Abstract. We show that all stationary probabilities of a finite irreducible Markov chain react essentially in the same way to perturbations in the transition probabilities. In particular, if at least one stationary probability is insensitive in a relative sense, then all stationary probabilities must be insensitive in an absolute sense.

Uniform Stability Of Markov Chains

Ilse C.F. Ipsen[1], Carl D. Meyer[2]

Research Report YALEU/DCS/RR-918
September 1992

[1] Department of Computer Science, Yale University, New Haven, CT 06520. The work of this author was supported in part by the National Science Foundation under grant CCR-9102853.

[2] Mathematics Department, North Carolina State University Raleigh, NC 27695-8205. The work of this author was supported in part by the National Science Foundation under grants DMS-9020915 and DDM-8906248.

1 Introduction

The purpose of this paper is to analyse the sensitivity of *individual* stationary probabilities to perturbations in the transition probabilities of finite irreducible Markov chains.

In addition to providing perturbation bounds that are much sharper than the traditional bounds, our analysis demonstrates that all stationary probabilities in an irreducible chain react in a somewhat uniform manner to perturbations in the transition probabilities. This property of uniform sensitivity distinguishes Markov problems markedly from general linear systems. Section 2 presents examples that illustrate why a Markov problem should not be treated as just another linear system.

Most of the work on perturbations of irreducible Markov chains has focused on the derivation of norm-based bounds of the following kind. Let P and $\tilde{P} = P + E$ be probability transition matrices of irreducible Markov chains with respective stationary probability vectors π^T and $\tilde{\pi}^T$ satisfying

$$\pi^T P = \pi^T, \quad \tilde{\pi}^T \tilde{P} = \tilde{\pi}^T, \quad \sum_i \pi_i = 1 = \sum_i \tilde{\pi}_i.$$

Then

$$\|\pi^T - \tilde{\pi}^T\| \leq \kappa \|E\|$$

for suitable vector and matrix norms. The condition numbers κ can be produced in various ways: from entries of Kemeny and Snell's "fundamental matrix" [13] as in [20], from entries of the group inverse [4] $A^{\#}$ of $A = I - P$ as in [2, 7, 9, 15, 17, 18, 22], and from the coefficient of ergodicity as in [21].

These norm-based bounds, however, are not satisfying for two reasons. First, there exist irreducible chains for which the bounds are not tight, so the condition number κ may seriously overestimate the sensitivity to perturbations. Secondly, the bounds generally provide little information about the relative error $|\pi_j - \tilde{\pi}_j|/\pi_j$ in individual stationary probabilities.

We remedy this situation in Section 3 by deriving perturbation bounds for individual stationary probabilities. The bounds are tight because for each Markov chain we can exhibit perturbations that satisfy the bounds with equality. Our analysis then leads to the uniform sensitivity of the stationary probabilities. In particular, if at least one large probability has low absolute sensitivity then all probabilities have low absolute sensitivity, and the chain is absolutely stable.

In Section 4 we relate our measure of sensitivity to the traditional condition numbers for the Markov problem. We prove that all relevant condition numbers for the problem $\pi^T A = 0$ are small multiples of each other.

After discussing the ramifications of the perturbation results on direct methods for computing the stationary probabilities in Section 5, we consider the case of nearly transient chains in Sections 6 and 7. We show that under special perturbations even small probabilities may have low relative sensitivity. In addition, we give conditions under which a nearly transient chain is absolutely stable under general perturbations. A summary of the results in Section 8 concludes the paper.

Notation

Unless otherwise specified, the infinity-norm is used for matrices and column vectors, and the one-norm for row vectors. The jth column of the identity matrix I is denoted by e_j, and the column vector of all ones is denoted by e.

The matrix P of order n denotes the matrix of transition probabilities of a n-state irreducible Markov chain with stationary distribution π^T. We define $A = I - P$. The matrix $\tilde{P} = P + E$ is a perturbation of P that represents the transition matrix of an irreducible chain with stationary distribution $\tilde{\pi}^T$.

2 Why Markov Chains Are Different

In this section we illustrate why the computation of the stationary distribution of an irreducible Markov chain should not be treated as just another linear system. It turns out that the rich structure of the Markov matrices regulates the sensitivity of the probabilities to perturbations in the matrix.

In general, the solution components of an ill-conditioned linear system are not necessarily all sensitive to perturbations. Some components may be quite insensitive as the example below from [5] demonstrates.

Example 1 Consider the linear system $Bx = b$, where

$$B = \begin{pmatrix} 0.4919 & 0.1112 & -0.6234 & -0.6228 \\ -0.5050 & -0.6239 & 0.0589 & 0.0595 \\ 0.5728 & -0.0843 & 0.7480 & 0.7483 \\ -0.4181 & 0.7689 & 0.2200 & 0.2204 \end{pmatrix}, \qquad b = \begin{pmatrix} 0.4351 \\ -0.1929 \\ 0.6165 \\ -0.8022 \end{pmatrix}.$$

The first three columns of B are nearly orthogonal while the last two columns are almost identical. The two-norm condition number of B is on the order of 10^3.

In [5] condition numbers for individual solution components are derived that indicate the relative change of a component to perturbations in the matrix. The first two components of x above have condition numbers of about one while the condition numbers for the last two components are on the order of 10^3. Hence the first two components are insensitive to perturbations while the last two are very sensitive to perturbations.

This is also evident when the "exact" solution x computed with 16-digit arithmetic is compared with the solution $\tilde{x}$ computed with 4-digit arithmetic, which can be viewed as the solution to a perturbed problem,

$$x = \begin{pmatrix} 1.0000754142405776 \\ -.5000879795933286 \\ -.0242511388797165 \\ .0262451395500586 \end{pmatrix} \qquad \tilde{x} = \begin{pmatrix} 1.000 \\ -.5003 \\ -.0589 \\ .06090 \end{pmatrix}$$

The first two components of $\tilde{x}$ are accurate to almost four digits, whereas the last two have no accuracy whatsoever.

In the example above, no component of x is particularly small, so the sensitivity of the first two components is not a result of their size. Hence, a general ill-conditioned linear system may give rise to both, large insensitive probabilities and large sensitive probabilities. We will show in Section 3 that this cannot happen for Markov chains.

Regarding the effects of perturbations on Markov chains it is also important to distinguish between absolute sensitivity and relative sensitivity, as the following example illustrates.

Example 2 For the 3-state chain whose transition matrix is

$$P(\epsilon) = \begin{pmatrix} 0 & 1-\epsilon & \epsilon \\ 1-\epsilon & 0 & \epsilon \\ 1 & 0 & 0 \end{pmatrix},$$

the associated stationary distribution is

$$\pi^T(\epsilon) = \left(\tfrac{1}{(2-\epsilon)(1+\epsilon)} \quad \tfrac{1-\epsilon}{(2-\epsilon)(1+\epsilon)} \quad \tfrac{\epsilon}{1+\epsilon} \right).$$

If $P = P(10^{-8})$ is perturbed to become $\tilde{P} = P(10^{-4})$, then the magnitude of the perturbation $E = \tilde{P} - P$ is

$$\|E\| = 2(10^{-4} - 10^{-8}).$$

Consider the change in the respective stationary distributions

$$\pi^T = \pi^T(10^{-8}) \quad \text{and} \quad \tilde{\pi}^T = \pi^T(10^{-4}).$$

As for the third component, its *absolute* change (the change relative to 1), is

$$|\pi_3 - \tilde{\pi}_3| = \left| \frac{10^{-8}}{1+10^{-8}} - \frac{10^{-4}}{1+10^{-4}} \right| = \frac{10^{-4} - 10^{-8}}{(1+10^{-4})(1+10^{-8})} \approx 10^{-4} - 10^{-8} = \frac{\|E\|}{2},$$

but its *relative* change (the change relative to the original value) is

$$\left| \frac{\pi_3 - \tilde{\pi}_3}{\pi_3} \right| = \left| 1 - \frac{10^{-4}(1+10^{-8})}{10^{-8}(1+10^{-4})} \right| \approx 10^4.$$

If the change in probabilities is assessed by comparing it to 1, then π_3 is as insensitive to perturbations as can be expected because the change of magnitude $\|E\|$ in the transition probabilities produces a change in π_3 of only $\|E\|/2$. But if the change in probabilities is assessed in a relative sense then the change in π_3 is large, and π_3 must be considered to be sensitive to perturbations.

As for the sensitivity of the other two probabilities π_1 and π_2, if $a^{\#}_{i,j}$ is element (i,j) in the group inverse $A^{\#}$ of $A = I - P$, then the absolute error in the jth stationary probability is bounded by [7]

$$|\pi_j - \tilde{\pi}_j| \leq \kappa_j \|E\|, \qquad \kappa_j = \max_i |a^{\#}_{i,j}|.$$

In this example, $\max_{i,j} |a^{\#}_{i,j}| < 1$, so all three stationary probabilities are insensitive in the absolute sense. Because π_1 and π_2 are both very close to 1/2, they are insensitive in the relative sense as well.

The preceding example motivates the following definition.

Definition 1 *An irreducible chain is said to be* absolutely stable *whenever each π_j is insensitive to perturbations in P in the absolute sense, i.e., whenever there is a small constant κ such that for all perturbations E,*

$$|\pi_j - \tilde{\pi}_j| \leq \kappa \|E\|, \qquad 1 \leq j \leq n,$$

where the term "small" is to be interpreted in the context of the underlying application.

Sufficient conditions for absolute stability are well-known. The results in [2, 7, 9, 15, 17, 18, 22], for instance, use the fact that a chain is absolutely stable if the group inverse $A^{\#}$ of $A = I - P$ has no large entries (relative to 1).

3 Component-Wise Analysis

In this section we derive tight upper bounds on the relative change in individual stationary probabilities, and we prove that all stationary probabilities show essentially the same sensitivity to perturbations in the transition probabilities.

We make use of the following properties of M–matrices, cf. [3]. If P is an irreducible stochastic matrix of order n then $A = I-P$ is a singular M–matrix of rank $n-1$. Moreover, if A_j is the principal submatrix of A obtained by deleting the jth row and column from A, then A_j is a nonsingular M–matrix. Hence $A_j^{-1} > 0$, and if e is the column vector of all ones, then $\|A_j^{-1}e\| = \|A_j^{-1}\|$. The following theorem demonstrates that the entries in A_j^{-1} determine the relative sensitivity of the jth stationary probability to perturbations in the transition probabilities.

Theorem 1 *If E_j denotes the matrix obtained by deleting the jth column of E, then*

$$\frac{\pi_j - \tilde{\pi}_j}{\pi_j} = \tilde{\pi}^T E_j A_j^{-1} e.$$

Furthermore,

$$\left|\frac{\pi_j - \tilde{\pi}_j}{\pi_j}\right| \leq \|E_j\| \, \|A_j^{-1}\|$$

and, for a particular j, there always exists a perturbation E for which equality is attained.

Proof: By applying a symmetric permutation to P, the states may be reordered so that a particular stationary probability occurs in the last position of π^T. Thus it suffices to prove the theorem for $j = n$.

With the partitioning

$$\pi^T = (\,\bar{\pi}^T \quad \pi_n\,), \qquad A = \begin{pmatrix} A_n & b \\ c^T & \delta \end{pmatrix},$$

$\pi^T A = 0^T$ implies $\bar{\pi}^T = -\pi_n c^T A_n^{-1}$. Replacing the last column by the vector of all ones produces a nonsingular matrix

$$N = \begin{pmatrix} A_n & e \\ c^T & 1 \end{pmatrix}$$

with inverse

$$N^{-1} = \begin{pmatrix} A_n^{-1}(I - e\bar{\pi}^T) & -\pi_n A_n^{-1} e \\ \bar{\pi}^T & \pi_n \end{pmatrix}.$$

The stationary distribution of the original chain is the solution of the nonsingular system

$$\pi^T N = e_n^T \quad \text{where} \quad e_n^T = (\,0 \quad \cdots \quad 0 \quad 1\,),$$

and the stationary distribution for the perturbed chain is the solution of another nonsingular system

$$\tilde{\pi}^T (N - F) = e_n^T \quad \text{where} \quad F = (\,E_n \quad 0\,).$$

Consequently,

$$\pi^T - \tilde{\pi}^T = -\tilde{\pi}^T F N^{-1},$$

so

$$\pi_n - \tilde{\pi}_n = -\tilde{\pi}^T (\,E_n \quad 0\,) \begin{pmatrix} -\pi_n A_n^{-1} e \\ \pi_n \end{pmatrix} = \pi_n \left(\tilde{\pi}^T E_n A_n^{-1} e\right),$$

and therefore

$$\frac{\pi_n - \tilde{\pi}_n}{\pi_n} = \tilde{\pi}^T E_n A_n^{-1} e.$$

Applying Hölder's inequality and $||A_n^{-1}e|| = ||A_n^{-1}||$ yields

$$\left|\frac{\pi_n - \tilde{\pi}_n}{\pi_n}\right| \leq ||\tilde{\pi}|| \, ||E_n A_n^{-1} e|| \leq ||E_n|| \, ||A_n^{-1}||.$$

To see that equality is always attainable, let k be the position where the largest component of $A_n^{-1}e$ occurs,

$$e_k^T A_n^{-1} e = ||A_n^{-1} e|| = ||A_n^{-1}||,$$

and let $E = \epsilon e(e_k - e_n)^T$. Then

$$\tilde{\pi}^T E_n = \epsilon e_k^T, \qquad ||E_n|| = \epsilon$$

implies

$$\frac{\pi_n - \tilde{\pi}_n}{\pi_n} = \tilde{\pi}^T E_n A_n^{-1} e = \epsilon e_k^T A_n^{-1} e = \epsilon ||A_n^{-1}|| = ||E_n|| \, ||A_n^{-1}||. \qquad \blacksquare$$

Corollary 1 *An irreducible chain is* absolutely *stable if and only if $\pi_j ||A_j^{-1}||$ is small for every $1 \leq j \leq n$.*

The results of Theorem 1 and its corollary suggest the following definitions.

Definition 2 *Let A_j be the principal submatrix obtained by deleting the jth row and column from A, and let π_j denote the jth stationary probability.*

The relative *condition number for π_j is defined to be*

$$\rho_j = ||A_j^{-1}||.$$

The absolute *condition number for π_j is defined to be*

$$\alpha_j = \pi_j ||A_j^{-1}||,$$

and the absolute condition number for the entire chain is defined to be

$$\alpha = \max_j \{\alpha_j\}.$$

Applying the new notation to Theorem 1 gives

$$\left|\frac{\pi_j - \tilde{\pi}_j}{\pi_j}\right| \leq \rho_j \, ||E_j||, \qquad |\pi_j - \tilde{\pi}_j| \leq \alpha_j \, ||E||.$$

Notice that if π_j is relatively well-conditioned, then it is absolutely well-conditioned, but not conversely, cf. Example 2.

Now we prove that the sensitivity of the stationary distribution is uniform in the sense that some π_j is relatively well-conditioned if and only if the entire chain is absolutely stable.

Theorem 2 *For every $1 \le j \le n$,*

$$|\pi_j - \tilde{\pi}_j| \le \min_i \rho_i \, \|E\|.$$

Consequently, some π_j is relatively well-conditioned if and only if every π_j is absolutely well-conditioned.

Proof: As in the proof of Theorem 1 assume that the states have been permuted so the best conditioned stationary probability is in the last position, $\rho_n = \min_j\{\rho_j\}$. If

$$N = \begin{pmatrix} A_n & e \\ c^T & 1 \end{pmatrix}$$

is the matrix obtained by replacing the last column of A by ones then

$$\pi^T - \tilde{\pi}^T = -\tilde{\pi}^T F N^{-1}, \qquad F = (\, E_n \quad 0\,).$$

From

$$N^{-1} = \begin{pmatrix} A_n^{-1}(I - e\pi^T) & -\pi_n A_n^{-1} e \\ \pi^T & \pi_n \end{pmatrix} = \begin{pmatrix} A_n^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I - e\pi^T & -\pi_n e \\ \pi^T & \pi_n \end{pmatrix},$$

it follows that

$$\pi_j - \tilde{\pi}_j = \tilde{\pi}^T F N^{-1} e_j = \begin{cases} \tilde{\pi}^T E_n A_n^{-1}(e_j - \pi_j e) & \text{if} \quad j < n. \\ \tilde{\pi}^T E_n A_n^{-1}(-\pi_j e) & \text{if} \quad j = n. \end{cases}$$

Since $\|e_j - \pi_j e\| = \max\{\pi_j,\, 1 - \pi_j\} < 1$ and $\|A_n^{-1} e\| = \|A_n^{-1}\| = \rho_n$, we have that

$$|\pi_j - \tilde{\pi}_j| \le \rho_n \|E_n\| \le \rho_n \|E\|, \qquad 1 \le j \le n. \qquad \blacksquare$$

Corollary 2 *If at least one stationary probability is relatively well-conditioned, then all* large *stationary probabilities must be relatively well-conditioned.*

Corollary 3 *If there is at least one* large *stationary probability that is absolutely well-conditioned, then all stationary probabilities must be absolutely well-conditioned and the chain is absolutely stable.*

In passing, it is of interest to note that the existence of a small ρ_j means that the $(n-1)$st singular value of A is large, cf. [1]. The following example from [24] shows that Theorem 2 can be too pessimistic with regard to the relative conditioning of small π_j for restricted classes of perturbations.

Example 3 The transition matrix

$$P = \begin{pmatrix} 1-\epsilon & \epsilon \\ \alpha & 1-\alpha \end{pmatrix}, \qquad \epsilon \ll \alpha,$$

gives rise to

$$A = \begin{pmatrix} \epsilon & -\epsilon \\ -\alpha & \alpha \end{pmatrix}, \qquad \pi^T = \frac{1}{\alpha+\epsilon}(\,\alpha \quad \epsilon\,).$$

If only the elements in the second row of P are perturbed, say each by magnitude $\eta > 0$, so the perturbed matrix remains stochastic, then

$$\frac{\pi_2 - \tilde{\pi}_2}{\pi_2} = \frac{\eta}{\alpha + \epsilon + \eta} \le \frac{\eta}{\alpha}.$$

Thus the tiny probability π_2 is relatively well-conditioned. This is because the chain is absolutely stable, as α is close to one, and because the perturbation leaves untouched the transient part, which is represented by the small element ϵ in position (1,2) of P. Section 6 discusses the sensitivity of nearly transient chains in greater detail.

Theorem 2 shows that the existence of one *relatively* well-conditioned π_j implies that all π_i are absolutely well-conditioned. This raises the question whether the existence of one ***absolutely*** well-conditioned π_j implies that all π_i must be absolutely well-conditioned. The example below shows that the answer is "no".

Example 4 For small $0 < \epsilon < 1$, let

$$P = \begin{pmatrix} 1-\epsilon & \epsilon/2 & \epsilon/2 \\ \epsilon/2 & 1-\epsilon & \epsilon/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}, \qquad \pi^T = \frac{1}{2+\epsilon}\begin{pmatrix} 1 & 1 & \epsilon \end{pmatrix}.$$

Since

$$A = \begin{pmatrix} \epsilon & -\epsilon/2 & -\epsilon/2 \\ -\epsilon/2 & \epsilon & -\epsilon/2 \\ -1/2 & -1/2 & 1 \end{pmatrix},$$

the relative condition numbers are

$$\rho_1 = \rho_2 = \frac{2}{3} + \frac{4}{3\epsilon} \quad \text{and} \quad \rho_3 = \frac{2}{\epsilon},$$

and the absolute condition numbers are

$$\alpha_1 = \alpha_2 = \frac{1}{2+\epsilon}\left(\frac{2}{3} + \frac{4}{3\epsilon}\right) \quad \text{and} \quad \alpha_3 = \frac{2}{2+\epsilon}.$$

As $\epsilon \to 0$, $\alpha_1 = \alpha_2 \to \infty$, but $\alpha_3 \to 1$. Although π_3 is absolutely well-conditioned, it is not relatively well-conditioned because it is small. The other two probabilities π_1 and π_2 are large but not absolutely well-conditioned. Thus, none of the probabilities is relatively well-conditioned and the chain is not absolutely stable.

Small stationary probabilities π_j are the ones that appear least likely to be relatively well-conditioned. Therefore it makes sense to try to determine those features of the matrix that may be responsible for the small size. The following theorem shows that a relatively well-conditioned π_j cannot be small. It also shows that a nearly reducible matrix A that is far from being uncoupled produces small π_j.

Theorem 3 *If Q is a permutation matrix such that $Q^T A Q = \begin{pmatrix} A_j & b_j \\ c_j^T & \delta_j \end{pmatrix}$, then*

$$\frac{1}{1+\rho_j} \le \pi_j \le \frac{\|b_j\|}{\|c_j^T\| + \|b_j\|}.$$

Proof: Let $\pi^T Q = \psi^T = \begin{pmatrix} \overline{\psi}^T & \pi_j \end{pmatrix}$. Since $\psi^T A = 0$ implies $\overline{\psi}^T = -\pi_j c_j^T A_j^{-1}$, Hölder's inequality gives the lower bound

$$1 - \pi_j = \overline{\psi}^T e = \pi_j |c_j^T A_j^{-1} e| \le \pi_j \rho_j.$$

To obtain the upper bound, use $\|c_j^T\| = -c_j^T e = \delta_j$ and $\delta_j \pi_j = -\overline{\psi}^T b_j$, and again apply Hölder's inequality,

$$\pi_j \|c_j^T\| = \pi_j \delta_j = -\overline{\psi}^T b_j \le \|\overline{\psi}^T\| \, \|b_j\| = (1 - \pi_j)\|b_j\|. \qquad \blacksquare$$

4 Condition Numbers

We showed in the previous section that the sensitivity of stationary probabilities is determined by the smallest ρ_j. In this section we relate this measure of sensitivity to the condition numbers that are traditionally used in this context: the norm of the group inverse and the norm of the pseudo-inverse. We will show that all these condition numbers are equivalent in the sense that any two of them are related by small multiplicative constants.

We first show that the sensitivity of the probabilities is proportional to the condition number of a certain non-singular linear system. If

$$\pi^T = (\bar{\pi}^T \quad \pi_n), \qquad A = \begin{pmatrix} A_n & b \\ c^T & \delta \end{pmatrix}, \qquad N = \begin{pmatrix} A_n & e \\ c^T & 1 \end{pmatrix},$$

then π represents the solution to the non-singular system $\pi^T N = e_n^T$.

Theorem 4 *If* $\rho_n = \|A_n^{-1}\|$ *then*

$$1 \leq \|N^{-1}\| \leq 2\rho_n.$$

Proof: An upper bound on the norm of the inverse

$$N^{-1} = \begin{pmatrix} A_n^{-1}(I - e\bar{\pi}^T) & -\pi_n A_n^{-1} e \\ \bar{\pi}^T & \pi_n \end{pmatrix} = \begin{pmatrix} A_n^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I - e\bar{\pi}^T & -\pi_n e \\ \bar{\pi}^T & \pi_n \end{pmatrix}$$

is given by

$$\|N^{-1}\| \leq 2\max\{1, \rho_n\} = 2\rho_n$$

because $Ae = 0$ implies $e = -A_n^{-1}b$, so $\rho_n = \|A_n^{-1}\| \geq 1$. The lower bound comes from the last row of N^{-1}. ■

Thus $\|N^{-1}\| \leq 2\rho_n$ guarantees that N is well-conditioned if π_n is relatively well-conditioned. As a consequence, any stable algorithm can accurately solve the system $\pi^T N = e_n^T$.

It is not clear, though, that π should be computed from $\pi^T N = e_n^T$ when the chain is absolutely stable but ρ_n is large. Although Theorem 1 insures that some ρ_j must be small, it need not be ρ_n as Example 2 demonstrates. It appears that instead of the last column one should insert e into the column associated with the smallest ρ_j. However, the determination of the smallest ρ_j is expensive, and this may be the reason why this is dismissed as "naïve" in [19] and not included in a comparison with other methods.

Surprisingly, with regard to the conditioning of the system, it does not matter which column of A is replaced by e. The following theorem demonstrates that the norm of N^{-1} and the norm of the group inverse $A^{\#}$ [14, 4] are small multiples of the smallest ρ_j. The group inverse matters because its norm indicates the absolute stability of the chain [15],

$$\pi - \tilde{\pi} = \tilde{\pi} E A^{\#}.$$

Hence, $|\pi_j - \tilde{\pi}_j| \leq \|E\| \, \|A^{\#}\|$, and the results of the previous section hold for all perturbations E that preserve the stochastic nature.

Theorem 5 *If $A^{\#}$ is the group inverse of A and if α is the absolute condition number for the chain then*

$$\frac{1}{2}\,\|A^{\#}\| \leq \|N^{-1}\| \leq 2\,\|A^{\#}\| + 1$$

and

$$\frac{1}{2}\,\alpha \leq \|A^{\#}\| \leq 4 \min_{1\leq i\leq n} \{\rho_i\}.$$

Proof: We derive the upper bounds first. The group inverse of A can be written as [4, 7, 14, 17],

$$\begin{aligned} A^{\#} &= (I - e\pi^T)\begin{pmatrix} A_n^{-1} & 0 \\ 0 & 0 \end{pmatrix}(I - e\pi^T) \\ &= \begin{pmatrix} (I - e\bar{\pi}^T)A_n^{-1}(I - e\bar{\pi}^T) & -\pi_n(I - e\bar{\pi}^T)A_n^{-1}e \\ -\bar{\pi}^T A_n^{-1}(I - e\bar{\pi}^T) & \pi_n\bar{\pi}^T A_n^{-1}e \end{pmatrix}. \end{aligned}$$

A symmetric permutation can bring any principal submatrix A_j of A to the upper left-hand corner of $Q^T AQ$. Then $(Q^T AQ)^{\#} = Q^T A^{\#} Q$, and

$$\psi^T = \pi^T Q = (\bar{\psi}^T \quad \pi_j)$$

imply

$$\begin{aligned} Q^T A^{\#} Q &= (I - e\psi^T)\begin{pmatrix} A_j^{-1} & 0 \\ 0 & 0 \end{pmatrix}(I - e\psi^T) \\ &= \begin{pmatrix} (I - e\bar{\psi}^T)A_j^{-1}(I - e\bar{\psi}^T) & -\pi_j(I - e\bar{\psi}^T)A_j^{-1}e \\ -\bar{\psi}^T A_j^{-1}(I - e\bar{\psi}^T) & \pi_j\bar{\psi}^T A_j^{-1}e \end{pmatrix}. \end{aligned}$$

The penultimate equality gives the second upper bound $\|A^{\#}\| = \|Q^T A^{\#} Q\| \leq 4\rho_j$ for any j. The first upper bound follows from

$$N^{-1} = \begin{pmatrix} I & -e \\ -c^T & -\delta \end{pmatrix} A^{\#} + e_n e_n^T,$$

which can be verified by using the second expression for $A^{\#}$, so that

$$\|N^{-1}\| \leq 2\|A^{\#}\| + 1.$$

To establish the lower bounds, use the expressions for $Q^T A^{\#} Q$ and $A^{\#}$ to write

$$\pi_j A_j^{-1} = (\,I \quad -e\,)\; Q^T A^{\#} Q \begin{pmatrix} \pi_j I \\ -\bar{\psi}^T \end{pmatrix}$$

and

$$A^{\#} = \begin{pmatrix} I - e\bar{\pi}^T & 0 \\ -\bar{\pi}^T & 0 \end{pmatrix} N^{-1}.$$

Hence $\|A^{\#}\| \leq 2\|N^{-1}\|$ and, for every j,

$$\alpha_j = \pi_j \|A_j^{-1}\| \leq 2\|A^{\#}\|. \qquad \blacksquare$$

This theorem establishes three facts: the conditioning of N is proportional to that of the best conditioned submatrix of order $n-1$ of A; the norm of the group inverse is proportional to the norm of this best conditioned submatrix; and the condition number used traditionally for linear system solution and the condition number used traditionally for Markov chains are equivalent.

So far we have viewed the stationary probabilities π as a solution to two different linear systems: the singular system $\pi^T A = 0$ and the non-singular system $\pi^T N = e_n^T$. There is a yet a third linear system of which π is a solution,

$$\pi^T M = e_{n+1}^T, \qquad M = (A \quad e).$$

The augmented matrix M is of order $n \times (n+1)$ and has full row rank. If the perturbed system is $\tilde{\pi}^T(M+E) = e_{n+1}^T$ then

$$\pi^T - \tilde{\pi}^T = \tilde{\pi}^T E M^\dagger,$$

where $M^\dagger$ is the Moore-Penrose pseudo-inverse [4, 10], and

$$|\pi_j - \tilde{\pi}_j| \leq \|\tilde{\pi}\| \, \|E\| \, \|M^\dagger\|.$$

The next theorem shows that $M^\dagger$ is a small multiple of the smallest ρ_j.

Theorem 6

$$\frac{\|A^\#\|}{4} \leq \|M^\dagger\| \leq 2\,\|A^\#\|.$$

Proof: Properties of the Moore-Penrose and group pseudo-inverses [4] allow us to write

$$M^\dagger = M^\dagger M M^\dagger = M^\dagger M \begin{pmatrix} A^\# \\ \pi^T \end{pmatrix} M M^\dagger = M^\dagger M \begin{pmatrix} A^\# \\ \pi^T \end{pmatrix}$$

and

$$\begin{pmatrix} A^\# \\ 0 \end{pmatrix} = \begin{pmatrix} A^\# \\ 0 \end{pmatrix} M \begin{pmatrix} A^\# \\ 0 \end{pmatrix} = \begin{pmatrix} A^\# \\ 0 \end{pmatrix} M M^\dagger M \begin{pmatrix} A^\# \\ 0 \end{pmatrix} = \begin{pmatrix} A^\# A & 0 \\ 0 & 0 \end{pmatrix} M^\dagger A A^\#.$$

Combining these identities with

$$M M^\dagger = I, \qquad M^\dagger M = \begin{pmatrix} I - ee^T/n & 0 \\ 0 & 1 \end{pmatrix}$$

and $AA^\# = A^\# A = I - e\pi^T$ produces

$$\frac{\|A^\#\|}{4} \leq \|M^\dagger\| \leq 2 \max\{\|A^\#\|, 1\}.$$

To show that $\|A^\#\| \geq 1$, use the last equality to derive a lower bound on the maximal row sum,

$$\|A^\# A\| \geq 1 - \pi_j + \sum_{i \neq j} \pi_i = 2 \sum_{i \neq j} \pi_i, \qquad 1 \leq j \leq n.$$

Since there exists at least one $\pi_j \geq 1/n$,

$$\|A^\# A\| \geq 2\frac{n-1}{n} \geq 1, \qquad n > 1,$$

and $\|A^\#\| \geq 1/\|A\| \geq 1$. ∎

Therefore, the condition number of the augmented system is proportional to the group inverse.

Below is a summary of the results on condition numbers of irreducible chains.

Corollary 4 *For a n-state irreducible Markov chain, the following statements are equivalent.*

- *At least one stationary probability π_j is relatively well-conditioned.*
- *The chain is absolutely stable.*
- *The matrix N and the system $\pi^T N = e_n^T$ are well-conditioned, regardless of the size of ρ_n.*
- *All entries of the group inverse $A^{\#}$ are small.*
- *The matrix $M = (A \quad e)$ and the system $\pi^T M = e_{n+1}^T$ are well conditioned.*
- *All entries of Kemeny and Snell's "fundamental matrix" $Z = (A + e\pi^T)^{-1}$ are small.*

The last statement is derived from the identity $Z = A^{\#} + e\pi^T$ [14], so

$$\|A^{\#}\| - 1 \leq \|Z\| \leq \|A^{\#}\| + 1.$$

Corollary 5 *An irreducible chain can only have stationary probabilities that are all absolutely sensitive if all probabilities are relatively sensitive. This, in turn, can only happen, if the matrices N^{-1}, $A^{\#}$, $M^{\dagger}$, and Z have large norms.*

5 Algorithms

In light of the perturbation results from the previous sections, we now discuss direct methods for computing the stationary probabilities.

It turns out that the algorithms in [1, 2, 6, 12, 17, 16, 23, 25] all amount to solving the nonsingular system $N^T\pi = e_n$ by Gaussian elimination. Let $A_n^T = L_n U_n$ be the LU factorization of A_n^T (this is the LU factorization of A^T without the last row of L). Pivoting is not necessary because A^T is column diagonally dominant and the growth factor is at most one [8]. Although the pivots are positive in exact arithmetic, as A_n^T is a nonsingular M–matrix, finite precision arithmetic may produce a zero or negative pivot [6]. However, this can be avoided with a diagonal adjustment scheme [11, 25, 2].

If at least one stationary probability is relatively well-conditioned, i.e., if the chain is absolutely stable, then the results of the last section obviate the problem addressed in [1, 2, 12] of having to locate a well-conditioned principal submatrix A_j in A because $N^T\pi = e_n$ is a well-conditioned system. Hence one can compute the LU factorization

$$N^T = \begin{pmatrix} A_n^T & c \\ e^T & 1 \end{pmatrix} = LU = \begin{pmatrix} L_n & 0 \\ e^T U_n^{-1} & 1 \end{pmatrix} \begin{pmatrix} U_n & L_n^{-1}c \\ 0 & 1/\pi_n \end{pmatrix}.$$

The solution of the lower triangular system $Lz = e_n$ is simply $z = e_n$, and the solution of $U\pi = e_n$ amounts to solving $U_n x = -L_n^{-1}c$ and setting

$$\pi^T = \frac{1}{1 + e^T x} (x^T \quad 1).$$

Accordingly, the linear system $N^T\pi = e_n$ can be solved in three steps:

1. Factor $A_n^T = L_n U_n$, possibly with diagonal adjustment.
2. Solve $L_n y = c$ and $U_n x = -y$.
3. Normalize $\pi^T = \frac{1}{1+e^T x} \begin{pmatrix} x^T & 1 \end{pmatrix}$.

The diagonal elements of U_n produced in step 1 can be used to estimate the absolute stability [17]. A backward error analysis [12, 2] insures that this algorithm is stable, and numerical experiments suggest it is as accurate as the methods recommended in [19]. A symmetric pivoting strategy combined with a diagonal adjustment scheme is proposed in [11, 25] to solve nearly uncoupled Markov chains, which are inherently ill-conditioned.

Alternatively, the full column-rank system

$$M^T \pi = e_{n+1}, \qquad M = \begin{pmatrix} A & e \end{pmatrix}$$

can be solved either by applying the QR factorization to M, which requires $(4/3)n^3$ floating point operations [10], or else by applying the QR factorization to A [9].

6 Sensitivity of Nearly Transient Chains

In this section we examine the sensitivity of stationary probabilities in the special case of irreducible chains with nearly transient states, i.e. chains in which the states can be ordered so the transition matrix is almost block triangular. We prove two results, one for special classes of perturbations and one for general perturbations.

The first theorem establishes a result similar to the one in [24]. It says that *small* stationary probabilities of an absolutely stable chain are *relatively* well-conditioned if only the states corresponding to these probabilities are perturbed and all other states remain unaffected.

Theorem 7 *If E can be symmetrically permuted so that*

$$E = \begin{pmatrix} 0 \\ E_2 \end{pmatrix}, \qquad \|E\| = \epsilon,$$

and $\pi^T = \begin{pmatrix} \pi_1^T & \pi_2^T \end{pmatrix}$ is partitioned conformably then

$$\frac{|\pi_j - \tilde{\pi}_j|}{\|\tilde{\pi}_2\|} \leq 4\epsilon \min_i \rho_i, \qquad 1 \leq j \leq n.$$

Proof: According to [15] and Theorem 5,

$$\pi^T - \tilde{\pi}^T = \tilde{\pi}^T E A^{\#} = \tilde{\pi}_2^T E_2 A^{\#}$$

and $\|A^{\#}\| \leq 4 \min_i \rho_i$. ■

The second theorem concerns nearly transient chains whose matrix is almost block upper triangular,

$$P = \begin{pmatrix} \overbrace{P_{11}}^{r} & \overbrace{P_{12}}^{s} \\ P_{21} & P_{22} \end{pmatrix} \begin{matrix} \}\, r \\ \}\, s \end{matrix}, \qquad \|P_{21}\| = \epsilon << 1,$$

and makes no restrictions on the structure of the perturbations. We show the following: If $A_{11} = I - P_{11}$ has an inverse with small norm; if $A_{22} = I - P_{22}$ has a principal submatrix of order $s-1$ whose inverse has small norm; and if ϵ is small then π_n is relatively well-conditioned and the chain is absolutely stable.

Theorem 8 *Let*

$$A = \begin{pmatrix} A_n & b \\ c^T & \delta \end{pmatrix}, \qquad A_n = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

where A_n is the principal submatrix obtained by deleting the nth row and column of A. If $\|A_{21}\| = \epsilon$ and $\epsilon\|A_{22}^{-1}\| < 1$ then

$$\rho_n \leq \frac{2\max\{\|A_{11}^{-1}\|, \|A_{22}^{-1}\|\}}{1 - \epsilon\|A_{22}^{-1}\|}.$$

Proof:

$$A_n = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ A_{21} & 0 \end{pmatrix} = T + K = T(I + T^{-1}K).$$

If $\|T^{-1}K\| < 1$ then, §2.3.4 in [10],

$$\rho_n = \|A_n^{-1}\| \leq \|T^{-1}\| \, \|(I + T^{-1}K)^{-1}\| \leq \frac{\|T^{-1}\|}{1 - \|T^{-1}K\|}.$$

Since A is a M–matrix, $A_{ii}^{-1} > 0$, $A_{ij} < 0$ and $b < 0$. Let $b^T = \begin{pmatrix} b_1^T & b_2^T \end{pmatrix}$ be conformably partitioned with A_n. Then $Ae = 0$ implies

$$0 \leq -A_{11}^{-1}A_{12}e = e + A_{11}^{-1}b_1 \leq e$$

and $\|A_{11}^{-1}A_{12}\| = \|A_{11}^{-1}A_{12}e\| \leq 1$. An analogous derivation shows $\|A_{22}^{-1}A_{21}\| \leq 1$. Thus,

$$T^{-1}K = \begin{pmatrix} -A_{11}^{-1}A_{12}A_{22}^{-1}A_{21} & 0 \\ A_{22}^{-1}A_{21} & 0 \end{pmatrix}$$

with

$$\|T^{-1}K\| \leq \max\{\|A_{11}^{-1}A_{12}A_{22}^{-1}A_{21}\|, \|A_{22}^{-1}A_{21}\|\} \leq \|A_{22}^{-1}A_{21}\| \leq \epsilon\|A_{22}^{-1}\|.$$

Moreover,

$$T^{-1} = \begin{pmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ 0 & A_{22}^{-1} \end{pmatrix}$$

and

$$\|T^{-1}\| \leq (1 + \|A_{11}A_{12}\|)\max\{\|A_{11}^{-1}\|, \|A_{22}^{-1}\|\} \leq 2\max\{\|A_{11}^{-1}\|, \|A_{22}^{-1}\|\}.$$

So

$$\rho_n \leq \frac{\|T^{-1}\|}{1 - \|T^{-1}K\|} \leq \frac{2\max\{\|A_{11}^{-1}\|, \|A_{22}^{-1}\|\}}{1 - \epsilon\|A_{22}^{-1}\|}. \qquad \blacksquare$$

7 Small Probabilities In Nearly Transient Chains

Again, let P be the transition matrix of a nearly transient chain and let

$$A = I - P = \begin{array}{c} \begin{matrix} \overbrace{\hphantom{A_{11}}}^{r} & \overbrace{\hphantom{A_{12}}}^{s} \end{matrix} \\ \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \end{array} \begin{matrix} \}\, r \\ \}\, s \end{matrix}, \qquad \|A_{21}\| = \epsilon << 1.$$

If $\pi^T = (\bar{\pi}_1^T \;\; \bar{\pi}_2^T)$, then due to the transient nature $\bar{\pi}_1^T \rightarrow 0$ as $\epsilon \rightarrow 0$. In particular, $\bar{\pi}_1^T = -\bar{\pi}_2^T A_{21} A_{11}^{-1}$ implies

$$\|\bar{\pi}_1^T\| \leq \epsilon \|A_{11}^{-1}\|.$$

This means, the smaller the norm of the inverse of the leading principal submatrix, the smaller the stationary probabilities for the corresponding states.

For nearly transient chains with a finer block structure, say

$$A = \begin{pmatrix} A_{11} & * & * & \dots & * & * \\ F_{21} & A_{22} & * & \dots & * & * \\ F_{31} & F_{32} & A_{33} & \dots & * & * \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & & A_{k-1,k-1} & * \\ F_{k1} & F_{k2} & F_{k3} & \dots & F_{k,k-1} & A_{kk} \end{pmatrix}, \qquad \pi = \begin{pmatrix} \bar{\pi}_1 \\ \bar{\pi}_2 \\ \bar{\pi}_3 \\ \vdots \\ \vdots \\ \bar{\pi}_k \end{pmatrix}$$

and

$$\left\| \begin{pmatrix} F_{j+1,j} \\ F_{j+2,j} \\ \vdots \\ F_{k,j} \end{pmatrix} \right\| = \epsilon_j, \qquad 1 \leq j \leq k-1,$$

the same should be true: the trailing stationary probabilities tend to be larger than the leading ones. We will quantify this statement by providing bounds in terms of ϵ_j on the probabilities $\bar{\pi}_j$ associated with each block.

To this end, we present a lemma that makes it possible to proceed inductively by applying the above 2×2 case to successive diagonal blocks, moving from top to bottom. First apply the 2×2 case to the probabilities associated with A_{11}, giving $\|\bar{\pi}_1\| \leq \epsilon_1 \|A_{11}^{-1}\|$. The lemma provides a perturbation of size ϵ that essentially uncouples A_{11} from the remaining blocks. As a consequence one can apply the same considerations recursively to a perturbation of the remaining blocks.

In particular, the lemma shows that the remaining probabilities are the exact probabilities of a perturbed problem of the same form (the only difference being that the sum of the probabilities is less than one).

Lemma 1 *Let*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \qquad \|A_{21}\| \leq \epsilon.$$

If

$$\Delta = \frac{A_{21} e \bar{\pi}_1^T A_{12}}{\bar{\pi}_2^T A_{21} e},$$

then $A_{22} + \Delta$ *is a singular M-matrix with*

$$\bar{\pi}_2^T (A_{22} + \Delta) = 0, \qquad (A_{22} + \Delta) e = 0, \qquad \|\Delta\| \leq \epsilon.$$

Proof: We first verify that Δ satisfies the required equations. From $\pi^T A = 0$ and $Ae = 0$ we get

$$r_1^T = \bar{\pi}_2^T A_{22} = -\bar{\pi}_1^T A_{12}, \qquad r_2 = A_{22} e = -A_{21} e,$$

so one can write

$$\Delta = -\frac{r_2 r_1^T}{\bar{\pi}_2^T r_2}.$$

Since $\bar{\pi}_2^T \Delta = -r_1^T$, it follows that $\bar{\pi}_2^T(A_{22} + \Delta) = 0$, and thus Δ satisfies the first equation.

To prove that Δ satisfies the second equation, observe that $\pi^T A = 0$ and $Ae = 0$ imply

$$\bar{\pi}_2^T r_2 = -\bar{\pi}_2^T A_{21} e = \bar{\pi}_1^T A_{11} e = -\bar{\pi}_2^T A_{12} e = r_1^T e.$$

Thus,

$$\Delta = -\frac{r_2 r_1^T}{r_1^T e},$$

so $\Delta e = -r_2$ and $(A_{22} + \Delta)e = 0$.

As for the bound on the norm of Δ, notice that r_1 and r_2 both consist entirely of non-negative elements since A is a M-matrix, so Δ consists entirely of non-positive elements. This means

$$\|\Delta\| = \left\| \frac{r_2 r_1^T}{r_1^T e} e \right\| = \|r_2\| \leq \epsilon.$$

Moreover, since all elements of Δ are non-positive, the off-diagonal elements in $A_{22} + \Delta$ are more negative than those of A_{22}. This implies with $(A_{22} + \Delta)e = 0$ that the diagonal elements must be non-negative. From $\pi > 0$ it follows that $A_{22} + \Delta$ must be irreducible, for otherwise a component of $\bar{\pi}_2$ would be zero. According to Corollary 1 in Section 3.5 of [26], the signs of the matrix elements and the irreducibility imply that every principal submatrix of $A_{22} + \Delta$ is a M-matrix. Therefore $A_{22} + \Delta$ is a singular M-matrix. ■

Now we can prove the following theorem which says that in a nearly transient chain, the size of the π_i in the jth block is controlled by the smallness of the preceding off-diagonal columns 1, ..., $j-1$, and by the condition of a perturbed jth diagonal block. The size of this perturbation is again determined by the smallness of the off-diagonal columns 1, ..., j. This implies that the trailing solution components tend to be larger than the leading ones.

Theorem 9 *Let*

$$A = \begin{pmatrix} A_{11} & * & * & \dots & * & * \\ F_{21} & A_{22} & * & \dots & * & * \\ F_{31} & F_{32} & A_{33} & \dots & * & * \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & & A_{k-1,k-1} & * \\ F_{k1} & F_{k2} & F_{k3} & \dots & F_{k,k-1} & A_{kk} \end{pmatrix}, \qquad \pi = \begin{pmatrix} \bar{\pi}_1 \\ \bar{\pi}_2 \\ \bar{\pi}_3 \\ \vdots \\ \vdots \\ \bar{\pi}_k \end{pmatrix}$$

with

$$\left\| \begin{pmatrix} F_{j+1,j} \\ F_{j+2,j} \\ \vdots \\ F_{k,j} \end{pmatrix} \right\| = \epsilon_j, \qquad 1 \leq j \leq k-1.$$

Then

$$\|\bar{\pi}_1^T\| \leq \epsilon_1 \kappa_1, \qquad \kappa_1 = \|A_{11}^{-1}\|,$$

and there exist matrices $X_{j+1,j+1}$, $1 \leq j \leq k-1$, *that satisfy*

$$\|A_{j+1,j+1} - X_{j+1,j+1}\| \leq \epsilon_1 + \dots + \epsilon_j,$$

and

$$\|\bar{\pi}_{j+1}^T\| \leq (\epsilon_1 + \dots + \epsilon_{j+1})\kappa_{j+1}, \qquad \kappa_{j+1} = \|X_{j+1}^{-1}\|.$$

Proof: The statements for $\bar{\pi}_1$ follow from the 2×2 block partitioning. Now apply the same argument recursively to the matrix

$$\bar{A}_{22} = \begin{pmatrix} A_{22} & * & \cdots & * & * \\ F_{32} & A_{33} & \cdots & * & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & & A_{k-1,k-1} & * \\ F_{k2} & F_{k3} & \cdots & F_{k,k-1} & A_{kk} \end{pmatrix} + \Delta,$$

where Δ is given by Lemma 1. For instance, X_2 is the leading diagonal block of $\bar{A}_{22}$ with $\|X_2^{-1}\| = \kappa_2$. Lemma 1 insures $\|\Delta\| \leq \epsilon_1$ and $\|A_{22} - X_2\| \leq \Delta \leq \epsilon_1$. Since the norm of the first off-diagonal column is bounded above by $\epsilon_1 + \epsilon_2$, Lemma 1 gives $\|\bar{\pi}_2^T\| \leq (\epsilon_1 + \epsilon_2)\kappa_2$. ∎

8 Concluding Remarks

The goal of this paper was to better understand how individual stationary probabilities change as the transition probabilities are perturbed.

Because most of our results were not intended to exploit any underlying structure, we measured all perturbations relative to 1 rather than relative to the magnitude of $A = I - P$ or relative to the structure of P. In other words, if $\tilde{P} = P + E$ is the transition matrix of a perturbed chain, then the relative perturbation is $\|E\|/\|P\| = \|E\|$, instead of $\|E\|/\|A\|$ or $\max_{ij} |e_{ij}|/p_{ij}$. The latter two measures result in a significantly different interpretation of the notions of condition and sensitivity to perturbations. For example, we regard

$$P(\epsilon) = \begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix} \quad \text{and} \quad A(\epsilon) = I - P = \begin{pmatrix} \epsilon & -\epsilon \\ -\epsilon & \epsilon \end{pmatrix}$$

for small ϵ as a sensitive Markov chain because small perturbations (relative to 1) in the transition probabilities can greatly affect the stationary probabilities. In the terminology of this paper, the relative condition numbers for π_1 and π_2 are $\|A_1^{-1}\| = \|A_2^{-1}\| = 1/\epsilon$. However, if we had instead decided to measure perturbations relative to the magnitude of A or relative to the structure of P, then we would have to view the stationary probabilities as insensitive to perturbations [15, 27].

The main results in our paper are the following: Given are two transition matrices P and $P+E$ of irreducible n-state Markov chains. Note that E is not necessarily constrained to be "small." The relative sensitivity of a stationary probability π_j is

$$\frac{\pi_j - \tilde{\pi}_j}{\pi_j} = \tilde{\pi}^T E_j A_j^{-1} e.$$

where E_j is the matrix obtained by deleting the jth column of E, and A_j is the principal submatrix obtained by deleting the jth row and column from $A = I - P$. Moreover,

$$\left| \frac{\pi_j - \tilde{\pi}_j}{\pi_j} \right| \leq \|E_j\| \, \|A_j^{-1}\|$$

with equality possible for each j. The absolute sensitivity of π_j is governed by the fact that

$$|\pi_j - \tilde{\pi}_j| \leq \min_i \|A_i^{-1}\| \, \|E\|.$$

In particular, if at least one π_j has low relative sensitivity, or if at least one *large* π_j has low absolute sensitivity, then all stationary probabilities have low absolute sensitivity, and the chain is absolutely stable. Furthermore, all other relevant condition numbers for the problem $\pi^T A = 0$ are small multiples of $\min_i \|A_i^{-1}\|$.

In the case of nearly transient chains we showed that under special perturbations even small probabilities may have low relative sensitivity. In addition, we gave conditions under which a nearly transient chain is absolutely stable under more general perturbations.

References

[1] J. Barlow, *On the smallest positive singular value of a singular M-matrix with applications to ergodic Markov chains*, SIAM J. Alg. Disc. Methods, 7 (1986), pp. 414–24.

[2] ——, *Error bounds and condition estimates for the computation of null vectors with applications to Markov chains*, SIAM J. Matrix Anal. Appl., (1992).

[3] A. Berman and R. Plemmons, *Nonnegative Matrices In The Mathematical Sciences*, Academic Press, New York, 1979.

[4] S. Campbell and C. Meyer, *Generalized Inverses Of Linear Transformations*, Dover Publications, New York (1979 Edition by Pitman Pub. Ltd., London), 1991.

[5] S. Chandrasekaran and I. Ipsen, *On the sensitivity of solution components in linear systems of equations.* 1992.

[6] R. Funderlic and J. Mankin, *Solution of homogeneous systems of linear equations arising from compartmental models*, SIAM J. Sci. Stat. Comput., 2 (1981), pp. 375–83.

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

[8] R. Funderlic, M. Neumann, and R. Plemmons, *LU decompositions of generalized diagonally dominant matrices*, Numer. Math., 40 (1982), pp. 57–69.

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

[10] G. Golub and C. van Loan, *Matrix Computations*, The Johns Hopkins Press, 1989.

[11] W. Grassmann, M. Taksar, and D. Heyman, *Regenerative analysis ans steady state distributions for Markov chains*, Operations Research, 33 (1985), pp. 1107–1116.

[12] W. Harrod and R. Plemmons, *Comparison of some direct methods for computing stationary distributions of Markov chains*, SIAM J. Sci. Stat. Comput., 5 (1984), pp. 453–69.

[13] J. Kemeny and J. Snell, *Finite Markov Chains*, D. Van Nostrand, New York, 1960.

[14] C. Meyer, *The role of the group inverse in the theory of finite Markov chain*, SIAM Review, 17 (1975), pp. 443–464.

[15] ——, *The condition of a finite Markov chain and perturbations bounds for the limiting probabilities*, SIAM J. On Alg. Dis Meth., 3 (1980), pp. 273–283.

[16] ——, *The character of a finite Markov chain*, in Proceedings of the IMA Workshop on Linear Algebra, Markov Chains, and Queueing Models, C. Meyer and R. Plemmons, eds., Springer Verlag, 1992.

[17] ——, *Sensitivity of Markov chains*, SIAM J. Matrix Anal., To appear (1992).

[18] C. MEYER AND G. STEWART, *Derivatives and perturbations of eigenvectors*, SIAM J. Num. Anal., 25 (1988), pp. 679–691.

[19] C. PAIGE, G. STYAN, AND P. WACHTER, *Computation of the stationary distribution of a Markov chain*, J. Statist. Comput. Simul., 4 (1975), pp. 173–86.

[20] P. SCHWEITZER, *Perturbation theory and finite Markov chains*, J. Applied Prob., 5 (1968), pp. 401–413.

[21] E. SENETA, *Sensitivity analysis, ergodicity coefficients, and rank-one updates for finite Markov chains*, in Numerical Solution of Markov Chains, Probability: Pure and Applied, Vol. 8, W. Stewart, ed., Marcel Dekker, 1991, pp. 121–129.

[22] G. STEWART, *On the sensitivity of nearly uncoupled Markov chains*, in Numerical Solution of Markov Chains, Probability: Pure and Applied, Vol. 8, W. Stewart, ed., 1991, pp. 105–119.

[23] ——, *Gaussian elimination, perturbation theory and Markov chains*, in Proceedings of the IMA Workshop on Linear Algebra, Markov Chains, and Queueing Models, C. Meyer and R. Plemmons, eds., Springer Verlag, 1992.

[24] ——, *On the perturbation of Markov chains with nearly transient states*, tech. report, University Of Maryland, Comp. Sci., 1992.

[25] G. STEWART AND G. ZHANG, *On a direct method for the solution of nearly uncoupled Markov chains*, Numer. Math., 59 (1991), pp. 1–11.

[26] R. VARGA, *Matrix Iterative Analysis*, Prentice-Hall, 1962.

[27] G. ZHANG, *On the sensitivity of the solution of nearly uncoupled Markov chains*, tech. report, University Of Maryland, Comp. Sci., CS TR 2601, 1991.