A Nontrivial Solution to a Stochastic Matrix Equation

J. Ding[1,*] and N. H. Rhee[2]

[1] *Department of Mathematics, The University of Southern Mississippi, Hattiesburg, MS 39406-5045, USA.*

[2] *Department of Mathematics and Statistics, University of Missouri - Kansas City, Kansas City, MO 64110-2499, USA.*

Abstract. If A is a nonsingular matrix such that its inverse is a stochastic matrix, the classic Brouwer fixed point theorem implies that the matrix equation $AXA = XAX$ has a nontrivial solution. An explicit expression of this nontrivial solution is found via the mean ergodic theorem, and fixed point iteration is considered to find a nontrivial solution.

Key words: Matrix equation, Brouwer's fixed point theorem.

1. Introduction

The matrix equation

$$ABA = BAB, \tag{1.1}$$

where both A and B are square matrices of the same size, is closely related to the parameter-independent Yang-Baxter equation (independently introduced by C. N. Yang in 1968 and Rodney Baxter in 1971 in statistical mechanics) and the theory of braid groups. The Yang-Baxter equation and braid groups, together with knot theory, have been extensively studied by physicists and mathematicians in the past decades — cf. Ref. [4] for more details on the Yang-Baxter equation and related topics. However, to our knowledge even the relevant simple matrix equation (1.1) has not been seriously explored in matrix theory.

Finding all pairs of matrices (A, B) that satisfy Eq. (1.1) is no trivial task. Thus given one of the two $n \times n$ matrices A and B (say A), finding B to satisfy the matrix equation (1.1) is equivalent to solving a system of n^2 quadratic equations. The solution to a system

of polynomial equations is a major topic in algebraic geometry, and it is not easy to find all of the solutions of the system even with 3×3 matrices.

In this paper, we are interested in finding all solutions of the following matrix equation

$$AXA = XAX \tag{1.2}$$

for a given invertible matrix A such that A^{-1} is a stochastic matrix. A matrix is called *stochastic* if it is a nonnegative matrix such that each of its row sums equals 1.

Most matrix equations that have been studied are linear in the unknown matrix, and there are many established tools and methods to solve linear matrix equations. For example, the vector space structure of the solution set to the equation $AX = XA$ (for all matrices that commute with A) is determined by the Jordan form structure of the matrix A [1]. However, the matrix equation (1.2) is nonlinear, and there is no general algebraic theory to assist in finding its solution. Eq. (1.2) obviously has two trivial solutions, the zero solution and the solution of $X = A$. In this paper, we show that Eq. (1.2) has nontrivial solutions when A^{-1} is a stochastic matrix, one of which is a stochastic solution. We may also regard Eq. (1.2) as a stochastic matrix equation, for which we want to find a stochastic solution. For another example of a stochastic matrix equation, see Refs. [5,6,8] and other references therein.

In the next section, we use the classical Brouwer fixed point theorem to establish the existence of a nontrivial solution to Eq. (1.2). This famous theorem says that if f is a continuous mapping from a compact convex set D into itself, then f has a fixed point — i.e. $f(x^*) = x^*$, where $x^* \in D$. An explicit form of this nontrivial solution is found in Section 3, by invoking the mean ergodic theorem for stochastic matrices. Another explicit nontrivial solution is also obtained in Section 4, due to the special structure of the matrix equation. A discussion on computing nontrivial solutions is presented in Section 5, and we summarize our results in Section 6.

2. The Existence of a Nontrivial Solution

The following Theorem proves the existence of a solution to Eq. (1.2).

Theorem 2.1. *Suppose $A \in R^{n \times n}$ is invertible such that its inverse A^{-1} is a stochastic matrix. Then the equation* (1.2) *has a solution $X^* = Z^* A^{-1}$ where Z^* is a stochastic matrix, and hence X^* is a stochastic matrix.*

Proof. Write the equation (1.2) as

$$XA = A^{-1}XAXAA^{-1} .$$

Let $Z = XA$. Then the above equation is the following fixed point equation for Z:

$$Z = A^{-1}Z^2 A^{-1} .$$

Denote by D the set of all $n \times n$ stochastic matrices. Then D is a closed, bounded, and convex subset of $R^{n \times n}$ [3]. For each $Z \in D$ let

$$f(Z) = A^{-1} Z^2 A^{-1} \, . \tag{2.1}$$

Since $A^{-1} \in D$ and since the set D is closed under multiplication, f maps D into itself. Clearly $f : D \to D$ is continuous. Hence, by Brouwer's fixed point theorem, there is $Z^* \in D$ such that $f(Z^*) = Z^*$ — i.e.

$$Z^* = A^{-1} (Z^*)^2 A^{-1} \, .$$

Define $X^* = Z^* A^{-1}$. Then X^* is a stochastic matrix and a solution of Eq. (1.2). $\qquad\square$

This theorem does not guarantee that a nontrivial solution can be obtained via the fixed point approach. However, if A is not a stochastic matrix then the solution X^* cannot be A, so it must be a nontrivial solution. Thus we have the following result.

Corollary 2.1. *Under the same condition as in Theorem* 2.1*, if in addition A is not a stochastic matrix then a nontrivial solution of Eq.* (1.2) *can be obtained via the fixed point approach.*

Remark 2.1. From [2], if A^{-1} is a stochastic matrix, then A is not a stochastic matrix if and only if A is not a permutation matrix.

We have the following result as a special case of Theorem 2.1 and Corollary 2.1.

Corollary 2.2. *Let A be an upper (or lower) triangular matrix. If A is invertible such that its inverse A^{-1} is a stochastic matrix, then the equation* (1.2) *has a solution $X^* = Z^* A^{-1}$ where Z^* is an upper (or lower) triangular stochastic matrix. Consequently, the solution X^* is also an upper (or lower) triangular stochastic matrix. If in addition A is not a stochastic matrix, then the solution is nontrivial.*

Proof. The proof is the same, on choosing D to be the set of all upper (or lower) triangular stochastic matrices and noting that the inverse of an upper (or lower) triangular matrix is also upper (or lower) triangular, and that D is closed under multiplication. $\qquad\square$

3. Explicit Nontrivial Solution

We can find an explicit nontrivial solution, as envisaged under Theorem 2.1 and Corollary 2.2, and to do so we introduce the following definition.

Definition 3.1. For any $n \times n$ matrix C, define

$$C_m = \frac{1}{m} \sum_{i=0}^{m-1} C^i \, .$$

Then by the *Cesaro limit* we mean

$$P = \lim_{m \to \infty} C_m \, ,$$

if this limit exists.

We recall the well-known result that the Cesaro limit exists for an $n \times n$ matrix C if and only if its spectral radius $\rho(C) < 1$, or else $\rho(C) = 1$ with the condition that each eigenvalues on the unit circle in the complex plane is semisimple [7]. We have the following properties for any stochastic matrix [3, 7].

Lemma 3.1. *If $C \in R^{n \times n}$ is a stochastic matrix, then the following statements are true.*

(a) Each eigenvalue on the unit circle in the complex plane is semisimple — i.e. its algebraic and geometric multiplicities are equal.

(b) The Cesaro limit $P = \lim_{m \to \infty} C_m$ exists.

(c) P is a stochastic matrix.

(d) P is idempotent — i.e. $P^2 = P$. (In fact, P is a projection matrix onto $N(C - I)$ along $R(C - I)$.)

(e) $CP = PC = P$.

(f) If C is invertible, then $C^{-1}P = PC^{-1} = P$.

(g) There exist bases $\{x_1, \cdots, x_k\}$ of $N(C - I)$ and $\{y_1, \cdots, y_k\}$ of $N(C^T - I)$ such that $x_i^T y_j = \delta_{ij}$ for all i and j, and $P = \sum_{i=1}^{k} x_i y_i^T$ where k is the multiplicity of the eigenvalue 1 of C.

We now proceed to find the explicit nontrivial solution.

Theorem 3.1. *If $C = A^{-1}$ is a stochastic matrix, then $P = \sum_{i=1}^{k} x_i y_i^T$ is the matrix in Lemma 3.1 — a fixed point matrix to Eq. (2.1), and hence a nontrivial solution of Eq. (1.2) unless $P = A$.*

Proof. From Lemma 3.1 (d) and (e),

$$f(P) = CP^2C = CPC = PC = P,$$

so P is a fixed point matrix of Eq. (2.1) and hence a solution to Eq. (1.2). $\square$

Now we show that the matrix $C = A^{-1}$ in Corollary 2.2 is a stochastic matrix such that 1 is the only eigenvalue on the unit circle in the complex plane.

Lemma 3.2. *Let A be an upper (or lower) triangular matrix. If A is invertible such that its inverse $C = A^{-1}$ is a stochastic matrix, then the eigenvalue 1 is the only eigenvalue of C on the unit circle in the complex plane.*

Proof. Since C is a stochastic matrix, 1 is an eigenvalue of C. Since C is an upper (or lower) triangular matrix, all of its eigenvalues appear in its diagonal entries; and since C is also a nonnegative matrix, no eigenvalue of C can be complex or negative. Consequently, 1 is the only eigenvalue of C on the unit circle in the complex plane. $\square$

For a stochastic matrix C such that 1 is the only eigenvalue on the unit circle in the complex plane, we have the following result that is stronger than Lemma 3.1 [7].

Lemma 3.3. *Suppose that an n × n matrix C is a stochastic matrix such that* 1 *is the only eigenvalue on the unit circle in the complex plane. Then all the conclusions of Lemma* 3.1 *are valid, with part (b) strengthened to*

(b) $P = \lim_{n \to \infty} C^n$ *exists.*

Thus when C is a stochastic matrix such that 1 is the only eigenvalue on the unit circle in the complex plane, the fixed point matrix P can be found from either Lemma 3.1 (g) or Lemma 3.3 (b).

4. Another Nontrivial Solution

Prior to presenting another solution of the matrix equation (1.2), we give the following simple motivating example.

Example 4.1. Let

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

so that its inverse (i.e. itself) is stochastic. From direct computation, we find all four solutions to Eq. (1.2) — viz. the zero matrix and the A matrix, as the trivial solutions, and the projection matrix

$$P = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

and the row-sum-zero matrix

$$B = \frac{1}{2} \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

as nontrivial solutions. The observation that the last solution B is just $A - P$ leads to the following result.

Theorem 4.1. *If A^{-1} is a stochastic matrix, then a solution to Eq. (1.2) is the matrix $X^* = A - P$, where P is as in Theorem 3.1.*

Proof. Since $AP = PA = P$ and $P^2 = P$, we have

$$A(A - P)A = A^3 - APA = A^3 - P$$

and

$$(A - P)A(A - P) = A^3 - PA^2 - A^2 P + PAP$$
$$= A^3 - P - P + P = A^3 - P \, ,$$

hence $A - P$ is a solution of Eq. (1.2). □

5. Discussion on How to Compute P

Suppose we want to compute a fixed point matrix P of (2.1) by fixed point iteration — i.e. we set $Z_0 = I$ where I is the identity matrix, and

$$Z_m = f(Z_{m-1}) \text{ for } m = 0, 1, \cdots,$$

where

$$f(Z) = A^{-1} Z^2 A^{-1} = C Z^2 C.$$

Since

$$Z_m = C^{2^{m+1}-2} \text{ for } m = 0, 1, 2, \cdots, \tag{5.1}$$

the fixed point iteration is an efficient way to compute C^j for $j = 2^{m+1} - 2$ where $m = 0, 1, 2, \cdots$. Thus by Lemma 3.3, if 1 is the only eigenvalue of C on the unit circle in the complex plane then $Z_m \to P$ as $m \to \infty$. In such a case, it is notable from Eq. (5.1) that $Z_m \to P$ quadratically as $m \to \infty$, which is a very fast convergence. We illustrate this further with two numerical examples.

Example 5.1. Consider

$$C = \begin{bmatrix} 0.9 & 0.1 \\ 0.025 & 0.975 \end{bmatrix},$$

where 1 is the only eigenvalue of C on the unit circle in the complex plane with multiplicity 1. From Lemma 3.1 (g), one can compute

$$P = \begin{bmatrix} 0.2 & 0.8 \\ 0.2 & 0.8 \end{bmatrix}.$$

In Table 1, we have compiled $\|Z_m - P\|_\infty$:

Table 1: Infinity-norm errors in Example 5.1.

m	$\|Z_m - P\|_\infty$
1	1.23×10^0
2	2.47×10^{-1}
3	2.91×10^{-2}
4	4.06×10^{-4}
5	7.89×10^{-8}
6	6.13×10^{-15}

Example 5.2. Consider

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.4 & 0.3 & 0 & 0.3 \\ 0.1 & 0.5 & 0.2 & 0.2 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where 1 is the only eigenvalue of C on the unit circle in the complex plane, and semisimple with multiplicity 2. From Lemma 3.1 (g), one can compute

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 4/7 & 0 & 0 & 3/7 \\ 27/56 & 0 & 0 & 29/56 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

In Table 2, we have compiled $\|Z_m - P\|_\infty$:

Table 2: Infinity-norm errors in Example 5.2.

m	$\|Z_m - P\|_\infty$
1	5.80×10^0
2	6.78×10^{-3}
3	4.77×10^{-7}
4	2.14×10^{-15}

If C has eigenvalues other than 1 on the unit circle in the complex plane, the convergence of the fixed point iteration (2.1) is not guaranteed; and even if it converges, it may not converge to the fixed point matrix P in Theorem 3.1. We illustrate this through the following two examples.

Example 5.3. Let

$$C = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

The matrix C has -1 as one eigenvalue, and $A = C$. The fixed point iteration converges to I, a trivial fixed point matrix of Eq. (2.1) that gives a trivial solution $X^* = A$. Thus the fixed point iteration fails to find the fixed point matrix P in Theorem 3.1. From Lemma 3.1 (g), we have

$$P = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Example 5.4. Let

$$C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

The matrix C has two complex eigenvalues $-1/2 \pm i\sqrt{3}/2$ on the unit circle in the complex plane, and hence the fixed point iteration does not converge. In fact, we have $Z^m = I$ when m is even and $Z^m = C^{-1} = A$ when m is odd, so the fixed point iteration fails to find a fixed point matrix for Eq. (2.1). Again using Lemma 3.1 (g), one finds

$$P = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

6. Conclusions

We have obtained two nontrivial solutions of the Yang-Baxter-type matrix equation (1.2) when A^{-1} is stochastic. When 1 is the only eigenvalue of C on the unit circle in the complex plane, the fixed point iteration (2.1) provides a fast algorithm to compute the fixed point matrix P, without the necessity to find the left and right (orthonormal) eigenvectors of C associated with the eigenvalue 1. When C has at least one eigenvalue on the unit circle in the complex plane other than 1, the fixed point iteration may not converge — and even if it converges, it may not locate P. In such cases, Lemma 3.1 (b) or (g) can be used to compute P. Of course, another nontrivial solution $A - P$ is available once P is available. A residual question is whether there are nontrivial solutions other than those given in this paper, and *inter alia* we intend to consider this. Indeed, our future goal is to find all relevant solutions and the structure of the solution manifold.

References

[1] C. Cullen, *Matrices and Linear Transformations*, Dover, 1990.

[2] J. Ding and N. Rhee, When a matrix and its inverse are stochastic, *College Math. J.*, to appear.

[3] J. Ding and A. Zhou, *Nonnegative Matrices, Positive Operators, and Applications*, World Scientific, 2009.

[4] F. Felix, *Nonlinear Equations, Quantum Groups and Duality Theorems: A Primer on the Yang-Baxter Equation*, VDM Verlag, 2009.

[5] C. He, B. Meini, and N. Rhee, A shifted cyclic reduction algorithm for quasi-birth-death problems, *SIAM J. Matrix Anal. Appl.* 23 (2002), 673-691.

[6] C. He, B. Meini, N. Rhee, and K. Sohraby, A quadratically convergent Bernoulli-like algorithm for solving matrix polynomial equation in Markov chains, *Elec. Trans. Numer. Anal.* 17 (2004), 151-167.

[7] C. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, 2000.

[8] N. Rhee, Note on functional iteration technique for M/G/1 type Markov chains, *Linear Algb. Its Appl.* 432 (2010), 1042-1048.