Available online at www.sciencedirect.com
ScienceDirect
Linear Algebra and its Applications 424 (2007) 118–131
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

Transition matrices for well-conditioned Markov chains

S.J. Kirkland [a,1], Michael Neumann [b,*], Jianhong Xu [c]

[a] Department of Mathematics and Statistics, University of Regina, Regina, Saskatchewan, Canada S4S 0A2
[b] Department of Mathematics, University of Connecticut, Storrs, CT 06269-3009, USA
[c] Department of Mathematics, Southern Illinois University at Carbondale, IL 62901-4408, USA

Received 16 January 2006; accepted 4 June 2006
Available online 9 August 2006
Submitted by M. Tsatsomeros

Abstract

Let $T \in \mathbb{R}^{n\times n}$ be an irreducible stochastic matrix with stationary distribution vector π. Set $A = I - T$, and define the quantity $\kappa_3(T) \equiv \frac{1}{2}\max_{j=1,\ldots,n} \pi_j \|A_j^{-1}\|_\infty$, where A_j, $j = 1, \ldots, n$, are the $(n-1)\times(n-1)$ principal submatrices of A obtained by deleting the jth row and column of A. Results of Cho and Meyer, and of Kirkland show that κ_3 provides a sensitive measure of the conditioning of π under perturbation of T. Moreover, it is known that $\kappa_3(T) \geqslant \frac{n-1}{2n}$.

In this paper, we investigate the class of irreducible stochastic matrices T of order n such that $\kappa_3(T) = \frac{n-1}{2n}$, for such matrices correspond to Markov chains with desirable conditioning properties. We identify some restrictions on the zero–nonzero patterns of such matrices, and construct several infinite classes of matrices for which κ_3 is as small as possible.

AMS classification: 15A51; 15A18; 65F35; 60J10

Keywords: Markov chain; Stationary distribution; Stochastic matrix; Doubly stochastic matrix; Group inverse; Condition number

1. Introduction

Consider a finite, ergodic, homogeneous, Markov chain with transition matrix $T \in \mathbb{R}^{n\times n}$. For such a chain, its *stationary distribution vector* is the unique positive vector $\pi \in \mathbb{R}^n$ satisfying that

* Corresponding author. Tel.: +1 860 486 1290; fax: +1 860 486 4238.
E-mail address: neumann@math.uconn.edu (M. Neumann).
1 NSERC Grant No. OGP0138251.

0024-3795/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2006.06.003

$\pi^{\mathrm{T}}T = \pi^{\mathrm{T}}$ and $\|\pi\|_1 = 1$. The stationary distribution vector is a key parameter of the chain since it gives the long-term probabilities for the chain to be in each of the various states.

There is a good deal of interest in the literature on the question of the sensitivity of π to perturbations in the transition matrix T, see, for example, [15,8,16,6,7,12]. Specifically, let $\widetilde{T} = T + E$ be the transition matrix of another finite, irreducible, homogeneous, Markov chain with stationary distribution vector $\tilde{\pi}$. The problem is then to find an upper bound on the difference between π and $\tilde{\pi}$, measured under some suitable vector norm.

Perturbation bounds of that type are typically of the form

$$\|\pi - \tilde{\pi}\|_p \leqslant \kappa_l \|E\|_q, \tag{1.1}$$

where $(p, q) = (\infty, \infty)$ or $(1, \infty)$ and where κ_l is some scalar depending on T. Such a κ_l is known as a *condition number* for the chain, and an excellent survey and comparison of various condition numbers can be found in Cho and Meyer [7].

Among the condition numbers that Cho and Meyer discuss in [7] is the following:

$$\kappa_3(T) := \frac{1}{2} \max_{1 \leqslant j \leqslant n} \pi_j \|A_j^{-1}\|_\infty, \tag{1.2}$$

which satisfies $\|\pi - \tilde{\pi}\|_\infty \leqslant \kappa_3 \|E\|_\infty$ whenever T and $\widetilde{T} = T + E$ are irreducible and stochastic.

The comparisons between condition numbers made in [7], coupled with one of the main results in [12], show that κ_3 is the smallest among the eight condition numbers surveyed in [7]. Further, it is shown in [13] that for all sufficiently small $\epsilon > 0$, there is a perturbing matrix E such that $\|E\|_\infty < \epsilon$, $\widetilde{T} = T + E$ is irreducible and stochastic, and $\|\pi - \tilde{\pi}\|_\infty \geqslant \frac{1}{2}\kappa_3 \|E\|_\infty$. Thus κ_3 provides a tight measure of the conditioning of the stationary distribution under perturbation of T.

The paper [12] also provides a lower bound on κ_3, which we state (in somewhat modified form) in the following theorem.

Theorem 1 [12, Theorem 2.9]. *Let $T \in \mathbb{R}^{n \times n}$ be an irreducible stochastic matrix. Let $\pi = [\pi_1, \ldots, \pi_n]^{\mathrm{T}}$ be the stationary distribution vector for T. Put $A = I - T$. Then*

$$\kappa_3(T) = \frac{1}{2} \max_{1 \leqslant j \leqslant n} \pi_j \|A_j^{-1}\|_\infty \geqslant \frac{n-1}{2n}, \tag{1.3}$$

where, to recall, A_j, $j = 1, \ldots, n$, are the $(n-1) \times (n-1)$ principal submatrices of A obtained by deleting the jth row and column of A. Equality holds in (1.3) if and only if each of the following holds:

(i) *T is a doubly stochastic matrix with zero diagonal.*
(ii) *$\|A_j^{-1}\|_\infty = n - 1$, for each $j = 1, \ldots, n$.*
(iii) *If $i \neq j$ are indices such that $t_{j,i} > 0$, then the entry that corresponds to the index i of $A_j^{-1}e$ is equal to $n - 1$, where $e \in \mathbb{R}^{n-1}$ is the all ones vector.*

Theorem 1 shows that there are limits on how well-conditioned a Markov chain can be. For instance, picking up on the result in [13] quoted above, we see that for any irreducible stochastic matrix T of order n with stationary vector π^{T}, and all sufficiently small $\epsilon > 0$, there is a perturbing matrix E such that $\|E\|_\infty < \epsilon$, $\widetilde{T} = T + E$ is irreducible and stochastic, and $\|\pi - \tilde{\pi}\|_\infty \geqslant \frac{n-1}{4n} \|E\|_\infty$.

In light of that fact, it is natural to wonder which irreducible stochastic matrices T can yield equality in the lower bound (1.3), for such matrices will correspond to Markov chains having optimal conditioning properties. In principal, conditions (i)–(iii) of Theorem 1 characterize the Markov chains that minimize κ_3, but the conditions themselves impart little intuition as to the nature of the transition matrix, and may be tedious to check for a given stochastic matrix T.

It is straightforward to check (see [12]) that if T is a permutation matrix whose directed graph is an n-cycle, or if $T = \frac{1}{n-1}(J - I)$, where J is the all ones matrix of the appropriate size, then equality holds in (1.3). Indeed, it is conjectured in [12] (erroneously, as we shall show) that those classes of examples are the only ones to minimize κ_3.

In this paper we are concerned with constructing classes of irreducible stochastic matrices attaining equality in (1.3). As a byproduct of our investigation, we will refute the conjecture mentioned above.

In Section 2 we develop our main results. In particular, we provide a useful recasting of the equality conditions of Theorem 1. We then show that if an irreducible stochastic matrix T has either a row or a column having all off-diagonal entries positive, and if T yields equality in (1.3), then necessarily $T = \frac{1}{n-1}(J - I)$. We also show that the only symmetric, irreducible, stochastic matrix that yields equality in (1.3) is $T = \frac{1}{n-1}(J - I)$. These results help to narrow the field of search for matrices yielding equality in (1.3), since they imply that if $T \neq \frac{1}{n-1}(J - I)$ yields equality in that bound, then there are restrictions on the pattern of the entries in T.

Section 3 provides some constructions of classes of stochastic matrices yielding equality in (1.3). These constructions include both primitive stochastic matrices and irreducible periodic stochastic matrices, and suggest that the class of matrices yielding equality in (1.3) may be quite rich. In particular, we characterize the irreducible stochastic matrices yielding equality in (1.3), and having one off-diagonal zero entry in each row and column.

The techniques of Sections 2 and 3 rely in part on the notion of the *mean first passage time from state $\mathscr{S}_i$ to state $\mathscr{S}_j$* of an ergodic Markov chain with transition matrix T, that is,

$$m_{i,j} = \sum_{k=1}^{\infty} k Pr\{X_k = \mathscr{S}_j,\ X_\mu \neq \mathscr{S}_j,\ \mu = 1, \ldots, k-1 | X_0 = \mathscr{S}_i\}. \tag{1.4}$$

Standard results on Markov chains (see [10], for example) assert that

$$m^{(j)} := [m_{1,j}, \ldots, m_{j-1,j}, m_{j+1,j}, \ldots, m_{n,j}]^{\mathrm{T}} = A_j^{-1} e, \tag{1.5}$$

while $m_{i,i} = \frac{1}{\pi_i}$, $i = 1, \ldots, n$. Thus a discussion of $\|A_j^{-1}\|_\infty$ can be recast in terms of mean first passage times.

Due to the skip of the index i over the column number j of $m_{i,j}$ in (1.5), it will be convenient to adopt the notation that

$$\phi(i,j) = \begin{cases} i, & \text{if } 1 \leqslant i < j, \\ i-1, & \text{if } i > j. \end{cases}$$

Note that $1 \leqslant \phi(i,j) \leqslant n-1$, for all $i, j = 1, \ldots, n$. From (1.5) we see that

$$\|A_j^{-1}\|_\infty = \max_{1 \leqslant i \leqslant n} m^{(j)}_{\phi(i,j)} \tag{1.6}$$

and that

$$\max_{1 \leqslant j \leqslant n} \|A_j^{-1}\|_\infty = \max_{1 \leqslant j \leqslant n} \max_{1 \leqslant i \leqslant n} m^{(j)}_{\phi(i,j)}.$$

We mention that the quantity $\max_{1 \leqslant j \leqslant n} \max_{1 \leqslant i \leqslant n} m^{(j)}_{\phi(i,j)}$ has been used in various applications of mean first passage times. In particular, it has been applied to the quantification of the small-world properties of ring networks, see, for example, Catral et al. [5].

2. Equality in (1.3) and mean first passage times

We begin with a few useful preliminary results concerning the mean first passage matrix, i.e. the $n \times n$ matrix $M = [m_{i,j}]$ consisting of the mean first passage times.

Lemma 2.1 [14, Eq. (3.3)]. *Let $T \in \mathbb{R}^{n\times n}$ be irreducible and stochastic. Let M be its mean first passage matrix. Then M satisfies the matrix equation*

$$(I - T)M = J - TM_{\text{diag}}, \tag{2.1}$$

where M_{diag} is the diagonal matrix obtained from M by setting its off-diagonal entries to zero.

Remark 2.2. From (2.1) we easily find that for any $1 \leqslant i, j \leqslant n$,

$$m_{i,j} = 1 + \sum_{1\leqslant k\leqslant n; k\neq j} t_{i,k} m_{k,j}. \tag{2.2}$$

Observe that in the special case that T is an irreducible doubly stochastic matrix, taking $i = j$ in (2.2), we find that $n = m_{j,j} = 1 + \sum_{k\neq j} t_{j,k} m_{k,j} \leqslant (1 - t_{jj}) \max_{k\neq j} m_{k,j}$, so that

$$\max_{1\leqslant k\leqslant n} m^{(j)}_{\phi(k,j)} = \|A_j^{-1}\|_\infty \geqslant \frac{n-1}{1-t_{j,j}} \tag{2.3}$$

(note that $t_{j,j} < 1$ since T is irreducible). Observe that equality is attainable in that lower bound on $\|A_j^{-1}\|_\infty$; for instance, if $T = J/n \in \mathbb{R}^{n\times n}$, then $\|A_j^{-1}\|_\infty = n = (n-1)/(1-t_{j,j})$, for each $j = 1, \ldots, n$.

Lemma 2.3 [9, Theorem 2.4]. *Let $T \in \mathbb{R}^{n\times n}$ be an irreducible stochastic matrix. Let $M = [m_{i,j}]$ and $\pi = [\pi_1, \ldots, \pi_n]^{\mathrm{T}}$ be the mean first passage matrix and the stationary distribution vector for T, respectively. Then there exists a constant $K > 0$ such that for any $1 \leqslant i \leqslant n$,*

$$\sum_{k=1}^{n} \pi_k m_{i,k} = 1 + \operatorname{trace}(A^{\#}) = K. \tag{2.4}$$

The constant K is called the *Kemeny constant*. Here $A^{\#}$ is the group inverse[2] of $A = I - T$. In particular, when T is irreducible and doubly stochastic, (2.4) reduces to

$$\sum_{k=1}^{n} m_{i,k} = nK, \quad i = 1, \ldots, n. \tag{2.5}$$

Lemma 2.4 [1]. *Let $A \in \mathbb{R}^{n\times n}$, with $\det(A) = 0$, be positive semidefinite and let $A^\dagger$ be the Moore–Penrose inverse of A. Then in the ordering of the cone of positive semidefinite matrices, i.e. the Lowner ordering,*

$$A \circ A^\dagger \geqslant P \circ P, \tag{2.6}$$

where P is the orthogonal projection onto $R(A)$ and where $\circ$ denotes the entrywise (i.e. Hadamard) product of matrices.

[2] See the books by Ben-Israel and Greville [2] and Campbell and Meyer [4].

It is known that when $A = I - T$, where $T \in \mathbb{R}^{n\times n}$ is irreducible and stochastic, P is given by $I - e\pi^{\mathrm{T}}$, where π is the stationary distribution vector for T. Thus in particular, $P = I - J/n$ when T is doubly stochastic. Note that if T is also symmetric, then $A^{\dagger} = A^{\#}$, the group inverse of A.

Lemma 2.5 [18, Theorem 2.2]. *Let $T \in \mathbb{R}^{n\times n}$ be an irreducible doubly stochastic matrix. Put $A = I - T$. Let $M = [m_{i,j}]$ be the mean first passage matrix for T. Define the matrix $L = [\ell_{i,j}] \in \mathbb{R}^{(n-1)\times(n-1)}$ by*

$$\ell_{i,j} = m_{i,n} + m_{n,j} - m_{i,j}, \quad \text{for } i \neq j; \;\; \ell_{i,i} = m_{i,n} + m_{n,i}. \tag{2.7}$$

Then

$$nA_n^{-1} = L. \tag{2.8}$$

In our first result we show that condition (ii) in the case of equality in Theorem 1, coupled with the fact that T is doubly stochastic, is sufficient to imply that T has zero diagonal.

Theorem 2. *Let $T = [t_{i,j}] \in \mathbb{R}^{n\times n}$ be irreducible and doubly stochastic. If $\|A_j^{-1}\|_\infty \leqslant n - 1$, for each $j = 1, \ldots, n$, then $t_{j,j} = 0$, for each $j = 1, \ldots, n$.*

Proof. From the hypothesis, we find that $\max_{1\leqslant j\leqslant n} \pi_j \|A_j^{-1}\|_\infty \leqslant (n-1)/n$. That inequality, coupled with (1.3), then yields that $\max_{1\leqslant j\leqslant n} \pi_j \|A_j^{-1}\|_\infty = (n-1)/n$. From the equality characterization of Theorem 1, we see that $t_{j,j} = 0$, for each $j = 1, \ldots, n$. □

Next, we show that condition (iii) of Theorem 1 can be deduced from condition (ii) and the fact that T is doubly stochastic.

Theorem 3. *Let $T = [t_{i,j}] \in \mathbb{R}^{n\times n}$ be irreducible and doubly stochastic. Then*:

(a) *if $\|A_j^{-1}\|_\infty \leqslant n - 1$, for each $j = 1, \ldots, n$, then for any $i \neq j$, $t_{j,i} > 0$ implies that $(A_j^{-1}e)_{\phi(i,j)}$ is equal to $n - 1$.*
(b) *$\|A_j^{-1}\|_\infty \leqslant n - 1$, for all $j = 1, \ldots, n$, if and only if $\|A_j^{-1}\|_\infty = n - 1$, $j = 1, \ldots, n$.*

Proof. (a) Suppose that T is doubly stochastic and that $\|A_j^{-1}\|_\infty \leqslant n - 1$, for each $j = 1, \ldots, n$. As in the proof of Theorem 2, those conditions imply that equality holds in (1.3). The conclusion now follows from condition (iii) of Theorem 1.

(b) Suppose that T is doubly stochastic. If $\|A_j^{-1}\|_\infty \leqslant n - 1$, for each $j = 1, \ldots, n$, then equality holds in (1.3), so that $\|A_j^{-1}\|_\infty = n - 1$, for each $j = 1, \ldots, n$, by condition (ii) of Theorem 1. The converse implication is immediate. □

Theorems 2 and 3 lead us to the following recasting of Theorem 1.

Theorem 4. *Let $T \in \mathbb{R}^{n\times n}$ be an irreducible stochastic matrix and put $A = I - T$. Let $\pi = [\pi_1, \ldots, \pi_n]^{\mathrm{T}}$ be the stationary distribution vector for T. Then*

$$\max_{1\leqslant j\leqslant n} \pi_j \|A_j^{-1}\|_\infty \geqslant \frac{(n-1)}{n}. \tag{2.9}$$

Equality holds in (2.9) (*equivalently*, *in* (1.3)) *if and only if*

(i) T *is a doubly stochastic matrix*
and
(ii) $\|A_j^{-1}\|_\infty = n-1$, *for each* $j = 1, \ldots, n$.
In addition, *conditions* (i) *and* (ii) *yield that*:
(iii) T *has a zero diagonal*
and
(iv) *for any* $i \neq j$, $t_{j,i} > 0$ *implies that* $m^{(j)}_{\phi(i,j)} = (A_j^{-1}e)_{\phi(i,j)} = n-1$.

Our next two results discuss equality in (1.3) under certain pattern restrictions on T.

Theorem 5. *Let* $T = [t_{i,j}] \in \mathbb{R}^{n\times n}$ *be an irreducible*, *stochastic matrix that yields equality in* (1.3). *Suppose that for some* $1 \leqslant j \leqslant n$, *the off-diagonal entries in the* j*th column of* T *are all positive*. *Then* $T = \frac{1}{n-1}(J - I)$.

Proof. By Theorem 3 we know that all the off-diagonal entries in the jth row of M are equal to $n-1$ and that further, as T is doubly stochastic, the jth diagonal entry of M is equal to n. It follows that $\sum_{k=1}^n m_{j,k} = n + (n-1)^2$. Using Lemma 2.3 concerning the Kemeny constant K, we conclude that for any $1 \leqslant i \leqslant n$,

$$\sum_{k=1}^n m_{i,k} = n + (n-1)^2. \tag{2.10}$$

Since, for any index i we have that $m_{i,i} = n$ and $m_{i,j} \leqslant n-1$, $j \neq i$, it follows from (2.10) that we must have that $m_{i,j} = n-1$ whenever $i \neq j$. Hence the mean first passage matrix of T is given by $M = (n-1)J + I$, which, according to (2.1), leads to $T = \frac{1}{n-1}(J - I)$. □

Theorem 6. *Let* $T = [t_{i,j}] \in \mathbb{R}^{n\times n}$ *be an irreducible*, *stochastic matrix that yields equality in* (1.3). *Suppose that for some* $1 \leqslant i \leqslant n$, *the off-diagonal entries in the* i*th row of* T *are all positive*. *Then* $T = \frac{1}{n-1}(J - I)$.

Proof. By part (iv) of Theorem 4, $m_{j,i} = n-1$, for all $j \neq i$. Moreover, from (2.2) we have for any index $j \neq i$,

$$m_{j,i} = 1 + \sum_{1\leqslant k\leqslant n;\, k\neq i} t_{j,k} m_{k,i},$$

from which it follows that

$$n-1 = 1 + (n-1) \sum_{1\leqslant k\leqslant n;\, k\neq i} t_{j,k} = 1 + (n-1)(1 - t_{j,i}).$$

Thus

$$t_{j,i} = \frac{1}{n-1},$$

for any $j \neq i$, so that all off-diagonal entries in the ith column of T are positive. By Theorem 5, we conclude that $T = \frac{1}{n-1}(J - I)$. □

We note that Theorems 5 and 6 generalize a result in [12], which asserts that if the stochastic matrix T of order n has all off-diagonal entries positive and yields equality in (1.3), then necessarily $T = \frac{1}{n-1}(J - I)$.

Next, we characterize the irreducible symmetric stochastic matrices that yield equality in (1.3).

Theorem 7. *Let $T \in \mathbb{R}^{n\times n}$ be a symmetric, irreducible, and stochastic matrix yielding equality in (1.3). Then $T = \frac{1}{n-1}(J - I)$.*

Proof. Put $A = I - T$. Clearly A is positive semidefinite. As mentioned earlier, $A^{\dagger} = A^{\#}$ and the orthogonal projection P of $\mathbb{R}^{n\times n}$ onto $R(A)$ is given by $I - J/n$. By Lemma 2.4, we have that

$$A \circ A^{\#} \geqslant (I - J/n) \circ (I - J/n), \tag{2.11}$$

where the inequality is in the ordering of the cone of positive semidefinite matrices.

Since the matrix formed from the difference between the left and right sides of (2.11) is positive semidefinite, its diagonal entries are necessarily nonnegative. Thus, by inspection, for all $1 \leqslant j \leqslant n$, we have that:

$$a^{\#}_{j,j} \geqslant \frac{(n-1)^2}{n^2},$$

which implies the Kemeny constant

$$K = 1 + \mathrm{tr}(A^{\#}) \geqslant 1 + \frac{(n-1)^2}{n}.$$

On the other hand, the condition that $\|A_j^{-1}\|_\infty \leqslant n - 1$, for all $1 \leqslant j \leqslant n$, implies that

$$K \leqslant 1 + \frac{(n-1)^2}{n}.$$

Thus K must be exactly $1 + \frac{(n-1)^2}{n}$. The conclusion now follows from an argument similar to that in the proof of Theorem 5. □

Next, we investigate the structure of an $(n - 1) \times (n - 1)$ reducible submatrix of $A = I - T$ when $T \in \mathbb{R}^{n\times n}$ yields equality in (1.3).

Theorem 8. *Let $T \in \mathbb{R}^{n\times n}$ be an irreducible stochastic matrix yielding equality in (1.3). Suppose that T_n is reducible and, without loss of generality, assume that T_n has the form*

$$T_n = \begin{bmatrix} B_{1,1} & B_{1,2} \\ 0^{(n-k-1,k)} & B_{2,2} \end{bmatrix}, \tag{2.12}$$

where $B_{1,1} \in \mathbb{R}^{k\times k}$, for some k, $1 \leqslant k < n - 1$, and $B_{2,2} \in \mathbb{R}^{(n-k-1)\times(n-k-1)}$. Then T has the following form:

$$T = \begin{bmatrix} B_{1,1} & B_{1,2} & 0^{(k,1)} \\ 0^{(n-k-1,k)} & B_{2,2} & B_{2,3} \\ B_{3,1} & 0^{(1,n-k-1)} & 0^{(1,1)} \end{bmatrix}, \tag{2.13}$$

where for any two positive integers μ and ν, $0^{(\mu,\nu)}$ denotes the zero matrix in $\mathbb{R}^{\mu\times\nu}$.

Proof. Let $A = I - T$. From (2.12) we know that A_n^{-1} is of the form

$$A_n^{-1} = \begin{bmatrix} * & * \\ 0^{(n-k-1,k)} & * \end{bmatrix}$$

Hence, by (2.7) and (2.8),

$$m_{i,n} + m_{n,j} - m_{i,j} = 0, \quad \text{for } i = k+1, \ldots, n-1, \text{ and for } j = 1, \ldots, k. \tag{2.14}$$

But then, as $0 < m_{i,j} \leqslant n-1$, it follows that

$$n - 1 > \begin{cases} m_{i,n}, & \text{for } k+1 \leqslant i \leqslant n-1, \\ m_{n,j}, & \text{for } 1 \leqslant j \leqslant k. \end{cases}$$

We find from Theorem 3 that

$$0 = \begin{cases} t_{n,i}, & \text{for } k+1 \leqslant i \leqslant n-1, \\ t_{j,n}, & \text{for } 1 \leqslant j \leqslant k. \end{cases}$$

The conclusion now follows. □

3. Constructions for optimally-conditioned Markov chains

Theorems 5 and 6 show that if the irreducible stochastic matrix $T \neq \frac{1}{n-1}(J - I)$ is to yield equality in (1.3), then necessarily T must have at least one off-diagonal zero entry in each row and column. Our next result characterizes those matrices yielding equality in (1.3) that have exactly one off-diagonal zero entry in each row and column.

Theorem 9. *T is an $n \times n$ irreducible stochastic matrix having one off-diagonal zero entry in each row and each column and yielding equality in* (1.3) *if and only if*

$$T = \frac{1}{n-1+x} J + I - \frac{n(n-1)^2}{(n-1)^3 + x^3} \left[I - \frac{x}{n-1} P + \frac{x^2}{(n-1)^2} P^2 \right], \tag{3.1}$$

where P is permutationally similar to a direct sum of 3×3 cyclic permutation matrices, and where x is the positive root of the equation $x^3 + nx^2 - (n-1)^2 = 0$.

Proof. First, suppose that (3.1) is satisfied. From the fact that $x^3 + nx^2 - (n-1)^2 = 0$, it follows that $\frac{1}{n-1+x} + 1 - \frac{n}{n-1} \frac{(n-1)^3}{(n-1)^3+x^3} = 0$, and that $\frac{1}{n-1+x} - \frac{n}{n-1} \frac{(n-1)^3}{(n-1)^3+x^3} \frac{x^2}{(n-1)^2} = 0$, so that T has off-diagonal zeros precisely in the positions corresponding to the non-zero entries of P^2. In particular, T is irreducible and nonnegative, and a straightforward computation shows that its row and column sums are all 1.

Further, it is readily checked that the matrix $M = (n-1)J + I - xP$ is the (unique) matrix satisfying (2.1), and thus M is the mean first passage matrix for T. Since T is doubly stochastic, and the maximum off-diagonal entry in each row of M is $n-1$, we see from Theorem 4 that T yields equality in (1.3).

Now suppose that T is an irreducible stochastic matrix with one off-diagonal zero in each row and column, and that T yields equality in (1.3). Necessarily T is doubly stochastic. Let Q^{T} denote the permutation matrix with ones in the positions where T has off-diagonal zeros. Let M be the mean first passage matrix for T. Since equality holds in (1.3), the only off-diagonal positions where an entry of M can be less than $n-1$ correspond to positions where Q has ones. Further, since

M has constant row sums, it follows that M is of the form $M = (n-1)J + I - yQ$, for some $n-1 > y > 0$. Substituting that expression into (2.1) and solving for T, we find that necessarily $T = ((n-1)I + yQ)^{-1}(J - I + yQ) = \frac{1}{n-1+y}J + I - \frac{n}{n-1}(I + \frac{y}{n-1}Q)^{-1}$ (observe that the appropriate inverses exist since $\frac{|y|}{n-1} < 1$).

Consider an index i that is on a cycle of length l in the directed graph of Q. We find (by expanding the inverse as a geometric series) that the ith diagonal entry of $(I + \frac{y}{n-1}Q)^{-1}$ is given by $\frac{(n-1)^l}{(n-1)^l-(-y)^l}$. Since T has constant diagonal, it follows that for some integer k, each vertex in the directed graph of Q is on a cycle of length k. We now deduce that the permutation matrix Q is permutationally similar to a direct sum of $k \times k$ cyclic permutation matrices.

Since $Q^k = I$, a straightforward computation now yields the fact that

$$T = \frac{1}{n-1+y}J + I - \frac{n}{n-1}\frac{(n-1)^k}{(n-1)^k-(-y)^k} \times \left[I - \frac{y}{n-1}Q + \frac{y^2}{(n-1)^2}Q^2 + \cdots + (-1)^{k-1}\frac{y^{k-1}}{(n-1)^{k-1}}Q^{k-1}\right]. \quad (3.2)$$

Note that because Q is permutationally similar to a direct sum of k-cyclic permutation matrices, no pair of the powers $I, Q, Q^2, \ldots, Q^{k-1}$ contains a 1 in a common position.

Recall that the off-diagonal zeros of T correspond to the nonzero entries of $Q^{\mathrm{T}} = Q^{k-1}$. From (3.2), we see that the off-diagonal entries of T of smallest size correspond to the positions where Q^2 has nonzero entries. Thus we find that in fact $k-1 = 2$, i.e. $k = 3$. Further, in order for T to have zero entries in those positions, we must also have

$$\frac{1}{n-1+y} = \frac{n}{n-1}\frac{(n-1)^3}{(n-1)^k-(-y)^3}\frac{y^2}{(n-1)^2}. \quad (3.3)$$

This last is readily seen to simplify as $(n-1)^2 = ny^2 + y^3$. Eq. (3.1) is now easily established. □

Remark 3.1. Note that in Theorem 9, necessarily n must be divisible by 3 if there is to be a matrix T satisfying (3.1).

Our next two examples provide primitive circulant matrices yielding equality in (1.3).

Example 1. Consider the 5×5 circulant, doubly stochastic matrix

$$T = \begin{bmatrix} 0 & 0 & 0 & a & 1-a \\ 1-a & 0 & 0 & 0 & a \\ a & 1-a & 0 & 0 & 0 \\ 0 & a & 1-a & 0 & 0 \\ 0 & 0 & a & 1-a & 0 \end{bmatrix}, \quad (3.4)$$

where $0 < a < 1$. Put $A = I - T$ and note that A is also a circulant matrix. Consequently, $\|A_1^{-1}\|_\infty = \|A_j^{-1}\|_\infty$, $j = 2, \ldots, 5$, so it suffices to consider $\|A_1^{-1}\|_\infty$. On computing A_1^{-1} we find that

$$A_1^{-1}e = \frac{1}{d}\begin{bmatrix} 1+3a-2a^2+a^3 \\ 2+a-4a^2+2a^3 \\ 3-a-a^2+3a^3 \\ 4-3a+2a^2-a^3 \end{bmatrix},$$

where $d = 1 - a + a^2 - a^3 + a^4$. Taking

$$a = \frac{1}{4}\left[1 - \frac{5^{2/3}}{(27 + 12\sqrt{6})^{1/3}} + \frac{(45 + 20\sqrt{6})^{1/3}}{3^{2/3}}\right] \approx 0.6058,$$

we obtain that

$$A_1^{-1}e = \begin{bmatrix} 3.42332 \\ 2.34937 \\ 4.00000 \\ 4.00000 \end{bmatrix} \Rightarrow \|A_1^{-1}\|_\infty = 4.$$

It now follows from Theorem 4 that T yields equality in (1.3).

Example 2. Consider a 7×7 circulant doubly stochastic matrix

$$T = \begin{bmatrix} 0 & 0 & 0 & 0 & a & b & c \\ c & 0 & 0 & 0 & 0 & a & b \\ b & c & 0 & 0 & 0 & 0 & a \\ a & b & c & 0 & 0 & 0 & 0 \\ 0 & a & b & c & 0 & 0 & 0 \\ 0 & 0 & a & b & c & 0 & 0 \\ 0 & 0 & 0 & a & b & c & 0 \end{bmatrix},$$

where $0 < a < 1$, $0 < b < 1$, $a + b < 1$, and $c = 1 - a - b$. Put $A = I - T$. Again, it suffices to consider $\|A_1^{-1}\|_\infty$. It can be checked that the vector $A_1^{-1}e \in \mathbb{R}^6$ is given by

$$\frac{1}{d}\begin{bmatrix} 1 - 3a^5 + a^4(10 - 11b) + 5b - 4b^2 + 3b^3 - 2b^4 + b^5 + a^3(-4 + 13b - 11b^2) \\ - 2a^2(1 - 4b + 3b^2 + b^3) + a(3 - 9b + 11b^2 - 9b^3 + 3b^4) \\ \\ 2 + a^5 + 3b - 8b^2 + 6b^3 - 4b^4 + 2b^5 - a^4(1 + b) - a^3(-6 + 2b + b^2) \\ + a^2(-11 + 23b - 19b^2 + 3b^3) + a(6 - 18b + 22b^2 - 18b^3 + 6b^4) \\ \\ 3 - 2a^5 + a^4(2 - 5b) + b - 5b^2 + 9b^3 - 6b^4 + 3b^5 + a^3(9 - 10b + 2b^2) \\ + a^2(-13 + 38b - 32b^2 + 8b^3) + a(2 - 20b + 33b^2 - 27b^3 + 9b^4) \\ \\ 4 + 2a^5 - b - 2b^2 + 5b^3 - 8b^4 + 4b^5 - 2a^4(1 + b) + a^3(12 - 18b + 5b^2) \\ + a^2(-8 + 39b - 45b^2 + 13b^3) + 2a(-1 - 4b + 15b^2 - 18b^3 + 6b^4) \\ \\ 5 - a^5 - 3b + b^2 + b^3 - 3b^4 + 5b^5 + a^4(1 + b) + a^3(1 - 12b + 8b^2) \\ + 2a^2(2 + 6b - 15b^2 + 9b^3) + a(-6 + 4b + 6b^2 - 24b^3 + 15b^4) \\ \\ 6 - 4a^5 + a^4(18 - 17b) - 5b + 4b^2 - 3b^3 + 2b^4 - b^5 + a^3(-24 + 50b - 31b^2) \\ + a^2(16 - 36b + 48b^2 - 26b^3) - 2a(5 - 8b + 9b^2 - 8b^3 + 5b^4) \end{bmatrix},$$

where

$$\begin{aligned} d = {} & 1 + a^6 - b + b^2 - b^3 + b^4 - b^5 + b^6 + a^5(-4 + 3b) + a^4(9 - 17b + 9b^2) \\ & + a^3(-8 + 25b - 31b^2 + 13b^3) + a^2(4 - 12b + 24b^2 - 26b^3 + 11b^4) \\ & + a(-2 + 4b - 6b^2 + 8b^3 - 10b^4 + 5b^5). \end{aligned}$$

Table 1
Coefficients c_i

$c_0 = -223{,}598{,}525{,}562{,}865{,}964$	$c_{11} = 480{,}792{,}853{,}550{,}130{,}281{,}169$
$c_1 = 857{,}421{,}319{,}719{,}767{,}996$	$c_{12} = -171{,}613{,}060{,}237{,}094{,}290{,}698$
$c_2 = -1{,}598{,}231{,}725{,}315{,}339{,}706$	$c_{13} = 3{,}846{,}178{,}988{,}995{,}107{,}181$
$c_3 = 178{,}549{,}507{,}128{,}775{,}107$	$c_{14} = 53{,}373{,}135{,}248{,}089{,}147{,}589$
$c_4 = 5{,}515{,}703{,}520{,}210{,}685{,}479$	$c_{15} = -32{,}867{,}818{,}557{,}984{,}624{,}140$
$c_5 = -10{,}988{,}929{,}087{,}694{,}835{,}990$	$c_{16} = -176{,}616{,}259{,}260{,}589{,}708{,}410$
$c_6 = 28{,}255{,}588{,}903{,}918{,}568{,}435$	$c_{17} = 525{,}655{,}742{,}514{,}692{,}341{,}968$
$c_7 = -53{,}149{,}536{,}722{,}526{,}281{,}625$	$c_{18} = -486{,}308{,}557{,}070{,}340{,}966{,}648$
$c_8 = 32{,}010{,}397{,}238{,}264{,}856{,}144$	$c_{19} = 182{,}601{,}989{,}324{,}789{,}126{,}688$
$c_9 = 178{,}214{,}490{,}985{,}905{,}048{,}116$	$c_{20} = -32{,}101{,}767{,}892{,}913{,}675{,}616$
$c_{10} = -525{,}834{,}292{,}021{,}821{,}117{,}075$	

Table 2
Coefficients s_i

$s_0 = -19{,}343{,}511{,}880{,}075{,}782$
$s_7 = 19{,}343{,}511{,}880{,}075{,}782$

We use Mathematica's symbolic manipulation capability to search for feasible values of the parameters a and b for which $\|A_1^{-1}\|_\infty \leqslant 6$ will hold. We set the last three entries of $A_1^{-1}e$ to 6, thus obtaining three equations in the variables a and b. We then first used Mathematica to reduce these three equations to two and subsequently solve them with Mathematica's root-finding mechanism. Mathematica produced pseudo-closed formulas for a and b. Specifically, a can be expressed as the smaller of the two real roots of the 10th degree polynomial

$$p(x) = 1 - 3x + 3x^2 + 6x^3 - 21x^4 + 21x^5 - 91x^6 + 132x^7 + 108x^8 - 864x^9 + 1296x^{10}$$

which gives that $a \approx 0.46624659695529764516590349787468590700884086202551$, while b can be expressed as a rational function in a of the form

$$\begin{aligned} b &= \frac{c_0 + c_1 a + c_2 a^2 + \cdots + c_{20} a^{20}}{s_0 + s_7 a^7} \\ &\approx 0.28514380149522220421751774962379305227592250250294, \end{aligned}$$

with the coefficients c_i and s_i determined by Mathematica to be the (large) integers given in Tables 1 and 2. In particular, $s_0 = -s_7$. With these formulas for a and b, we computed A_1^{-1} and obtained that

$$A_1^{-1}e = \begin{bmatrix} 5.50834239070311909630052748499087375570858018717 07 \\ 5.16690638868259922547341106749528352593866477728 96 \\ 3.85521222775621231174653881673084683151072075096 51 \\ 6.00000000000000000000000000000000000000000000000000 \\ 6.00000000000000000000000000000000000000000000000000 \\ 6.00000000000000000000000000000000000000000000000000 \end{bmatrix}.$$

Notice that the first three entries of $A_1^{-1}e$ are strictly less than 6. It now follows that T yields equality in (1.3).

Next, we provide constructions of irreducible, periodic matrices yielding equality in (1.3). We begin with a result of Kirkland [11, Theorem 5] on the mean first passage matrix arising from a periodic Markov chain.

Lemma 3.2. *Let $T \in \mathbb{R}^{n\times n}$ be irreducible, stochastic, and r-periodic, $r \geqslant 2$, in the form*

$$T = \begin{bmatrix} 0 & T_1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & T_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & \cdots & \ddots & T_{r-1} \\ T_r & 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}, \tag{3.5}$$

where the diagonal blocks are square. Let

$$P_1 = T_1 \cdots T_r \in \mathbb{R}^{n_1\times n_1} \tag{3.6}$$

and

$$P_j = T_j \cdots T_r T_1 \cdots T_{j-1} \in \mathbb{R}^{n_j\times n_j}, \quad j = 2, \ldots, r. \tag{3.7}$$

Denote the mean first passage matrix for T by M, partitioned in conformity with T, and denote the (i, j)th block of M by $M^{(i,j)}$, $1 \leqslant i, j \leqslant r$. Let $M^{(P_j)}$ be the mean first passage matrix for P_j. Then:

(i) *For $1 \leqslant j \leqslant r$,*

$$M^{(j,j)} = rM^{(P_j)}. \tag{3.8}$$

(ii) *For $1 \leqslant i < j \leqslant r$,*

$$M^{(i,j)} = rT_i \cdots T_{j-1}\left[M^{(P_j)} - (M^{(P_j)})_{\text{diag}}\right] + (j-i)J. \tag{3.9}$$

(iii) *For $1 \leqslant j < i \leqslant r$,*

$$M^{(i,j)} = rT_i \cdots T_r T_1 \cdots T_{j-1}\left[M^{(P_j)} - (M^{(P_j)})_{\text{diag}}\right] + (r+j-i)J. \tag{3.10}$$

Observe that if T is doubly stochastic, then $T_1, \ldots, T_r$ are all doubly stochastic matrices of the same size. Suppose that $T_j \in \mathbb{R}^{k\times k}$, $j = 1, \ldots, r$. Note too that each $P_j \in \mathbb{R}^{k\times k}$ is doubly stochastic and primitive.

Lemma 3.2 leads to the following result.

Theorem 10. *Let $T \in \mathbb{R}^{n\times n}$ be irreducible, doubly stochastic, and r-periodic, $r \geqslant 2$, as in (3.5), and with $T_j \in \mathbb{R}^{k\times k}$, where $k \geqslant 2$, for $j = 1, \ldots, r$. Suppose that for $j = 1, \ldots, r$, and for $\beta = 1, \ldots, k$,*

$$\|(I - (P_j)_{(\beta)})^{-1}\|_\infty \leqslant k - 1, \tag{3.11}$$

where $(P_j)_{(\beta)}$ is the principal submatrix obtained from P_j by deleting its βth row and column. Then $\|A_j^{-1}\|_\infty = n - 1$, for all $j = 1, \ldots, n$.

Proof. Let M be the mean first passage matrix for T given by (3.5) and partitioned in conformity with T; denote the (i, j)th block of M by $M^{(i,j)}$, $1 \leqslant i, j \leqslant r$. Also denote the mean first passage matrix of P_j by $M^{(P_j)}$, $j = 1, \ldots, r$.

We proceed by considering three possible cases for the indices i and j.

Case (i) $1 \leqslant i = j \leqslant r$. By Lemma 3.2, we have that $M^{(i,i)} = rM^{(P_i)}$, and hence

$$rM^{(P_i)} \leqslant r((k-1)J + I) \leqslant (n-1)J + I.$$

Case (ii) $1 \leqslant i < j \leqslant r$. This time by Lemma 3.2 we have that

$$M^{(i,j)} = rT_i \cdots T_{j-1}\left(M^{(P_j)} - (M^{(P_j)})_{\text{diag}}\right) + (j-i)J.$$

Note that $0 \leqslant M^{(P_j)} - (M^{(P_j)})_{\text{diag}} \leqslant (k-1)J$, $T_i \cdots T_{j-1}J = J$, and $j - i \leqslant r - 1$. Thus we obtain that

$$M^{(i,j)} \leqslant r(k-1)J + (r-1)J = (n-1)J.$$

Case (iii) $1 \leqslant j < i \leqslant r$. Again using Lemma 3.2, we have that

$$M^{(i,j)} = rT_i \cdots T_r T_1 \cdots T_{j-1}(M^{(P_j)} - (M^{(P_j)})_{\text{diag}}) + (r+j-i)J.$$

Similar to part (ii) we arrive at

$$M^{(i,j)} \leqslant r(k-1)J + (r-1)J = (n-1)J.$$

Combining parts (i)–(iii) we conclude that the off-diagonal entries of M are all bounded above by $n-1$. The conclusion now follows from Theorem 3. □

Theorem 10 suggests a method for constructing periodic matrices yielding equality in (1.3) by ensuring that each of the products P_j satisfies (3.11). The following examples illustrate that method.

Example 3. Fix natural numbers $k \geqslant 3$ and $r \geqslant 2$, and let

$$T = \begin{bmatrix} 0 & B & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & I \\ I & 0 & \cdots & 0 & 0 \end{bmatrix} \in \mathbb{R}^{rk \times rk},$$

where B is any $k \times k$ primitive stochastic matrix that yields equality in (1.3). (In particular, B can be $\frac{1}{k-1}(J - I)$ with $J, I \in \mathbb{R}^{k \times k}$, or the matrix of Example 1 if $k = 5$, or the matrix of Example 2 if $k = 7$, or the matrix of (3.1) if k is divisible by 3.) From the construction of T, each of the cyclic products P_j, $j = 1, \ldots, r$ of Theorem 10 is equal to B, and since B satisfies (3.11), we see that T yields equality in (1.3).

Remark 3.3. From Example 3, we see that if $n \geqslant 5$ is not a prime number, then there is an $n \times n$ periodic stochastic matrix that yields equality in (1.3).

Example 4. Fix a natural number k, and suppose that $k-1$ is not prime, with $k-1$ factored as $k - 1 = qr$, $q, r \geqslant 2$, say. Let C be the $k \times k$ circulant permutation matrix

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

Let $T_1 = \frac{1}{r}(C + C^2 + \cdots + C^r)$ and $T_2 = \frac{1}{q}(I + C^r + C^{2r} + \cdots + C^{(q-1)r})$, and consider the periodic doubly stochastic matrix $T = \begin{bmatrix} 0 & T_1 \\ T_2 & 0 \end{bmatrix}$.

It is straightforward to see that $T_1T_2 = T_2T_1 = \frac{1}{qr}(C + C^2 + \cdots + C^{qr}) = \frac{1}{k-1}(J - I)$. Hence the matrices P_1 and P_2 satisfy (3.11), from which we conclude that T yields equality in (1.3).

References

[1] T. Ando, Hadamard products of matrices and their Moore–Penrose inverse, Private communication, 1999.

[2] A. Ben-Israel, T.N.E. Greville, Generalized Inverses: Theory and Applications, second ed., Springer, New York, 2003.

[3] A. Berman, R.J. Plemmons, Nonnegative Matrices in the Mathematical Sciences, SIAM, Philadelphia, 1994.

[4] S.L. Campbell, C.D.J. Meyer, Generalized Inverses of Linear Transformations, Dover Publications, New York, 1991.

[5] M. Catral, M. Neumann, J. Xu, Matrix analysis of a Markov chain small-world model, Linear Algebra Appl. 409 (2005) 32–50.

[6] G.E. Cho, C.D. Meyer, Markov chain sensitivity measured by mean first passage times, Linear Algebra Appl. 316 (2000) 21–28.

[7] G.E. Cho, C.D. Meyer, Comparison of perturbation bounds for the stationary distribution of a Markov chain, Linear Algebra Appl. 335 (2001) 137–150.

[8] R. Funderlic, C.D. Meyer, Sensitivity of the stationary distribution vector for an ergodic Markov chain, Linear Algebra Appl. 76 (1986) 1–17.

[9] J.J. Hunter, Mixing times with applications to perturbed Markov chains, Preprint, Massey University, Auckland, New Zealand, 2003.

[10] J.G. Kemeny, J.L. Snell, Finite Markov Chains, Van Nostrand, Princeton, NJ, 1960.

[11] S.J. Kirkland, The group inverse associated with an irreducible periodic nonnegative matrix, SIAM J. Matrix Anal. Appl. 16 (1995) 1127–1134.

[12] S. Kirkland, On a question concerning condition numbers for Markov chains, SIAM J. Matrix Anal. Appl. 23 (2002) 1109–1119.

[13] S. Kirkland, Conditioning properties of the stationary distribution for a Markov chain, Electron. J. Linear Algebra 10 (2003) 1–15.

[14] C.D.J. Meyer, The role of the group generalized inverse in the theory of finite Markov chains, SIAM Rev. 17 (1975) 443–464.

[15] C.D. Meyer, The condition of a finite Markov chain and perturbation bounds for the limiting probabilities, SIAM J. Algebra Disc. Meth. 1 (1980) 273–283.

[16] C.D. Meyer, Sensitivity of the stationary distribution of a Markov chain, SIAM J. Matrix Anal. Appl. 15 (1994) 715–728.

[17] E. Seneta, Sensitivity analysis, ergodicity coefficients, and rank-one updates for finite Markov chains, in: W.J. Stewart (Ed.), Numerical Solution of Markov Chains, Marcel Dekker, New York, 1991.

[18] P. Tetali, Design of on-line algorithm using hitting times, SIAM J. Comput. 28 (1999) 1232–1246.