NORTH-HOLLAND

On the Structure of Stochastic Matrices With a Subdominant Eigenvalue Near 1

D. J. Hartfiel*
Mathematics Department
Texas A & M University
College Station, Texas 77843-3368

and

Carl D. Meyer†
Mathematics Department
North Carolina State University
Raleigh, North Carolina 27695-8205

Submitted by Uriel G. Rothblum

ABSTRACT

An $n \times n$ irreducible stochastic matrix $\mathbf{P}$ can possess a subdominant eigenvalue, say $\lambda_2(\mathbf{P})$, near $\lambda = 1$. In this article we clarify the relationship between the nearness of these eigenvalues and the near-uncoupling (some authors say "nearly completely decomposable") of $\mathbf{P}$. We prove that for fixed n, if $\lambda_2(\mathbf{P})$ is sufficiently close to $\lambda = 1$, then $\mathbf{P}$ is nearly uncoupled. We then provide examples which show that $\lambda_2(\mathbf{P})$ must, in general, be remarkably close to 1 before such uncoupling occurs. © 1998 Elsevier Science Inc.

*E-mail: hartfiel@math.tamu.edu

†The work of this author was supported in part by the National Science Foundation under grants DMS-9403224 and CCR-9413309. E-mail: meyer@math.ncsu.edu.

LINEAR ALGEBRA AND ITS APPLICATIONS 272:193–203 (1998)

655 Avenue of the Americas, New York, NY 10010

0024-3795/98/$19.00
PII S0024-3795(97)00333-9

1. INTRODUCTION

Markov chain techniques are often used to model the behavior of large irreducible nearly uncoupled evolutionary systems in which the states (assumed to be finite in number) naturally divide into say k-clusters such that the states within each cluster are strongly coupled, but the clusters themselves are only weakly coupled to each other. Such systems are commonly encountered in the analysis of queuing networks and computer systems, discrete economic models, and many stochastic models found in biological and social science. When the states of such a chain are suitably arranged, the transition probability matrix $\mathbf{P}$ is an irreducible nearly uncoupled (row) stochastic matrix which can be partitioned into $k + 1$ levels:

$$\mathbf{P}_{n\times n} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \cdots & \mathbf{P}_{1k} & \mathbf{P}_{1\,k+1} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \cdots & \mathbf{P}_{2k} & \mathbf{P}_{2\,k+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{P}_{k1} & \mathbf{P}_{k2} & \cdots & \mathbf{P}_{kk} & \mathbf{P}_{k\,k+1} \\ \mathbf{P}_{k+1\,1} & \mathbf{P}_{k+1\,2} & \cdots & \mathbf{P}_{k+1\,k} & \mathbf{P}_{k+1\,k+1} \end{pmatrix}, \tag{1.1}$$

such that the nearly uncoupled diagonal blocks $\mathbf{P}_{11}, \ldots, \mathbf{P}_{kk}$ are square, and the norm of each off-diagonal block, in block rows $1, \ldots, k$, is small relative to 1. Because the nearly uncoupled diagonal blocks are nearly stochastic matrices, continuity of the eigenvalues forces $\mathbf{P}$ to have at least k eigenvalues near the unit eigenvalue $\lambda = 1$. (See [4] for related material.) This is of some concern, because having subdominant eigenvalues near $\lambda = 1$ poses problems for the numerical computation of statistics for the associated Markov chain—e.g., there are sensitivity and conditioning problem [5, 7], and iterative methods converge slowly.

While it is clear that nearly uncoupled irreducible stochastic matrices must possess subdominant eigenvalues near $\lambda = 1$, the converse is less than clear. In other words, if an irreducible stochastic matrix $\mathbf{P}$ has a subdominant eigenvalue, say $\lambda_2(\mathbf{P})$, near $\lambda = 1$, must $\mathbf{P}$ be nearly uncoupled? We answer this question in the following section.

2. UNCOUPLING RESULT

Let $\mathbf{P}$ be an $n \times n$ stochastic matrix. Define the *uncoupling measure* of P as

$$\sigma(\mathbf{P}) = \min\left(\sum_{\substack{i \in M_1 \\ j \notin M_1}} p_{ij} + \sum_{\substack{i \in M_2 \\ j \notin M_2}} p_{ij} \right), \tag{2.1}$$

where the minimum is taken over all nonempty proper subsets M_1, M_2 of $\{1, \dots, n\}$ where $M_1 \cap M_2 = \varnothing$. Thus, if we suppose that $\mathbf{P}$ has already been rearranged so that $\sigma(\mathbf{P})$ is achieved at $M_1 = \{1, \dots, k_1\}$, $M_2 = \{k_1 + 1, \dots, k_2\}$, then $\mathbf{P}$ can be partitioned as

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \mathbf{P}_{13} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \mathbf{P}_{23} \\ \mathbf{P}_{31} & \mathbf{P}_{32} & \mathbf{P}_{33} \end{bmatrix} \tag{2.2}$$

where $\mathbf{P}_{11}$ is $k_1 \times k_1$, $\mathbf{P}_{22}$ is $(k_2 - k_1) \times (k_2 - k_1)$, and $\sigma(\mathbf{P})$ is the sum of the entries in $\mathbf{P}_{12}, \mathbf{P}_{13}, \mathbf{P}_{21}, \mathbf{P}_{23}$. From this, if $\sigma(\mathbf{P})$ is small, $\mathbf{P}$ is nearly uncoupled into two blocks $\mathbf{P}_{11}, \mathbf{P}_{22}$. (This definition can be extended, in an obvious way, to more than two blocks. For our work, however, two blocks are sufficient.)

Our uncoupling theorem follows.

THEOREM. *Let $n > 0$ be a fixed integer. For that integer, given $\epsilon > 0$ there is $\delta > 0$ such that if $\mathbf{P}$ is an $n \times n$ stochastic matrix with $|\lambda_2(\mathbf{P}) - 1| < \delta$ then $\sigma(\mathbf{P}) < \epsilon$.*

Proof. The proof is by contradiction. Thus, suppose there is an $\epsilon > 0$ such that for any $\delta > 0$ there is an $n \times n$ stochastic matrix $\mathbf{P}$ with $|\lambda_2(\mathbf{P}) - 1| < \delta$ and $\sigma(\mathbf{P}) > \epsilon$. For $\delta = 1/k$ let $\mathbf{P}_k$ be such a matrix. Let $\mathbf{P}_{i_1}, \mathbf{P}_{i_2}, \dots$ be a subsequence of $\mathbf{P}_1, \mathbf{P}_2, \dots$ which converges, say to $\mathbf{P}_0$. Then $\mathbf{P}_0$ must have $\lambda_2(\mathbf{P}_0) = 1$ and thus $\sigma(\mathbf{P}_0) = 0$. Yet, $\sigma(\mathbf{P}_0) = \lim_{k \to \infty} \sigma(\mathbf{P}_k) \geqslant \epsilon$, a contradiction. The result follows. ∎

In the next section we show how remarkably close, in general, $\lambda_2(\mathbf{P})$ must be to 1 in order to see the uncoupling described in this theorem.

3. EXAMPLES

We now look at three examples

EXAMPLE 1. Consider tridiagonal Toeplitz matrices of the form

$$\mathbf{T}_n = \begin{pmatrix} a & b & & & & & \\ c & a & b & & & & \\ & c & a & b & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & c & a & b & \\ & & & & c & a & b \\ & & & & & c & a \end{pmatrix}_{n\times n}. \tag{3.1}$$

It is a straightforward exercise to show that if $\Delta_n = \det \mathbf{T}_n$, then

$$\Delta_n = a\Delta_{n-1} - bc\Delta_{n-2} \qquad \text{with} \quad \Delta_0 = 1. \tag{3.2}$$

Moreover, it is known [8, pp. 59, 154] that the eigenvalues of $\mathbf{T}_n$ are given by

$$\tau_k = a + 2\sqrt{bc}\cos\frac{k\pi}{n+1} \qquad \text{for} \quad k = 1, 2, \ldots, n. \tag{3.3}$$

Matrices of the form (3.1) are not stochastic, so some alteration is needed. While it is not absolutely necessary, it is computationally convenient to set $a = 0$ and begin with the special class of tridiagonal Toeplitz matrices of the form

$$\mathbf{A}_n = \begin{pmatrix} 0 & b & & & & & \\ c & 0 & b & & & & \\ & c & 0 & b & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & c & 0 & b & \\ & & & & c & 0 & b \\ & & & & & c & 0 \end{pmatrix}_{n\times n} \tag{3.4}$$

whose eigenvalues are

$$\lambda_k = 2\sqrt{bc}\,\cos\frac{k\pi}{n+1} \qquad \text{for} \quad k = 1, 2, \ldots, n.$$

Now adjust the first and last row of $\mathbf{A}_n$ to force the resulting matrices to have constant row sums by setting

$$\mathbf{P}_n = \begin{pmatrix} 0 & b+c & & & & & \\ c & 0 & b & & & & \\ & c & 0 & b & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & c & 0 & b & \\ & & & & c & 0 & b \\ & & & & & b+c & 0 \end{pmatrix}_{n\times n}. \tag{3.5}$$

(Other options are also possible.) Of course, this adjustment will alter the eigenvalues—but in a predictable way. To see how, consider the characteristic equation $|\mathbf{P}_n - \lambda\mathbf{I}| = 0$, and use straightforward determinant expansion to find that

$$|\mathbf{P}_n - \lambda\mathbf{I}| = \lambda^2 D_{n-2} + (b+c)^2(\lambda D_{n-3} + bcD_{n-4}), \tag{3.6}$$

in which D_k is the $k \times k$ determinant

$$D_k = \begin{vmatrix} -\lambda & b & & & & & \\ c & -\lambda & b & & & & \\ & c & -\lambda & b & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & c & -\lambda & b & \\ & & & & c & -\lambda & b \\ & & & & & c & -\lambda \end{vmatrix}. \tag{3.7}$$

Since D_k is the same as $\Delta_k = \det \mathbf{T}_k$ with $a = -\lambda$, the relation (3.2) can be used in (3.6) to conclude that

$$|\mathbf{P}_n - \lambda\mathbf{I}| = \left[\lambda^2 - (b+c)^2\right] D_{n-2}.$$

But $D_{n-2} = |\mathbf{A}_{n-2} - \lambda \mathbf{I}|$ is the characteristic polynomial for $\mathbf{A}_{n-2}$, so the spectrum of $\mathbf{P}_n$, denoted spec($\mathbf{P}_n$), must be

$$\operatorname{spec}(\mathbf{P}_n) = \operatorname{spec}(\mathbf{A}_{n-2}) \cup \{(b+c), -(b+c)\}.$$

Consequently, (3.3) guarantees that the eigenvalues of $\mathbf{P}_n$ are

$$\operatorname{spec}(\mathbf{P}_n) = \left\{(b+c), -(b+c), 2\sqrt{bc}\cos\frac{k\pi}{n-1}\right\}$$

$$\text{for} \quad k = 1, 2, \dots, n-2.$$

While there are many choices for b and c which produce irreducible stochastic matrices that fail to be nearly uncoupled, the point is best made by setting $b = c = \frac{1}{2}$, so that $\mathbf{P}_n$ becomes

$$\mathbf{P}_n = \begin{pmatrix} 0 & 1 & & & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & & & \\ & \frac{1}{2} & 0 & \frac{1}{2} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \frac{1}{2} & 0 & \frac{1}{2} & \\ & & & & \frac{1}{2} & 0 & \frac{1}{2} \\ & & & & & 1 & 0 \end{pmatrix}_{n \times n} \tag{3.8}$$

with eigenvalues

$$\operatorname{spec}(\mathbf{P}_n) = \left\{1, -1, \cos\frac{k\pi}{n-1} \text{ for } k = 1, 2, \dots, n-2\right\}.$$

Clearly, the eigenvalue $\lambda_2 = \cos[\pi/(n-1)]$ can be made arbitrarily close to 1 by increasing the size of n. Yet, for $n \geqslant 4$, $\sigma(\mathbf{P}_n) = 1$. In other words, for no value of n can $\mathbf{P}_n$ in (3.8) be considered to be nearly uncoupled, because in every partition the sum of the entries in the off-diagonal blocks will always be at least 1. From this it is clear that a subdominant eigenvalue of $\mathbf{P}$ can be arbitrarily close to 1 without $\mathbf{P}$ being nearly uncoupled.

Note in our example that to get $\lambda_2(\mathbf{P})$ close to 1, without changing $\sigma(\mathbf{P})$, requires the increase of n. Studying this further, let $\mathbf{I}_n$ denote the $n \times n$

identity matrix and, for all $n \geqslant 4$,

$$C_n = \{a\mathbf{I}_n + b\mathbf{P}_n : a + b = 1 \text{ and } a \geqslant 0,\, b \geqslant 0\}. \tag{3.9}$$

Note that if $\mathbf{P} \in C_n$ then

$$\sigma(\mathbf{P}) = b, \tag{3.10a}$$

$$\begin{aligned} \lambda_2(\mathbf{P}) &= a + b\cos\frac{\pi}{n-1} \\ &= 1 - \sigma(\mathbf{P})\left(1 - \cos\frac{\pi}{n-1}\right). \end{aligned} \tag{3.10b}$$

It is clear that if we choose any $n \geqslant 4$ and any number λ_2 ($\cos[\pi/(n-1)] \leqslant \lambda_2 \leqslant 1$), then with the right choice of b, there is a $\mathbf{P} \in C_n$ such that $\lambda_2(\mathbf{P}) = \lambda_2$.

Using (3.10b), write

$$\frac{1 - \lambda_2(\mathbf{P})}{1 - \cos\dfrac{\pi}{n-1}} = \sigma(\mathbf{P}). \tag{3.11}$$

Some data, choosing n and λ_2, and computing σ from this equation, are given in Table 1. These data show that $\lambda_2(\mathbf{P})$ must be close to 1 in order to assure $\sigma(\mathbf{P}) = 0.1$. And this is true even for small size matrices.

TABLE 1

DATA SHOWING THE RELATIONSHIP BETWEEN λ_2 AND σ FOR C_n

n	$\cos\frac{\pi}{n-1}$	$\leqslant$	λ_2	σ
4	.50000		.95000	.1
5	.70712		.97071	.1
6	.80902		.98090	.1

EXAMPLE 2. For this example, define

$$S_2 = \{\mathbf{P} : \mathbf{P} \text{ is a } 2 \times 2 \text{ stochastic matrix}\}. \tag{3.12}$$

If $\mathbf{P} \in S_2$ and $\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$, then

$$1 + \lambda_2(\mathbf{P}) = p_{11} + p_{22} = (1 - p_{12}) + (1 - p_{21}).$$

Thus, since $\sigma(\mathbf{P}) = p_{12} + p_{21}$,

$$1 - \lambda_2(\mathbf{P}) = \sigma(\mathbf{P}).$$

We point out that this expression can be written in the form

$$\frac{1 - \lambda_2(\mathbf{P})}{1 - \cos \dfrac{\pi}{2}} = \sigma(\mathbf{P}). \tag{3.13}$$

(In the conclusion we explain the significance of this form.)

EXAMPLE 3. For this example, define

$$T_n = \{\mathbf{P} : \mathbf{P} \text{ is an } n \times n \text{ symmetric stochastic matrix}\}. \tag{3.14}$$

As used by Fiedler [2], define for any $\mathbf{P} \in T_n$

$$\mu(\mathbf{P}) = \min \sum_{\substack{i \in M \\ j \notin M}} p_{ij}, \tag{3.15}$$

where M is a nonempty proper subset of $\{1, \dots, n\}$. Note that if the minimum is achieved at M, and we set $M_1 = M$, $M_2 = \{1, \dots, n\} \setminus M$, then since $\mathbf{P}$ is symmetric, $\sigma(\mathbf{P}) = 2\mu(\mathbf{P})$.

Fiedler proved [2, Theorem 3.2] that if $\mu(\mathbf{P}) \leqslant \frac{1}{2}$ then

$$\begin{aligned} 1 - \lambda_2(\mathbf{P}) &\geqslant \left(1 - \cos\frac{\pi}{n}\right) 2\mu(\mathbf{P}) \\ &\geqslant \left(1 - \cos\frac{\pi}{n}\right) \sigma(\mathbf{P}), \end{aligned}$$

TABLE 2

DATA SHOWING THE RELATIONSHIP BETWEEN λ_2 AND $\sigma(\mathbf{P})$

n	$\frac{n}{2(n-1)}$	$\geqslant$	$1-\lambda_2$	$\geqslant$	$\left(1-\cos\frac{\pi}{n}\right)\sigma$	λ_2	σ
2	.1		.1		.1	.9	.1
3	.075		.05		.05	.95	.1
4	.067		.03		.03	.97	.1
5	.063		.02		.02	.98	.1

or

$$\frac{1-\lambda_2(\mathbf{P})}{1-\cos\frac{\pi}{n}} \geqslant \sigma(\mathbf{P}). \tag{3.16}$$

Fiedler shows that if σ and λ_2 are chosen such that

$$0 \leqslant \sigma \leqslant 2 \quad \text{and} \quad \frac{n}{2(n-1)}\sigma \geqslant 1-\lambda_2 \geqslant \left(1-\cos\frac{\pi}{n}\right)\sigma,$$

then there is a $\mathbf{P} \in T_n$ such that $\sigma(\mathbf{P}) = \sigma$ and $\lambda_2(\mathbf{P}) = \lambda_2$.

We use the inequality (3.16), choosing n, σ, and λ_2 so equality holds, to provide the data in Table 2. Some larger values of n are shown in Table 3. Again, it is interesting to see just how close $\lambda_2(\mathbf{P})$ needs to be to 1 to obtain $\sigma(\mathbf{P}) = 0.1$.

TABLE 3

RELATIONSHIP DATA FOR T_n, LARGER n

n	$\frac{n}{2(n-1)}\sigma$	$\geqslant$	$1-\lambda_2$	$\geqslant$	$\left(1-\cos\frac{\pi}{n}\right)\sigma$	λ_2	σ
10	.05556		.00489		.00489	.99511	.1
20	.052632		.0012312		.0012312	.99877	.1
100	.052910		.000049344		.000049344	.99995	.1
1000	.050050		.00000049348		.00000049348	.99999	.1

4. CONCLUDING REMARKS AND A CONJECTURE

It is easily shown that

$$\frac{1-\lambda_2}{1-\cos\dfrac{\pi}{n}} \geqslant \frac{1-\lambda_2}{1-\cos\dfrac{\pi}{n-1}} \tag{4.1}$$

for all n. Thus, assuming the hypotheses of our examples, we have

$$f(n, \lambda_2(\mathbf{P})) = \frac{1+\lambda_2(\mathbf{P})}{1-\cos(\pi/n)} \tag{4.2}$$

as an upper bound on $\sigma(\mathbf{P})$ over S_2, as well as T_n and C_n. With this function in mind we conclude this paper with a conjecture.

CONJECTURE. Let S_n denote the set of $n \times n$ stochastic matrices. Then there is a function $f(n, \lambda_2)$ such that for fixed n, $f(n, \lambda_2) \to 0$ as $\lambda_2 \to 1$. Further, for all $\mathbf{P} \in S_n$, $f(n, \lambda_2(\mathbf{P})) \geqslant \sigma(\mathbf{P})$, and for each n, equality holds for some $\mathbf{P}$. (Reasonable bounds would also be interesting.)

In addition, by extending $\sigma(\mathbf{P})$ to k nearly uncoupled blocks $\mathbf{P}_{11}, \dots, \mathbf{P}_{kk}$ in

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \cdots & \mathbf{P}_{1k} & \mathbf{P}_{1\,k+1} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \cdots & \mathbf{P}_{2k} & \mathbf{P}_{2\,k+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{P}_{k1} & \mathbf{P}_{k2} & \cdots & \mathbf{P}_{kk} & \mathbf{P}_{k\,k+1} \\ \mathbf{P}_{k+1} & \mathbf{P}_{k+1\,2} & \cdots & \mathbf{P}_{k+1\,k} & \mathbf{P}_{k+1\,k+1} \end{bmatrix}, \tag{4.3}$$

f can be extended to cover $\lambda_2, \dots, \lambda_{k+1}$ eigenvalues near 1. (See [3] for one such bound.)

For somewhat related work, on the inverse eigenvalue problem for stochastic matrices, the reader may want to see [1], [5], [6], and [8].

REFERENCES

1 N. Dimitriev and E. Dynkin, Eigenvalues of a stochastic matrix, *Izv. Akad. Nauk SSSR Ser. Math.* 10:167–184 (1946).

2 Miroslav Fiedler, Bounds for eigenvalues of doubly stochastic matrices, *Linear Algebra Appl.* 5:299–310 (1972).
3 D. J. Hartfiel, Distribution of entries in a substochastic matrix having eigenvalues near 1, submitted for publication.
4 Moshe Haviv and Uriel Rothblum, Bounds on distances between eigenvectors, *Linear Algebra Appl.* 63:101–118 (1984).
5 I. C. F. Ipsen and C. D. Meyer, Uniform stability of Markov chains, *SIAM J. Matrix Anal. Appl.* 15:1061–1074 (1994).
6 F. I. Karpelevich, On the eigenvalues of a matrix with nonegative elements, *Izv. Akad. Nauk SSSR Ser. Math.* 15:361–383 (1951).
7 C. D. Meyer, Sensitivity of Markov chains, *SIAM J. Matrix Anal. Appl.* 15:715–728 (1994).
8 G. D. Smith, *Numerical Solution of Partial Differential Equiations: Finite Difference Methods*, 3rd Ed., Clarendon, Oxford, 1985.

Received 10 May 1996; final manuscript accepted 30 June 1997