\}}\} \leq 2E\{q_i(\omega)^2\}, \end{aligned}$$

where the last inequality follows from $0 > \lambda[M(\omega)x]_i + q_i(\omega) \geq q_i(\omega)$, implying $(\lambda[M(\omega)x]_i + q_i(\omega))^2 \leq q_i(\omega)^2$. So, we obtain

$$G(\lambda x) \leq E\{\|q(\omega)\|^2\} \text{ for any } \lambda > 0.$$

Since $x \geq 0$ with $x \neq 0$, this particularly implies that G is bounded above on a nonnegative ray in $\mathbb{R}^n_+$. This completes the proof of the "only if " part. ∎

The solution set of $\mathrm{ERM}(M(\cdot), q(\cdot))$ may be bounded even if $M(\cdot)$ is not a stochastic R_0 matrix. It depends on the distribution of $q(\omega)$, as shown in the following two propositions.

PROPOSITION 3.2. *If $M(\cdot)$ is not a stochastic R_0 matrix, $P\{\omega : q_i(\omega) > 0\} > 0$ for some $i \in J(x)$, and $P\{\omega : q_i(\omega) \geq 0\} = 1$ for all $i \in I(x)$, where $x \neq 0$ is any nonnegative vector at which the conditions (a) and (b) in Theorem 2.2 (ii) fail to hold, then the solution set of ERM$(M(\cdot), q(\cdot))$ is bounded.*

Proof. Note that

$$G(0) = E\{\|\Phi(0,\omega)\|^2\} = \sum_{i=1}^{n} E\{q_i(\omega)^2 1_{\{q_i(\omega) < 0\}}\}. \tag{3.6}$$

For any nonnegative vector $x \neq 0$ satisfying the conditions (a) and (b) in Theorem 2.2 (ii), the proof of Theorem 3.1 indicates that

$$G(\lambda x) \to \infty \text{ as } \lambda \to 0. \tag{3.7}$$

Let $x \neq 0$ be any nonnegative vector which does not satisfy the conditions (a) and (b) in Theorem 2.2 (ii), i.e., $[M(\omega)x]_i = 0$ for $i \in J(x)$, and $[M(\omega)x]_i \geq 0$ for $i \in I(x)$, a.e. Then by (3.1), we have

$$\begin{aligned} G(\lambda x) &= \sum_{i \in J(x)} E\{[(1 - \mathrm{sign}(q_i(\omega) - \lambda x_i))q_i(\omega)^2 + (1 + \mathrm{sign}(q_i(\omega) - \lambda x_i))(\lambda x_i)^2]/2\} \\ &= \sum_{i \in J(x)} \{E\{q_i(\omega)^2\} - E\{1_{\{q_i(\omega) - \lambda x_i > 0\}}[q_i(\omega)^2 - (\lambda x_i)^2]\}\}, \end{aligned} \tag{3.8}$$

where the first equality follows from the assumption that $P\{\omega : q_i(\omega) \geq 0\} = 1$ for $i \in I(x)$ and hence $[M(\omega)x]_i + q_i(\omega) \geq 0$, a.e., for $i \in I(x)$. Note that

$$\begin{aligned} 0 &\leq E\{1_{\{q_i(\omega)-\lambda x_i>0\}}[q_i(\omega)^2 - (\lambda x_i)^2]\} = E\{1_{\{q_i(\omega)>\lambda x_i\}}[q_i(\omega)^2 - (\lambda x_i)^2]\} \\ &\leq E\{1_{\{q_i(\omega)>\lambda x_i\}} q_i(\omega)^2\} \to 0 \text{ as } \lambda \to \infty, \end{aligned}$$

which together with (3.8) implies

$$\lim_{\lambda\to\infty} G(\lambda x) = \sum_{i\in J(x)} E\{q_i(\omega)^2\}. \tag{3.9}$$

On the other hand, for any nonzero $x \geq 0$, we have

$$\sum_{i=1}^{n} E\{q_i(\omega)^2 1_{\{q_i(\omega)<0\}}\} = \sum_{i\in J(x)} E\{q_i(\omega)^2 1_{\{q_i(\omega)<0\}}\} < \sum_{i\in J(x)} E\{q_i(\omega)^2\}, \tag{3.10}$$

where the equality follows from the assumption that $P\{\omega : q_i(\omega) \geq 0\} = 1$ for all $i \in I(x)$ and the strict inequality follows from the assumption that $P\{\omega : q_i(\omega) > 0\} > 0$ for some $i \in J(x)$. Combining (3.6), (3.9), and (3.10), we have

$$G(0) < \lim_{\lambda\to+\infty} G(\lambda x). \tag{3.11}$$

Let $\Lambda := \{x \in \mathbb{R}^n_+ : G(x) \leq G(0)\}$. From (3.7) and (3.11), we have $\sup_{x\in\Lambda} \|x\| < +\infty$. Since any solution belongs to Λ, this implies that the solution set is bounded. □

PROPOSITION 3.3. *If $M(\cdot)$ is not a stochastic R_0 matrix and, for any i, $P\{\omega : -b \leq q_i(\omega) < 0\} = 1$ for some $b > 0$ and $P\{\omega : q_i(\omega) \neq 0$ and $[M(\omega)x^0]_i = 0\} = 0$, where $x^0 \neq 0$ is any nonnegative vector at which the conditions (a) and (b) in Theorem 2.2 (ii) fail to hold, then the solution set of ERM$(M(\cdot), q(\cdot))$ is empty or unbounded.*

Proof. Let $x^0 \neq 0$ be any nonnegative vector which does not satisfy the conditions (a) and (b) in Theorem 2.2 (ii). From (1.2) and (3.1), we have

$$\begin{aligned} G(\lambda x^0) = \sum_{i=1}^{n} E\{[(1 - \mathrm{sign}(\lambda[M(\omega)x^0]_i + q_i(\omega) - \lambda x_i^0))(\lambda[M(\omega)x^0]_i + q_i(\omega))^2 \\ + (1 + \mathrm{sign}(\lambda[M(\omega)x^0]_i + q_i(\omega) - \lambda x_i))(\lambda x_i^0)^2]/2\}. \end{aligned} \tag{3.12}$$

For every $i \in J(x^0)$, we have $[M(\omega)x^0]_i = 0$ and $q_i(\omega) = 0$, a.e., and hence the ith term of the right-hand side of (3.12) is zero for any $\lambda > 0$. For every $i \in I(x^0)$, we have $[M(\omega)x^0]_i \geq 0$ and $q_i(\omega) < 0$, a.e., which implies

$$\begin{aligned} &E\{(1 - \mathrm{sign}(\lambda[M(\omega)x^0]_i + q_i(\omega)))(\lambda[M(\omega)x^0]_i + q_i(\omega))^2\} \\ &= 2E\{(\lambda[M(\omega)x^0]_i + q_i(\omega))^2 1_{\{\lambda[M(\omega)x^0]_i<-q_i(\omega),[M(\omega)x^0]_i>0\}}\} \\ &\quad + 2E\{q_i^2(\omega)1_{\{[M(\omega)x^0]_i=0\}}\}. \end{aligned} \tag{3.13}$$

By assumption, the second term on the right-hand side of (3.13) is zero for any $\lambda > 0$, and

$$\begin{aligned} &E\{(\lambda[M(\omega)x^0]_i + q_i(\omega))^2 1_{\{\lambda[M(\omega)x^0]_i<-q_i(\omega),[M(\omega)x^0]_i>0\}}\} \\ &\leq b^2 P\{\omega : 0 < \lambda[M(\omega)x^0]_i < b\} \to 0 \text{ as } \lambda \to \infty. \end{aligned}$$

Therefore, we obtain

$$\lim_{\lambda \to +\infty} G(\lambda x^0) = 0,$$

but for any $x \in \mathbb{R}^n_+$, $G(x) \geq 0$. So for any $\gamma > 0$, the level set $\Lambda_\gamma := \{x : G(x) \leq \gamma\}$ is unbounded, which means the solution set is unbounded if it is not empty. ∎

From Theorem 3.1, we have the following necessary and sufficient condition for the solution set of ERM$(M(\cdot), q(\cdot))$ to be bounded for any $q(\cdot)$.

THEOREM 3.4. *The solution set of ERM$(M(\cdot), q(\cdot))$ is nonempty and bounded for any $q(\cdot)$ if and only if $M(\cdot)$ is a stochastic R_0 matrix.*

4. Differentiability of G. The objective function G of ERM$(M(\cdot), q(\cdot))$ is, in general, not convex. If G is differentiable at x, then $\min(\nabla G(x), x) = 0$ implies that x is a stationary point of ERM$(M(\cdot), q(\cdot))$. The differentiability of G is studied in [3] for the special case where $M(\omega) \equiv M$, $q(\omega) = \bar{q} + T\omega$ with $M \in \mathbb{R}^{n\times n}$, $\bar{q} \in \mathbb{R}^n$, $T \in \mathbb{R}^{n\times m}$ being constants and T having at least one nonzero element in each row.

In this section, we will give a condition for the function G to be differentiable under a general setting. The continuity of $M(\cdot)$ and $q(\cdot)$ is not assumed.

DEFINITION 4.1. *We say that the strict complementarity condition holds at x with probability one if*

$$P\{\omega : [M(\omega)x]_i + q_i(\omega) = x_i\} = 0, \quad i = 1, \cdots, n.$$

Obviously this definition is a generalization of the strict complementarity condition for the LCP. The proof for the differentiability of G at x under the strict complementarity condition with probability one is not trivial.

For any fixed ω, if $[M(\omega)x]_i + q_i(\omega) - x_i \neq 0$ for all i, then $\|\Phi(x,\omega)\|^2$ is differentiable at x and

$$\nabla_x \|\Phi(x,\omega)\|^2 = M(\omega)^T(I - D(x,\omega))(M(\omega)x + q(\omega)) + (I + D(x,\omega))x.$$

To simplify the notation, we define

$$f(x,\omega) := M(\omega)^T(I - D(x,\omega))(M(\omega)x + q(\omega)) + (I + D(x,\omega))x. \tag{4.1}$$

THEOREM 4.2. *The function $g(x) := \int_\Omega f(x,\omega)dF(\omega)$ is continuous at x if the strict complementarity condition holds at x with probability one.*

Proof. We will show that $\|g(x+h) - g(x)\| \to 0$ as $h \to 0$. Since

$$\begin{aligned} f(x,\omega) - f(x+h,\omega) = {} & (M(\omega)^T(I - D(x,\omega))M(\omega) + I + D(x,\omega))h \\ & + M(\omega)^T(D(x+h,\omega) - D(x,\omega))(M(\omega)(x+h) + q(\omega)) \\ & - (D(x+h,\omega) - D(x,\omega))(x+h), \end{aligned}$$

there exist some constants c_1, $c_2 > 0$ such that

$$\begin{aligned} \|g(x+h) - g(x)\| &= \|\int_\Omega [f(x+h,\omega) - f(x,\omega)]dF(\omega)\| \\ &\leq c_1\|h\| + c_2 \int_\Omega \|D(x+h,\omega) - D(x,\omega)\|dF(\omega). \end{aligned}$$

Then we just need to show that

$$\int_\Omega \|D(x+h,\omega) - D(x,\omega)\|dF(\omega) \to 0 \text{ as } h \to 0.$$

Note that

$$\{\omega : \|D(x+h,\omega) - D(x,\omega)\| \neq 0\} \subset \cup_{i=1}^n \{A_i \cup B_i\},$$

where

$$\begin{aligned} A_i &:= \{\omega : [M(\omega)x]_i + q_i(\omega) - x_i \geq 0,\ [M(\omega)(x+h)]_i + q_i(\omega) - x_i - h_i \leq 0\}, \\ B_i &:= \{\omega : [M(\omega)x]_i + q_i(\omega) - x_i \leq 0,\ [M(\omega)(x+h)]_i + q_i(\omega) - x_i - h_i \geq 0\}. \end{aligned}$$

For any $\varepsilon > 0$, since the strict complementarity condition holds at x with probability one, there is a $\delta > 0$ such that

$$P\{\omega : |[M(\omega)x]_i + q_i(\omega) - x_i| < \delta\} < \varepsilon/2. \tag{4.2}$$

Let

$$C_i := \{\omega : [M(\omega)x]_i + q_i(\omega) - x_i \geq \delta,\ [M(\omega)(x+h)]_i + q_i(\omega) - x_i - h_i \leq 0\}.$$

Then, we have

$$\begin{aligned} A_i &\subset C_i \cup \{\omega : |[M(\omega)x]_i + q_i(\omega) - x_i| < \delta\}, \\ C_i &\subset \{\omega : [M(\omega)h]_i - h_i \leq -\delta\}. \end{aligned}$$

Applying a similar procedure to B_i, we have

$$P\{A_i \cup B_i\} \leq P\{\omega : |[M(\omega)h]_i - h_i| \geq \delta\} + P\{\omega : |[M(\omega)x]_i + q_i(\omega) - x_i| < \delta\}.$$

By the Chebychev inequality, there is an $h_0 > 0$ such that for any h with $\|h\| < h_0$,

$$P\{\omega : |[M(\omega)h]_i - h_i| \geq \delta\} < \varepsilon/2.$$

This together with (4.2) implies that g is continuous at x. $\square$

THEOREM 4.3. *If the strict complementarity condition holds at any x in an open set $U \subset \mathbb{R}^n$ with probability one, then G is Fréchet differentiable at $x \in U$ and*

$$\nabla G(x) = \int_\Omega f(x,\omega) dF(\omega). \tag{4.3}$$

Proof. First, we will show that for almost all ω, $\mu\{x \in U : \ [M(\omega)x]_i + q_i(\omega) = x_i\} = 0$ for any i, where μ is Lebesgue measure. If it were no true, then for some i

$$P\{\omega : \mu\{x \in U : \ [M(\omega)x]_i + q_i(\omega) = x_i\} > 0\} > 0,$$

which implies

$$\int_\Omega \int_U 1_{\{[M(\omega)x]_i + q_i(\omega) = x_i\}} dx dF(\omega) > 0. \tag{4.4}$$

But from the assumption and the Fubini Theorem [11], we obtain

$$\int_\Omega \int_U 1_{\{[M(\omega)x]_i + q_i(\omega) = x_i\}} dx dF(\omega) = \int_U \int_\Omega 1_{\{[M(\omega)x]_i + q_i(\omega) = x_i\}} dF(\omega) dx = 0.$$

This contradicts (4.4), and hence for almost all ω, $\mu\{x \in U : \ [M(\omega)x]_i + q_i(\omega) = x_i\} = 0$ for any i.

Note that, for any $\omega \in \Omega$, $\|\Phi(x,\omega)\|^2$ is locally Lipschitz and hence absolutely continuous with respect to x. For any (x,ω) such that $[M(\omega)x]_i + q_i(\omega) \neq x_i$, $\|\Phi(x,\omega)\|^2$ is differentiable with respect to x. Therefore by the Fundamental Theorem of Calculus for Lebesgue Integrals [11], for any x, we have

$$\|\Phi(x+h_ie_i,\omega)\|^2 - \|\Phi(x,\omega)\|^2 = \int_0^{h_i} [f(x+se_i,\omega)]_i ds \tag{4.5}$$

for almost all ω, where $e_i = (0,\cdots,0,\underset{i}{1},0,\cdots,0)^T$. Thus

$$G(x+h) - G(x) = \sum_{i=1}^n \int_\Omega (\|\Phi(x+\sum_{k=i}^n h_ke_k,\omega)\|^2 - \|\Phi(x+\sum_{k=i+1}^n h_ke_k,\omega)\|^2) dF(\omega). \tag{4.6}$$

By (4.5), (4.6), and the Fubini Theorem, we deduce that

$$\int_\Omega \int_0^{h_i} [f(y+se_i,\omega)]_i ds dF(\omega) = \int_0^{h_i} \int_\Omega [f(y+se_i,\omega)]_i dF(\omega) ds$$

for any i and $y \in B(x,\|h\|) \subset U$, and hence

$$\begin{aligned}
&G(x+h) - G(x) - h^T \int_\Omega f(x,\omega)dF(\omega)\\
&= \sum_{i=1}^n \int_0^{h_i} \int_\Omega [f(x+\sum_{k=i+1}^n h_ke_k + se_i,\omega)]_i dF(\omega)ds - \sum_{i=1}^n \int_0^{h_i} \int_\Omega [f(x,\omega)]_i dF(\omega)ds\\
&= \sum_{i=1}^n \int_0^{h_i} \int_\Omega ([f(x+\sum_{k=i+1}^n h_ke_k + se_i,\omega)]_i - [f(x,\omega)]_i) dF(\omega)ds\\
&= \sum_{i=1}^n \int_0^{h_i} (g_i(x+\sum_{k=i+1}^n h_ke_k + se_i) - g_i(x))ds,
\end{aligned}$$

where g is defined in Theorem 4.2. From Theorem 4.2, for any $\varepsilon > 0$, there exists a sufficiently small $h_0 > 0$ such that for any h with $\|h\| < h_0$,

$$|G(x+h) - G(x) - h^T \int_\Omega f(x,\omega)dF(\omega)| < \varepsilon\|h\|,$$

which implies

$$\frac{|G(x+h) - G(x) - h^T \int_\Omega f(x,\omega)dF(\omega)|}{\|h\|} \to 0 \quad \text{as} \quad \|h\| \to 0.$$

Therefore, G is Fréchet differentiable at x and (4.3) holds. □

REMARK. When $M(\omega) \equiv M$ and $q(\omega) = \bar{q} + T\omega$, if $[T\omega]_i$ has no mass at any point for each i, i.e., $P\{\omega : [T\omega]_i = a\} = 0$ for any $a \in \mathbb{R}$, then $P\{\omega : [M(\omega)x]_i + q_i(\omega) = x_i\} = 0$, $i = 1,\cdots,n$. Therefore, if T has at least one nonzero element in each row [3], then for all $x \in \mathbb{R}^n$, the strict complementarity condition holds with probability one, and G is differentiable in $\mathbb{R}^n_+$. This indicates that the result shown in Theorem 4.3 contains the results established in [3].

Let $F_{q_i}(s)$ be the distribution function of $q_i(\omega)$, i.e., $F_{q_i}(s) = P\{\omega : q_i(\omega) \le s\}$. Suppose $M(\omega) \equiv M$. Then, we have

$$\begin{aligned} G(x) &= \int_{-\infty}^{+\infty} \sum_{i=1}^{n} (\min([Mx]_i + s, x_i)^2 dF_{q_i}(s) \\ &= \sum_{i=1}^{n} \int_{-\infty}^{[(I-M)x]_i} ([Mx]_i + s)^2 dF_{q_i}(s) + \sum_{i=1}^{n} x_i^2 (1 - F_{q_i}([(I-M)x]_i)). \end{aligned} \tag{4.7}$$

It is shown in [3] that for some special distribution functions, $G(x)$ can be computed without using discrete approximation. The following proposition shows that, under some conditions, we can also compute $\nabla G(x)$ without using discrete approximation.

PROPOSITION 4.4. *If $M(\omega) \equiv M$ and $F_{q_i}(s)$ is a continuous function for all i, then*

$$\nabla G(x) = 2M^T H(x)x + 2(I - H(x))x - 2M^T v(x), \tag{4.8}$$

where

$$\begin{aligned} H(x) &:= \mathrm{diag}(F_{q_1}([(I-M)x]_1), \cdots, F_{q_n}([(I-M)x]_n), \\ v(x) &:= (\int_{-\infty}^{[(I-M)x]_1} F_{q_1}(s)ds, \cdots, \int_{-\infty}^{[(I-M)x]_n} F_{q_n}(s)ds)^T. \end{aligned}$$

Proof. If $M(\omega) \equiv M$ and $F_{q_i}(s)$ is continuous for all i, then $P\{\omega : q_i(\omega) = a\} = 0$ for any $a \in \mathbb{R}$, and hence $P\{\omega : [Mx]_i + q_i(\omega) = x_i\} = 0$ for each $x \in \mathbb{R}^n_+$. Then, by Theorem 4.3, $G(x)$ is differentiable at any $x \in \mathbb{R}^n_+$ and

$$\begin{aligned} \nabla G(x) &= \int_\Omega [M^T(I - D(x,\omega))(Mx + q(\omega)) + (I + D(x,\omega))x] dF(\omega) \\ &= M^T[\int_\Omega (I - D(x,\omega))dF(\omega)Mx + \int_\Omega (I - D(x,\omega))q(\omega)dF(\omega)] \\ &\quad + (\int_\Omega (I + D(x,\omega))dF(\omega))x \\ &= 2M^T H(x)Mx + 2M^T R(x) + 2(I - H(x))x, \end{aligned} \tag{4.9}$$

where

$$R(x) := (\int_{-\infty}^{[(I-M)x]_1} s dF_{q_1}(s), \cdots, \int_{-\infty}^{[(I-M)x]_n} s dF_{q_n}(s))^T.$$

By integration by parts, we have

$$\int_{-\infty}^{[(I-M)x]_i} s dF_{q_i}(s) = [(I-M)x]_i F_{q_i}([(I-M)x]_i) - \int_{-\infty}^{[(I-M)x]_i} F_{q_i}(s)ds.$$

This implies that

$$R(x) = (\int_{-\infty}^{[(I-M)x]_1} s dF_{q_1}(s), \cdots, \int_{-\infty}^{[(I-M)x]_n} s dF_{q_n}(s))^T = H(x)(I-M)x - v(x).$$

Combining this with (4.9), we have the desired formula (4.8). ☐

From (4.8), we see that the smoothness of $G(\cdot)$ depends on the smoothness of $F_{q_i}(\cdot)$, $i = 1, \cdots, n$. If for all i, $F_{q_i}(\cdot)$ is differentiable at $[(I-M)x]_i$ and $\rho_i(\cdot)$, the derivative of $F_{q_i}(\cdot)$, is continuous at $[(I-M)x]_i$, then the Hessian matrix of $G(x)$ can be written as

$$\begin{aligned}\nabla^2 G(x) &= 2M^T H(x)M + 2(M^T S(x) + S(x)M) - 2M^T S(x)M + 2(I - S(x) - H(x)) \\ &= 2M^T H(x)M + 2(I - H(x)) - 2(I-M)^T S(x)(I-M), \qquad (4.10)\end{aligned}$$

where

$$S(x) := \operatorname{diag}(x_1\rho_1([(I-M)x]_1), \cdots, x_n\rho_n([(I-M)x]_n)).$$

5. Optimality conditions and error bounds. In numerical algorithms, residual functions play an important role in terminating iterations and verifying accuracy of a computed solution. The following theorem shows the basic properties of the residual function defined by

$$r(x) = \|\min(\nabla G(x), x)\|.$$

THEOREM 5.1. *Suppose that the strict complementarity condition holds at any x in an open set U with probability one. Then the following statements are true.*

(1) *If $\bar{x} \in U$ is a local solution of ERM$(M(\cdot), q(\cdot))$, then $r(\bar{x}) = 0$.*

(2) *If $G(\cdot)$ is twice continuously differentiable at $\bar{x} \in \mathbb{R}^n_+$ where $r(\bar{x}) = 0$ and the Hessian matrix $\nabla^2 G(\bar{x})$ is positive definite, then there are an open set $\bar{U} \subset U$ and a constant $\tau > 0$ such that $\bar{x}$ is a unique local solution of ERM$(M(\cdot), q(\cdot))$ in $\bar{U}$, and for all $x \in \bar{U}$*

$$\|x - \bar{x}\| \le \tau r(x). \qquad (5.1)$$

Proof. From Theorem 4.3, we can write the first order optimality condition for the ERM problem (1.3) as $r(x) = 0$. Now we show (5.1). Since $G(\cdot)$ is twice continuously differentiable at $\bar{x}$ and $\nabla^2 G(\bar{x})$ is positive definite, there is an open set $\bar{U} \subset U$ such that $\nabla G(x)$ is a locally Lipschitz continuous and uniform P function in $\bar{U}$. Applying Proposition 6.3.1 in [7] to the nonsmooth equation $\min(x, \nabla G(x)) = 0$, we obtain (5.1). ☐

COROLLARY 5.2. *If $\Omega = \{\omega_1, \cdots, \omega_N\}$, and the strict complementarity condition holds at $x \in \mathbb{R}^n_+$ with probability one, then $r(x) = 0$ implies x is a local solution of ERM$(M(\cdot), q(\cdot))$.*

Proof. Since the strict complementarity condition holds at $x \in \mathbb{R}^n_+$ with probability one, by (3.1), for each i, $\|\Phi(x, \omega_i)\|^2$ is twice continuously differentiable and

$$\nabla^2 G(x) = \sum_{i=1}^N [M(\omega_i)^T (I - D(x, \omega_i))M(\omega_i) + (I + D(x, \omega_i))].$$

Since the Hessian matrix $\nabla^2 G(x)$ is positive semidefinite, and $G(x)$ is a quadratic function in $B(x, \nu)$ for a sufficiently small $\nu > 0$, x is a local solution. ☐

Now we consider error bounds for the case where G is not necessarily differentiable. Let

$$s(x) = G(x) - \min_{x \in \mathbb{R}^n_+} G(x).$$

When $\Omega = \{\omega_1, \cdots, \omega_N\}$, we can write

$$G(x) = \sum_{j=1}^{N}\sum_{i=1}^{n} |\min([M(\omega_j)x + q(\omega_j)]_i, x_i)|^2 p(\omega_j),$$

where $p(\omega_j)$ is the probability of ω_j. Clearly, there exist finitely many convex polyhedra such that G is a convex quadratic function on each polyhedron, i.e., G is a piecewise convex quadratic function. By Theorem 2.5 in [15], we have the following local error bound result:

PROPOSITION 5.3. *If $\Omega = \{\omega_1, \cdots, \omega_N\}$, then there exist constants $\tau > 0$ and $\varepsilon > 0$ such that for any $x \in \mathbb{R}^n_+$ with $s(x) \le \varepsilon$*

$$\|x - x^*(x)\| \le \tau s(x)^{1/2},$$

where $x^*(x)$ is a global solution of ERM$(M(\cdot), q(\cdot))$ closest to x under the norm $\|\cdot\|$.

Let us denote $s_\gamma(x) = s(x)^\gamma$ for $\gamma > 0$. For a general continuous distribution of ω, G may not be a piecewise convex function. The following example shows that the function s_γ provides a local error bound for the ERM$(M(\cdot), q(\cdot))$ with various values of γ depending on the distribution of ω.

EXAMPLE 5.1. Consider the SLCP$(M(\omega), q(\omega))$ with

$$M(\omega) \equiv \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & 0 & -1 \end{pmatrix}, \quad q(\omega) = \begin{pmatrix} -1 \\ -1 \\ \omega \end{pmatrix},$$

where ω is a random variable with supp$\Omega \subset [-1, 0]$. It is easy to check that $M(\omega)$ is an R_0 matrix. For any ω, the solution set of LCP$(M(\omega), q(\omega))$ is $\{(x_1, 1 - x_1, 0)^T : x_1 \in [-\omega, 1]\} \cup \{(-\frac{1}{2}\omega + \frac{1}{2}, 0, \frac{1}{2}\omega + \frac{1}{2})^T\}$. Let $\rho(\omega)$ be the density function of ω. We consider the following two cases: $\rho(\omega) \equiv 1$ and $\rho(\omega) = 2(\omega + 1)$. Clearly, $x^* = (1, 0, 0)^T$ is the unique global solution of ERM$(M(\cdot), q(\cdot))$ and $r(x^*) = 0$ for these two cases. But for any $x = (x_1, 1 - x_1, 0)^T$ with $x_1 \in [0, 1]$, if $\rho(\omega) \equiv 1$, then $s(x) = (1 - x_1)^{3/2}/\sqrt{3}$, but if $\rho(\omega) = 2(\omega + 1)$, then $s(x) = (1 - x_1)^2/\sqrt{6}$. Noticing that $\|x - x^*\| = \sqrt{2}(1 - x_1)$, we have $\|x - x^*\| \le \tau s_\gamma(x)$, where γ depends on the distribution of ω. So the general form of local error bound for ERM$(M(\cdot), q(\cdot))$ with continuous random variables is difficult to obtain unless the information on the distribution of ω is known.

THEOREM 5.4. *Let $M(\cdot)$ be a stochastic R_0 matrix. Then for any $\varepsilon > 0$, there exists $\tau > 0$ such that for each $x \in \mathbb{R}^n_+$ with $s(x) > \varepsilon$*

$$\|x - x^*(x)\| \le \tau s(x)^{1/2},$$

where $x^*(x)$ is a global solution of ERM$(M(\cdot), q(\cdot))$ closest to x under the norm $\|\cdot\|$.

Proof. If the assertion were not true, then for any positive integer k, there exists x^k with $s(x^k) > \varepsilon$ such that

$$\|x^k - x^*(x^k)\| > ks(x^k)^{1/2} > k\varepsilon^{1/2}.$$

Since $M(\cdot)$ is a stochastic R_0 matrix, by Theorem 3.4, the global solution set of ERM$(M(\cdot), q(\cdot))$ is nonempty and bounded. Therefore $\|x^k\| \to \infty$ as $k \to \infty$, and

$$\frac{s(x^k)^{1/2}}{\|x^k\|} \le \frac{\|x^k - x^*(x^k)\|}{k\|x^k\|} \to 0 \text{ as } k \to \infty. \tag{5.2}$$

Let $\{x^{n_k}/\|x^{n_k}\|\}$ be a convergent subsequence of $\{x^k/\|x^k\|\}$. Note that

$$\lim_{k\to\infty}\frac{s(x^{n_k})^{1/2}}{\|x^{n_k}\|}=\lim_{k\to\infty}\frac{(G(x^{n_k})-\min_{x\in\mathbb{R}^n_+}G(x))^{1/2}}{\|x^{n_k}\|}=\lim_{k\to\infty}\frac{G(x^{n_k})^{1/2}}{\|x^{n_k}\|}$$
$$=\lim_{k\to\infty}(\int_\Omega\sum_{i=1}^n|\min(\frac{[M(\omega)x^{n_k}]_i+q_i(\omega)}{\|x^{n_k}\|},\frac{x_i^{n_k}}{\|x^{n_k}\|})|^2dF(\omega))^{1/2}.$$

Since for any x with $\|x\|=1$

$$\int_\Omega\sum_{i=1}^n|\min([M(\omega)x]_i+q_i(\omega),x_i)|^2dF(\omega)\le\int_\Omega(\|M(\omega)\|^2+\|q(\omega)\|^2)dF(\omega)+1<\infty,$$

by the dominated convergence theorem, we obtain

$$\lim_{k\to\infty}\frac{s(x^{n_k})^{1/2}}{\|x^{n_k}\|}=(\int_\Omega\sum_{i=1}^n|\min([M(\omega)\hat{x}]_i,\hat{x}_i)|^2dF(\omega))^{1/2}<\infty,$$

where $\hat{x}$ is an accumulation point of $\{x^{n_k}/\|x^{n_k}\|\}$. This, together with (5.2) and $s(x^{n_k})^{1/2}/\|x^{n_k}\|\ge 0$, yields

$$\int_\Omega\sum_{i=1}^n|\min([M(\omega)\hat{x}]_i,\hat{x}_i)|^2dF(\omega)=0,$$

which implies that $\hat{x}$ is a solution of the ERM$(M(\cdot),0)$. Since $\|\hat{x}\|=1$, this contradicts the assumption that $M(\cdot)$ is a stochastic R_0 matrix from Theorem 2.2 (iii). □

REMARK If Ω contains only one element ω and LCP$(M(\omega),q(\omega))$ has a solution, then error bounds in Theorem 5.1 and Theorem 5.4 reduce to the local and global error bounds for the R_0 matrix LCP given in [16]. Hence the two theorems are extensions of error bounds for the R_0 matrix LCP given in [16] to the stochastic R_0 matrix LCP in the ERM formulation.

6. Examples and numerical results. In this section, we report numerical results of four examples of the stochastic R_0 matrix LCP in the ERM formulation.

Let the measure of feasibility of $x\in\mathbb{R}^n_+$ with tolerance $\varepsilon\ge 0$ be defined by

$$rel_\varepsilon(x)=P\{\omega:[M(\omega)x]_i+q_i(\omega)\ge-\varepsilon,\ i=1,\cdots,n\}. \tag{6.1}$$

This measure indicates how much we may expect that x satisfies the constraints $M(\omega)x+q(\omega)\ge 0$ (with some tolerance).

EXAMPLE 6.1. We use $\tilde{M}$ and $M_0(\omega_0)$ given in Example 2.2, and

$$M_1(\omega')=\begin{pmatrix}0&0&0&-\omega_1&0\\0&0&0&0&-0.4-0.4\ln\omega_2\\0&0&0&0&0\\\omega_1&0&0&-2\sqrt{3}\omega_3&-2\sqrt{3}\omega_3\\0&0.4+0.4\ln\omega_2&0&-3\omega_4&3\omega_4\end{pmatrix},$$

where $\omega'=(\omega_1,\cdots,\omega_4)^T$ with the distributions of ω_1, ω_2, ω_3, ω_4 being $\mathcal{U}[-0.8,0.8]$, $\mathcal{U}[0,1]$, $\mathcal{N}(0,1)$ and $\mathcal{N}(0,1)$, respectively. Let $\omega=(\omega_0,\omega_1,\cdots,\omega_4)^T$ and $M(\omega)=\tilde{M}+M_0(\omega_0)+M_1(\omega')$. From Example 2.2, we know that $\tilde{M}+M_0(\omega_0)$ is a stochastic

R_0 matrix. It is easy to verify that $E\{M_1(\omega')\} = 0$. Hence by Proposition 2.7, $M(\cdot)$ is a stochastic R_0 matrix.

We set $q(\omega) = \tilde{q} + q_0(\omega)$, where $\tilde{q}$ is a constant vector and $E\{q_0(\omega)\} = 0$. In this example, we choose

$$q_0(\omega) = (0.1\omega_0, 0.1\omega_0, 0, -2\sqrt{3}\omega_3, -3\omega_4)^T$$

with three different cases for $\tilde{q}$,

$$\tilde{q}^1 = (2, 3, 100, -180, -162)^T,\ \tilde{q}^2 = (-5, -5, 0, 10, 10)^T,\ \tilde{q}^3 = (-5, -5, -5, -5, -5)^T.$$

The deterministic LCP$(\tilde{M}, \tilde{q}^i)$, $i = 1, 2, 3$, have a unique solution $(36, 18, 0, 0.25, 0.5)^T$, multiple solutions $(0, 0, \lambda, 0, 0)^T$ with $\lambda \geq 5$, and no feasible solution, respectively.

For all $q_i(\omega)$, we can check that for any $x = (x_1, \cdots, x_5)^T \in \mathbb{R}^5_+$ with $x_i \neq 0, i = 1, 2$, the strict complementarity condition holds at x with probability one, and so $\nabla G(x)$ exists at these points. Hence we can use a stochastic approximation algorithm [2, 14, 18] to find a minimizer of $G(x)$ in $\mathbb{R}^n_+$. The iterative formula is given by

$$x^{k+1} = \max(x^k - a_k f(x^k, \omega^k), 0), \tag{6.2}$$

where $f(x, \omega)$ is defined by (4.1), a_k is a stepsize satisfying $\sum_{k=1}^{\infty} a_k = \infty$ and $a_k \to 0$, and ω^k is the kth sample of ω. By the convergence theorems of stochastic approximation algorithms ([2, Theorem 2.2.1] and [14, Theorem 5.2.1]), the generated sequence $\{x^k\}$ will converge to a connected set S such that every $\bar{x} \in S$ satisfies $\min(g(\bar{x}), \bar{x}) = 0$ with $g(x)$ defined in Theorem 4.2. If $\bar{x}_i \neq 0, i = 1, 2$, then by Theorem 4.3, $\nabla G(\bar{x}) = g(\bar{x})$. In this example, a_k is chosen as

$$a_k = \begin{cases} 0.003, & k \leq 10^4, \\ 0.0025, & 10^4 < k \leq 10^5, \\ 0.002, & 10^5 < k \leq 5 \times 10^5, \\ \frac{1}{k^{0.6}}, & 5 \times 10^5 < k \leq 2 \times 10^6. \end{cases}$$

When $k \geq 5 \times 10^5$, we use the averaging technique proposed by [18] to accelerate the convergence.

The stochastic approximation algorithm is a local optimization algorithm. To avoid being trapped in a local minimum, for each $\tilde{q}^i$, $i = 1, 2, 3$, we executed 36 times simulation from different initial points $x^0 = (10l,\ 10l',\ 0,\ 0,\ 0)^T$, $l, l' \in \{0,\ 1,\ \cdots,\ 5\}$. The step size a_k and initial points were chosen based on suggestions for stochastic approximation algorithms in [14].

For each $\tilde{q}^i$, the information on the last iterate x^{kmax}, where $kmax = 2 \times 10^6$, obtained by (6.2) is shown in Table 6.1. The columns labeled as "$G(x)$" and "$r(x)$" show the respective values obtained by the Monte Carlo method with 10^6 samples. The row labeled as "average" shows the average of the values obtained from 36 different initial points. The rows labeled as "min" and "max" indicate the interval of those values, which represents the variability of the values obtained from 36 different initial points.

Recall that the EV method solves the deterministic LCP$(\tilde{M}, \tilde{q})$. Let $\tilde{x}$ be a solution of LCP$(\tilde{M}, \tilde{q})$. For $\tilde{q}^2 = (-5, -5, 0, 10, 10)^T$, since there are multiple solutions $(0, 0, \lambda, 0, 0)^T$ with $\lambda \geq 5$, $G(x)$ and $r(x)$ are evaluated at $\tilde{x} = (0, 0, 5, 0, 0)^T$. There is no feasible solution of LCP$(\tilde{M}, \tilde{q})$ with $\tilde{q}^3 = (-5, -5, -5, -5, -5)^T$.

TABLE 6.1
Simulation results for Example 6.1 where $E\{M(\omega)\}$ is not an R_0 matrix.

	x_1	x_2	x_3	x_4	x_5	$G(x)$	$r(x)$
$\tilde{q}^1 = (2,\ 3,\ 100,\ -180,\ -162)^T$							
min	39.5439	23.298	0	0.2079	0.345	7.2405	0.0115
max	40.1396	23.5793	0.0096	0.3804	0.5413	7.5741	0.6486
average	39.7865	23.4563	0.0014	0.2610	0.4635	7.413	0.15826
$\tilde{x}$	36	18	0	0.25	0.5	197.03	12025
$\tilde{q}^2 = (-5,\ -5,\ 0,\ 10,\ 10)^T$							
min	0.0004	0.0044	11.4092	0	0	1.8518	0
max	0.0030	0.0154	11.6959	0	0	1.9037	0.002
average	0.0008	0.0068	11.5410	0	0	1.8747	4.44×10^{-5}
$\tilde{x}$	0	0	5	0	0	3.1428	0.5718
$\tilde{q}^3 = (-5,\ -5,\ -5,\ -5,\ -5)^T$							
min	0.004	1.3915	5.7993	0.0005	0.1137	51.652	0.0440
max	0.0377	1.5017	5.896	0.0186	0.2343	51.943	5.2424
average	0.011	1.4347	5.8414	0.003	0.1555	51.734	1.2536

TABLE 6.2
$rel_\varepsilon(\tilde{x})$ and average of $rel_\varepsilon(x^{kmax})$ for Example 6.1 with $\tilde{q}^1$ in Table 6.1

ε	0.0	0.1	0.2	0.5	1
$rel_\epsilon(\tilde{x})$	0.0018	0.0581	0.2971	0.3236	0.3417
ave.$rel_\epsilon(x^{kmax})$	0.3084	0.7190	0.9007	0.9488	0.9518

By using the Monte Carlo method with 10^6 samples, we evaluated the measure of feasibility rel_ε defined by (6.1) for the case of $\tilde{q}^1 = (2, 3, 100, -180, -162)^T$, at $\tilde{x} = (36, 18, 0, 0.25, 0.5)^T$ and at the last iterates obtained by the iterative formula (6.2) from 36 different initial points. The results are presented in Table 6.2. The row labeled as "ave.rel_ϵ" shows the average values of $rel_\varepsilon(x^{kmax})$ obtained at the 36 last iterates x^{kmax}. For each $\tilde{q}^i$ and initial points, the computational time for obtaining the values in Table 6.1 is about 185 second by Matlab 7.0 at computer with P4 3.06 GHz CPU.

EXAMPLE 6.2. In this example, we consider the case where $M(\omega) \equiv \tilde{M}$ is a P matrix and $q(\omega)$ has continuous distribution. In this case, the EV formulation LCP$(\tilde{M}, \tilde{q})$ has a unique solution $\tilde{x}$. The objective function G of the ERM formulation is twice continuously differentiable and the values of $G(x)$, $\nabla G(x)$, and $\nabla^2 G(x)$ can be computed by (4.7), (4.8), (4.10), respectively, without resorting to stochastic approximation.

Let $q(\omega) = \tilde{q} + q_0(\omega)$, where $\tilde{q} = E\{q(\omega)\}$, $q_0(\omega) = B\omega$, $\omega = (\omega_1, \omega_2, \omega_3)^T \in \mathcal{N}(0, I)$, $B \in \mathbb{R}^{n\times 3}$ is 100% dense, and the elements of B are randomly generated with the uniform distribution $\mathcal{U}(0, 5)$. We use Example 4.4 of [4] to generate $\tilde{M}$ and $\tilde{q}$. First, we randomly generate 100% dense $A \in \mathbb{R}^{n\times n}$ and $\bar{q} \in \mathbb{R}^n$ whose elements are uniformly distributed in $(-5, 5)$. Then we use the QR decomposition of A to get an upper triangular matrix N, and obtain a triangular matrix $\tilde{M}$ by replacing the diagonal elements of N by their absolute values.

We first use Lemke's method [8] to find a solution $\tilde{x}$ of LCP$(\tilde{M}, \tilde{q})$, and then take $\tilde{x}$ as an initial point to find a local solution $\bar{x}$ of the ERM formulation by applying the semismooth Newton method [7] to the equation $\min(\nabla G(x), x) = 0$.

The numerical experiments were carried out for $n = 20$, 50, and 100. For each n,

TABLE 6.3
Simulation results for Example 6.2 where $E\{M(\omega)\}$ is a P matrix

	$n = 20$		$n = 50$		$n = 100$	
	$r(x)$	$G(x)$	$r(x)$	$G(x)$	$r(x)$	$G(x)$
$\tilde{x}$	97.39	180.77	168.04	427.58	307.42	823.8
$\bar{x}$	6.17×10^{-8}	75.78	5.14×10^{-8}	167.07	1.34×10^{-7}	293.72
$\|\tilde{x} - \bar{x}\|$	7.51		21.03		38.18	

TABLE 6.4
$rel_\varepsilon(\tilde{x})$ and $rel_\varepsilon(\bar{x})$ for Example 6.2

	$n = 20$			$n = 50$			$n = 100$		
ε	0	1	5	0	1	5	0	1	5
$rel_\varepsilon(\tilde{x})$	0.2507	0.3387	0.6615	0.2072	0.2930	0.6152	0.1856	0.2709	0.5934
$rel_\varepsilon(\bar{x})$	0.3863	0.4955	0.8049	0.2712	0.3713	0.7225	0.2208	0.3193	0.6723

we generated 100 problems and solved them by the above-mentioned procedure. The figures presented in Table 6.3 are the average of the results obtained in this manner.

The measures rel_ε of feasibility at $\tilde{x}$ and $\bar{x}$ obtained by the EV method and the ERM method, respectively, are presented in Table 6.4.

EXAMPLE 6.3. To illustrate the application of stochastic R_0 matrix linear complementarity problems, we use a simple transportation network shown in Figure 1, which is based on an example of the deterministic traffic equilibrium network model in [6].

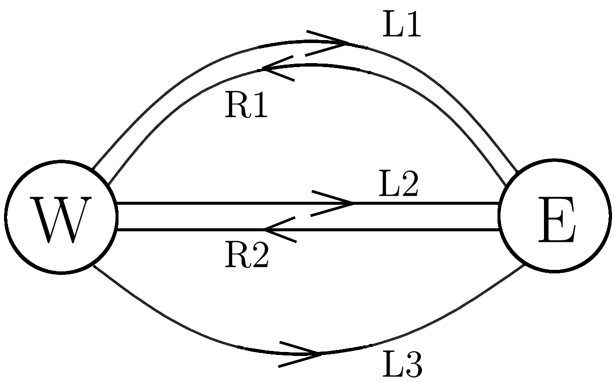

FIG. 6.1. *Road Network*

In the network, two cities West and East are connected by two two-way roads and one one-way road. More specifically, the network consists of five links, L1, L2, L3, R1, R2, where L1, L2, L3 are directed from West to East, and R1 and R2 are the returns of L1 and L2, respectively. L1-R1 is a mountain road, and L2-R2 and L3 are sea-side roads. We are interested in the traffic flow between the two cities. The Wardrop equilibrium principle states that each driver will choose the minimum cost route between the origin-destination pair, and through this process the routes that are used will have equal cost; routes with costs higher than the minimum will have no flow. In a deterministic model, the parameters in the demand and cost function are fixed, and the problem can be formulated as a (deterministic) LCP based on the Wardrop equilibrium principle.

In practice, however, the traffic condition will significantly be affected by some uncertain factors such as weather. So we want to estimate the traffic flow and the travel time that are most likely to occur, before we know such uncertain factors.[1]

We suppose that there are three possible uncertain weather conditions; sunny, windy and rainy. On a sunny day, the network is free from traffic congestion, and the travel times of all roads are constant, which are given as $(c_1, c_2, c_3, c_4, c_5)^T = (1000, 950, 3000, 1000, 1300)^T$, where c_1, c_2, c_3, c_4, c_5 denote the travel times of roads L1, L2, L3, R1, R2, respectively. On a windy day, the sea-side roads suffer from traffic jams due to congestions and the travel times of the roads in the whole network are given by

$$\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 60 & 0 & 0 & 20 \\ 0 & 0 & 80 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 100 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{pmatrix} + \begin{pmatrix} 1000 \\ 950 \\ 3000 \\ 1000 \\ 1300 \end{pmatrix},$$

where v_1, v_2, v_3, v_4, v_5 denote the traffic volumes of roads L1, L2, L3, R1, R2, respectively. On the other hand, on a rainy day, the mountain roads suffer from traffic jams and the travel times of the roads in the whole network are given by

$$\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix} = \begin{pmatrix} 40 & 0 & 0 & 20 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 8 & 0 & 0 & 80 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{pmatrix} + \begin{pmatrix} 1000 \\ 950 \\ 3000 \\ 1000 \\ 1300 \end{pmatrix}.$$

Moreover, trip demands between the two cities are higher on a sunny day than on a windy or rainy day. Specifically, $(d_1, d_2)^T = (260, 170)^T$ on a sunny day and $(d_1, d_2)^T = (160, 70)^T$ on a windy day and a rainy day, where d_1 and d_2 are trip demands from West to East and from East to West, respectively.

It is convenient to represent the travel cost functions and trip demands in a unified manner as follows:

$$c(v, \omega) = H(\omega)v + h,$$

where

$$c(v, \omega) = (c_1(v, \omega), \dots, c_5(v, \omega))^T,$$

$$H(\omega) = \begin{pmatrix} 40\alpha(\omega) & 0 & 0 & 20\alpha(\omega) & 0 \\ 0 & 60\beta(\omega) & 0 & 0 & 20\beta(\omega) \\ 0 & 0 & 80\beta(\omega) & 0 & 0 \\ 8\alpha(\omega) & 0 & 0 & 80\alpha(\omega) & 0 \\ 0 & 4\beta(\omega) & 0 & 0 & 100\beta(\omega) \end{pmatrix},$$

$$\alpha(\omega) = \frac{1}{2}\omega(\omega - 1), \quad \beta(\omega) = \omega(2 - \omega), \quad h = (1000, 950, 3000, 1000, 1300)^T.$$

[1]It should be noted that we do not intend to construct a traffic equilibrium model in which the drivers choose their routes under uncertainty.

Here $\Omega = \{\omega^1, \omega^2, \omega^3\}$ with $\omega^1 = 0$, $\omega^2 = 1$, $\omega^3 = 2$ represents the set of uncertain events of the weather, {sunny, windy, rainy}, with probabilities $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{4}$, $p_3 = \frac{1}{4}$, respectively. Also, the traffic flow $v = (v_1, v_2, v_3, v_4, v_5)^T$ should satisfy[2]

$$v \geq 0, \quad Bv \geq d(\omega),$$

where

$$B = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}, \quad d(\omega) = \begin{pmatrix} 260 - 100(\alpha(\omega) + \beta(\omega)) \\ 170 - 100(\alpha(\omega) + \beta(\omega)) \end{pmatrix}.$$

By Wardrop's principle, for each event $\omega \in \Omega$, the traffic equilibrium problem can be formulated as LCP$(M(\omega), q(\omega))$ with

$$M(\omega) = \begin{pmatrix} H(\omega) & -B^T \\ B & 0 \end{pmatrix}, \; q(\omega) = \begin{pmatrix} h \\ -d(\omega) \end{pmatrix}.$$

The solutions $x(\omega^i)$ of LCP$(M(\omega^i), q(\omega^i))$, $i = 1, 2, 3$ express the equilibrium traffic flow on each link as well as the minimum travel time between each origin-destination pair, on a sunny day, a windy day and a rainy day, respectively. The average traffic flow is given by $(E\{x_1(\omega)\}, \ldots, E\{x_5(\omega)\})$, and the average travel time on each direction is given by $(E\{x_6(\omega)\}, E\{x_7(\omega)\})$.

On the other hand, the average travel costs and demands are given by

$$E\{c(v,\omega)\} = \begin{pmatrix} 10 & 0 & 0 & 5 & 0 \\ 0 & 15 & 0 & 0 & 5 \\ 0 & 0 & 20 & 0 & 0 \\ 2 & 0 & 0 & 20 & 0 \\ 0 & 1 & 0 & 0 & 25 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{pmatrix} + \begin{pmatrix} 1000 \\ 950 \\ 3000 \\ 1000 \\ 1300 \end{pmatrix},$$

$$E\{d(\omega)\} = \begin{pmatrix} 210 \\ 120 \end{pmatrix},$$

which are exactly the same as those of the five-link example in [6].

Below we compare the estimates of the traffic flows and travel time obtained by the EV formulation and the ERM formulation.

The solution of the EV formulation, LCP$(E\{M(\omega)\}, E\{q(\omega)\})$, is denoted by $\tilde{x}$. The ERM formulation for this example is the problem of minimizing the function

$$G(x) = \sum_{i=1}^{3} p_i || \min(x, M(\omega^i)x + q)||^2.$$

We denote the solution by $\bar{x}$. In Table 6.5, we report numerical results.

It is observed from $\tilde{x}_3 = x(\omega^i) = 0, i = 1, 2, 3$ in Table 6.5 that the user-optimal load pattern estimated from the EV formulation has no flow on L3, which is the same as the user-optimal traffic pattern estimated from the LCPs for a sunny day, a windy day and a rainy day, respectively. However, the estimated total travel time $\tilde{x}_6 + \tilde{x}_7 = 5190$ from the EV formulation is larger than the total travel time obtained from the LCP for any day. On the other hand, the user-optimal traffic pattern estimated from

[2]For the purpose of our presentation, we may replace the equality constraint $Bv = d(\omega)$ by the inequality constraint. In practice, this change will not affect the solution of the problem.

TABLE 6.5
Traffic flow and travel time for Example 6.3

EV solution $\tilde{x}$	(120, 90, 0, 70, 50, 2550, 2640)
ERM solution $\bar{x}$	(84, 84, 21, 80, 20, 975, 1000)
$x(\omega^1)$	(0, 260, 0, 170, 0, 950, 1000)
$x(\omega^2)$	(955/6, 5/6, 0, 70, 0, 1000, 1000)
$x(\omega^3)$	(0, 160, 0, 3.75, 66.75, 950, 1300)
$E\{x(\omega)\}$	(39.8, 170.2, 0, 103.4, 16.6, 962.5, 1075)
$E\{\|x(\omega)-\tilde{x}\|\}$, $\|E\{x(\omega)\}-\tilde{x}\|$	2239.66, 2232.60
$E\{\|x(\omega)-\bar{x}\|\}$, $\|E\{x(\omega)\}-\bar{x}\|$	222.42, 127.16

the ERM formulation has light flow on L3 and the total travel time $\bar{x}_6 + \bar{x}_7 = 1975$ is close to $x_6(\omega^i) + x_7(\omega^i), i = 1, 2, 3$.

The two formulations yield different estimates of the user-optimal traffic pattern and the travel time, and both solutions, $\bar{x}$ and $\tilde{x}$, try to explain the phenomenon in the real world. The EV formulation uses the average of data to estimate the user-optimal traffic pattern. The ERM formulation uses the least square method to find a traffic pattern which has minimum total error to each user-optimal traffic pattern for each day. It is worth mentioning that, as far as this example is concerned, $\bar{x}$ may be considered closer to the realized traffic patterns than $\tilde{x}$ because $E\{\|x(\omega) - \tilde{x}\|\} > E\{\|x(\omega) - \bar{x}\|\}$ and $\|E\{x(\omega)\} - \tilde{x}\| > \|E\{x(\omega)\} - \bar{x}\|$.

Now, we use this example to show that the theoretical results given in this paper substantially extend the results in [3]. It is easy to verify that the matrix $E\{M(\omega)\}$ is an R_0 matrix. By Proposition 2.5, $M(\cdot)$ is a stochastic R_0 matrix. Hence by Theorem 3.1 the solution set of ERM$(M(\cdot), q(\cdot))$ is nonempty and bounded. However, for each ω^i, $M(\omega^i)$ is not an R_0 matrix. Hence the statement on the solution set cannot be obtained by using the results in [3].

EXAMPLE 6.4. The last example is a simplified control problem: Let $\hat{\omega} \in \mathbb{R}^n$ be the system parameter. Based on prior experience, we assume that $\hat{\omega}$ is generated from $\mathcal{N}(a, B)$. At each time t, we have the following observer:

$$y_{t+1} = X_t\hat{\omega} + F_t v_t, \tag{6.3}$$

where $X_t \in \mathbb{R}^{m\times n}$ is a known input, $F_t \in \mathbb{R}^{m\times r}$ is a known matrix, and $v_t \in \mathbb{R}^r$ is an unknown noise which is independent identically and normally distributed with $E\{v_t\} = 0$, $E\{v_t v_t^T\} = I$.

Suppose $B \succ 0$ and $F_t F_t^T \succeq 0$. By the Kalman filter theory [1], we have the following recursive estimation for the parameter $\hat{\omega}$:

$$\begin{aligned} \omega_{t+1} &= \omega_t + K_{t+1}(y_{t+1} - X_t\omega_t) \\ K_{t+1} &= B_t X_t^T (X_t B_t X_t^T + F_t F_t^T)^+ \\ B_{t+1} &= B_t - B_t X_t^T K_{t+1}^T \\ \omega_0 &= a,\ B_0 = B, \end{aligned} \tag{6.4}$$

where A^+ denotes the pseudo-inverse of matrix A. Then the posterior distribution of $\hat{\omega}$ is given by $N(\omega_t, B_t)$. The control law u_t is obtained as a solution of the following convex quadratic program:

$$\begin{aligned} \min\ & c_t(\hat{\omega})^T u + \frac{1}{2}u^T Q_t(\hat{\omega})u \\ \text{s.t.}\ & A_t(\hat{\omega})u \le b_t(\hat{\omega}) \\ & u \ge 0, \end{aligned} \tag{6.5}$$

where $Q_t(\hat{\omega})$, $A_t(\hat{\omega})$ are matrices and $c_t(\hat{\omega})$, $b_t(\hat{\omega})$ are vectors. The first order optimality condition of (6.5) is equivalent to the $\mathrm{LCP}(M_t(\hat{\omega}), q_t(\hat{\omega}))$ with

$$M_t(\hat{\omega}) = \begin{pmatrix} Q_t(\hat{\omega}) & A_t(\hat{\omega})^T \\ -A_t(\hat{\omega}) & 0 \end{pmatrix}, \ q_t(\hat{\omega}) = \begin{pmatrix} c_t(\hat{\omega}) \\ b_t(\hat{\omega}) \end{pmatrix}.$$

In traditional adaptive control, we replace the unknown parameter $\hat{\omega}$ by its estimate ω_t in the quadratic program (6.5) to obtain an approximation $\check{u}_t$ of the control law u_t for each t, that is, $\check{u}_t$ is the vector whose elements are the first n components of the solution of the $\mathrm{LCP}(M_t(\omega_t), q_t(\omega_t))$.

If ω_t is far away from the parameter $\hat{\omega}$, the error of $\check{u}_t$ is big and will cause trouble in some situations. Hence we take the variance of the estimate into account by using the solution $\bar{u}_t$ of the ERM formulation for $\mathrm{SLCP}(M_t(\omega), q_t(\omega))$ with $\omega \sim \mathcal{N}(\omega_t, B_t)$. Here we report numerical results for a tracking problem with the ARX model $y_{t+1} = \hat{\omega}^{(1)} y_t + \hat{\omega}^{(2)} u_t + v_t$. The controller u_t would be designed so that y_{t+1} can track a given trajectory $\exp(0.5t)$. Let the performance function be $p(u_t, \omega) := (\omega^{(1)} y_t + \omega^{(2)} u_t - \exp(0.5t))^2$. Then, from

$$p(u_t, \omega) = (\omega^{(2)})^2 u_t^2 - 2(\exp(0.5t) - \omega^{(1)} y_t)\omega^{(2)} u_t + (\exp(0.5t) - \omega^{(1)} y_t)^2,$$

we have $c_t(\omega) = -2(\exp(0.5t) - \omega^{(1)} y_t)\omega^{(2)}$, $Q_t(\omega) = 2(\omega^{(2)})^2$. We set $X_t = (y_t, u_t)$ and choose $a = (0,1)^T$, $B = \begin{pmatrix} 0.25 & 0 \\ 0 & 4 \end{pmatrix}$, $F_t = 1$, $A_t(\omega) \equiv 1$, $b_t(\omega) = 4 + 2(\omega^{(2)})^2$.

For $k \geq 1$, we generate a true parameter $\hat{\omega}^k$ from $\mathcal{N}(1,1)$ and noise $\{v_t\}$ from $\mathcal{N}(0,1)$. We solve the ERM formulation for $\mathrm{SLCP}(M_t(\omega), q_t(\omega))$ with $\omega \sim \mathcal{N}(\omega_t, B_t)$ to obtain $\bar{u}_t^k$. We then set $X_t = (y_t^k, \bar{u}_t^k)$ and use (6.3) and (6.4) to obtain y_{t+1}^k, ω_{t+1} and B_{t+1}. We also solve $\mathrm{LCP}(M_t(\omega_t), q_t(\omega_t))$ and the EV formulation of $\mathrm{SLCP}(M_t(\omega), q_t(\omega))$ with $\omega \sim \mathcal{N}(\omega_t, B_t)$ to get $\check{u}_t^k$ and $\tilde{u}_t^k$, respectively.

For the purpose of comparison, we define the average performance (for $k = 1, 2, \ldots, 100$) of these formulations by

$$\bar{\sigma}_t := \frac{1}{100} \sum_{k=1}^{100} (\bar{u}_t^k - u_t^*(\hat{\omega}^k))^2, \quad t = 1,2,3,4,5$$

$$\check{\sigma}_t := \frac{1}{100} \sum_{k=1}^{100} (\check{u}_t^k - u_t^*(\hat{\omega}^k))^2, \quad t = 1,2,3,4,5$$

$$\tilde{\sigma}_t := \frac{1}{100} \sum_{k=1}^{100} (\tilde{u}_t^k - u_t^*(\hat{\omega}^k))^2, \quad t = 1,2,3,4,5,$$

where $u_t^*(\hat{\omega}^k)$ is obtained by solving $\mathrm{LCP}(M_t(\hat{\omega}^k), q_t(\hat{\omega}^k))$ with true parameter $\hat{\omega}^k$.

From the results shown in Table 6.6, we find that the ERM formulation has better performance than $\mathrm{LCP}(M_t(\omega_t), q_t(\omega_t))$ and the EV formulation in the sense that $\bar{\sigma}_t < \check{\sigma}_t$ and $\bar{\sigma}_t < \tilde{\sigma}_t$ hold for all $t = 1,2,3,4,5$. This suggests that $\bar{u}_t^k$ is a better control law for $\hat{\omega}^k$ than $\tilde{u}_t^k$ and $\bar{u}_t^k$ in all cases.

7. Final remark. This paper proves that a necessary and sufficient condition for the $\mathrm{ERM}(M(\cdot), q(\cdot))$ having a nonempty and bounded solution set is that $M(\cdot)$ is a stochastic R_0 matrix. Proposition 2.5 shows that if the matrix $E\{M(\omega)\}$ is an R_0 matrix, then $M(\cdot)$ is a stochastic R_0 matrix. Moreover, Example 6.1 shows that there are many cases where $M(\cdot)$ is a stochastic R_0 matrix, but $E\{M(\omega)\}$ is not an

TABLE 6.6
Average performance for Example 6.4

t	1	2	3	4	5
$\bar{\sigma}_t$	1.0103	1.9764	2.286	1.6755	1.8693
$\check{\sigma}_t$	1.7053	3.0257	2.3345	1.7385	1.8918
$\tilde{\sigma}_t$	1.2425	2.5505	2.3852	1.8015	2.0903

R_0 matrix, and the EV formulation LCP$(E\{M(\omega)\}, E\{q(\omega)\})$ either has no solution or has an unbounded solution set. Therefore, the condition for the ERM$(M(\cdot), q(\cdot))$ having a nonempty and bounded solution set is weaker than the condition for the EV formulation having a nonempty and bounded solution set. Furthermore, when ERM$(M(\cdot), q(\cdot))$ has a solution $\bar{x}$ and LCP$(E\{M(\omega)\}, E\{q(\omega)\})$ has a solution $\tilde{x}$, the residuals always satisfy

$$G(\bar{x}) = E\{\|\min(M(\omega)\bar{x} + q(\omega), \bar{x})\|^2\} \leq E\{\|\min(M(\omega)\tilde{x} + q(\omega), \tilde{x})\|^2\} = G(\tilde{x}).$$

Example 6.1 shows that $G(\bar{x})$ can be much smaller than $G(\tilde{x})$. Moreover, the values of rel_ε shown in Table 6.2 reveal that, for each tolerance level $\varepsilon \geq 0$, the number of ω^i at which $M(\omega^i)\bar{x} + q(\omega^i) < -\varepsilon$ holds is much less than the number of ω^i at which $M(\omega^i)\tilde{x} + q(\omega^i) < -\varepsilon$ holds. Example 6.2 shows that a local solution $\bar{x}$ of ERM$(M(\cdot), q(\cdot))$ may conveniently be obtained from a solution $\tilde{x}$ of the EV formulation. Example 6.3 and Example 6.4 show that the EV formulation and ERM formulation express different concerns in our real life. Solving the EV formulation is usually less expensive computationally than solving the ERM formulation. Nevertheless, since $\bar{x}$ is generally expected to have better reliability than $\tilde{x}$, we may recommend the ERM method to those decision makers who do not want to take high risk of violating the conditions $M(\omega)x + q(\omega) \geq 0$.

Acknowledgements

We are very grateful to Professor Jong-Shi Pang and the referees of this paper for their helpful comments.

REFERENCES

[1] P.E. CAINES, *Linear Stochastic Systems*, Wiley, New York, 1988.
[2] H.F. CHEN, *Stochastic Approximation and Its Applications*, Kluwer Academic Publishers, Dordrecht, 2002.
[3] X. CHEN AND M. FUKUSHIMA, *Expected residual minimization method for stochastic linear complementarity problems*, Math. Oper. Res., 30 (2005) pp. 1022–1038.
[4] X. CHEN AND Y. YE, *On homotopy-smoothing methods for box-constrained variational inequalities*, SIAM J. Control Optim., 37 (1999), pp. 589–616.
[5] R.W. COTTLE, J.S. PANG, AND R.E. STONE, *The Linear Complementarity Problem*, Academic Press, Inc. New York, 1992.
[6] S. DAFERMOS, *Traffic equilibrium and variational inequalities*, Transportation Science, 14(1980), pp. 42–54.
[7] F. FACCHINEI AND J.S. PANG, *Finite-Dimensional Variational Inequalities and Complementarity Problems, I and II*, Springer-Verlag, New York, 2003.
[8] M.C. FERRIS, ftp://ftp.cs.wisc.edu/math-prog/matlab/lemke.m., Department of Computer Sciences, University of Wisconsin, 1998.
[9] M.C. FERRIS AND J.S. PANG, *Engineering and economic applications of complementarity problems*, SIAM Rev., 39 (1997), pp. 669–713.
[10] A. FISCHER, *A special Newton-type optimization method*, Optim., 24 (1992), pp. 269–284.
[11] G.B. FOLLAND, *Real Analysis: Modern Techniques and Their Applications*, John Wiley & Sons, New York, 1984.

[12] G. Gürkan, A.Y. Özge and S.M. Robinson, *Sample-path solution of stochastic variational inequalities*, Math. Programming, 84 (1999), pp. 313–333.
[13] P. Kall and S.W. Wallace, *Stochastic Programming*, John Wiley & Sons, Chichester, 1994.
[14] H.J. Kushner and G. Yin, *Stochastic Approximation Algorithms and Applications*, Springer, New York, 1997.
[15] W. Li, *Error bounds for piecewise convex quadratic programs and applications*, SIAM J. Control Optim., 33 (1994), pp. 1510–1529.
[16] O.L. Mangasarian and J. Ren, *New improved error bounds for the linear complementarity problem*, Math. Programming, 66 (1994), pp. 241–257.
[17] J.S. Pang, *Error bounds in mathematical programming*, Math. Programming, 79 (1997), pp. 299–332.
[18] B.T. Polyak and A.B. Juditsky, *Acceleration of stochastic approximation by averaging*, SIAM J. Control Optim., 30 (1992), pp. 838–855.