On Optimal Condition Numbers For Markov Chains

Stephen J. Kirkland[*], Michael Neumann [†], and Nung–Sing Sze [‡]

July 7, 2008

Abstract

Let T and $\tilde{T} = T - E$ be arbitrary nonnegative, irreducible, stochastic matrices corresponding to two ergodic Markov chains on n states. A function κ is called a *condition number for Markov chains* with respect to the (α, β)–norm pair if $\|\pi - \tilde{\pi}\|_\alpha \leq \kappa(T)\|E\|_\beta$. Here π and $\tilde{\pi}$ are the stationary distribution vectors of the two chains, respectively.

Various condition numbers, particularly with respect to the $(1, \infty)$ and (∞, ∞)–norm pairs have been suggested in the literature. They were ranked according to their size by Cho and Meyer in a paper from 2001. In this paper we first of all show that what we call the *generalized ergodicity coefficient* $\tau_p(A^{\#}) = \sup_{y^t e=0} \frac{\|y^t A^{\#}\|_p}{\|y\|_1}$, where e is the n–vector of all 1's, is the smallest of the condition number of Markov chains with respect to the (p, ∞)–norm pair. We use this result to identify the smallest of the condition numbers of Markov chains among the (∞, ∞) and $(1, \infty)$–norm pairs. These are, respectively, κ_3 and κ_6 in the Cho–Meyer list of 8 condition numbers.

Kirkland has studied $\kappa_3(T)$. He has shown that $\kappa_3(T) \geq \frac{n-1}{2n}$ and he has characterized transition matrices for which equality holds. We prove here again that $2\kappa_3(T) \leq \kappa(6)$ which appears in the Cho–Meyer paper and we characterize the transition matrices T for which $\kappa_6(T) = \frac{n-1}{n}$. There is actually only one such matrix: $T = (J_n - I)/(n-1)$, where J_n is the $n \times n$ matrix of all 1's.

Keywords: Markov chains, stationary distribution, stochastic matrix, group inverses, sensitivity analysis, perturbation theory, condition numbers.

AMS Classification: 65F35, 60J10, 15A51, 15A12, 15A18.

1 Introduction

Recall that a homogeneous finite *Markov chain* is a sequence of random variables $X_0, X_1, X_2, \ldots$ with the *Markov property*, namely, that given the present state, the future and past states are independent. Formally, this can be expressed as follows.

$$Pr\left(X_{k+1} = \mathcal{S}_j \,\middle|\, X_k = \mathcal{S}_{i_k}, X_{k-1} = \mathcal{S}_{i_{k-1}}, \ldots, X_0 = \mathcal{S}_{i_0}\right) \;=\; Pr\left(X_{k+1} = \mathcal{S}_j \,|X_k = \mathcal{S}_{i_k}\right),$$

[*]Department of Mathematics and Statistics, University of Regina, Regina, Saskatchewan, CANADA S4S 0A2, Research supported in part by NSERC under grant OGP0138251.

[†]Department of Mathematics, University of Connecticut, Storrs, CT 06269–3009 (neumann@math.uconn.edu). Research supported in part by NSA Grant No. 06G–232.

[‡]Department of Mathematics, University of Connecticut, Storrs, CT 06269–3009 (sze@math.uconn.edu)

where the values of the $\mathcal{S}_i$'s come from a finite set $\{\mathcal{S}_1, \ldots, \mathcal{S}_n\}$, called the *state space*. Markov's property is sometimes referred to in the literature as a *memoryless process*, see Meyer [24].

Finite Markov chains arise in many applications and disciplines, for example, in physics in the area of statistical mechanics. The probabilities are used to represent unknown or unmodelled details of the system when it can be assumed that the dynamics of the system are time–invariant and that no relevant history need be considered which is not already included in the state description. Another area in which finite Markov chains commonly arise is in **queuing theory**. More examples of areas in which finite Markov chains arise can be found in Feller [8], Kemney and Snell [16] , Seneta [28, 30], Wikipidia, and many more. We mention that the Markov process is called *ergodic* if there is a positive probability of reaching any state in the system from any other state.

Given a finite homogeneous Markov chain on the n states $\mathcal{S}_1, \ldots, \mathcal{S}_n$, for $i, j = 1, \ldots, n$, let $t_{i,j} = Pr\left(X_{k+1} = \mathcal{S}_j \,|X_k = \mathcal{S}_i\right)$. The stochastic and nonnegative matrix $T = (t_{i,j}) \in \mathbb{R}^{n,n}$ thus obtained is called the *transition matrix* of the chain. If, in addition, the chain is ergodic, then it is well known that T is (also) an irreducible matrix. For such a chain, its *stationary distribution vector* is the unique positive vector $\pi \in \mathbb{R}^n$ satisfying that $\pi^T T = \pi^T$ and $\|\pi\|_1 = 1$. In particular π is the left Perron vector[1] of T. The stationary distribution vector is a key parameter for 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, [22, 9, 23, 5, 6, 17]. Specifically, let $\tilde{T} = T - E$ 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. In 1968 Schweitzer [27] introduced the following error analysis for Markov chains:

$$\pi^t - \tilde{\pi}^t \;=\; \pi^t T - \tilde{\pi}^t (T - E) \;=\; (\pi^t - \tilde{\pi}^t) T + \tilde{\pi}^t E \tag{1.1}$$

and hence

$$(\pi^t - \tilde{\pi}^t) A \;=\; \tilde{\pi}^t E \;\Rightarrow\; \pi^t - \tilde{\pi}^t \;=\; \tilde{\pi}^t E A^{\#},^{2} \tag{1.2}$$

where $A = I - T$ and $A^{\#}$ is the group inverse of A. The implication follows because the fact that $(\pi^t - \tilde{\pi}^t)e = 0$ puts the vector $\pi^t - \tilde{\pi}^t$ in $R(A^t)$, namely, in the row space of A onto which $AA^{\#}$ is a projection matrix.

If there is a scalar $\kappa = \kappa(T)$ such that

$$\|\pi - \tilde{\pi}\|_\alpha \le \kappa \|E\|_\beta,$$

[1] For more background material on nonnegative matrices and the Perron–Frobenius theory see Berman and Plemmons [3] and Varga [32].

[2] It should be commented that Schweitzer derived this equation with the right hand side given by $\tilde{\pi}^t E Z$, where $Z = (A + e\pi^t)^{-1}$ is the so called *fundamental matrix of the chain.*

for some suitable compatible matrix–vector norms, then κ is called a *condition number of Markov chains for the* (α, β)*–norm pair.*

In [6], Cho and Meyer survey most of the existing condition numbers which have been found by researchers and determine some rankings between them. To describe these results, we have to introduce some notations and definitions.

For a given transition matrix $T \in \mathbb{R}^{n,n}$ with stationary vector π, let $A = I - T$. Denote by $A_j \in \mathbb{R}^{n-1,n-1}$ the principal submatrix of A obtained by deleting its j–th row and column, $j = 1, \ldots, n$. Next, recall that the *group inverse* of a matrix $B \in \mathbb{R}^{n,n}$, if it exists, is the unique matrix denoted by $B^{\#}$ satisfying

$$BB^{\#}B \;=\; B, \quad B^{\#}BB^{\#} \;=\; B^{\#}, \;\text{ and }\; BB^{\#} \;=\; B^{\#}B.$$

It is well known, see, for example, Ben–Israel and Greville [2] and Campbell and Meyer [4], that the group inverse of a matrix exists if and only if its range and nullspaces are complimentary subspaces in $\mathbb{R}^n$, a property which is possessed by the M–matrix A obtained as above. Another notion that we shall need is that of the *mean first passage time from state* $\mathcal{S}_i$ *to state* $\mathcal{S}_j$ which is the quantity

$$m_{i,j} \;=\; \sum_{k=1}^{\infty} kPr\left(X_k \;=\; \mathcal{S}_j,\; X_\mu \;\neq\; \mathcal{S}_j,\; \mu = 1, \ldots, k-1 | X_0 \;=\; \mathcal{S}_i\right). \tag{1.3}$$

According to Meyer [21], the mean first passage matrix is given by

$$M \;=\; (m_{i,j}) \;=\; \left[I - A^{\#} + J_n A_d^{\#}\right] \Pi^{-1}, \tag{1.4}$$

where $J_n \in \mathbb{R}^{n,n}$ is the all 1's matrix, where $A_d^{\#}$ denotes the $n \times n$ diagonal matrix whose diagonal entries are the corresponding diagonal entries of $A^{\#}$, and where Π is the diagonal matrix whose diagonal entries are the corresponding entries of π. Finally, for a matrix $B \in \mathbb{C}^{n,n}$ with a constant row sum we follow Senta [28] closely and call the quantity

$$\tau_1(B) \;:=\; \sup_{y^t e = 0} \frac{\|y^t B\|_1}{\|y^t\|_1}$$

the *ergodicity coefficient* of B, where $e \in \mathbb{R}^n$ is the vector of all ones. It is essentially a result of Dobrushin [7] that

$$\tau_1(B) \;=\; \frac{1}{2} \max_{i,j} \sum_{k=1}^{n} |b_{ik} - b_{jk}|\,.$$

Following the notations in [6], we define by

$$\kappa_1 = \|Z\|_\infty, \qquad \kappa_2 = \|A^\#\|_\infty,$$

$$\kappa_3 = \frac{\max_j (a^\#_{j,j} - \min_i a^\#_{i,j})}{2}, \qquad \kappa_4 = \max_{i,j} |a^\#_{ij}|,$$

$$\kappa_5 = \frac{1}{1-\tau_1(T)}, \qquad \kappa_6 = \tau_1(A^\#), \tag{1.5}$$

$$\kappa_7 = \frac{\min_j \|A_j^{-1}\|_\infty}{2}, \qquad \kappa_8 = \frac{1}{2}\max_j \left(\frac{\max_{i\neq j} m_{ij}}{m_{jj}} \right).$$

We comment that κ_1 is due to Schweitzer [27], κ_2 is due to Meyer [22], κ_3 is due to Haviv and van Heyden [13] and Kirkland, Neumann, and Shader [19], κ_4 is due to Funderlic and Meyer [9], κ_5 is due to Seneta [29], κ_6 is due to Seneta [30], κ_7 is due to Ipsen and Meyer [15] and Kirkland, Neumann, and Shader [19], and, finally, κ_8 is due to Cho and Meyer [5].

We mention that according to the above papers,

$$\|\pi - \tilde{\pi}\|_1 \ \leq \ \kappa_i \|E\|_\infty, \quad \text{for } i = 1, 2, 5, \text{ and } 6 \tag{1.6}$$

and

$$\|\pi - \tilde{\pi}\|_\infty \ \leq \ \kappa_i \|E\|_\infty \quad \text{for } i = 3, 4, 7, \text{ and } 8. \tag{1.7}$$

In the paper by Cho and Meyer [6] and in references cited therein which motivated our present work, it shown that

$$\kappa_3 \ = \ \kappa_8 \ \leq \ \kappa_4 \ < \ 2\kappa_3 \ \leq \ \kappa_6 \ \leq \ \kappa_i, \quad \text{for } i = 1, 2, \text{ and } 5. \tag{1.8}$$

Cho and Meyer conjectured that $\kappa_3 \leq \kappa_7$ also. Later, in [17], Kirkland proved the conjecture and in fact showed that

$$\frac{n-1}{2n} \ \leq \ \kappa_3 \ \leq \ \kappa_7. \tag{1.9}$$

Furthermore, he showed that the left inequality is sharp on the total class of the $n \times n$ stochastic and irreducible matrices.

Notice that $\kappa_{1,2,5,6}$ are condition numbers with respect to the $(1, \infty)$–norm pair and $\kappa_{3,4,7,8}$ are condition numbers with respect to the (∞, ∞)–norm pair. From the above inequalities, we see that κ_3 is the smallest among all the four condition numbers satisfying (1.6), while κ_6 is the smallest among all the four condition numbers satisfying (1.7). Furthermore, we see that κ_3 is the smallest value among all the eight condition numbers. **These properties of κ_3 and κ_6 lead to the central questions of this paper: Determine the condition numbers that are the smallest among the (∞, ∞)–norm pair and the smallest**

among the $(1,\infty)$–norm pair, respectively.

For any $1 \le p \le \infty$ and matrix B with fixed row sum, we define

$$\tau_p(B) \;:=\; \sup_{y^t e=0} \frac{\|y^t B\|_p}{\|y^t\|_1} \tag{1.10}$$

as the *generalized ergodicity coefficient* of B with respect to the p–norm[3] Let $T \in R^{n,n}$ be a transition matrix for an ergodic Markov chain and $A = I - T$. We shall show in Proposition 2.1 in Section 2 that $\tau_p(A^{\#})$ is actually a condition number and in Theorem 2.3 that it is the smallest one with respect to the (p,∞)–norm pair, that is,

$$\|\pi - \tilde{\pi}\|_p \le \kappa(T)\|E\|_\infty \;\Rightarrow\; \tau_p(A^{\#}) \;\le\; \kappa(T).$$

Thus as $\kappa_6(T) = \tau_1(A^{\#})$, it will immediately follow that κ_6 is the minimal $(1,\infty)$–norm pair condition number for Markov chains. Next, in Proposition 2.5 we shall show that $\tau_\infty(A^{\#}) = \kappa_3(T)$ and so we shall be able to deduce immediately that κ_3 is the minimal (∞,∞)–norm pair condition number for Markov chains. Actually, using the vector norm inequality that for $x \in \mathbb{R}^n$ with $x^t e = 0$,

$$2^{\frac{1}{p}-1}\|x\|_1 \;\ge\; \|x\|_p \;\ge\; 2^{\frac{1}{p}}\|x\|_\infty,$$

it will follow immediately from the definition of $\tau_p(B)$ that

$$2^{\frac{1}{p}-1}\tau_1(B) \;\ge\; \tau_p(B) \;\ge\; 2^{\frac{1}{p}}\tau_\infty(B)$$

for matrices B with a constant row sum. Hence for such matrices, $\tau_1(B) \ge 2\tau_\infty(B)$. Substituting $A^{\#}$ for B in this inequality yields then another proof that $2\kappa_3 \le \kappa_6$ holds as displayed in (1.8).

Finally, we mentioned earlier that Kirkland [17] showed that $\kappa_3(T) \ge (n-1)/2n$. In his paper Kirkland characterizes the properties that T has to satisfy in order for $\kappa_3(T) = (n-1)/2n$. From (1.8) it thus follows that for all nonnegative, stochastic, and irreducible matrices $T \in \mathbb{R}^{n,n}$,

$$\frac{n-1}{n} \;\le\; \kappa_6(T). \tag{1.11}$$

We devote Section 3 of this paper to showing that equality holds in this inequality if and only if $T = (J_n - I)/(n-1)$. This is achieved by studying, in Theorem **??**, the cycle structure of a doubly stochastic matrix with a zero diagonal. From results in [17] and the paper [20] we see that the class of stochastic matrices T for which $\kappa_3(T) = (n-1)/2n$ is much wider than the class of stochastic matrices T for which $\kappa_6(T) = (n-1)/n$. We make

[3] It should be commented that for nonnegative and stochastic matrices $T \in \mathbb{R}^{n,n}$, Rothblum and Tan [26] show that coefficients of ergodicity are upper bounds on their eigenvalues other than 1.

some closing comments and give an example to illustrate the results of this paper in Section 4.

We close this introduction by mentioning that in a seminal paper on Markov chains Meyer [21] has shown that the group generalized inverse of $A = I - T$ plays an important role in the theory and application of Markov chains and that virtually everything that one would want to know about the chain can be either computed or deduced from $A^{\#}$. Algorithms for computing $A^{\#}$ have been suggested, among others, by Anstreicher and Rothblum [1], Golub and Meyer [10], and by Hartwig [12]. We have found that the algorithms suggested in [1] and [12], both of which are based on the shuffle algorithm, generally do very well in the computation of the group inverse of a singular and irreducible M–matrix. Further perturbation analysis for Markov chains, the use of the group inverse in such analysis, and stability analysis for the computation of the group inverse can be found in Meyer and Shoaf [25], Stewart [31], and Wilkinson [33].

2 Optimal Condition Numbers For Markov Chains

Let $T \in \mathbb{R}^{n,n}$ be a nonnegative, stochastic, and irreducible transition matrix for a finite ergodic Markov process on n states. Put $A = I - T$. We begin by showing that the generalized ergodicity coefficient $\tau_p(A^{\#})$, obatined by substituting $A^{\#}$ for B in (1.10), is actually a condition number for the chains with respect to the (p, ∞)–norm pair.

Proposition 2.1 *Let T and $\tilde{T} = T - E$ be $n \times n$ nonnegative, stochastic, and irreducible, and let π and $\tilde{\pi}$ be their stationary distribution vectors, respectively. Put $A = I - T$. Then*

$$\|\pi - \tilde{\pi}\|_p \;\leq\; \tau_p(A^{\#})\|E\|_\infty.$$

Hence $\tau_p(A^{\#})$ is a condition number for Markov chains with respect to the (p, ∞)–norm pair.

Proof. Recall first that by (1.10),

$$\tau_p(A^{\#}) \;=\; \sup_{y^t e = 0} \frac{\|y^t A^{\#}\|_p}{\|y^t\|_1}. \tag{2.12}$$

Combining (1.2) and (2.12) we obtain that:

$$\|\pi^t - \tilde{\pi}^t\|_p \;=\; \|\tilde{\pi}^t E A^{\#}\|_p \;\leq\; \|\tilde{\pi}^t E\|_1 \tau_p(A^{\#}) \;\leq\; \|\tilde{\pi}\|_1 \|E\|_\infty \tau_p(A^{\#}) \;=\; \tau_p(A^{\#})\|E\|_\infty.$$

Hence $\kappa(T) = \tau_p(A^{\#})$ is a condition number. □

Remark 2.2 Recall that we mentioned earlier that it is already known from the literature that when $p = 1$, $\tau_1(A^{\#}) = \kappa_6(T)$ is a condition number with respect to the $(1, \infty)$–norm pair.

We now present the main result of this section.

Theorem 2.3 *Suppose κ is a condition number of Markov chains with respect to the (p,∞)–norm pair, viz.,*

$$\|\pi - \tilde{\pi}\|_p \;\leq\; \kappa(T)\|E\|_\infty. \tag{2.13}$$

Then

$$\tau_p(A^\#) \leq \kappa(T).$$

Thus $\tau_p(A^\#)$ is the smallest condition number for Markov chains with respect to the (p,∞)–norm pair.

Proof. Suppose κ is a condition number satisfying (2.13). Then we can write that

$$\kappa(T) \;\geq\; \sup_{Ee=0} \frac{\|\pi - \tilde{\pi}\|_p}{\|E\|_\infty} \;=\; \sup_{Ee=0} \frac{\|\tilde{\pi}EA^\#\|_p}{\|E\|_\infty} \;\geq\; \sup_{y^te=0} \frac{\|\tilde{\pi}ey^tA^\#\|_p}{\|ey^t\|_\infty} \;=\; \sup_{y^te=0} \frac{\|y^tA^\#\|_p}{\|y^t\|_1} \;=\; \tau_p(A^\#).$$

□

Since $\tau_1(A^\#) = \kappa_6(T)$, we have thus arrived at the following corollary.

Corollary 2.4 *The condition number $\kappa_6(T)$ is the smallest among the condition numbers for Markov chains with respect to $(1,\infty)$–norm pair.*

Now we turn to the study of the condition number $\tau_\infty(A^\#)$.

Proposition 2.5 *Let $T \in \mathbb{R}^{n,n}$ be a nonnegative, irreducible, and stochastic matrix and put $A = I - T$. Then*

$$\tau_\infty(A^\#) \;=\; \kappa_3(T).$$

Proof. By Proposition 2.1 with $p = \infty$, we have that $\tau_\infty(A^\#) \leq \kappa_3(T)$. Consider the expression for $\kappa_3(T)$ given in (1.5) and suppose that $1 \leq u, v \leq n$ is a pair of indices for which the value of $\kappa_3(T)$ is attained, that is:

$$\kappa_3(T) \;=\; \frac{\max_j (a^\#_{j,j} - \min_i a^\#_{i,j})}{2} \;=\; \frac{a^\#_{v,v} - a^\#_{u,v}}{2}.$$

Then we can write that:

$$\tau_\infty(A^\#) \;=\; \sup_{y^te=0} \frac{\|y^tA^\#\|_\infty}{\|y^t\|_1} \;\geq\; \frac{\|(e_v - e_u)^tA^\#\|_\infty}{\|(e_v - e_u)^t\|_1} \;=\; \frac{\max_w |a^\#_{vw} - a^\#_{uw}|}{2} \;\geq\; \frac{a^\#_{vv} - a^\#_{uv}}{2} \;=\; \kappa_3(T).$$

□

Proposition 2.5 has the immediately corollary:

Corollary 2.6 *The condition number κ_3 is the smallest among the condition numbers for Markov chains with respect to the (∞, ∞)–norm pair.*

Before moving to the next section we provide a further proof for the inequality within (1.8) which says that $2\kappa_3 \leq \kappa_6$. The proof is immediate from (1.10), Proposition 2.1, and by the fact that $\|x^t B\|_p = \|x^t(B + \alpha J_n)\|_p$, for any $\alpha \in \mathbb{R}$ and $x \in \mathbb{R}^n$ with $x^t e = 0$.

Proposition 2.7 *For any $x \in \mathbb{R}^n$ with $x^t e = 0$,*

$$2^{\frac{1}{p}-1}\|x\|_1 \ \geq \ \|x\|_p \ \geq \ 2^{\frac{1}{p}}\|x\|_\infty.$$

Consequencely, we have that

$$2^{\frac{1}{p}-1}\tau_1(B) \ \geq \ \tau_p(B) \ \geq \ 2^{\frac{1}{p}}\tau_\infty(B).$$

In particular,

$$\kappa_6(T) \ = \ \tau_1(A^\#) \ \geq \ 2\tau_\infty(A^\#) \ = \ 2\kappa_3(T).$$

3 Investigation Of The Optimal Lower Bound For κ_6

In the introduction we mentioned that Kirkland [17] has shown that for any nonnegative, irreducible, and stochastic matrix $T \in \mathbb{R}^{n,n}$,

$$\kappa_3(T) \ \geq \ \frac{n-1}{2n}.$$

Thus from (1.8) we immediately see that for any such T,

$$\kappa_6(T) \ \geq \ \frac{n-1}{n}. \tag{3.1}$$

Continuing, in [17] the author characterized transition matrices $T \in \mathbb{R}^{n,n}$ for which $\kappa_3(T)$ attains the lower bound $(n-1)/2n$. Let us cite his result:

Theorem 3.1 (Kirkland [17, Theorem 2.9]) *Suppose that $T \in \mathbb{R}^{n,n}$ is a nonnegative and irreducible stochastic matrix. Then $\kappa_3(T) = (n-1)/2n$ if and only if the following hold:*

(i) *T is a doubly stochastic matrix with zero diagonal,*

(ii) *$\|A_i^{-1}\|_\infty = n-1$ for all $i = 1, \ldots, n$, and*

(iii) *the j–th entry of $A_i^{-1}e$ equals $n-1$ whenever $t_{i,j} > 0$.*

We comment that Kirkland's characterization was somewhat condensed in the paper [20] in which the authors, Kirkland, Neumann, and Xu, studied specific classes of transition matrices for which the characterization holds.

From Theorem 3.1 and from (1.8) we see that a **necessary** condition for an irreducible, nonnegative, and stochastic matrix $T \in \mathbb{R}^{n,n}$ to satisfy that:

$$\kappa_6(T) \;=\; \frac{n-1}{n} \tag{3.2}$$

is that T fulfills the conditions of Theorem 3.1. In the remainder of this paper we shall show that the **only** stochastic matrix for which (3.2) holds is $T = (J_n - I)/(n-1)$. In the course of proving this fact several results of interest in their own right will be proved. We begin with a theorem which is in the spirit of Kirkland's Theorem 3.1. Theorem 3.2 will include a fresh proof of (3.1), but this will not be the main purpose of its statement.

Theorem 3.2 *Let $T \in \mathbb{R}^{n,n}$ be a nonnegative and irreducible stochastic matrix and set $A = I - T$. Then*

$$\kappa_6(T) \;\geq\; \frac{n-1}{n}. \tag{3.3}$$

Moreover, equality holds in (3.3) if and only if the following three conditions hold:

(i) *T is a doubly stochastic matrix with zero diagonal,*

(ii) *If $t_{i,j} > 0$, then $a^{\#}_{i,k} \geq a^{\#}_{j,k}$, whenever $k \neq i$, that is,*

$$a^{\#}_{j,k} \;=\; \min_{k\neq i} a^{\#}_{i,k},$$

(iii) *If $t_{i,j} > 0$, then* $\displaystyle \frac{1}{2}\sum_{k=1}^{n}\left|a^{\#}_{i,k} - a^{\#}_{j,k}\right| \;=\; \frac{n-1}{n}.$

Proof. We begin by noting that $AA^{\#} = I - e\pi^t$. On equating the (i,k)–th position of both sides of this equality we find that:

$$a^{\#}_{i,k} - \sum_{j=1}^{n} t_{ij} a^{\#}_{j,k} \;=\; \begin{cases} 1-\pi_i & \text{if } k = i, \\ -\pi_k & \text{Otherwise.} \end{cases}$$

But then, for any $i = 1, \ldots, n$,

$$\begin{aligned} 2(1-\pi_i) &= |1-\pi_i| + \sum_{k\neq i} |-\pi_k| \\ &= \sum_{k=1}^{n}\left|a^{\#}_{i,k} - \sum_{j=1}^{n} t_{i,j} a^{\#}_{j,k}\right| \end{aligned}$$

$$\leq \sum_{k=1}^{n}\sum_{j\neq i} t_{ij}\left|a^{\#}_{i,k}-a^{\#}_{j,k}\right|$$
$$= 2\sum_{j\neq i} t_{i,j}\left(\frac{1}{2}\sum_{k=1}^{n}\left|a^{\#}_{i,k}-a^{\#}_{j,k}\right|\right)$$
$$\leq 2\sum_{j\neq i} t_{i,j}\tau_1(A^{\#}) \leq 2\tau_1(A^{\#}) = 2\kappa_6(T).$$

We know that of necessity $\min_i \pi_i \leq \frac{1}{n}$ and so from the above inequalities we can conclude that

$$\kappa_6(T) \geq 1-\min_i \pi_i \geq 1-\frac{1}{n} = \frac{n-1}{n}.$$

Furthermore, the above inequalities become equalities only when

(a) $\min_i \pi_i = \dfrac{1}{n}$ and $t_{j,j}=0$, for all $j=1,\ldots,n$,

and if $t_{i,j}>0$, then

(b) $a^{\#}_{j,k} \geq a^{\#}_{i,k}$, whenever $k\neq i$, and

(c) $\dfrac{1}{2}\sum_{k=1}^{n}\left|a^{\#}_{i,k}-a^{\#}_{j,k}\right| = \dfrac{n-1}{n}$.

We see that (a) implies that T is a doubly stochastic matrix with zero diagonal. □

As mentioned above, in [20] the authors investigated classes of stochastic matrices T which satisfy the conditions of Theorem 3.1 and hence for which $\kappa_3(T)=(n-1)/2n$ holds. Generally what can be said about the classes which have been found to date is that they comprise of doubly stochastic matrices, with zero diagonal, and which possess some form of a circulant or block circulant structure, but that no overall class which encompasses all has yet been determined. In contrast, we shall now begin to show that for any $n\neq 3$, there is a single stochastic matrix T for which $\kappa_6(T)=(n-1)/n$, while when $n=3$, there are precisely three such stochastic matrices.

Theorem 3.3 *An $n\times n$ nonnegative irreducible stochastic matrix T satisfying that $\kappa_6(T)=\frac{n-1}{n}$ must be either the matrix $\frac{1}{n-1}(J_n-I)$ or, when $n=3$, one of the two cyclic permutation matrices.*

Recall first that for an $n\times n$ nonnegative matrix T, the *directed graph of* T is the directed graph D on vertices $1,\ldots,n$ such that $i\to j$ is an arc in D if and only if $t_{i,j}>0$. A *simple directed cycle* $C=\{i_1,\ldots,i_p\}$ of D is a finite distinct sequence of indices $\{i_1,\ldots,i_p\}$ in $\{1,\ldots,n\}$ such that D contains the arcs $i_1\to i_2\to\ldots\to i_{p-1}\to i_p\to i_1$. In that case, the *length* of C is p.

To prove Theorem 3.3 we need to develop several auxiliary results, some of which are of interest in their own right. We begin with the following lemma:

Lemma 3.4 *Suppose $T \in \mathbb{R}^{n,n}$ is a nonnegative, irreducible, and stochastic matrix for which $\kappa_6(T) = \frac{n-1}{n}$. Suppose that $C = \{i_1, \ldots, i_p\}$ is a simple directed cycle in the directed graph of T. If $r \notin C$, then*

$$a^{\#}_{i_1,r} \;=\; a^{\#}_{i_2,r} \;=\; \cdots \;=\; a^{\#}_{i_p,r}. \tag{3.4}$$

Proof. For any permutation matrix P, $PA^{\#}P^t$ is the group inverse of PAP^t and so, clearly, $\kappa_6(PTP^t) = \kappa_6(T)$. Thus we may assume that the simple directed cycle of T consists of the indices $\{1, \ldots, p\}$. Under this assumption, we will focus on the first p rows of $A^{\#}$.

As

$$t_{1,2}, t_{2,3}, \ldots, t_{p-1,p}, t_{p,1} \;>\; 0,$$

by Theorem 3.2 (ii), if $t \notin \{1, \ldots, p\}$, then

$$a^{\#}_{1,t} \;\leq\; a^{\#}_{2,t} \;\leq\; \cdots \;\leq\; a^{\#}_{p-1,t} \;\leq\; a^{\#}_{p,t} \;\leq\; a^{\#}_{1,t}.$$

Thus, all inequalities are equalities which yields (3.4). □

Our next lemma is:

Lemma 3.5 *Suppose $T \in \mathbb{R}^{n,n}$ is a nonnegative, irreducible, and stochastic matrix for which $\kappa_6(T) = \frac{n-1}{n}$. Suppose that the directed graph of T contains a simple directed cycle of length $p \leq n-1$, without loss of generality on vertices $1, \ldots, p$. Partition the first p rows and columns of T as*

$$T \;=\; \left[\begin{array}{c|c} C & X \\ \hline Y & Z \end{array}\right].$$

Then C has constant row sums and X can be written as ev^t for some nonnegative vector v^t.

Proof. Partition $A^{\#}$ conformally with T as

$$A^{\#} \;=\; \left[\begin{array}{c|c} M_1 & M_2 \\ \hline M_3 & M_4 \end{array}\right].$$

From (3.4) we find that $M_2 = ew^t$ for some vector w^t. Since $A^{\#}(I-T) = I - \frac{1}{n}J$, we have $M_1(I-C) - ew^tY \;=\; I - \frac{1}{n}J$, so that $M_1 \;=\; (I-C)^{-1} + eu^t$, for some vector u^t. Since $M_1e + (w^te)e = 0$, we see that M_1 has constant rows sums. We conclude that necessarily $(I-C)^{-1}$ has constant row sums, and hence so does C.

Again using the fact that $A^{\#}(I-T) = I - \frac{1}{n}J$, we have $-M_1X + ew^t(I-Z) = -\frac{1}{n}J$, so that for some vector r^t, we have $(I-C)^{-1}X + eu^tX = er^t$. Thus $(I-C)^{-1}X$ can be written as $(I-C)^{-1}X = es^t$ for some vector s^t, and since C has constant row sums, it follows that for some vector v^t, we have $X = ev^t$. □

We are now ready to prove the main theorem of this section, Theorem 3.3.

Proof of Theorem 3.3. Let D be the directed graph of T. Suppose that D has a simple directed cycle of length p and that the directed cycle contains no chords, that is, the subgraph of D induced by the vertices of the directed cycle contains only the arcs of the directed cycle, and no others. Without loss of generality we suppose that the directed cycle is $1 \to 2 \to \ldots \to p \to 1$. Let C denote the principal submatrix of T on the first p rows and columns.

If $p = n$, then $T = C = Q$, where Q is the cyclic permutation matrix corresponding to the directed cycle of length n. It is readily verified that in that case,

$$A^{\#} = \sum_{j=0}^{n-1} \left(\frac{n-2j-1}{2n} \right) Q^j,$$

, from which it follows that $\kappa_6(T) = \frac{n-1}{n}$ only in the cases $n = 2$ and $n = 3$. In both of those cases, T has the desired form.

Suppose now that the length p of the chord–free directed cycle satisfies $p \leq n-1$. From Lemma 3.5, we find that for some scalar $a > 0$, $C = aQ$, where Q is the cyclic permutation matrix corresponding to the directed cycle on indices $1, \ldots, p$. Hence

$$(I - C)^{-1} = \frac{1}{1-a^p}(I + aQ + a^2Q^2 + \ldots a^{p-1}Q^{p-1}).$$

Referring to the partitioned form for $A^{\#}$ in Lemma 3.5, we have $M_1 = (I - C)^{-1} + eu^t$, for some vector u^t, while from (3.4) we have $M_2 = ew^t$, for some vector w^t. Thus, for any indices i, j with $1 \leq i, j \leq p$, we see that

$$\left\| (e_i - e_j)^t A^{\#} \right\|_1 = \left\| (e_i - e_j)^t (I - C)^{-1} \right\|_1 .$$

Since $t_{1,2} = a > 0$, we must have that

$$\left\| (e_1 - e_2)^t (I - C)^{-1} \right\|_1 = 2 \left(\frac{n-1}{n} \right)$$

and so, on employing the explicit expression for $(I - C)^{-1}$, it follows that

$$\frac{1 - a^{p-1}}{1 - a^p} = \frac{n-1}{n}.$$

Note that if $p \geq 4$, then

$$\left\| (e_1 - e_4)^t A^{\#} \right\|_1 = \left\| (e_1 - e_4)^t (I - C)^{-1} \right\|_1 = 2 \left(\frac{1 - a^{p-1} + a - a^{p-2}}{1 - a^p} \right) > 2 \left(\frac{1 - a^{p-1}}{1 - a^p} \right) = 2 \left(\frac{n-1}{n} \right),$$

a contradiction. We conclude that either $p = 2$ and $a = \frac{1}{n-1}$ or $p = 3$ and $a = \frac{1+\sqrt{4n-3}}{2(n-1)} \equiv a_0$.

Next, fix an arc $i \to j$ in D, and let ℓ denote the length of a shortest path from j back to i in D (so that $i \to j$ is an arc on a chord–free simple directed cycle of length $\ell + 1$). From the above considerations, we see that either: $\ell = 1$ and $t_{i,j} = t_{j,i} = \frac{1}{n-1}$ or $\ell = 2, t_{i,j} = a_0$, and $t_{j,i} = 0$. In particular, T only has entries in $\{0, \frac{1}{n-1}, a_0\}$.

Suppose that, contrary to the conclusion of Theorem 3.3, $T \neq \frac{1}{n-1}(J-I)$, that D contains the chord–free simple directed cycle $1 \to 2 \to 3 \to 1$ and that $n \geq 4$. Since T is irreducible, we may assume without loss of generality that $1 \to 4$ in D. From Lemma 3.5 we find that in fact $1, 2, 3 \to 4$ in that directed graph. There are two cases to consider: either $4 \to 1$ in D or the shortest path from 4 back to 1 has length 2.

Suppose first that $4 \to 1$. Then $t_{1,4} = t_{4,1} = \frac{1}{n-1}$, and hence it follows from Lemma 3.5 that $t_{i,4} = t_{4,i} = \frac{1}{n-1}$, for $i = 1, 2, 3$. Thus, D contains the directed cycle $2 \to 3 \to 4 \to 2$, but the principal submatrix of T on rows and columns $2, 3, 4$ has row sums $a_0 + \frac{1}{n-1}, \frac{1}{n-1}, \frac{2}{n-1}$, contradicting Lemma 3.5.

Now suppose that the shortest path in D from 4 to 1 has length 2. Observe that this shortest path cannot be $4 \to 3 \to 1$, otherwise we have $3 \to 4$ and $4 \to 3$, so that $t_{3,4} = \frac{1}{n-1}$; if that were the case, then by Lemma 3.5, $t_{1,4} = \frac{1}{n-1}$ and hence $t_{4,1} = \frac{1}{n-1}$, so that in fact the shortest path from 4 to 1 has length 1.

Thus, a shortest path from 4 to 1 passes through a vertex in $\{5, \ldots, n\}$, and without loss of generality, we take the path to be $4 \to 5 \to 1$. Hence D contains the directed cycle $4 \to 5 \to 1 \to 4$, and so by Lemma 3.5, $2 \to 4$ as well. Observe that since $t_{5,1} = a_0$, we must have $t_{1,5} = 0$. Similarly, since $t_{1,4} = t_{2,4} = a_0$, we have $t_{4,1} = t_{4,2} = 0$. Note that D contains the directed cycle $1 \to 2 \to 4 \to 5 \to 1$. Considering the corresponding principal submatrix C of T, we see that the row sum of C corresponding to vertex 1 is $2a_0$, while the row sum of C corresponding to vertex 4 is a_0, contradicting Lemma 3.5. □

4 Concluding Remarks And An Example

Let $T \in \mathbb{R}^{n,n}$ be a transition matrix for an ergodic homogeneous Markov chain on n states and let $E \in \mathbb{R}^{n,n}$ be a zero row sums matrix such that $\hat{T} = T - E$ is a transition matrix of a second Markov chain. Finally, let π and $\hat{\pi}$ be the stationary distribution vectors of the Markov chains whose transition matrices are T and $\hat{T}$, respectively.

From (1.8) we know that $2\kappa_3(T) \leq \kappa_6(T)$. However, $\kappa_3(T)\|E\|_\infty$ bounds $\|\pi - \hat{\pi}\|_\infty$ while $\kappa_6(T)\|E\|_\infty$ bounds $\|\pi - \hat{\pi}\|_1$ and $\|\pi - \hat{\pi}\|_\infty \leq \|\pi - \hat{\pi}\|_1$. Thus a comparison between the quality of these two optimal condition numbers can be not meaningful just based on their respective sizes. One advantage of using $\kappa_6(T)$ over $\kappa_3(T)$ is that $\kappa_6(T)$ also furnishes us with information on the eigenvalues of the transition matrix T. This fact is well known, but let us describe it briefly.

Let $B \in \mathbb{C}^{n,n}$ have constant row sums b so that b is an eigenvalue of B. Then it is known, see, for example, Seneta [28, p.63, Theorem 2.10], that

$$\gamma(B) \ := \ \max\{|\lambda| \mid \lambda \in \sigma(B), \ \lambda \neq b\} \ \leq \ \tau_1(B).$$

Thus, in the context of Markov chains, $\tau_1(T)$ bounds $\gamma(T)$ which is the asymptotic convergence rate of the iteration $x_i^t = x_{i-1}^t T$ to the stationary distribution vector π of the chain. It is further known from (1.8) or can easily be proved that:

$$\max_{\lambda\in\sigma(T),\lambda\neq 1} \frac{1}{|1-\lambda|} \ \leq \ \underbrace{\tau_1(A^{\#})}_{=\kappa_6(T)} \ \leq \ \underbrace{\frac{1}{1-\tau_1(T)}}_{=\kappa_5(T)}. \tag{4.5}$$

We see then that $1/\tau_1(A^{\#}) = 1/\kappa_6(T)$ is a lower bound on

$$\min_{\lambda\in\sigma(T),\lambda\neq 1} |1-\lambda|$$

In [11], Hartfiel and Meyer showed that the closer this quantity is to 0, the more *nearly–uncoupled* the chain becomes. We further mention that in Kirkland and Neumann [18], conditions are studied on T under which equality holds throughout (4.5).

As an illustration for our results here we borrow and example from Funderlic and Meyer [9], which models mammillary systems in compartmental analysis, and which was further discussed in Kirkland, Neumann, and Shader [19]. Let

$$T \ = \ \begin{bmatrix} 0.74 & 0.11 & 0 & 0 & 0 & 0 & 0 & 0.15 \\ 0 & 0.689 & 0 & 0 & 0.011 & 0 & 0 & 0.3 \\ 0 & 0 & 0 & 0.4 & 0 & 0 & 0 & 0.6 \\ 0 & 0 & 0 & 0.669 & 0.011 & 0 & 0 & 0.32 \\ 0 & 0 & 0 & 0 & 0.912 & 0 & 0 & 0.088 \\ 0 & 0 & 0 & 0 & 0 & 0.74 & 0 & 0.26 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.87 & 0.13 \\ 0.15 & 0 & 0.047 & 0 & 0 & 0.055 & 0.27 & 0.478 \end{bmatrix}, \tag{4.6}$$

in which case,

$$A^{\#} = \begin{bmatrix} 3.276 & 1.003 & -0.01465 & -0.05851 & 0.02992 & -0.2091 & -3.952 & -0.07398 \\ -0.3289 & 2.943 & 0.005007 & -0.03475 & 0.2754 & -0.1206 & -3.084 & 0.3443 \\ -0.1564 & -0.2113 & 1.019 & 1.191 & 0.03432 & -0.05735 & -2.462 & 0.6434 \\ -0.2990 & -0.2617 & 0.007450 & 2.989 & 0.2529 & -0.1096 & -2.976 & 0.3963 \\ -1.392 & -0.6482 & -0.08156 & -0.1394 & 11.18 & -0.5102 & -6.909 & -1.497 \\ -0.3603 & -0.2835 & 0.002450 & -0.03784 & -0.1283 & 3.714 & -3.196 & 0.2899 \\ -0.8879 & -0.4701 & -0.04053 & -0.08978 & -0.1581 & -0.3256 & 2.597 & -0.6246 \\ 0.1673 & -0.09685 & 0.04543 & 0.01410 & -0.09844 & 0.06133 & -1.297 & 1.204 \end{bmatrix}.$$

Also let E be the following matrix with zero row sums:

$$E = \begin{bmatrix} 0.01 & 0.01 & 0 & 0 & 0 & 0 & -0.005 & -0.015 \\ 0 & 0.005 & 0 & 0 & 0 & -0.01 & 0 & 0.005 \\ 0 & 0 & -0.02 & 0.02 & 0 & 0 & -0.015 & 0.015 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.07 & 0 & 0 & 0 & 0 & 0 & 0 & 0.07 \\ 0.03 & 0 & 0 & 0 & 0 & 0.03 & -0.07 & 0.01 \end{bmatrix}.$$

Then $\|E\|_\infty = 0.14$. Next put $\hat{T} := T - E$. Then for the Markov chains whose transition matrices are T and $\hat{T}$, the stationary distribution vectors are given by:

$$\pi = \begin{bmatrix} 0.1372 \\ 0.04852 \\ 0.01117 \\ 0.01350 \\ 0.007753 \\ 0.05030 \\ 0.4938 \\ 0.2378 \end{bmatrix} \quad \text{and} \quad \hat{\pi} = \begin{bmatrix} 0.2093 \\ 0.06624 \\ 0.008842 \\ 0.01015 \\ 0.009548 \\ 0.02028 \\ 0.4913 \\ 0.1844 \end{bmatrix},$$

respectively. We next compute for T given in (4.6), the 8 condition numbers given in (1.5) and find that:

$\kappa_1(T)$	21.3697
$\kappa_2(T)$	22.3542
$\kappa_3(T)$	5.6676
$\kappa_4(T)$	11.1771
$\kappa_5(T)$	11.3636
$\kappa_6(T)$	11.3352
$\kappa_7(T)$	5.6818
$\kappa_8(T)$	5.6676

On checking (1.7) and (1.6), for $i = 3$ and $i = 6$, respectively, we find that

$$\|\pi - \hat{\pi}\|_\infty \;=\; 0.0721 \;<\; 0.7935 \;=\; \|E\|_\infty \kappa_3(T)$$

and

$$\|\pi - \hat{\pi}\|_1 \;=\; 0.1833 \;<\; 1.5869 \;=\; \|E\|_\infty \kappa_6(T).$$

We further find the

$$1/\kappa_6(T) \;=\; .0882 \;<\; 0.089 \;=\; \min_{\lambda \in \sigma(T), \lambda \neq 1} |1 - \lambda|$$

and so, in this example, $\kappa_6(T)$ furnishes a close estimate to an upper bound on the measure for near–uncoupling introduced by Hartfiel and Meyer in [11].

References

[1] K. Anstreicher and U. Rothblum. Using Gauss–Jordan elimination to compute the index, generalized nullspace, and Drazin inverse. *Lin. Alg. Appl.*, 85:221–239, 1987.

[2] A. Ben–Israel and T. N. E. Greville. Generalized Inverses: Theory and Applications, 2nd Edition. *Springe*r, New York, 2003.

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

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

[5] G. E. Cho and C. D. Meyer. Markov chain sensitivity measured by mean first passage times. *Lin. Alg. Appl.*, 316:21–28, 2000.

[6] G.E. Cho and C. D. Meyer. Comparison of perturbation bounds for the stationary distribution of a Markov chain. *Lin. Alg. Appl.*, 335:137–150, 2001.

[7] R. L. Dobrushin. Central limit theorem for non–stationary Markov chains, I, II. *Theory of Probability and Applications*, 1:65–80, 329–383, 1956.

[8] W. Feller. An Introduction to Probability Theory and its Applications, Vol. I. *John Wiley*, New York, 1960.

[9] R. Funderlic and C. Meyer. Sensitivity of the stationary distribution vector for an ergodic Markov chain. *Lin. Alg. Appl.*, 76:1–17, 1986.

[10] G. H. Golub and C. D. Meyer. Using the QR–factorization and group inversion to compute, differentiate, and estimate the sensitivity of stationary probabilities for Markov chains. *SIAM J. Alg, Discrete Meth.*, 7:273–281, 1986 (1986).

[11] D. J. Hartfiel and C. D. Meyer. On the structure of stochastic matrices with a subdominant eigenvalue near 1. *Lin. Alg. Appl.*, 272:193–203, 1998.

[12] R. Hartwig. A method for calculating A^d. *Math. Japon.*, 26:37–43, 1981.

[13] M. Haviv and L. Van der Heyden. Perturbation bounds for the stationary probabilities of a finite Markov chain. *Adv. Appl. Prob.*, 16:804–818, 1984.

[14] N. J. Higham and P. A. Night. Componentwise error analysis for stationary iterative methods, in Linear algebra, Markov chains, and queueing models. *Inst. Math. Appl. Minnieapolis*, 48:29–46, Springer, New York, 1993.

[15] I. Ipsen and C. Meyer. Uniform stability of Markov chains. *SIAM J. Matrix Anal. Appl.*, 15:1061–1074, 1994.

[16] J. G. Kemeny and J. L. Snell. Finite Markov Chains. *Van Nostrand*, Princeton, New Jersey, 1960.

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

[18] S. J. Kirkland and M. Neumann. The case of equality in the Dobrushin–Deutsch–Zenger bound. Lin. Alg. Appl., submitted, 2008.

[19] S. Kirkland, M. Neumann, and B. Shader. Applications of Pazs inequality to perturbation bounds for Markov chains. *Lin. Alg. Appl.*, 268:183–196, 1998.

[20] S. Kirkland, M. Neumann, and J. Xu. Transition matrices for well–conditioned Markov chains. *Lin. Alg. Appl.*, 424:118–131, 2007.

[21] C. D. Meyer, The role of the group generalized inverse in the theory of finite Markov chains. *SIAM Rev.* 17:443–464, 1975.

[22] C. D. Meyer. The condition of a finite Markov chain and perturbation bounds for the limiting probabilities. *SIAM J. Alg. Disc. Meth.* 1: 273–283, 1980.

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

[24] C. D. Meyer. Matrix Analysis and Applied Linear Algebra. *SIAM Publications*, Philadelphia, 2000.

[25] C. D. Meyer and J. M. Shoaf Updating finite Markov chains by using techniques of group matrix inversion. *J. Statist. Comput. Simulation*, 11:163–181, 1980.

[26] U. Rothblum and C–P Tan. Upper bounds on the maximum modulus of subdominant eigenvlaues. of nonnegative matrices. *Lin. Alg. Appl.*, 66:45–86, 1985.

[27] P. Schweitzer. Perturbation theory and finite Markov chains. *J. Appl. Prob.*, 5:410–413, 1968.

[28] E. Seneta. Non–Negative Matrices and Markov Chains. *Springer*, New York, 1981.

[29] E. Seneta. Perturbation of the stationary distribution measured by ergodicity coefficients. *Adv. Appl. Prob.* , 20:228–230, 1988.

[30] 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*, N.Y., 1991.

[31] G. W. Stewart. Gaussian elimination, perturbation theory, and Markov chains, in Linear algebra, Markov chains, and queueing models. *Inst. Math. Appl. Minneapolis*, 48:59–69, Springer, New York, 1993.

[32] R. S. Varga. Matrix Iterative Analysis, Second Revised and Expanded Edition. *Springer–Verlag*, Heidelberg, 2000.

[33] J. H. Wilkinson. Note on the practical significance of the Drazin inverse, in Recent applications of generalized inverses. *Res. Notes in Math.* 66:pp. 82–99, Pitman, Boston, Mass., 1982.