SENSITIVITY OF MARKOV CHAINS

By

Carl D. Meyer

IMA Preprint Series # 919

February 1992
SENSITIVITY OF MARKOV CHAINS

Carl D. Meyer

Mathematics Department
North Carolina State University
Raleigh, NC 27695-8205

NCSU Math Tech. Report #1271992709

October, 1991
ABSTRACT

It is well-known that if an irreducible Markov chain has a subdominant eigenvalue which is close to 1, then the chain is ill-conditioned. However, the converse of this statement has heretofore been unresolved. The purpose of this article is to address the following question—if the subdominant eigenvalues of an irreducible chain are well-separated from 1, can we be sure that the chain is well-conditioned? In other words, do the subdominant eigenvalues somehow provide complete information about the sensitivity of the stationary probabilities—or do we really need to include singular values in the discussion?
SENSITIVITY OF MARKOV CHAINS

Carl D. Meyer†

1. INTRODUCTION

The problem under consideration is that of analyzing the effects of small perturbations to the transition probabilities of a finite, irreducible, homogeneous Markov chain. More precisely, if $P_{n \times n}$ is the transition probability matrix for such a chain, and if $\pi^T = (\pi_1, \pi_2, \ldots, \pi_n)$ is the stationary distribution vector satisfying $\pi^T P = \pi^T$ and $\sum_{i=1}^{n} \pi_i = 1$, the goal is to describe the effect on $\pi^T$ when $P$ is perturbed by a matrix $E$ such that $\tilde{P} = P + E$ is the transition probability matrix of another irreducible Markov chain.

Schweitzer (1968) provided the first perturbation analysis in terms of Kemeny & Snell’s (1960) “fundamental matrix” $Z = (A + e\pi^T)^{-1}$ in which $A = I - P$ and $e$ is a column of 1’s. If $A^#$ denotes the group inverse of $A$ [see Meyer (1975) or Campbell & Meyer (1991)], then

$$Z = (A + e\pi^T)^{-1} = A^# + e\pi^T.$$ 

But in virtually all applications involving $Z$, the term $e\pi^T$ is redundant—i.e., all relevant information is contained in $A^#$. In particular, if $\tilde{\pi}^T = (\tilde{\pi}_1, \tilde{\pi}_2, \ldots, \tilde{\pi}_n)$ is the stationary distribution for $\tilde{P} = P + E$, then

$$\tilde{\pi}^T = \pi (I + EA^#)^{-1}$$  \hspace{1cm} (1.1)

and

$$\|\pi^T - \tilde{\pi}^T\| \leq \|E\| \|A^#\|$$  \hspace{1cm} (1.2)

† Department of Mathematics, North Carolina State University, Box 8205, Raleigh, NC 27695-8205. This work was supported in part by the National Science Foundation under grants DMS-9020915 and DDM-8906248.
in which \( \| \cdot \| \) can be either the 1-, 2-, or \( \infty \)-norm. If the \( j \)'th column and the \((i,j)\)-entry of \( A^\# \)

are denoted by \( A^\#_{ij} \) and \( a^\#_{ij} \), respectively, then

\[
|\pi_j - \bar{\pi}_j| \leq \|E\| \|A^\#_{ij}\| \tag{1.3}
\]

and

\[
\max_j |\pi_j - \bar{\pi}_j| \leq \|E\|_\infty \max_{i,j} |a^\#_{ij}|. \tag{1.4}
\]

There are chains for which equality in (1.4) is realized—e.g., consider

\[
P = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \quad \text{with} \quad E = \begin{pmatrix} \epsilon & -\epsilon \\ \epsilon & -\epsilon \end{pmatrix}.
\]

Moreover, if the transition probabilities are analytic functions of a parameter \( t \) so that \( P = P(t) \), then

\[
\frac{d\pi}{dt} = \pi \frac{dP}{dt} A^\# \quad \text{and} \quad \frac{d\pi_j}{dt} = \pi \frac{dP}{dt} A^\#_{ij}. \tag{1.5}
\]

The results (1.1) and (1.2) are due to Meyer (1980), and (1.3) appears in Golub & Meyer (1986). The inequality (1.4) was given by Funderlic & Meyer (1986), and the formulas (1.5) are derived in Golub & Meyer (1986) and Meyer & Stewart (1988).

These facts make it absolutely clear that the entries in \( A^\# \) determine the extent to which \( \pi^T \)

is sensitive to small changes in \( P \), so, on the basis of (1.4), it is natural to adopt the following definition.

**Definition 1.1.** [Funderlic & Meyer (1986)] The **condition** of a Markov chain with a transition matrix \( P \) is measured by the size of its **condition number** which is defined to be

\[
\kappa = \max_{i,j} |a^\#_{ij}|
\]

where \( a^\#_{ij} \) is the \((i,j)\)-entry in the group inverse \( A^\# \) of \( A = I - P \). It is an elementary fact that \( \kappa \) is invariant under permutations of the states of the chain.
Sensitivity Of Markov Chains

C. D. Meyer

It has been known for some time, and it is easy to prove (see the proof of Theorem 2.1), that if there exists a subdominant eigenvalue of \( P \) which is close to 1, then \( \kappa \) must be large and the chain exhibits sensitivities. However, the converse of this statement has heretofore been unresolved, and our purpose is to focus on this issue. More precisely, we address the following question.

*If the subdominant eigenvalues of an irreducible Markov chain are well-separated from 1, can we be sure that the chain is well-conditioned? In other words, do the subdominant eigenvalues of \( P \) (or equivalently, the nonzero eigenvalues of \( A \)) somehow provide complete information about the sensitivity of the chain—or do we really need to know something about the singular values of \( A \)?*

The conjecture that \( \max_{i,j} |a_{ij}^\#| \) is somehow controlled by the nonzero eigenvalues of \( A \) is contrary to what is generally true—a standard example is the triangular matrix

\[
T_{n \times n} = \begin{pmatrix}
1 & -2 & 0 & \cdots & 0 & 0 \\
0 & 1 & -2 & \cdots & 0 & 0 \\
0 & 0 & 1 & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 1 & -2 \\
0 & 0 & 0 & \cdots & 0 & 1
\end{pmatrix}
\text{ with } T^{-1} = \begin{pmatrix}
1 & 2 & 4 & \cdots & 2^{n-2} & 2^{n-1} \\
0 & 1 & 2 & \cdots & 2^{n-3} & 2^{n-2} \\
0 & 0 & 1 & \ddots & 2^{n-4} & 2^{n-3} \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1 & 2 \\
0 & 0 & 0 & \cdots & 0 & 1
\end{pmatrix}
\]

(1.6)

for which \( \max_{i,j} |T^{-1}|_{ij} \) is immense for even moderate values of \( n \), but the eigenvalues of \( T \) provide no clue whatsoever that this occurs. The fact that the eigenvalues are repeated or that \( T \) is nonsingular is irrelevant—consider a small perturbation of \( T \) or the matrices

\[ \tilde{T} = \begin{pmatrix} 0 & 0 \\ 0 & T \end{pmatrix} \text{ and } \tilde{T}^\# = \begin{pmatrix} 0 & 0 \\ 0 & T^{-1} \end{pmatrix}. \]

We will prove that, unlike the situation illustrated above, irreducible stochastic matrices \( P \) possess enough structure to guarantee that growth of the entries in \( A^\# \) is controlled by the nonzero eigenvalues of \( A = I - P \). As a consequence, it will follow that the sensitivity of an irreducible Markov chain is governed by the location of its subdominant eigenvalues.

2. THE MAIN RESULT

In the sequel, it is convenient to adopt the following terminology and notation.
DEFINITION 2.1. For an irreducible Markov chain whose eigenvalues are \( \{1, \lambda_2, \lambda_3, \ldots, \lambda_n\} \), the character of the chain is defined to be the (necessarily real) number

\[
\chi = (1 - \lambda_2)(1 - \lambda_3) \cdots (1 - \lambda_n).
\]

It will follow from later developments that

\[
0 < \chi < n. \tag{2.1}
\]

A chain is said to be of “weak character” when \( \chi \) is close to 0, and the chain is said to have a “strong character” when \( \chi \) is significantly larger than 0.

If \( P = T^{-1} \begin{pmatrix} 1 & 0 \\ 0 & C \end{pmatrix} T \) where the spectral radius of \( C \) is less than 1, then

\[
A = T^{-1} \begin{pmatrix} 0 & 0 \\ 0 & I - C \end{pmatrix} T \quad \text{and} \quad A^# = T^{-1} \begin{pmatrix} 0 & 0 \\ 0 & (I - C)^{-1} \end{pmatrix} T,
\]

so \( \chi = \det (I - C) \) and \( \chi^{-1} = \det (I - C)^{-1} \). In other words, \( \chi \) and \( \chi^{-1} \) are the respective determinants of the nonsingular parts of \( A \) and \( A^# \) in the sense that

\[
\chi = A/\mathbb{R}(A) \quad \text{and} \quad \chi^{-1} = \det \left( A^#/\mathbb{R}(A) \right).
\]

It is also true that \( \chi^{-1} = \det (Z) \) where \( Z \) is Kemeny & Snell’s “fundamental matrix.”

The main result of this paper is the following theorem which establishes the connection between the condition of an irreducible chain and its character.

THEOREM 2.1. For an irreducible stochastic matrix \( P_{n \times n} \), let \( A = I - P \), and let \( \delta_{ij}(A) \) denote the deleted product of diagonal entries

\[
\delta_{ij}(A) = \prod_{k \neq i,j} a_{kk} = \prod_{k \neq i,j} (1 - p_{kk}).
\]

If \( \delta = \max_{i,j} \delta_{ij}(A) \) (the product of all but the two smallest diagonal entries), then the condition number \( \kappa \) is bounded by

\[
\frac{1}{n \min_{i \neq 1} |1 - \lambda_i|} \leq \kappa < \frac{2\delta(n - 1)}{\chi} \leq \frac{2(n - 1)}{\chi}. \tag{2.2}
\]
The proof of this theorem depends on exploiting the rich structure of \( A \), some of which is apparent, and some of which requires illumination. Before giving a formal argument, it is necessary to detail the various components of this structure, so the important facets are first laid out in §3 as a sequence of lemmas. After the necessary framework is in place, it will be a simple matter to connect the lemmas together in order to construct a proof—this is contained in §4.

By combining Theorem 2.1 with (1.4) and the other facts listed in §1, we arrive at the following conclusion.

**THEOREM 2.2.** The condition of an irreducible Markov chain is primarily governed by how close the subdominant eigenvalues of the chain are to 1. More precisely, if an irreducible chain is well-conditioned, then all subdominant eigenvalues must be well-separated from 1, and if all subdominant eigenvalues are well-separated from 1 in the sense that the chain has a strong character, then it must be well-conditioned.

It is a corollary of Theorem 2.1 that if \( \max_{\lambda_i \neq 1} |\lambda_i| << 1 \), then the chain is not overly sensitive, but it is important to underscore the point that the issue of sensitivity is not equivalent to the question of how close \( \max_{\lambda_i \neq 1} |\lambda_i| \) is to 1. Knowing that some \( |\lambda_i| \approx 1 \) is not sufficient to guarantee that the chain is sensitive—e.g., consider the well-conditioned periodic chain (or any small perturbation thereof) for which

\[
P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad A^\# = \frac{1}{3} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix}.
\]

### 3. THE UNDERLYING STRUCTURE

The purpose of this section is to organize relevant properties of \( A = I - P \) into a sequence of lemmas from which the formal proof of Theorem 2.1 can be constructed. Some of the more transparent or well-known features of \( A \) are stated in the first lemma.
**Lemma 3.1.** If $A = I - P$ where $P_{n \times n}$ is an irreducible stochastic matrix, then the following facts are either self-evident, or they are direct consequences of well known results found in Berman & Plemmons (1979) or Horn & Johnson (1991).

(3.1) $A$ as well as each principal submatrix of $A$ has strictly positive diagonal entries, and the off-diagonal entries are non-positive.

(3.2) $A$ is a singular M-matrix of rank $n - 1$.

(3.3) If $B_{k \times k}$ $(k < n)$ is a principal submatrix of $A$, then each of the following statements is true.

(a) $B$ is a nonsingular M-matrix.

(b) $B^{-1} \geq 0$.

(c) $\det(B) > 0$.

(d) $B$ is diagonally dominant.

(e) $\det(B) \leq b_{11}b_{22}\cdots b_{kk} \leq 1$.

Some of the less transparent structure of $A$ is illuminated in the following sequence of lemmas.

**Lemma 3.2.** If $P_{n \times n}$ is an irreducible stochastic matrix and if $A_i$ denotes the principal submatrix of $A = I - P$ obtained by deleting the $i^{th}$ row and column from $A$, then

$$\chi = \sum_{i=1}^{n} \det(A_i).$$

**Proof.** Suppose that the eigenvalues of $A$ are denoted by $\{\mu_1, \mu_2, \cdots, \mu_n\}$, and write characteristic equation for $A$ as

$$x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_1x + \alpha_0 = 0.$$ 

Each coefficient $\alpha_{n-k}$ is given by $(-1)^{k}$ times the sum of the product of the eigenvalues of $A$ taken $k$ at a time. That is,

$$\alpha_{n-k} = (-1)^{k} \sum_{1 \leq i_1 < \cdots < i_k \leq n} \mu_{i_1}\mu_{i_2}\cdots\mu_{i_k}. \quad (3.4)$$
Sensitivity Of Markov Chains

C. D. Meyer

But it is also a standard result from elementary matrix theory that each coefficient \( \alpha_{n-k} \) can be described as

\[
\alpha_{n-k} = (-1)^k \sum_{(all \; k \times k \; principal \; minors \; of \; A)}.
\]

Since 0 is a simple eigenvalue for \( A \), there is only one nonzero term in the sum (3.4) when \( k = n-1 \), and hence

\[
\alpha_1 = (-1)^{n-1}\mu_2\mu_3\cdots\mu_n = (-1)^{n-1}(1-\lambda_2)(1-\lambda_3)\cdots(1-\lambda_n) = (-1)^{n-1}\sum_{i=1}^{n} \text{det}(A_i).
\]

Therefore, \( \sum_{i=1}^{n} \text{det}(A_i) = \prod_{k=2}^{n}(1-\lambda_k) = \chi. \]

**Lemma 3.3.** If \( A_i \) denotes the principal submatrix of \( A = I - P \) obtained by deleting the \( i \)th row and column from \( A \), and if \( \pi_i \) is the \( i \)th stationary probability, then the character of the chain is given by

\[
\chi = \frac{\text{det}(A_i)}{\pi_i}.
\]

**Proof.** This result follows directly from Lemma 3.2 and the fact that the stationary distribution \( \pi^T \) is given by the formula

\[
\pi^T = \frac{1}{\sum_{i=1}^{n} \text{det}(A_i)} (\text{det}(A_1), \text{det}(A_2), \ldots, \text{det}(A_n))
\]
[see Golub & Meyer (1986) or Iosifescu (1980), pg. 123].

The mean return time for the \( k \)th state is \( R_k = 1/\pi_k \) [see Kemeny & Snell (1960)], and, since not all of the \( \pi_k \)'s can be less than \( 1/n \), there must exist a state such that \( R_k \leq n \). By combining this with (3.3)–(c) and (3.3)–(e), an interesting corollary—which proves (2.1)—is produced.

**Corollary 3.1.** If \( R_k \) denotes the mean return time for the \( k \)th state then

\[
0 < \text{det}(A_i) < \chi \leq \min_k R_k \leq n
\]

for each \( i = 1, 2, \ldots, n \).
LEMMA 3.4. If $A = I - P$ where $P_{n \times n}$ is an irreducible stochastic matrix, and if $B_{k \times k}$ ($k < n$) is a principal submatrix of $A$, then the largest entry in each column of $B^{-1}$ is the diagonal entry. That is, for $j = 1, 2, \ldots, k$, it must be the case that $[B^{-1}]_{jj} \geq [B^{-1}]_{ij}$ for each $i \neq j$.

Proof. Without loss of generality, we may assume that $B$ is the leading $k \times k$ principal submatrix of $A$. Rearrange the states so that the $j^{th}$ state is listed first and the $i^{th}$ state is listed second, and prove that $[B^{-1}]_{11} \geq [B^{-1}]_{21}$. Because

$$[B^{-1}]_{11} = \frac{\det(B_{11})}{\det(B)} \quad \text{and} \quad [B^{-1}]_{21} = \frac{-\det(B_{12})}{\det(B)}$$

where $B_{ij}$ denotes the submatrix of $B$ obtained by deleting the $i^{th}$ row and $j^{th}$ column from $B$, and because (3.3)-part (c) guarantees that $\det(B) > 0$, it suffices to prove that

$$\det(B_{11}) + \det(B_{12}) \geq 0.$$ 

Denote the first unit vector by $e_1^T = (1, 0, \ldots, 0)$, and let

$$b_1 = \begin{pmatrix} -p_{21} \\
-p_{31} \\
\vdots \\
-p_{k1} \end{pmatrix}, \quad b_2 = \begin{pmatrix} 1 - p_{22} \\
-p_{32} \\
\vdots \\
-p_{k2} \end{pmatrix}, \quad \cdots, \quad b_k = \begin{pmatrix} -p_{2k} \\
-p_{3k} \\
\vdots \\
1 - p_{kk} \end{pmatrix}$$

so that

$$B = \begin{pmatrix} 1 - p_{11} & -p_{12} & \cdots & -p_{1k} \\
-p_{21} & 1 - p_{22} & \cdots & -p_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
-p_{k1} & -p_{k2} & \cdots & 1 - p_{kk} \end{pmatrix} \begin{pmatrix} 1 - p_{11} & -p_{12} & \cdots & -p_{1k} \\
b_1 & b_2 & \cdots & b_k \end{pmatrix}. \quad (3.5)$$

In terms of these quantities, $\det(B_{11}) + \det(B_{12})$ is given by

$$\det(B_{11}) + \det(B_{12}) = \det(b_2 | b_3 | \cdots | b_k) + \det(b_1 | b_3 | \cdots | b_k)$$

$$= \det(b_2 + b_1 | b_3 | \cdots | b_k)$$

$$= \det(B_{11} + b_1 e_1^T)$$

$$= \det(B_{11}) \det(I + B_{11}^{-1} b_1 e_1^T)$$

$$= \det(B_{11}) (1 + e_1^T B_{11}^{-1} b_1).$$
Part (c) of (3.3) also insures that \( \det(B_{11}) > 0 \), so the proof can be completed by arguing that 
\[ 1 + e_1^T B_{11}^{-1} b_1 \geq 0. \]
To do so, modify the chain by making state 1 as well as states \( k+1, k+2, \ldots, n \) absorbing states so that the transition matrix has the form

\[
\hat{P} = \begin{pmatrix}
1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
p_{21} & p_{22} & p_{23} & \cdots & p_{2k} & p_{2,k+1} & \cdots & p_{2n} \\
p_{31} & p_{32} & p_{33} & \cdots & p_{3k} & p_{3,k+1} & \cdots & p_{3n} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\
p_{k1} & p_{k2} & p_{k3} & \cdots & p_{kk} & p_{k,k+1} & \cdots & p_{kn} \\
0 & 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\
0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 1
\end{pmatrix}
= \begin{pmatrix}
1 & 0 & 0 \\
-b_1 & Q & R \\
0 & 0 & I_{n-k}
\end{pmatrix}.
\]

It follows from the elementary theory of absorbing chains given by Kemeny & Snell (1960) that the entries in the matrix

\[
(I - Q)^{-1}(-b_1 | R) = B_{11}^{-1}(-b_1 | R)
\]
represent the various absorption probabilities, and consequently all entries in \( -B_{11}^{-1}b_1 \) are between 0 and 1 so that

\[ 0 \leq 1 + e_1^T B_{11}^{-1} b_1 \leq 1. \]

**Lemma 3.5.** If \( A = I - P \) where \( P_{n \times n} \) is an irreducible stochastic matrix, and if \( B_{k \times k} \) \((k < n)\) is a principal submatrix of \( A \), then

\[ 0 < \det(B) \leq \frac{\max_i \delta_i(B)}{\max_{i,j} [B^{-1}]_{ij}} \leq \frac{1}{\max_{i,j} [B^{-1}]_{ij}} \]

where \( \delta_r(B) \) denotes the deleted product \( \delta_r(B) = b_{11}b_{22}\cdots b_{kk}/b_{rr} \).

**Proof.** Lemma 3.4 insures that there is some diagonal entry \( [B^{-1}]_{rr} \) of \( B^{-1} \) such that

\[ [B^{-1}]_{rr} = \max_{i,j} [B^{-1}]_{ij}. \]  

(3.6)

If \( B_{rr} \) is the principal submatrix of \( B \) obtained by deleting the \( r^{th} \) row and column from \( B \), then (3.3)–part (e) together with (3.6) produces

\[ \det(B) = \frac{\det(B_{rr})}{[B^{-1}]_{rr}} \leq \frac{\delta_r(B)}{[B^{-1}]_{rr}} = \frac{\delta_r(B)}{\max_{i,j} [B^{-1}]_{ij}} \leq \frac{\max_i \delta_i(B)}{\max_{i,j} [B^{-1}]_{ij}} \leq \frac{1}{\max_{i,j} [B^{-1}]_{ij}}. \]
LEMMA 3.6. For an irreducible stochastic matrix \( P_{n \times n} \), let \( A_j \) be the principal submatrix of \( A = I - P \) obtained by deleting the \( j^{th} \) row and column from \( A \), and let \( Q \) be the permutation matrix such that

\[
Q^T A Q = \begin{pmatrix} A_j & c_j \\ d_j^T & a_{jj} \end{pmatrix}.
\]

If the stationary distribution for \( Q^T P Q \) is written as \( \psi^T = \pi^T Q = (\pi_j, \pi) \), then the group inverse of \( A \) is given by

\[
A^# = Q \begin{pmatrix} (I - e\pi^T)A_j^{-1}(I - e\pi^T) & -\pi_j(I - e\pi^T)A_j^{-1}e \\ -\pi^T A_j^{-1}(I - e\pi^T) & \pi_j \pi^T A_j^{-1}e \end{pmatrix} Q^T
\]

where \( e \) is a column of 1's whose size is determined by the context in which it appears.

Proof. The group inverse possesses the property that \( (T^{-1}AT)^# = T^{-1}A^#T \) for all nonsingular matrices \( T \) [see Campbell & Meyer (1991)], so

\[
A^# = Q \begin{pmatrix} A_j & c_j \\ d_j^T & a_{jj} \end{pmatrix}^# Q^T.
\]

Since \( \text{rank}(Q^T A Q) = n - 1 \), it follows that \( a_{jj} - d_j^T A_j^{-1} c_j = 0 \), and this is used to verify that

\[
\begin{pmatrix} A_j & c_j \\ d_j^T & a_{jj} \end{pmatrix}^# = (I - e\psi^T) \begin{pmatrix} A_j^{-1} & 0 \\ 0 & 0 \end{pmatrix} (I - e\psi^T)
\]

\[
= \begin{pmatrix} (I - e\pi^T)A_j^{-1}(I - e\pi^T) & -\pi_j(I - e\pi^T)A_j^{-1}e \\ -\pi^T A_j^{-1}(I - e\pi^T) & \pi_j \pi^T A_j^{-1}e \end{pmatrix}.
\]

4. PROOF OF THEOREM 2.1

The preceding sequence of lemmas are now connected together to construct a proof of the main result stated in Theorem 2.1.

The Upper Bound. To derive the inequalities

\[
\max_{i,j} |a_{ij}^#| < \frac{2\delta(n - 1)}{\chi} \leq \frac{2(n - 1)}{\chi},
\]

(4.1)
begin by letting $Q$ be the permutation matrix given in Lemma 3.6 so that for $i \neq j$, the $(i,j)$–entry of $A^\#$ is the $(k,n)$–entry of $Q^T A^\# Q$ where $k \neq n$. In succession, use the formula of Lemma 3.6 and Hölder's inequality followed by the results of Lemma 3.5 and Lemma 3.3 to write

$$|a^\#_{ij}| = \pi_j |e_k^T (I - e\pi^T) A_j^{-1} e| \leq \pi_j \| e_k - \bar{\pi} \|_1 \| A_j^{-1} e \|_\infty$$

$$< 2\pi_j \| A_j^{-1} \|_\infty \leq 2\pi_j (n-1) \max_{r,s} [A^{-1}_j]_{rs}$$

$$\leq \frac{2\pi_j (n-1) \max_r \delta(A)}{\text{det}(A_j)} \leq \frac{2\pi_j (n-1) \delta}{\text{det}(A_j)}$$

$$= \frac{\delta(n-1)}{\chi} \leq \frac{(n-1)}{\chi},$$

Now consider the diagonal elements. The $(j,j)$–entry of $A^\#$ is the $(n,n)$–entry of $Q^T A^\# Q$, so proceeding in a manner similar to that above produces

$$|a^\#_{jj}| = \pi_j |\pi^T A_j^{-1} e| \leq \pi_j \| \pi \|_1 \| A_j^{-1} e \|_\infty$$

$$< \pi_j \| A_j^{-1} \|_\infty \leq \pi_j (n-1) \max_{r,s} [A^{-1}_j]_{rs}$$

$$\leq \frac{\pi_j (n-1) \max_r \delta(A)}{\text{det}(A_j)} \leq \frac{\pi_j (n-1) \delta}{\text{det}(A_j)}$$

$$= \frac{\delta(n-1)}{\chi} \leq \frac{(n-1)}{\chi},$$

thus proving (4.1).

**The Lower Bound.** To establish that

$$\frac{1}{n \min_{\lambda \neq 1} |1-\lambda|} \leq \max_{i,j} |a^\#_{ij}|, \quad (4.2)$$

make use of the fact that if $Ax = \mu x$ for $\mu \neq 0$, then $A^\# x = \mu^{-1} x$ [see Meyer & Campbell [1991, pg. 129]]. In particular, if $\lambda \neq 1$ is an eigenvalue of $P$, and if $x$ is a corresponding eigenvector, then $Ax = (1-\lambda)x$ implies that $A^\# x = (1-\lambda)^{-1} x$, and thus

$$\frac{1}{1-\lambda} \leq \| A^\# \|_\infty \leq n \max_{i,j} |a^\#_{ij}|. \quad \blacksquare$$
5. USING AN LU FACTORIZATION

Except for chains which are too large to fit into a computer's main memory, the stationary distribution $\pi^T$ is generally computed by direct methods—i.e., either an LU or QR factorization of $A = I - P$ (or $A^T$) is computed [Harrod & Plemmons (1984), Grassmann et al. (1985), Funderlic & Meyer (1986), Golub & Meyer (1986), or Barlow (1991)]. Even for very large chains which are nearly uncoupled, direct methods are usually involved—they can be the basis of the main algorithm [Stewart & Zhang (1991)], or they can be used to solve the aggregated and coupling chains in iterative aggregation/disaggregation algorithms [Chatelin & Miranker (1982) or Haviv (1987)]. In the conclusion of their paper, Golub & Meyer (1986) make the following observation.

*Computational experience suggests that when a triangular factorization of $A_{n \times n}$ is used to solve an irreducible chain, then the condition of the chain seems to be a function of the size of the nonzero pivots, and this means that it should be possible to estimate $\kappa$ with little or no extra cost beyond that incurred in computing $\pi^T$. For large chains, this can be a significant savings over the $O(n^2)$ operations demanded by traditional condition estimators.*

Of course, this is contrary to situation which exists for general nonsingular matrices because the absence of small pivots (or the existence of a large determinant) is not a guarantee of a well-conditioned matrix—consider the matrix in (1.6). A mathematical formulation and proof (or even an intuitive explanation) of Golub & Meyer's observation has heretofore not been given, but the results of §2 and §3 now make it possible to give a more precise statement and a rigorous proof of the Golub-Meyer observation. The arguments hinge on the fact that whenever $\pi^T$ is computed by means of a triangular factorization of $A$ (or $A^T$), the character of the chain is always an immediate by-product. The results for an LU factorization are given below, and the analogous theory for a QR factorization is given in the next section.

Suppose that the LU factorization of $A = I - P$ is computed to be

$$ A = LU = \begin{pmatrix} L_n & 0 \\ r^T & 1 \end{pmatrix} \begin{pmatrix} U_n & c \\ 0 & 0 \end{pmatrix}. $$

**Remark.** Regardless of whether $A$ or $A^T$ is used, gaussian elimination with finite-precision arithmetic can prematurely produce a zero pivot, and this can happen for well-conditioned chains. Consequently, practical implementation demands a strategy to deal with this situation. Funderlic
& Meyer (1986) and Stewart & Zhang (1991) discuss this problem along with possible remedies. Although practical algorithms may involve reordering schemes—such as diagonal pivoting—which introduce permutation matrices, such permutations are irrelevant in our discussion, and they are suppressed.

If $A_n$ is the principal submatrix of $A$ obtained by deleting the last row and column from $A$, then $A_n$ is a nonsingular $M$-matrix, and its LU factorization is $A_n = L_n U_n$. Since the LU factors of a nonsingular $M$-matrix are also nonsingular $M$-matrices [Berman & Plemmons (1979) or Horn & Johnson (1991)], it follows that $L_n$ and $U_n$ are nonsingular $M$-matrices, and hence $L_n^{-1} \geq 0$ and $U_n^{-1} \geq 0$. Consequently, $r^T \leq 0$, so the solution (obtained by a simple substitution process with no divisions) of the nonsingular triangular system

$$x^T L_n = -r^T$$

is nonnegative. This together with the result of Lemma 3.3 and Theorem 2.1 produces the following conclusion.

**Theorem 5.1.** For an irreducible Markov chain whose transition matrix is $P$, let the LU factorization of $A = I - P$ be given by

$$A = LU = \begin{pmatrix} L_n & 0 \\ r^T & 1 \end{pmatrix} \begin{pmatrix} U_n & 0 \\ 0 & 0 \end{pmatrix}.$$

If $x^T$ is the solution of $x^T L_n = -r^T$, then:

- $\pi^T = \frac{1}{1 + \|x\|_1} (x^T, 1)$ is the stationary distribution of the chain.
- $\chi = \frac{\det(U_n)}{\pi_n} = (1 + \|x\|_1) \det(U_n)$ is the character of the chain.
- The condition number for the chain is bounded above by

$$\kappa \leq \frac{2\delta(n-1)\pi_n}{\det(U_n)} = \frac{2\delta(n-1)}{\det(U_n)} \leq \frac{2(n-1)}{(1 + \|x\|_1) \det(U_n)}.$$
The condition number for the chain is bounded below by
\[ \pi_n \sum_{i=1}^{n-1} \frac{\pi_i}{u_{ii}} = \frac{1}{(1 + \|x\|_1)^2} \sum_{i=1}^{n-1} \frac{x_i}{u_{ii}} \leq \kappa(A) \]
where \( u_{ii} \) denotes the \( i^{th} \) pivot in \( U_n \).

Proof. The first three points are straightforward consequences of the previous discussion. To establish the lower bound for \( \kappa \), first recall from Lemma 3.6 that
\[ a_{nn}^\# = \pi_n \pi^T A_{n}^{-1} e = \pi_n \pi^T U_{n}^{-1} L_{n}^{-1} e > 0. \]
Since \( U_{n}^{-1} \geq 0 \) and \( L_{n}^{-1} \geq 0 \), it follows that \( \pi^T U_{n}^{-1} \) and \( L_{n}^{-1} e \) can be written as
\[ \pi^T U_{n}^{-1} = \left( \frac{\pi_1}{u_{11}}, \frac{\pi_2}{u_{22}} + \alpha_2, \ldots, \frac{\pi_{n-1}}{u_{n-1,n-1}} + \alpha_{n-1} \right) \]
\[ L_{n}^{-1} e = (1, 1 + \beta_2, \ldots, 1 + \beta_{n-1})^T \]
where each \( \alpha_i \) and \( \beta_i \) is nonnegative, and consequently (setting \( \alpha_0 = \beta_0 = 0 \))
\[ \pi^T A_{n}^{-1} e = \pi^T U_{n}^{-1} L_{n}^{-1} e = \sum_{i=1}^{n-1} \frac{(\pi_i + \alpha_i)(1 + \beta_i)}{u_{ii}} \geq \sum_{i=1}^{n-1} \frac{\pi_i}{u_{ii}}. \]
Therefore,
\[ \kappa \geq a_{nn}^\# = \pi_n \pi^T U_{n}^{-1} L_{n}^{-1} e \geq \pi_n \sum_{i=1}^{n-1} \frac{\pi_i}{u_{ii}} = \frac{1}{(1 + \|x\|_1)^2} \sum_{i=1}^{n-1} \frac{x_i}{u_{ii}}. \]

As mentioned before, the pivots or the determinant need not be indicators of the condition of a general nonsingular matrix—in particular, the absence of small pivots (or the existence of a large determinant) is not a guarantee of a well-conditioned matrix. However, for our special matrices \( A = I - P \), the bounds in Theorem 5.1 allow the pivots to be used as condition estimators.

**Corollary 5.1.** For an irreducible Markov chain whose transition matrix is \( P \), suppose that the LU factorization of \( A = I - P \) and the stationary distribution \( \pi^T \) have been computed as described in Theorem 5.1.

- If the pivots \( u_{ii} \) are large relative to \( \pi_n \) in the sense that \( \pi_n/det(U_n) \) is not too small, then the chain must be well-conditioned.
- If there are pivots \( u_{ii} \) which are small relative to \( \pi_n \pi_i \) in the sense that \( \pi_n \sum_{i=1}^{n-1} \pi_i / u_{ii} \) is large, then the chain must be ill-conditioned.
6. USING A QR FACTORIZATION

The utility of orthogonal triangularization is well-documented in the vast literature on matrix computations, and the use of a QR factorization to solve and analyze Markov chains is discussed in the paper by Golub & Meyer (1986). The following theorem shows that the character of an irreducible chain can be directly obtained from the diagonal entries of $R$, and the last column of $Q$, and this will establish an upper bound using a QR factorization which is analogous to that in Theorem 5.1 for an LU factorization. A lower bound analogous to the one in Theorem 5.1 is not readily available.

**THEOREM 6.1.** For an irreducible Markov chain whose transition matrix is $P$, the QR factorization of $A = I - P$ is given by

$$
A = QR = \begin{pmatrix}
Q_n & c \\
0 & -R_n e
\end{pmatrix}
\begin{pmatrix}
R_n & -Q_n R_n e \\
d^T R_n & -d^T R_n e
\end{pmatrix}.
$$

If $q$ denotes the last column of $Q$, then:

- $\pi^T = \frac{q^T}{\sum_{i=1}^n q_{in}}$ is the stationary distribution of the chain. \hfill (6.1)
- $\chi(A) = \|q\|_1 \det(R_n)$ is the character of the chain. \hfill (6.2)
- The condition number for the chain is bounded above by

$$
\kappa(A) < \frac{2\delta(n - 1)}{\|q\|_1 \det(R_n)} \leq \frac{2(n - 1)}{\|q\|_1 \det R_n}. \hfill (6.3)
$$

**Proof.** The formula (6.1) for $\pi^T$ is derived in Golub & Meyer (1986). To prove (6.2), first recall the result of Lemma 3.3, and observe that

$$
\chi^2 = \left(\frac{\det A_n}{\pi_n}\right)^2 = \frac{(\det Q_n R_n)^2}{\pi_n^2} = \frac{(\det Q_n)^2 (\det R_n)^2}{q_{nn}^2/\|q\|_1^2}.
$$

Use the fact that $QQ^T = I$ implies $Q_n Q_n^T + cc^T = I$ to obtain

$$
(\det Q)^2 = \det(Q_n Q_n^T) = \det(I - cc^T) = 1 - c^T c = q_{nn}^2,
$$

and substitute this into the previous expression to obtain (6.2). The bound (6.3) is now a consequence of the result of Theorem 2.1. 

---15---
7. CONCLUDING REMARKS

It has been argued that the sensitivity of an irreducible chain is primarily governed by how close the subdominant eigenvalues are to 1 in the sense that the condition number of the chain is bounded by

\[ \frac{1}{n \min_{\lambda_i \neq 1} |1 - \lambda_i|} \leq \kappa < \frac{2\delta(n - 1)}{\chi}. \] (7.1)

Although the upper bound explicitly involves \( n \), it is generally not the case that \( 2\delta(n - 1)/\chi \) grows in proportion to \( n \). Except in the special case when the diagonal entries of \( P \) are 0, the term \( \delta \) somewhat mitigates the presence of \( n \) because as \( n \) becomes larger, \( \delta \) becomes smaller.

Computational experience suggests that \( 2\delta(n - 1)/\chi \) is usually a rather conservative estimate of \( \kappa \), and although it is not always an upper bound for \( \kappa \), the term \( \delta/\chi \) by itself is often of the same order of magnitude as \( \kappa \). However, there exist pathological cases for which even \( \delta/\chi \) severely over estimates \( \kappa \). This seems to occur for chains which are not too badly conditioned and no single eigenvalue is extremely close to 1, but enough eigenvalues are within range of 1 to force \( \chi^{-1} \) to be too large. This suggests that for the purposes of bounding \( \kappa \) above, perhaps not all of the subdominant eigenvalues need to be taken into account.

When direct methods are used to solve an irreducible chain, standard condition estimators can be used to produce reliable estimates for \( \kappa \), but the cost of doing so is \( O(n^2) \) operations beyond the solution process. The results of Theorems 5.1 and 6.1 make it possible to estimate \( \kappa \) with the same computations which produce \( \pi^T \). Although the bounds for \( \kappa \) produced by Theorem 5.1 are sometimes rather loose, they are nevertheless virtually free. One must balance the cost of obtaining condition estimates against the information which one desires to obtain from these estimates.
8. REFERENCES


Sensitivity Of Markov Chains

Cambridge.


Recent IMA Preprints

#   Author/s                                      Title
839  Oscar P. Bruno and Fernando Reitich,       Numerical solution of diffraction problems: a method of variation of boundaries
840  Oscar P. Bruno and Fernando Reitich,       Solution of a boundary value problem for Helmholtz equation via variation of the boundary into the complex domain
841  Victor A. Galaktionov and Juan L. Vazquez, Asymptotic behaviour for an equation of superslow diffusion. The Cauchy problem
842  Josephus Hulshof and Juan Luis Vazquez,     The Dipole solution for the porous medium equation in several space dimensions
843  Shoshana Kamin and Juan Luis Vazquez,      The propagation of turbulent bursts
844  Miguel Escobedo, Juan Luis Vazquez and Enrike Zuazua, Source-type solutions and asymptotic behaviour for a diffusion-convection equation
845  Marco Biroli and Umberto Mosco,            Discontinuous media and Dirichlet forms of diffusion type
846  Stathis Filippas and Jong-Shenq Guo,       Quenching profiles for one-dimensional semilinear heat equations
847  H. Scott Dumas, A Nekhoroshev-like theory of classical particle channeling in perfect crystals
848  R. Natalini and A. Tesei,                  On a class of perturbed conservation laws
849  Paul K. Newton and Shinya Watanabe,        The geometry of nonlinear Schrödinger standing waves
850  S.S. Sritharan, On the nonsmooth verification technique for the dynamic programming of viscous flow
851  Mario Taboada and Yuncheng You,             Global attractor, inertial manifolds and stabilization of nonlinear damped beam equations
852  Shigeru Sakaguchi, Critical points of solutions to the obstacle problem in the plane
853  F. Abergel, D. Hilhorst and F. Issard-Roch, On a dissolution-growth problem with surface tension in the neighborhood of a stationary solution
854  Erasmus Langer, Numerical simulation of MOS transistors
855  Haim Brezis and Shoshana Kamin,            Sublinear elliptic equations in $\mathbb{R}^n$
856  Johannes C.C. Nitsche, Boundary value problems for variational integrals involving surface curvatures
857  Chao-Nien Chen, Multiple solutions for a semilinear elliptic equation on $\mathbb{R}^N$ with nonlinear dependence on the gradient
858  D. Brochet, X. Chen and D. Hilhorst,       Finite dimensional exponential attractor for the phase field model
859  Joseph D. Fehribach, Mullins-Sekerka stability analysis for melting-freezing waves in helium-4
860  Walter Schempp, Quantum holography and neurocomputer architectures
861  D.V. Anosov, An introduction to Hilbert's 21st problem
862  Herbert E Huppert and M Grae Worster,       Vigorous motions in magma chambers and lava lakes
863  Robert L. Pego and Michael I. Weinstein,   A class of eigenvalue problems, with applications to instability of solitary waves
864  Mahmoud Affouf, Numerical study of a singular system of conservation laws arising in enhanced oil reservoirs
865  Darin Beigie, Anthony Leonard and Stephen Wiggins, The dynamics associated with the chaotic tangles of two dimensional quasipersistent vector fields: theory and applications
866  Gui-Qiang Chen and Tai-Ping Liu, Zero relaxation and dissipation limits for hyperbolic conservation laws
867  Gui-Qiang Chen and Jian-Guo Liu, Convergence of second-order schemes for isentropic gas dynamics
868  Aleksander M. Simon and Zbigniew J. Grzywna, On the Larché-Cahn theory for stress-induced diffusion
869  Jerzy Lyczak, Adam Gadomski and Zbigniew J. Grzywna, Growth driven by diffusion
870  Mitchell Luskin and Tsong-Hwai Pan, Nonplanar shear flows for nonaligning nematic liquid crystals
871  Mahmoud Affouf, Unique global solutions of initial-boundary value problems for thermodynamic phase transitions
872  Richard A. Brualdi and Keith L. Chavey, Rectangular L-matrices
873  Xinfu Chen, Avner Friedman and Bei Hu, The thermistor problem with zero-one conductivity II
874  Raoul LePage, Controlling a diffusion toward a large goal and the Kelly principle
875  Raoul LePage, Controlling for optimum growth with time dependent returns
876  Marc Hallin and Madan L. Puri, Rank tests for time series analysis a survey
877  V.A. Solonnikov, Solvability of an evolution problem of thermocapillary convection in an infinite time interval
878  Horia I. Ene and Bogdan Vernescu, Viscosity dependent behaviour of viscoelastic porous media
879  Kaushik Bhattacharya, Self-accommodation in martensite
880  D. Lewis, T. Ratiu, J.C. Simo and J.E. Marsden, The heavy top: a geometric treatment
881  Leonid V. Kalachev, Some applications of asymptotic methods in semiconductor device modeling
882  David C. Dobson, Phase reconstruction via nonlinear least-squares
883  Patricio Aviles and Yoshikazu Giga, Minimal currents, geodesics and relaxation of variational integrals on mappings of bounded variation
884  Patricio Aviles and Yoshikazu Giga, Partial regularity of least gradient mappings
Charles R. Johnson and Michael Lundquist, Operator matrices with chordal inverse patterns
B.J. Bayly, Infinitely conducting dynamos and other horrible eigenproblems
Charles M. Elliott and Stefan Luckhaus, 'A generalised diffusion equation for phase separation of a multi-compon mixture with interfacial free energy'
Christian Schmeiser and Andreas Unterreiter, The derivation of analytic device models by asymptotic methods
LeRoy B. Beasley and Norman J. Pullman, Linear operators that strongly preserve the index of imprimitivity
Jerry Donato, The Boltzmann equation with lie and cartan
Thomas R. Hoffend Jr., Peter Smereka and Roger J. Anderson, A method for resolving the laser induced local heating of moving magneto-optical recording media
E.G. Kalnins, Willard Miller, Jr. and Sanchita Mukherjee, Models of q-algebra representations: the group of plane motions
T.R. Hoffend Jr. and R.K. Kaul, Relativistic theory of superpotentials for a nonhomogeneous, spatially isotropic medium
Reinhold von Schwerin, Two metal deposition on a microdisk electrode
Vladimir I. Oliker and Nina N. Uraltseva, Evolution of nonparametric surfaces with speed depending on curvature, III. Some remarks on mean curvature and anisotropic flows
Wayne Barrett, Charles R. Johnson, Raphael Loewy and Tamir Shalom, Rank incrementation via diagonal perturbations
Mingxiang Chen, Xu-Yan Chen and Jack K. Hale, Structural stability for time-periodic one-dimensional parabolic equations
Hong-Ming Yin, Global solutions of Maxwell’s equations in an electromagnetic field with the temperature-dependent electrical conductivity
Robert Grone, Russell Merris and William Watkins, Laplacian unimodular equivalence of graphs
Miroslav Fiedler, Structure-ranks of matrices
Miroslav Fiedler, An estimate for the nonstochastic eigenvalues of doubly stochastic matrices
Miroslav Fiedler, Remarks on eigenvalues of Hankel matrices
Charles R. Johnson, D.D. Olesky, Michael Tsatsomeros and P. van den Driessche, Spectra with positive elementary symmetric functions
Pierre-Alain Gremaud, Thermal contraction as a free boundary problem
K.L. Cooke, Janos Turi and Gregg Turner, Stabilization of hybrid systems in the presence of feedback delays
Robert P. Gilbert and Yongzhi Xu, A numerical transmutation approach for underwater sound propagation
LeRoy B. Beasley, Richard A. Brualdi and Bryan L. Shader, Combinatorial orthogonality
Richard A. Brualdi and Bryan L. Shader, Strong hall matrices
Håkan Wennerström and David M. Anderson, Difference versus Gaussian curvature energies; monolayer versus bilayer curvature energies applications to vesicle stability
Shmuel Friedland, Eigenvalues of almost all skew symmetric matrices and tournament matrices
Avner Friedman, Bei Hu and J.L. Velazquez, A Free Boundary Problem Modeling Loop Dislocations in Crystals
Ezio Venturino, The Influence of Diseases on Lotka-Volterra Systems
Steve Kirkland and Bryan L. Shader, On Multiparticle Tournament Matrices with Constant Team Size
Richard A. Brualdi and Jennifer J.Q. Massey, More on Structure-Ranks of Matrices
Douglas B. Meade, Qualitative Analysis of an Epidemic Model with Directed Dispersion
Kazuo Murota, Mixed Matrices Irreducibility and Decomposition
Richard A. Brualdi and Jennifer J.Q. Massey, Some Applications of Elementary Linear Algebra in Combinations
Carl D. Meyer, Sensitivity of Markov Chains
Hong-Ming Yin, Weak and Classical Solutions of Some Nonlinear Volterra Integrodifferential Equations
B. Leimkuhler and A. Riehl, Exploiting Symmetry and Regularity in Waveform Relaxation Convergence Estimation
Xinfu Chen and Charles M. Elliott, Asymptotics for a Parabolic Double Obstacle Problem
Yongzhi Xu and Yi Yan, An Approximate Boundary Integral Method for Acoustic Scattering in Shallow Oceans
Yongzhi Xu and Yi Yan, Source Localization Processing in Perturbed Waveguides
Kenneth L. Cooke and Janos Turi, Stability, Instability in Delay Equations Modeling Human Respiration
F. Bethuel, H. Brezis, B.D. Coleman and F. Hélein, Bifurcation Analysis of Minimizing Harmonic Maps Describing the Equilibrium of Nematic Phases Between Cylinders
Frank W. Elliott, Jr., Signed Random Measures: Stochastic Order and Kolmogorov Consistency Conditions
D.A. Gregory, S.J. Kirkland and B.L. Shader, Pick’s Inequality and Tournaments
J.W. Demmel, N.J. Higham and R.S. Schreiber, Block LU Factorization
Victor A. Galaktionov and Juan L. Vazquez, Regional Blow-Up in a Semilinear Heat Equation with Convergence to a Hamilton-Jacobi Equation
Bryan L. Shader, Convertible, Nearly Decomposable and Nearly Reducible Matrices
Dianne P. O’Leary, Iterative Methods for Finding the Stationary Vector for Markov Chains