STOCHASTIC COMPLEMENTATION, UNCOUPLING MARKOV CHAINS, AND THE THEORY OF NEARLY REDUCIBLE SYSTEMS*

C. D. MEYER†

Abstract. A concept called *stochastic complementation* is an idea which occurs naturally, although not always explicitly, in the theory and application of finite Markov chains. This paper brings this idea to the forefront with an explicit definition and a development of some of its properties. Applications of stochastic complementation are explored with respect to problems involving uncoupling procedures in the theory of Markov chains. Furthermore, the role of stochastic complementation in the development of the classical Simon–Ando theory of nearly reducible system is presented.

Key words. Markov chains, stationary distributions, stochastic matrix, stochastic complementation, nearly reducible systems, Simon–Ando theory

AMS(MOS) subject classifications. 65U05, 60-02, 60J10, 60J20, 15-02, 15A51, 90A14

1. Introduction. Although not always given an explicit name, a quantity which we shall refer to as a *stochastic complement* arises very naturally in the consideration of finite Markov chains. This concept has heretofore not been focused upon as an entity unto itself, and a detailed study of the properties of stochastic complementation has not yet been given. The purpose of the first part of this exposition is to explicitly publicize the utility of this concept and to present a more complete and unified discussion of the important properties of stochastic complementation.

A focal point of the development concerns the problem of determining the stationary distribution of an irreducible Markov chain by uncoupling the chain into several smaller independent chains. In particular, if $\mathbf{P}_{m\times m}$ is the transition matrix for an m-state, homogeneous, irreducible Markov chain $\mathcal{C}$, the stationary distribution problem concerns the determination of the unique vector $\boldsymbol{\pi}_{1\times m}$ which satisfies

$$\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}, \quad \pi_i > 0, \quad \sum_{i=1}^{m} \pi_i = 1.$$

For chains with relatively few states, this is not a difficult problem to solve using adaptations of standard techniques for solving systems of linear equations. However, there exist many applications for which the number of states is too large to be comfortably handled by standard methods. For large-scale problems, it is only natural to attempt to somehow uncouple the original m-state chain $\mathcal{C}$ into two or more smaller chains—say $\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_k$—containing $r_1, r_2, \cdots, r_k$ states, respectively, where $\sum_{i=1}^{k} r_i = m$. Ideally, this sequence of smaller chains should have the following properties:

- Each smaller chain $\mathcal{C}_i$ should be irreducible whenever the original chain $\mathcal{C}$ is irreducible so that each $\mathcal{C}_i$ has a unique stationary distribution vector $\mathbf{s}_i$.
- It should be possible to determine the $\mathbf{s}_i$'s completely independent of each other. For modern multiprocessor computer architectures it is desirable to be able to execute the computation of the $\mathbf{s}_i$'s in parallel.

* Received by the editors January 26, 1988; accepted for publication (in revised form) February 2, 1989. This work was supported by National Science Foundation grant DMS-8521154.

† North Carolina State University, Department of Mathematics; Center for Research in Scientific Computation, Box 8205, Raleigh, North Carolina 27695-8205.

- Finally, it must be possible to easily couple the smaller stationary distribution vectors $\mathbf{s}_i$ back together in order to produce the stationary distribution vector $\boldsymbol{\pi}$ for the larger chain $\mathcal{C}$.

The second part of this paper is dedicated to showing how to accomplish the above three goals, and it is indicated how this can lead to fully parallel algorithms for the determination of the stationary distribution vector of the original chain.

Finally, in the third part of this survey, the application of stochastic complementation to the classical Simon–Ando theory developed in [18] for nearly completely reducible chains is presented. It is demonstrated how to apply the concept of stochastic complementation in order to develop the theory for nearly completely reducible systems in a unified, clear, and simple manner while simultaneously sharpening some results and generalizing others.

2. Stochastic complementation. The purpose of this section is to introduce the concept of a stochastic complement in an irreducible stochastic matrix and to develop some of the basic properties of stochastic complementation. These ideas will be the cornerstone for all subsequent discussions.

Unless otherwise stated, $\mathbf{P}$ will denote an $m \times m$, irreducible, stochastic matrix which will be partitioned as

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

where all diagonal blocks are square. It is well known that $\mathbf{I}-\mathbf{P}$ is a singular M-matrix of rank $m-1$ and that every principal submatrix of $\mathbf{I}-\mathbf{P}$ of order $m-1$ or smaller is a nonsingular M-matrix (see [2, p. 156]). In particular, if $\mathbf{P}_i$ denotes the principal submatrix of $\mathbf{P}$ obtained by deleting the ith row and ith column of blocks from the partitioned form of $\mathbf{P}$, then each $\mathbf{I}-\mathbf{P}_i$ is a nonsingular M-matrix. Therefore,

$$(\mathbf{I}-\mathbf{P}_i)^{-1} \geq 0, \tag{2.1}$$

so the indicated inverses in the following definition are well defined.

DEFINITION 2.1. *Let $\mathbf{P}$ be an $m \times m$ irreducible stochastic matrix with a k-level partition*

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

in which all diagonal blocks are square. For a given index i, let $\mathbf{P}_i$ denote the principal block submatrix of $\mathbf{P}$ obtained by deleting the ith row and ith column of blocks from $\mathbf{P}$, and let $\mathbf{P}_{i*}$ and $\mathbf{P}_{*i}$ designate

$$\mathbf{P}_{i*} = \begin{pmatrix} \mathbf{P}_{i1} & \mathbf{P}_{i2} & \cdots & \mathbf{P}_{i,i-1} & \mathbf{P}_{i,i+1} & \cdots & \mathbf{P}_{ik} \end{pmatrix}$$

and

$$\mathbf{P}_{*i} = \begin{pmatrix} \mathbf{P}_{1i} \\ \vdots \\ \mathbf{P}_{i-1,i} \\ \mathbf{P}_{i+1,i} \\ \vdots \\ \mathbf{P}_{ki} \end{pmatrix}.$$

That is, $\mathbf{P}_{i*}$ *is the* ith *row of blocks with* $\mathbf{P}_{ii}$ *removed, and* $\mathbf{P}_{*i}$ *is the* ith *column of blocks with* $\mathbf{P}_{ii}$ *removed. The* ***stochastic complement*** *of* $\mathbf{P}_{ii}$ *in* $\mathbf{P}$ *is defined to be the matrix*

$$\mathbf{S}_{ii} = \mathbf{P}_{ii} + \mathbf{P}_{i*}(\mathbf{I} - \mathbf{P}_i)^{-1}\mathbf{P}_{*i}.$$

For example, the stochastic complement of $\mathbf{P}_{22}$ in

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

is given by

$$\mathbf{S}_{22} = \mathbf{P}_{22} + \begin{pmatrix} \mathbf{P}_{21} & \mathbf{P}_{23} \end{pmatrix} \begin{pmatrix} \mathbf{I} - \mathbf{P}_{11} & -\mathbf{P}_{13} \\ -\mathbf{P}_{31} & \mathbf{I} - \mathbf{P}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{P}_{12} \\ \mathbf{P}_{32} \end{pmatrix}.$$

The reason for the terminology "stochastic complement" stems from the fact that all stochastic complements are stochastic matrices (see Theorem 2.1) together with the observation that although stochastic complementation is not the same as the well known concept of Schur complementation, there is nevertheless a direct connection in the case of a two-level partition

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{pmatrix}.$$

If $\mathbf{P}$ is an irreducible stochastic matrix with square diagonal blocks, then the stochastic complement of $\mathbf{P}_{11}$ is given by

$$\mathbf{S}_{11} = \mathbf{P}_{11} + \mathbf{P}_{12}(\mathbf{I} - \mathbf{P}_{22})^{-1}\mathbf{P}_{21},$$

and the stochastic complement of $\mathbf{P}_{22}$ is

$$\mathbf{S}_{22} = \mathbf{P}_{22} + \mathbf{P}_{21}(\mathbf{I} - \mathbf{P}_{11})^{-1}\mathbf{P}_{12},$$

and it is easy to see that the stochastic complement of $\mathbf{P}_{11}$ is in fact

$$[\mathbf{I} - \text{Schur Complement}(\mathbf{I} - \mathbf{P}_{22})] \quad \text{in the matrix} \quad \mathbf{I} - \mathbf{P},$$

while the stochastic complement of $\mathbf{P}_{22}$ is

$$[\mathbf{I} - \text{Schur Complement}(\mathbf{I} - \mathbf{P}_{11})] \quad \text{in the matrix} \quad \mathbf{I} - \mathbf{P}.$$

The concept of stochastic complementation arises very naturally in the consideration of finite Markov chains, and more generally, in the theory of nonnegative matrices.

Consequently, it is not surprising that the concept of stochastic complementation—although not always known by that name—appears at least implicitly, if not explicitly, in a variety of places. For example, stochastic complementation is a special case of a more general concept known as *Perron complementation* which has been studied in [17]. Simple forms of stochastic complementation can be found either explicitly or implicitly in a variety of Markov chain applications such as those in [1], [3]–[6], [8]–[10], [12]–[16], [20], [21], in addition to several others. However, it seems that stochastic complementation has heretofore not been focused upon as an entity unto itself, and a detailed study of the properties of stochastic complement matrices has not yet been given. Part of the purpose of this paper is to explicitly publicize the utility of this concept and to present a more complete and unified discussion of the important properties of stochastic complementation.

The following technical lemma is needed to help develop the subsequent theory. Its proof is a straightforward application of the standard features of permutation matrices, and consequently the proof is omitted.

LEMMA 2.1. *For an $m \times m$ irreducible stochastic matrix*

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

in which all diagonal blocks are square, let $\mathbf{Q}$ be the permutation matrix associated with an interchange of the first and ith block rows (or block columns), and let $\tilde{\mathbf{P}}$ be the matrix

$$\tilde{\mathbf{P}} = \mathbf{QPQ}.$$

If $\tilde{\mathbf{P}}$ is partitioned into a 2×2 block matrix

$$\tilde{\mathbf{P}} = \begin{pmatrix} \tilde{\mathbf{P}}_{11} & \tilde{\mathbf{P}}_{12} \\ \tilde{\mathbf{P}}_{21} & \tilde{\mathbf{P}}_{22} \end{pmatrix} \quad \textit{where } \tilde{\mathbf{P}}_{11} = \mathbf{P}_{ii},$$

then the stochastic complement of $\mathbf{P}_{ii}$ in $\mathbf{P}$ is given by

$$\mathbf{S}_{ii} = \tilde{\mathbf{S}}_{11} = \tilde{\mathbf{P}}_{11} + \tilde{\mathbf{P}}_{12}\left(\mathbf{I} - \tilde{\mathbf{P}}_{22}\right)^{-1}\tilde{\mathbf{P}}_{21}. \tag{2.2}$$

We are now in a position to prove that every stochastic complement is indeed a stochastic matrix.

THEOREM 2.1. *If*

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

is an irreducible stochastic matrix, then each stochastic complement

$$\mathbf{S}_{ii} = \mathbf{P}_{ii} + \mathbf{P}_{i*}(\mathbf{I} - \mathbf{P}_i)^{-1}\mathbf{P}_{*i}$$

is also a stochastic matrix.

Proof. For a given index i, assume that $\mathbf{P}$ has been permuted and repartitioned as described in Lemma 2.1 so that we may consider

$$\tilde{\mathbf{P}} = \begin{pmatrix} \tilde{\mathbf{P}}_{11} & \tilde{\mathbf{P}}_{12} \\ \tilde{\mathbf{P}}_{21} & \tilde{\mathbf{P}}_{22} \end{pmatrix} \quad \text{where } \tilde{\mathbf{P}}_{11} = \mathbf{P}_{ii}.$$

According to (2.2), the stochastic complement of $\mathbf{P}_{ii}$ in $\mathbf{P}$ is the same as the stochastic complement of $\tilde{\mathbf{P}}_{11}$ in $\tilde{\mathbf{P}}$—that is, $\mathbf{S}_{ii} = \tilde{\mathbf{S}}_{11}$. Since each principal submatrix of $\mathbf{I} - \mathbf{P}$ (and $\mathbf{I} - \tilde{\mathbf{P}}$) of order $m-1$ or less is a nonsingular M-matrix, it follows from (2.1) that $\left(\mathbf{I} - \tilde{\mathbf{P}}_{22}\right)^{-1} \geq \mathbf{0}$, and hence

$$\tilde{\mathbf{S}}_{11} = \tilde{\mathbf{P}}_{11} + \tilde{\mathbf{P}}_{12}\left(\mathbf{I} - \tilde{\mathbf{P}}_{22}\right)^{-1}\tilde{\mathbf{P}}_{21} \geq 0.$$

Note. $\tilde{\mathbf{S}}_{11}$ need not be strictly positive—an example is given after this proof. To see that the row sums of $\tilde{\mathbf{S}}_{11}$ are each 1, let

$$\mathbf{e} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

and allow the dimension of $\mathbf{e}$ to be defined by the context in which it appears. The fact that the row sums in $\tilde{\mathbf{P}}$ are all 1 can be expressed by writing

$$\tilde{\mathbf{P}}_{11}\mathbf{e} + \tilde{\mathbf{P}}_{12}\mathbf{e} = \mathbf{e} \tag{2.3}$$

and

$$\tilde{\mathbf{P}}_{21}\mathbf{e} + \tilde{\mathbf{P}}_{22}\mathbf{e} = \mathbf{e}. \tag{2.4}$$

Equation (2.4) implies

$$\mathbf{e} = \left(\mathbf{I} - \tilde{\mathbf{P}}_{22}\right)^{-1}\tilde{\mathbf{P}}_{21}\mathbf{e},$$

and this together with (2.3) yields

$$\tilde{\mathbf{S}}_{11}\mathbf{e} = \tilde{\mathbf{P}}_{11}\mathbf{e} + \tilde{\mathbf{P}}_{12}\left(\mathbf{I} - \tilde{\mathbf{P}}_{22}\right)^{-1}\tilde{\mathbf{P}}_{21}\mathbf{e} = \tilde{\mathbf{P}}_{11}\mathbf{e} + \tilde{\mathbf{P}}_{12}\mathbf{e} = \mathbf{e}.$$

Consequently, $\tilde{\mathbf{S}}_{11} = \mathbf{S}_{ii}$ must be stochastic. ∎

To see that a stochastic complement need not be strictly positive, consider a 4×4 irreducible stochastic matrix whose partitions and zero pattern are shown below.

$$\mathbf{P} = \left(\begin{array}{ccc|c} + & + & 0 & 0 \\ + & + & + & + \\ + & + & + & + \\ \hline + & + & + & + \end{array}\right)$$

For this configuration,

$$\mathbf{S}_{11} = \mathbf{P}_{11} + \mathbf{P}_{12}(\mathbf{I} - \mathbf{P}_{22})^{-1}\mathbf{P}_{21} = \begin{pmatrix} + & + & 0 \\ + & + & + \\ + & + & + \end{pmatrix} + \begin{pmatrix} 0 \\ + \\ + \end{pmatrix}[+]\,(+ \quad + \quad +)$$

$$= \begin{pmatrix} + & + & 0 \\ + & + & + \\ + & + & + \end{pmatrix}.$$

Although stochastic complements can have zero entries, the zeros are always in "just the right places" so as to guarantee that each $\mathbf{S}_{ii}$ is an irreducible matrix. However, before this can be established, it is necessary to observe some additional facts concerning stochastic complementation.

THEOREM 2.2. *Suppose that*

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

is an irreducible stochastic matrix in which each $\mathbf{P}_{ii}$ is square, and let

$$\boldsymbol{\pi} = (\, \boldsymbol{\pi}^{(1)} \quad \boldsymbol{\pi}^{(2)} \quad \cdots \quad \boldsymbol{\pi}^{(k)} \,)$$

be the conformably partitioned stationary distribution vector for $\mathbf{P}$. If each $\boldsymbol{\pi}^{(i)}$ is normalized in order to produce the probability vector

$$\mathbf{s}_i = \frac{\boldsymbol{\pi}^{(i)}}{\boldsymbol{\pi}^{(i)}\mathbf{e}}, \tag{2.5}$$

then

$$\mathbf{s}_i\mathbf{S}_{ii} = \mathbf{s}_i \quad \textit{for each } i = 1, 2, \cdots, k.$$

That is, $\mathbf{s}_i$ is a stationary distribution vector for the stochastic complement $\mathbf{S}_{ii}$.

Proof. Assume that $\mathbf{P}$ has been permuted and repartitioned as described in Lemma 2.1 so that

$$\tilde{\mathbf{P}} = \mathbf{QPQ} = \begin{pmatrix} \tilde{\mathbf{P}}_{11} & \tilde{\mathbf{P}}_{12} \\ \tilde{\mathbf{P}}_{21} & \tilde{\mathbf{P}}_{22} \end{pmatrix}.$$

Use the fact that $\mathbf{0} = \boldsymbol{\pi}(\mathbf{I} - \mathbf{P})$ implies

$$\mathbf{0} = \boldsymbol{\pi}\mathbf{Q}^2(\mathbf{I} - \mathbf{P})\mathbf{Q} = (\, \boldsymbol{\pi}^{(i)} \quad \boldsymbol{\pi}^{(2)} \quad \cdots \quad \boldsymbol{\pi}^{(1)} \quad \cdots \quad \boldsymbol{\pi}^{(k)} \,)\,(\mathbf{I} - \tilde{\mathbf{P}})$$

together with the equation

$$\begin{pmatrix} \mathbf{I} - \tilde{\mathbf{P}}_{11} & -\tilde{\mathbf{P}}_{12} \\ -\tilde{\mathbf{P}}_{21} & \mathbf{I} - \tilde{\mathbf{P}}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ (\mathbf{I} - \tilde{\mathbf{P}}_{22})^{-1}\tilde{\mathbf{P}}_{21} & \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{I} - \tilde{\mathbf{S}}_{11} & -\tilde{\mathbf{P}}_{12} \\ \mathbf{0} & \mathbf{I} - \tilde{\mathbf{P}}_{22} \end{pmatrix}$$

in order to conclude that

$$\boldsymbol{\pi}^{(i)}(\mathbf{I} - \tilde{\mathbf{S}}_{11}) = \mathbf{0}.$$

The desired result now follows from Lemma 2.1 because $\mathbf{S}_{ii} = \tilde{\mathbf{S}}_{11}$. ∎

Although Theorem 2.2 establishes that each $\mathbf{s}_i$ is a stationary distribution vector for $\mathbf{S}_{ii}$, nothing proven to this point allows for the conclusion that the $\mathbf{s}_i$'s are unique. However, this will follow once it is established that each $\mathbf{S}_{ii}$ inherits the property of irreducibility from the original matrix $\mathbf{P}$.

THEOREM 2.3. *If*

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

is an irreducible stochastic matrix, then each stochastic complement

$$\mathbf{S}_{ii} = \mathbf{P}_{ii} + \mathbf{P}_{i*}(\mathbf{I} - \mathbf{P}_i)^{-1}\mathbf{P}_{*i}$$

is also an ***irreducible*** *stochastic matrix.*

Proof. Assume that $\mathbf{P}$ has been permuted and repartitioned as described in Lemma 2.1 so that

$$\tilde{\mathbf{P}} = \begin{pmatrix} \tilde{\mathbf{P}}_{11} & \tilde{\mathbf{P}}_{12} \\ \tilde{\mathbf{P}}_{21} & \tilde{\mathbf{P}}_{22} \end{pmatrix}$$

and

$$\mathbf{S}_{ii} = \tilde{\mathbf{S}}_{11}. \tag{2.6}$$

The fact that each $\mathbf{S}_{ii}$ is stochastic was established in Theorem 2.1. By virtue of (2.6), we need only to prove irreducibility of $\tilde{\mathbf{S}}_{11}$. By noting that

$$\begin{pmatrix} \mathbf{I} & \tilde{\mathbf{P}}_{12}\left(\mathbf{I} - \tilde{\mathbf{P}}_{22}\right)^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{I} - \tilde{\mathbf{P}}_{11} & -\tilde{\mathbf{P}}_{12} \\ -\tilde{\mathbf{P}}_{21} & \mathbf{I} - \tilde{\mathbf{P}}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \left(\mathbf{I} - \tilde{\mathbf{P}}_{22}\right)^{-1}\tilde{\mathbf{P}}_{21} & \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{I} - \tilde{\mathbf{S}}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} - \tilde{\mathbf{P}}_{22} \end{pmatrix},$$

it follows that

$$\text{Rank}\left(\mathbf{I} - \tilde{\mathbf{P}}\right) = \text{Rank}\left(\mathbf{I} - \tilde{\mathbf{S}}_{11}\right) + \text{Rank}\left(\mathbf{I} - \tilde{\mathbf{P}}_{22}\right).$$

Suppose that $\tilde{\mathbf{P}}$, $\tilde{\mathbf{P}}_{11}$, and $\tilde{\mathbf{P}}_{22}$ have dimensions $m \times m$, $r \times r$, and $q \times q$, respectively, with $r+q = m$. It is well known (see [2, p. 156]) that $\tilde{\mathbf{P}}$ being irreducible and stochastic implies that $\text{Rank}\left(\mathbf{I} - \tilde{\mathbf{P}}\right) = m - 1$ and $\text{Rank}\left(\mathbf{I} - \tilde{\mathbf{P}}_{22}\right) = q$. Consequently,

$$\text{Rank}\left(\mathbf{I} - \tilde{\mathbf{S}}_{11}\right) = m - 1 - q = r - 1,$$

and therefore $\mathbf{I} - \tilde{\mathbf{S}}_{11}$ has a one-dimensional nullspace. The left-hand nullspace of $\mathbf{I} - \tilde{\mathbf{S}}_{11}$ is spanned by the strictly positive row vector

$$\tilde{\mathbf{s}}_1 = \mathbf{s}_i$$

defined in (2.5), and the right-hand nullspace of $\mathbf{I} - \tilde{\mathbf{S}}_{11}$ is spanned by the strictly positive column vector $\mathbf{e}$ containing r 1's. Since $\tilde{\mathbf{s}}_1\mathbf{e} = 1$, the spectral projector associated with the eigenvalue $\lambda = 1$ for $\tilde{\mathbf{S}}_{11}$ must be

$$\mathbf{R}_{r\times r} = \mathbf{e}\tilde{\mathbf{s}}_1 > \mathbf{0}.$$

Because every stochastic matrix is Cesàro summable to the spectral projector associated with the unit eigenvalue (see [11]), it follows that

$$\lim_{n\to\infty} \frac{\mathbf{I} + \tilde{\mathbf{S}}_{11} + \tilde{\mathbf{S}}_{11}^2 + \cdots + \tilde{\mathbf{S}}_{11}^{n-1}}{n} = \mathbf{R} > \mathbf{0}. \tag{2.7}$$

It is now evident that $\tilde{\mathbf{S}}_{11}$ cannot be reducible—otherwise $\tilde{\mathbf{S}}_{11}$ could be permuted to a form

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}$$

and the limit in (2.7) would necessarily contain zero entries. □

The proof given above is algebraic in nature, but because irreducibility is a strictly combinatorial concept, one might wonder if a strictly combinatorial proof is possible. The answer is *yes,* and the details may be obtained by restricting the discussion given in [17] to the special case of stochastic matrices. Furthermore, it seems reasonable that there should be a probabilistic argument, and indeed there is. Irreducibility is a direct consequence of the probabilistic interpretation of stochastic complementation as explained in the next section.

3. The probabilistic interpretation of stochastic complementation. Assume again that for a given index i, the original transition matrix $\mathbf{P}$ has been permuted and repartitioned as described in Lemma 2.1 so that

$$\tilde{\mathbf{P}} = \begin{array}{c} \\ \mathcal{A} \\ \mathcal{B} \end{array} \begin{array}{c} \begin{array}{cc} \mathcal{A} & \mathcal{B} \end{array} \\ \begin{pmatrix} \tilde{\mathbf{P}}_{11} & \tilde{\mathbf{P}}_{12} \\ \tilde{\mathbf{P}}_{21} & \tilde{\mathbf{P}}_{22} \end{pmatrix} \end{array} \quad \text{where } \tilde{\mathbf{P}}_{11} = \mathbf{P}_{ii},$$

and

$$\mathbf{S}_{ii} = \tilde{\mathbf{S}}_{11} = \tilde{\mathbf{P}}_{11} + \tilde{\mathbf{P}}_{12}\left(\mathbf{I} - \tilde{\mathbf{P}}_{22}\right)^{-1}\tilde{\mathbf{P}}_{21}.$$

Consider a new process—called the ***reduced chain***—which is defined by observing the old process only when it is in a state belonging to the subclass $\mathcal{A}$. That is, transitions to states in $\mathcal{B}$ are masked out, so a direct path $\mathcal{A}_k \Longrightarrow \mathcal{A}_j$ in the reduced chain corresponds in the old process to either a direct path $\mathcal{A}_k \longrightarrow \mathcal{A}_j$ or a detour

$$\mathcal{A}_k \longrightarrow \mathcal{B} \longrightarrow \mathcal{A}_j$$

passing through $\mathcal{B}$. In loose terms, one can say that the reduced chain is derived from the original chain by "turning off the meter" whenever the old process is in $\mathcal{B}$. If the original chain is irreducible, then it is clear that the reduced chain is also irreducible. For example, if

$$\mathcal{A}_1 \longrightarrow \mathcal{B}_2 \longrightarrow \mathcal{A}_3 \longrightarrow \mathcal{B}_4 \longrightarrow \mathcal{B}_5 \longrightarrow \mathcal{A}_6 \longrightarrow \mathcal{A}_7$$

is a sequence of direct paths from $\mathcal{A}_1$ to $\mathcal{A}_7$ in the original chain, then

$$\mathcal{A}_1 \Longrightarrow \mathcal{A}_3 \Longrightarrow \mathcal{A}_6 \Longrightarrow \mathcal{A}_7$$

is a sequence of direct paths from $\mathcal{A}_1$ to $\mathcal{A}_7$ in the reduced chain.

For the reduced chain, the one-step transition probability of moving from $\mathcal{A}_k$ to $\mathcal{A}_j$ is the probability in the original process of moving directly from $\mathcal{A}_k$ to $\mathcal{A}_j$ plus the probability of moving directly from $\mathcal{A}_k$ to some state in $\mathcal{B}$, and then eventually moving back to $\mathcal{A}$, hitting $\mathcal{A}_j$ first upon return. The probability of moving directly from $\mathcal{A}_k$ to $\mathcal{A}_j$ in the original chain is

$$q_{kj} = \left[\tilde{\mathbf{P}}_{11}\right]_{kj},$$

and the probability of moving directly from $\mathcal{A}_k$ to $\mathcal{B}_h \in \mathcal{B}$ is

$$q_{kh} = \left[\tilde{\mathbf{P}}_{12}\right]_{kh}.$$

The probability of moving from $\mathcal{B}_h$ to $\mathcal{A}$ such that $\mathcal{A}_j$ is the first state entered upon return to $\mathcal{A}$ is

$$q_{hj} = \left[\left(\mathbf{I} - \tilde{\mathbf{P}}_{22}\right)^{-1} \tilde{\mathbf{P}}_{21}\right]_{hj}.$$

As explained in [12], this expression for q_{hj} is obtained by considering the states in $\mathcal{A}$ to be absorbing states and applying the theory of absorbing chains. Consequently, the one-step transition probabilities in the reduced chain are

$$q_{kj} + \sum_{\mathcal{B}_h \in \mathcal{B}} q_{kh} q_{hj} = \left[\tilde{\mathbf{P}}_{11} + \tilde{\mathbf{P}}_{12}\left(\mathbf{I} - \tilde{\mathbf{P}}_{22}\right)^{-1} \tilde{\mathbf{P}}_{21}\right]_{kj} = \left[\mathbf{S}_{ii}\right]_{kj}.$$

In other words, *the stochastic complement* $\mathbf{S}_{ii}$ *represents the transition matrix for the reduced chain which is obtained from the original chain by masking out transitions to states in* $\mathcal{B}$.

4. Uncoupling Markov chains. For an m-state irreducible Markov chain $\mathcal{C}$ with transition matrix $\mathbf{P}$, the object of uncoupling is to somehow decompose the chain $\mathcal{C}$ into two or more smaller chains—say $\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_k$—containing $r_1, r_2, \cdots, r_k$ states, respectively, where $\sum_{i=1}^{k} r_i = m$. This sequence of smaller chains is required to possess the following properties:

- Each smaller chain $\mathcal{C}_i$ should be irreducible whenever the original chain $\mathcal{C}$ is irreducible so as to guarantee that each $\mathcal{C}_i$ has a unique stationary distribution vector $\mathbf{s}_i$.
- It should be possible to determine the $\mathbf{s}_i$'s completely independent of each other. For modern multiprocessor computer architectures it is desirable to be able to execute the computation of the $\mathbf{s}_i$'s in parallel.
- Finally, it must be possible to easily couple the smaller stationary distribution vectors $\mathbf{s}_i$ back together in order to produce the stationary distribution vector $\boldsymbol{\pi}$ for the larger chain $\mathcal{C}$.

The purpose of this section is to show how the concept of stochastic complementation can be used to achieve these goals. As in the previous section, $\mathbf{P}$ will denote an $m \times m$ irreducible stochastic matrix with a k-level partition

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

in which all diagonal blocks $\mathbf{P}_{ii}$ are square, and the stationary distribution vector $\boldsymbol{\pi}$ for $\mathbf{P}$ will be partitioned as

$$\boldsymbol{\pi} = \begin{pmatrix} \boldsymbol{\pi}^{(1)} & \boldsymbol{\pi}^{(2)} & \cdots & \boldsymbol{\pi}^{(k)} \end{pmatrix}$$

where the size of $\boldsymbol{\pi}^{(i)}$ corresponds to the order of $\mathbf{P}_{ii}$. We know from the results of Theorem 2.3 that each stochastic complement

$$\mathbf{S}_{ii} = \mathbf{P}_{ii} + \mathbf{P}_{i*}(\mathbf{I} - \mathbf{P}_i)^{-1}\mathbf{P}_{*i}$$

is an irreducible stochastic matrix, and consequently, each $\mathbf{S}_{ii}$ possesses a unique stationary distribution vector $\mathbf{s}_i$. Theorem 2.2 shows that each $\mathbf{s}_i$ and $\boldsymbol{\pi}^{(i)}$ differ only by a positive scalar multiple so that it is indeed possible to combine the $\mathbf{s}_i$'s with "coupling factors" $\{\xi_1, \xi_2, \cdots, \xi_k\}$ in order to produce

$$\boldsymbol{\pi} = (\, \xi_1\mathbf{s}_1 \quad \xi_2\mathbf{s}_2 \quad \cdots \quad \xi_k\mathbf{s}_k \,).$$

However, the expressions for the coupling factors ξ_i given in Theorem 2.2 have the form

$$\xi_i = \sum_h \boldsymbol{\pi}_h^{(i)}.$$

At first glance, this might seem to place the issue in a hopeless circle because prior knowledge of $\boldsymbol{\pi}$, in the form of the sums $\sum_h \boldsymbol{\pi}_h^{(i)}$, is necessary in order to reconstruct $\boldsymbol{\pi}$ from the $\mathbf{s}_i$'s. Fortunately, there is an elegant way around this dilemma. The following theorem shows that the coupling factors ξ_i are easily determined without prior knowledge of the $\boldsymbol{\pi}^{(i)}$'s—this will become the key feature in the uncoupling-coupling technique.

THEOREM 4.1 (The coupling theorem). *If $\mathbf{P}$ is an $m \times m$ irreducible stochastic matrix partitioned as*

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

with square diagonal blocks, then the stationary distribution vector for $\mathbf{P}$ is given by

$$\boldsymbol{\pi} = (\, \xi_1\mathbf{s}_1 \quad \xi_2\mathbf{s}_2 \quad \cdots \quad \xi_k\mathbf{s}_k \,)$$

where $\mathbf{s}_i$ is the unique stationary distribution vector for the stochastic complement

$$\mathbf{S}_{ii} = \mathbf{P}_{ii} + \mathbf{P}_{i*}(\mathbf{I} - \mathbf{P}_i)^{-1}\mathbf{P}_{*i},$$

and where

$$\boldsymbol{\xi} = (\, \xi_1 \quad \xi_2 \quad \cdots \quad \xi_k \,)$$

is the unique stationary distribution vector for the $k \times k$ irreducible stochastic matrix $\mathbf{C}$ whose entries are defined by

$$c_{ij} \equiv \mathbf{s}_i\mathbf{P}_{ij}\mathbf{e}.$$

*The matrix $\mathbf{C}$ is hereafter referred to as the **coupling matrix,** and the scalars ξ_i are called the **coupling factors.***

Proof. First prove that the coupling matrix $\mathbf{C}$ is stochastic and irreducible. By its definition, it is clear that $\mathbf{C} \geq \mathbf{0}$. To see that $\mathbf{C}$ is stochastic, use the fact that $\sum_{j=1}^{k} \mathbf{P}_{ij}\mathbf{e} = \mathbf{e}$ and write

$$\sum_{j=1}^{k} c_{ij} = \sum_{j=1}^{k} \mathbf{s}_i\mathbf{P}_{ij}\mathbf{e} = \mathbf{s}_i \left(\sum_{j=1}^{k} \mathbf{P}_{ij}\mathbf{e} \right) = \mathbf{s}_i\mathbf{e} = 1.$$

To show that $\mathbf{C}$ is irreducible, note that because $\mathbf{P}_{ij} \geq \mathbf{0}$, $\mathbf{s}_i > \mathbf{0}$, and $\mathbf{e} > \mathbf{0}$, it must be the case that

$$c_{ij} = 0 \quad \text{if and only if } \mathbf{P}_{ij} = \mathbf{0}.$$

Since $\mathbf{P}$ is irreducible, this implies that $\mathbf{C}$ must also be irreducible—otherwise, if $\mathbf{C}$ could be permuted to a block triangular form, then so could $\mathbf{P}$. Let

$$\boldsymbol{\pi} = \begin{pmatrix} \boldsymbol{\pi}^{(1)} & \boldsymbol{\pi}^{(2)} & \cdots & \boldsymbol{\pi}^{(k)} \end{pmatrix}$$

denote the partitioned stationary distribution vector for $\mathbf{P}$ where the sizes of the $\boldsymbol{\pi}^{(i)}$'s correspond to the sizes of the $\mathbf{P}_{ii}$'s, respectively, and define $\boldsymbol{\xi}$ to be the vector

$$\boldsymbol{\xi} = \begin{pmatrix} \xi_1 & \xi_2 & \cdots & \xi_k \end{pmatrix}$$

where each component is defined by

$$\xi_i = \boldsymbol{\pi}^{(i)}\mathbf{e} = \sum_h \boldsymbol{\pi}_h^{(i)}.$$

According to Theorem 2.2,

$$\xi_i \mathbf{s}_i = \boldsymbol{\pi}^{(i)}, \tag{4.1}$$

and this together with the fact that $\sum_{i=1}^{k} \boldsymbol{\pi}^{(i)}\mathbf{P}_{ij} = \boldsymbol{\pi}^{(j)}$ yields

$$(\boldsymbol{\xi}\mathbf{C})_j = \sum_{i=1}^{k} \xi_i c_{ij} = \sum_{i=1}^{k} \boldsymbol{\pi}^{(i)}\mathbf{P}_{ij}\mathbf{e} = \left(\sum_{i=1}^{k} \boldsymbol{\pi}^{(i)}\mathbf{P}_{ij} \right) \mathbf{e} = \boldsymbol{\pi}^{(j)}\mathbf{e} = \xi_j.$$

Consequently, $\boldsymbol{\xi}\mathbf{C} = \boldsymbol{\xi}$ so that $\boldsymbol{\xi}$ is a stationary vector for $\mathbf{C}$. It is clear that $\boldsymbol{\xi}$ must also be a probability vector because

$$\sum_{i=1}^{k} \xi_i = \sum_{i=1}^{k} \boldsymbol{\pi}^{(i)}\mathbf{e} = \sum_{j=1}^{m} \pi_j = 1.$$

Therefore, since $\mathbf{C}$ is irreducible, $\boldsymbol{\xi}$ must be the unique stationary distribution vector for $\mathbf{C}$, and the desired conclusion that

$$\boldsymbol{\pi} = \begin{pmatrix} \xi_1\mathbf{s}_1 & \xi_2\mathbf{s}_2 & \cdots & \xi_k\mathbf{s}_k \end{pmatrix}$$

follows from (4.1). □

The results of Theorem 4.1 can be viewed as an "exact aggregation" technique. The case of a two-level partition is of special interest because, as the following corollary shows, it is particularly easy to uncouple and couple the stationary distribution vector for these situations.

COROLLARY 4.1. *If* $\mathbf{P}$ *is an* $m \times m$ *irreducible stochastic matrix partitioned as*

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{pmatrix}$$

where $\mathbf{P}_{11}$ *and* $\mathbf{P}_{22}$ *are square, then the stationary distribution vector for* $\mathbf{P}$ *is given by*

$$\boldsymbol{\pi} = \begin{pmatrix} \xi_1\mathbf{s}_1 & \xi_2\mathbf{s}_2 \end{pmatrix}$$

where $\mathbf{s}_1$ *and* $\mathbf{s}_2$ *are the unique stationary distribution vectors for the stochastic complements*

$$\mathbf{S}_{11} = \mathbf{P}_{11} + \mathbf{P}_{12}(\mathbf{I} - \mathbf{P}_{22})^{-1}\mathbf{P}_{21} \quad \text{and} \quad \mathbf{S}_{22} = \mathbf{P}_{22} + \mathbf{P}_{21}(\mathbf{I} - \mathbf{P}_{11})^{-1}\mathbf{P}_{12},$$

respectively, and where the coupling factors ξ_1 *and* ξ_2 *are given by*

$$\xi_1 = \frac{\mathbf{s}_2\mathbf{P}_{21}\mathbf{e}}{\mathbf{s}_1\mathbf{P}_{12}\mathbf{e} + \mathbf{s}_2\mathbf{P}_{21}\mathbf{e}} \quad \text{and} \quad \xi_2 = 1 - \xi_1 = \frac{\mathbf{s}_1\mathbf{P}_{12}\mathbf{e}}{\mathbf{s}_1\mathbf{P}_{12}\mathbf{e} + \mathbf{s}_2\mathbf{P}_{21}\mathbf{e}}.$$

In this corollary, the coupling factors ξ_i are described in terms of the off-diagonal blocks $\mathbf{P}_{12}$ and $\mathbf{P}_{21}$, but these quantities can also be expressed using only the diagonal blocks in $\mathbf{P}$ by replacing the terms $\mathbf{P}_{12}\mathbf{e}$ and $\mathbf{P}_{21}\mathbf{e}$ with $\mathbf{e} - \mathbf{P}_{11}\mathbf{e}$ and $\mathbf{e} - \mathbf{P}_{22}\mathbf{e}$, respectively, so as to produce

$$\xi_1 = \frac{1 - \mathbf{s}_2\mathbf{P}_{22}\mathbf{e}}{2 - \mathbf{s}_1\mathbf{P}_{11}\mathbf{e} - \mathbf{s}_2\mathbf{P}_{22}\mathbf{e}} \quad \text{and} \quad \xi_2 = 1 - \xi_1 = \frac{1 - \mathbf{s}_1\mathbf{P}_{11}\mathbf{e}}{2 - \mathbf{s}_1\mathbf{P}_{11}\mathbf{e} - \mathbf{s}_2\mathbf{P}_{22}\mathbf{e}}.$$

There is always a balancing act to be performed when uncoupling a Markov chain using stochastic complementation as described in Theorem 4.1. As k increases and the partition of $\mathbf{P}$ becomes finer, the sizes of the stochastic complements become smaller thus making it easier to determine each of the stationary distribution vectors $\mathbf{s}_i$, but the order of the matrix inversion embedded in each stochastic complement becomes larger, and the size of the coupling matrix $\mathbf{C}_{k\times k}$ becomes larger. In the two extreme cases when $k = m$ or $k = 1$, there is no uncoupling of the chain whatsoever—if $k = m$, then $\mathbf{C} = \mathbf{P}$, and if $k = 1$, then $\mathbf{S}_{11} = \mathbf{P}$. One must therefore choose the partition which best suits the needs of the underlying application. For example, if computation utilizing a particular multiprocessor computer is the goal, then the specific nature of the hardware and associated software may dictate the partitioning strategy.

Rather than performing a single uncoupling-coupling operation to a high level partition of $\mathbf{P}$, an alternate strategy is to execute a divide-and-conquer procedure using only two-level partitions at each stage. Starting with an irreducible stochastic matrix $\mathbf{P}$ of size $m \times m$, partition $\mathbf{P}$ roughly in half as

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{pmatrix}$$

to produce two stochastic complements, $\mathbf{S}_{11}$ and $\mathbf{S}_{22}$, which are each irreducible stochastic matrices of order approximately $m/2$. The stochastic complements $\mathbf{S}_{11}$ and $\mathbf{S}_{22}$ may in turn be partitioned roughly in half so as to produce four stochastic complements—say $(\mathbf{S}_{11})_{11}$, $(\mathbf{S}_{11})_{22}$, $(\mathbf{S}_{22})_{11}$, and $(\mathbf{S}_{22})_{22}$—each of which is of order approximately $m/4$. This process can continue until all stochastic complements are sufficiently small in size so that each easily yields a stationary distribution vector. The small stationary distribution vectors corresponding to the small stochastic complements are then successively coupled according to the rules given in Corollary 4.1 until the stationary distribution vector $\boldsymbol{\pi}$ for the original chain is produced. Using this divide-and-conquer strategy, there are only two coupling factors to determine at each coupling step, and they are almost trivial to compute. Furthermore, the order of the largest inversion imbedded in any stochastic complement never exceeds $m/2$.

For example, consider the following 8×8 irreducible stochastic matrix.

$$\mathbf{P} = \left(\begin{array}{cccc|cccc} 0 & .5 & 0 & .1 & .2 & 0 & .1 & .1 \\ .2 & 0 & .5 & 0 & 0 & .2 & 0 & .1 \\ .1 & .2 & 0 & .5 & .1 & 0 & .1 & 0 \\ 0 & .4 & .2 & 0 & .1 & .1 & 0 & .2 \\ \hline .1 & 0 & .5 & 0 & .1 & .1 & 0 & .2 \\ 0 & .5 & 0 & .2 & .1 & 0 & .2 & 0 \\ .5 & 0 & .2 & .1 & 0 & .1 & 0 & .1 \\ 0 & .2 & .4 & 0 & .2 & 0 & .1 & .1 \end{array}\right)$$

For the indicated partition, the stochastic complements of $\mathbf{P}$ are given by[1]

$$\mathbf{S}_{11} = \left(\begin{array}{cc|cc} .0899 & .5566 & .2329 & .1205 \\ .2327 & .1311 & .5885 & .0476 \\ \hline .1665 & .2194 & .0986 & .5155 \\ .0443 & .5156 & .4102 & .0299 \end{array}\right)$$

and

$$\mathbf{S}_{22} = \left(\begin{array}{cc|cc} .2832 & .2409 & .1110 & .3649 \\ .2576 & .2395 & .2819 & .2210 \\ \hline .2511 & .2823 & .1329 & .3337 \\ .3633 & .1683 & .2006 & .2678 \end{array}\right),$$

with coupling vector

$$\boldsymbol{\xi}^{(0)} = (\, .6852 \quad .3148 \,).$$

Now, the two stochastic complements of $\mathbf{S}_{11}$ are

$$(\mathbf{S}_{11})_{11} = \begin{pmatrix} .1744 & .8256 \\ .4041 & .5959 \end{pmatrix} \quad \text{and} \quad (\mathbf{S}_{11})_{22} = \begin{pmatrix} .4278 & .5722 \\ .9056 & .0944 \end{pmatrix}$$

with coupling vector

$$\boldsymbol{\xi}^{(1)} = (\, .4548 \quad .5452 \,),$$

while the two stochastic complements of $\mathbf{S}_{22}$ are

$$(\mathbf{S}_{22})_{11} = \begin{pmatrix} .5776 & .4224 \\ .5512 & .4488 \end{pmatrix} \quad \text{and} \quad (\mathbf{S}_{22})_{22} = \begin{pmatrix} .3468 & .6532 \\ .3955 & .6045 \end{pmatrix}$$

with coupling vector

$$\boldsymbol{\xi}^{(2)} = (\, .5219 \quad .4781 \,).$$

[1] Numbers have been rounded so as to display four digits behind the decimal point.

The stationary distribution vector for $(\mathbf{S}_{11})_{11}$ is given by

$$\mathbf{s}_1^{(1)} = (\,.3286 \quad .6714\,),$$

and the stationary distribution vector for $(\mathbf{S}_{11})_{22}$ is

$$\mathbf{s}_2^{(1)} = (\,.6128 \quad .3872\,),$$

so that the stationary distribution vector for $\mathbf{S}_{11}$ is

$$\begin{aligned}\mathbf{s}_1 &= \left(\boldsymbol{\xi}_1^{(1)}\mathbf{s}_1^{(1)} \;\middle|\; \boldsymbol{\xi}_2^{(1)}\mathbf{s}_2^{(1)}\right)\\ &= (\,.4548 \times (\,.3286 \quad .6714\,) \;\middle|\; .5452 \times (\,.6128 \quad .3872\,)\,)\\ &= (\,.1495 \quad .3054 \quad .3341 \quad .2111\,).\end{aligned}$$

Similarly, the stationary distribution vector for $(\mathbf{S}_{22})_{11}$ is given by

$$\mathbf{s}_1^{(2)} = (\,.5661 \quad .4339\,),$$

and the stationary distribution vector for $(\mathbf{S}_{22})_{22}$ is

$$\mathbf{s}_2^{(2)} = (\,.3771 \quad .6229\,),$$

so that the stationary distribution vector for $\mathbf{S}_{22}$ is

$$\begin{aligned}\mathbf{s}_2 &= \left(\boldsymbol{\xi}_1^{(2)}\mathbf{s}_1^{(2)} \;\middle|\; \boldsymbol{\xi}_2^{(2)}\mathbf{s}_2^{(2)}\right)\\ &= (\,.5219 \times (\,.5661 \quad .4339\,) \;\middle|\; .4781 \times (\,.3771 \quad .6229\,)\,)\\ &= (\,.2955 \quad .2264 \quad .1803 \quad .2978\,).\end{aligned}$$

Therefore, the stationary distribution vector for $\mathbf{P}$ must be

$$\begin{aligned}\boldsymbol{\pi} &= \left(\boldsymbol{\xi}_1^{(0)}\mathbf{s}_1 \;\middle|\; \boldsymbol{\xi}_2^{(0)}\mathbf{s}_2\right)\\ &= \Big(.6852 \times (\,.1495 \quad .3054 \quad .3341 \quad .2111\,) \;\Big|\\ &\qquad\qquad .3148 \times (\,.2955 \quad .2264 \quad .1803 \quad .2978\,)\Big)\\ &= (\,.1024 \quad .2092 \quad .2289 \quad .1446 \quad .0930 \quad .0713 \quad .0568 \quad .0938\,).\end{aligned}$$

In addition to the divide-and-conquer process illustrated above, there are several other variations and hybrid techniques (e.g., iterative methods) which are possible, and it is clear that the remarks of this section can serve as the basis for fully parallel algorithms for computing the stationary distribution vector for an irreducible chain. R. B. Mattingly has conducted some detailed experiments along these lines in which he implemented the divide-and-conquer technique described above on a Sequent Balance 21000—a shared memory machine with 24 tightly coupled processors. Exploring the numerical details and implementation of Mattingly's work would lead this exposition too far astray, but the interested reader can consult [14] or [15]. In brief, Mattingly was able to achieve moderate speedups with the stochastic complementation divide-and-conquer technique—e.g., a speedup of approximately 8.5 with 16 processors was obtained. Although the operation count for straightforward elimination is less than that for the divide-and-conquer technique, elimination methods do not parallelize as well as the divide-and-conquer technique, so, as the number of

processors was increased, the actual running times of the divide-and-conquer technique were competitive with optimized elimination schemes. But perhaps even more significant is the fact that the divide-and-conquer technique can be extremely stable for ill-conditioned chains—i.e., irreducible stochastic matrices which have a cluster of eigenvalues very near the unit eigenvalue. For Mattingly's ill-conditioned test matrices, standard elimination methods returned only one or two correct digits whereas the divide-and-conquer scheme based on stochastic complementation returned results which were correct to the machine precision. This in part is due to the spectrum splitting effect which stochastic complementation provides—these results are discussed in §6 of this exposition which concerns the spectral properties of stochastic complementation.

5. Primitivity issues. Primitivity is, of course, an important issue because for an irreducible stochastic matrix $\mathbf{P}$, the limit $\lim_{n\to\infty}\mathbf{P}^k$ exists if and only if $\mathbf{P}$ is primitive. If $\mathbf{P}$ is not primitive, then we are necessarily restricted to investigating limiting behavior in the weak (the Cesàro) sense, and consequently it is worthwhile to make some observations concerning the degree to which primitivity—or lack of it—in a partitioned stochastic matrix $\mathbf{P}$ is inherited by the smaller stochastic complements $\mathbf{S}_{ii}$.

The first observation to make is that $\mathbf{P}$ being a primitive matrix is not sufficient to guarantee that all stochastic complements are primitive. For example, the matrix

$$\mathbf{P} = \left(\begin{array}{cc|c} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \hline 1 & 0 & 0 \end{array}\right)$$

is irreducible and primitive because $\mathbf{P}^5 > 0$. However, for the indicated partition, the stochastic complement

$$\mathbf{S}_{11} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is not primitive. Although primitivity of a stochastic complement is not inherited from the matrix $\mathbf{P}$ itself, the following theorem shows that primitivity is inherited from the diagonal blocks of $\mathbf{P}$.

THEOREM 5.1. *If $\mathbf{P}_{ii}$ is primitive, then the corresponding stochastic complement $\mathbf{S}_{ii}$ must also be primitive.*

Proof. Since $\mathbf{P}_{ii} \geq \mathbf{0}$ and $\mathbf{P}_{i*}(\mathbf{I}-\mathbf{P}_i)^{-1}\mathbf{P}_{*i} \geq \mathbf{0}$, it follows that for each positive integer n,

$$\mathbf{S}_{ii}^n = \left[\mathbf{P}_{ii} + \mathbf{P}_{i*}(\mathbf{I}-\mathbf{P}_i)^{-1}\mathbf{P}_{*i}\right]^n = \mathbf{P}_{ii}^n + \mathbf{N}$$

where $\mathbf{N} \geq \mathbf{0}$. Therefore, $\mathbf{S}_{ii}^n > \mathbf{0}$ whenever $\mathbf{P}_{ii}^n > \mathbf{0}$. ∎

While stochastic complements need not be primitive, most of them are. The next theorem explains why those stochastic complements which are not primitive must come from rather special stochastic matrices.

THEOREM 5.2. *If $\mathbf{P}_{ii}$ has at least one nonzero diagonal entry, then the corresponding stochastic complement $\mathbf{S}_{ii}$ must be primitive.*

Proof. If $\mathbf{P}_{ii}$ has at least one nonzero diagonal entry, then so does $\mathbf{S}_{ii}$. Furthermore, Theorems 2.1 and 2.3 guarantee that each $\mathbf{S}_{ii}$ is always nonnegative and irreducible, and it is well known [2, p. 34] that an irreducible nonnegative matrix with a positive trace must be primitive, so each $\mathbf{S}_{ii}$ must be primitive. ∎

The converse of Theorem 5.1 as well as the converse of Theorem 5.2 is false. The stochastic matrix

$$\mathbf{P} = \left(\begin{array}{ccc|ccc} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \tag{5.1}$$

is irreducible, and the stochastic complements corresponding to the indicated partition are

$$\mathbf{S}_{11} = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{S}_{22} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}.$$

Notice that each stochastic complement is primitive, but $\mathbf{P}$, $\mathbf{P}_{11}$, and $\mathbf{P}_{22}$ are not primitive.

The preceding example indicates another advantage which stochastic complementation can provide. The matrix $\mathbf{P}$ in (5.1) is not primitive because it is the transition matrix of a periodic chain, and hence the associated stationary distribution vector $\boldsymbol{\pi}$ *cannot* be computed by the power method—i.e., by the simple iteration

$$\boldsymbol{\pi}_{n+1} = \boldsymbol{\pi}_n \mathbf{P} \quad \text{where } \boldsymbol{\pi}_0 \text{ is arbitrary.}$$

However, the chain uncouples into two aperiodic chains in the sense that $\mathbf{S}_{11}$ and $\mathbf{S}_{22}$ are both primitive, and therefore each $\mathbf{S}_{ii}$ yields a stationary distribution vector $\mathbf{s}_i$ by means of two straightforward iterations

$$\mathbf{s}_1^{(n+1)} = \mathbf{s}_1^{(n)} \mathbf{S}_{11} \quad \text{and} \quad \mathbf{s}_2^{(n+1)} = \mathbf{s}_2^{(n)} \mathbf{S}_{22} \quad \text{where each } \mathbf{s}_i^{(0)} \text{ is arbitrary.}$$

Take note of the fact that the two iterations represented here are completely independent of each other, and consequently they can be implemented simultaneously. By using the coupling factors described in Corollary 4.1, it is easy to couple the two limiting distributions $\mathbf{s}_1$ and $\mathbf{s}_2$ in order to produce the stationary distribution vector $\boldsymbol{\pi}$ for the larger periodic chain. When these observations are joined with the results of §6 of this exposition pertaining to the spectrum splitting effects which stochastic complementation can afford, it will become even more apparent that stochastic complementation has the potential to become a valuable computational technique for use with multiprocessor machines.

6. Nearly completely reducible chains and spectral properties. Consider an m-state irreducible chain $\mathcal{C}$ and k subclasses $\mathcal{Q}_1, \mathcal{Q}_2, \cdots, \mathcal{Q}_k$ which partition the state space. The chain $\mathcal{C}$ as well as an associated transition matrix is considered to be *nearly completely reducible* [2] when the $\mathcal{Q}_i$'s are only very weakly coupled together.

[2] Gantmacher's [7] terminology "completely reducible" and the associated phrase "nearly completely reducible" are adopted in this exposition. Some authors use the alternate terminology "nearly completely decomposable."

In this case, the states can be ordered so as to make the transition matrix $\mathbf{P}$ have a k-level partition

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \cdots & \mathbf{P}_{1k} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \cdots & \mathbf{P}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{k1} & \mathbf{P}_{k2} & \cdots & \mathbf{P}_{kk} \end{pmatrix} \tag{6.1}$$

in which the magnitude of each off-diagonal block $\mathbf{P}_{ij}$ is very small relative to 1.

We begin by investigating the limiting behavior of the chain as the off-diagonal blocks in (6.1) tend toward zero. To this end, allow the off-diagonal blocks to vary independently, and assume the diagonal block $\mathbf{P}_{ii}$ in each row depends continuously on the $\mathbf{P}_{ij}$'s (the off-diagonal blocks in the corresponding row) while always maintaining the irreducible and stochastic structure of the larger matrix $\mathbf{P}$. The following result gives an indication of why stochastic complementation is useful in understanding the nature of a nearly completely reducible chain. For vectors $\mathbf{x}$ and matrices $\mathbf{A}$, the norms defined by

$$\|\mathbf{x}\|_\infty = \max_i |x_i| \quad \text{and} \quad \|\mathbf{A}\|_\infty = \max_i \sum_j |a_{ij}|$$

will be employed throughout.

THEOREM 6.1. *If $\mathbf{P}$ is an irreducible stochastic matrix with a k-level partition as indicated in* (6.1)*, and if*

$$\mathbf{S}_{ii} = \mathbf{P}_{ii} + \mathbf{P}_{i*}(\mathbf{I} - \mathbf{P}_i)^{-1}\mathbf{P}_{*i}, \quad i = 1, 2, \cdots, k$$

are the associated stochastic complements, then

$$\|\mathbf{S}_{ii} - \mathbf{P}_{ii}\|_\infty = \|\mathbf{P}_{i*}\|_\infty, \tag{6.2}$$

and

$$\lim_{\mathbf{P}_{i*} \to \mathbf{0}} \mathbf{S}_{ii} = \mathbf{P}_{ii}. \tag{6.3}$$

Moreover, if $\mathbf{S}$ is the completely reducible stochastic matrix

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{S}_{kk} \end{pmatrix},$$

then

$$\|\mathbf{P} - \mathbf{S}\|_\infty = 2 \max_i \|\mathbf{P}_{i*}\|_\infty. \tag{6.4}$$

Proof. Let $\mathbf{e}$ denote a column of 1's whose size is determined by the context in which it appears, and begin by observing that all matrices which are involved are nonnegative so that

$$\|\mathbf{P}_{i*}(\mathbf{I} - \mathbf{P}_i)^{-1}\mathbf{P}_{*i}\|_\infty = \|\mathbf{P}_{i*}(\mathbf{I} - \mathbf{P}_i)^{-1}\mathbf{P}_{*i}\mathbf{e}\|_\infty \quad \text{and} \quad \|\mathbf{P}_{i*}\|_\infty = \|\mathbf{P}_{i*}\mathbf{e}\|_\infty.$$

Since $\mathbf{P}$ is stochastic, it must be the case that

$$\mathbf{P}_i\mathbf{e} + \mathbf{P}_{*i}\mathbf{e} = \mathbf{e},$$

which in turn implies

$$(\mathbf{I} - \mathbf{P}_i)^{-1}\mathbf{P}_{*i}\mathbf{e} = \mathbf{e}.$$

Therefore,

$$\|\mathbf{S}_{ii} - \mathbf{P}_{ii}\|_\infty = \|\mathbf{P}_{i*}(\mathbf{I} - \mathbf{P}_i)^{-1}\mathbf{P}_{*i}\|_\infty = \|\mathbf{P}_{i*}(\mathbf{I} - \mathbf{P}_i)^{-1}\mathbf{P}_{*i}\mathbf{e}\|_\infty = \|\mathbf{P}_{i*}\mathbf{e}\|_\infty = \|\mathbf{P}_{i*}\|_\infty,$$

which is the desired conclusion of (6.2), and the statement (6.3) that $\mathbf{S}_{ii} \to \mathbf{P}_{ii}$ as $\mathbf{P}_{i*} \to \mathbf{0}$ follows. To establish (6.4), use the fact that $\mathbf{P}_{ii} - \mathbf{S}_{ii} \leq \mathbf{0}$ in order to write

$$\begin{aligned} \|\mathbf{P} - \mathbf{S}\|_\infty &= \max_i \| \begin{pmatrix} \mathbf{P}_{i1} & \mathbf{P}_{i2} & \cdots & \mathbf{P}_{ii} - \mathbf{S}_{ii} & \cdots & \mathbf{P}_{ik} \end{pmatrix} \|_\infty \\ &= \max_i \|\mathbf{P}_{i1}\mathbf{e} + \mathbf{P}_{i2}\mathbf{e} + \cdots + (\mathbf{S}_{ii} - \mathbf{P}_{ii})\mathbf{e} + \cdots + \mathbf{P}_{ik}\mathbf{e}\|_\infty. \end{aligned} \tag{6.5}$$

Now observe that $\mathbf{P}_{ii}\mathbf{e} + \mathbf{P}_{i*}\mathbf{e} = \mathbf{e}$ implies

$$(\mathbf{S}_{ii} - \mathbf{P}_{ii})\mathbf{e} = \mathbf{e} - \mathbf{P}_{ii}\mathbf{e} = \mathbf{P}_{i*}\mathbf{e},$$

and use this in (6.5) together with the fact that

$$\sum_{\substack{j=1 \\ j\neq i}}^{k} \mathbf{P}_{ij}\mathbf{e} = \mathbf{P}_{i*}\mathbf{e}$$

to produce

$$\begin{aligned} \|\mathbf{P} - \mathbf{S}\|_\infty &= \max_i \|\mathbf{P}_{i1}\mathbf{e} + \cdots + \mathbf{P}_{i*}\mathbf{e} + \cdots + \mathbf{P}_{ik}\mathbf{e}\|_\infty \\ &= \max_i \|2\mathbf{P}_{i*}\|_\infty = 2\max_i \|\mathbf{P}_{i*}\|_\infty. \end{aligned}$$

□

If δ denotes the expression

$$\delta = 2\max_i \|\mathbf{P}_{i*}\|_\infty,$$

then it is clear that $0 \leq \delta \leq 2$, and $\mathbf{P}$ is completely reducible if and only if $\delta = 0$. This together with the result (6.4) motivates the following terminology.

DEFINITION 6.1. *For an $m \times m$ irreducible stochastic matrix with a k-level partition*

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

the number

$$\delta = 2\max_i \|\mathbf{P}_{i*}\|_\infty$$

is called the ***deviation from complete reducibility.***

It is worthwhile to explore the spectral properties of stochastic complementation as the deviation from complete reducibility tends toward 0. If $\mathbf{P}$ is nearly completely reducible, then each $\mathbf{P}_{ii}$ is nearly stochastic (in particular, Theorem 6.1 guarantees $\mathbf{P}_{ii} \approx \mathbf{S}_{ii}$), and continuity of the eigenvalues insures that $\mathbf{P}$ must necessarily have at least $k-1$ nonunit eigenvalues clustered near the simple eigenvalue $\lambda = 1$. This means that $\mathbf{P}$ is necessarily badly conditioned in the sense that if the sequence $\{\mathbf{P}^n\}_{n=1}^{\infty}$ converges, then it must converge very slowly. For the time being, suppose $\mathbf{P}$ has *exactly* $k-1$ nonunit eigenvalues clustered near $\lambda = 1$. Since each stochastic complement $\mathbf{S}_{ii}$ is an irreducible stochastic matrix, the Perron–Frobenius theorem guarantees that the unit eigenvalue of each $\mathbf{S}_{ii}$ is simple. By virtue of Theorem 6.1 and continuity of the eigenvalues, it follows that the nonunit eigenvalues of each $\mathbf{S}_{ii}$ must necessarily be rather far removed from the unit eigenvalue of $\mathbf{S}_{ii}$—otherwise the spectrum $\sigma(\mathbf{P})$ of $\mathbf{P}$ would contain a cluster of at least k nonunit eigenvalues positioned near $\lambda = 1$. In other words, $\mathbf{P} \to \mathbf{S}$ as $\delta \to 0$, and the cluster consisting of the $k-1$ nonunit eigenvalues together with $\lambda = 1$ itself in $\sigma(\mathbf{P})$ must "split" and map to the k unit eigenvalues in $\sigma(\mathbf{S})$—one unit eigenvalue in each $\sigma(\mathbf{S}_{ii})$. This splitting effect is pictorially illustrated below in Fig. 1 for a 12×12 matrix with the three-level partition

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \mathbf{P}_{13} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \mathbf{P}_{23} \\ \mathbf{P}_{31} & \mathbf{P}_{32} & \mathbf{P}_{33} \end{pmatrix} \quad \text{and} \quad \mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{33} \end{pmatrix}.$$

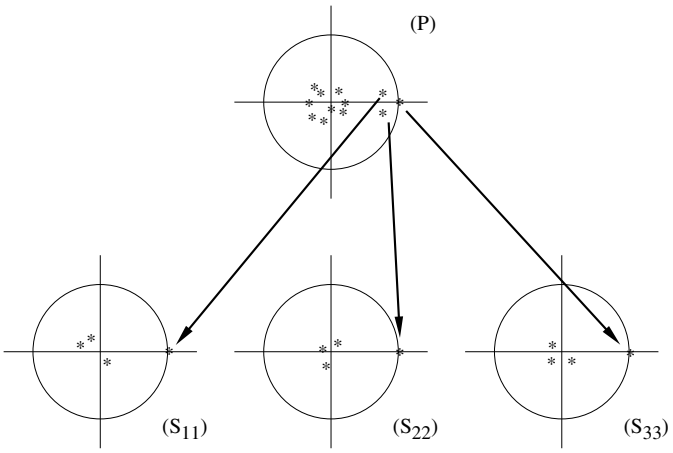

FIG. 1

The conclusion to be derived from this discussion is summarized in the following statement.

THEOREM 6.2. *If the underlying transition matrix*

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

of a nearly completely reducible Markov chain has exactly $k-1$ nonunit eigenvalues clustered near $\lambda = 1$, then the process of uncoupling the chain into k smaller chains by

using the method of stochastic complementation forces the spectrum of $\mathbf{P}$ *to naturally split apart so that each of the smaller chains has a well-conditioned transition matrix* $\mathbf{S}_{ii}$ *in the sense that each* $\mathbf{S}_{ii}$ *has no eigenvalues near its unit eigenvalue.*

When the states in a nearly completely reducible chain are naturally ordered, and when the transition matrix $\mathbf{P}$ is partitioned in the natural way according to the closely coupled subclasses, then the desirable spectrum splitting effect described above almost always occurs. However, there are pathological cases when this effect is not achieved. If, for a k-level partition of $\mathbf{P}$, the spectrum $\sigma(\mathbf{P})$ contains *more* than $k-1$ nonunit eigenvalues clustered near $\lambda = 1$, then continuity dictates that some $\mathbf{S}_{ii}$ must necessarily have a nonunit eigenvalue near its unit eigenvalue. This can occur even for naturally partitioned matrices. For example, consider

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{pmatrix}$$

in which both $\mathbf{P}_{12}$ and $\mathbf{P}_{21}$ are extremely small in magnitude, and assume neither $\mathbf{P}_{11}$ nor $\mathbf{P}_{22}$ are nearly completely reducible. Furthermore, suppose that $\mathbf{P}_{11}$ is a very small perturbation of

$$\mathbf{C}_{n\times n} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

where n is quite large. Since $\sigma(\mathbf{C})$ consists of the nth roots of unity

$$\omega_h = e^{(2\pi i/n)h}, \quad h = 0, 1, \cdots, n-1,$$

it is clear that $\omega_1 = e^{2\pi i/n}$ must be very close to $\omega_0 = 1$ whenever n is large. Therefore, $\mathbf{P}_{11}$ must have a nonunit eigenvalue near $\lambda = 1$ and, by virtue of Theorem 6.1, the same must hold for $\mathbf{S}_{11}$—thereby making $\mathbf{S}_{11}$ badly conditioned. Needless to say, pathological cases of this nature are rare in practical work.

As illustrated with the matrix given in (5.1), it is possible for stochastic complementation to uncouple a periodic chain into two aperiodic chains, thereby allowing the straightforward power method to be used where it ordinarily would not be applicable. A similar situation can hold for nearly completely reducible chains. The standard power method

$$\boldsymbol{\pi}_{n+1} = \boldsymbol{\pi}_n \mathbf{P} \quad \text{where } \boldsymbol{\pi}_0 \text{ is arbitrary} \tag{6.6}$$

applied to the transition matrix $\mathbf{P}$ of an aperiodic chain which is nearly completely reducible will fail—not in theory, but in practice—because the existence of eigenvalues of $\mathbf{P}$ which are near $\lambda = 1$ causes (6.6) to converge too slowly. However, if the chain is not pathological in nature (i.e., if $\mathbf{P}$ does not have more than $k-1$ eigenvalues near $\lambda = 1$), then the spectrum splitting effect described earlier insures that each of the k sequences

$$\mathbf{s}_i^{(n+1)} = \mathbf{s}_i^{(n)} \mathbf{S}_{ii} \quad \text{where } \mathbf{s}_i^{(0)} \text{ is arbitrary}, \quad i = 1, 2 \cdots k \tag{6.7}$$

will exhibit rapid convergence. Recall from Theorems 5.1 and 5.2 that the ith sequence in (6.7) converges if and only if $\mathbf{P}_{ii}$ is primitive, and this is guaranteed if $\mathbf{P}_{ii}$ has at

least one nonzero diagonal entry. Again take note of the fact that these k sequences in (6.7) are completely independent of each other, so each iteration can be executed in parallel—after which the limiting vectors

$$\mathbf{s}_i = \lim_{n\to\infty} \mathbf{s}_i^{(n)}$$

are easily coupled according to the rules of Theorem 4.1 in order to produce the stationary distribution vector for $\mathbf{P}$ as

$$\boldsymbol{\pi} = (\, \xi_1\mathbf{s}_1 \quad \xi_2\mathbf{s}_2 \quad \cdots \quad \xi_k\mathbf{s}_k \,).$$

7. The Simon–Ando theory for nearly completely reducible systems. Simon and Ando [18] provided the classical theory for nearly completely reducible systems, and Courtois [3] (along with others who followed his pioneering work) applied the theory and helped develop numerical aspects associated with queueing networks. The contribution of Simon and Ando [18] was to provide mathematical arguments for what had previously been a rather heuristic theory concerning nearly completely reducible systems. The major conclusion of this theory is that if the off-diagonal blocks $\mathbf{P}_{ij}$ in the transition matrix

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

of a finite aperiodic chain have sufficiently small magnitudes, then closely coupled subclasses associated with the diagonal blocks $\mathbf{P}_{ii}$ must each tend toward a local equilibrium long before a global equilibrium for the entire system is attained. In the short run, the chain behaves as though it were completely reducible with each subclass tending toward a local equilibrium independent of the evolution in other subclasses. After this short-run stabilization, the local equilibria are approximately maintained—in a relative sense—within each subclass while the entire chain evolves toward its global equilibrium. The approach and substance of the theorems of Simon and Ando have become accepted as the theoretical basis for aggregation techniques and algorithms.

Although the conclusions of Simon and Ando [18] are extremely important, their mathematical development utilizes some rather cumbersome notation and proofs. At times, the arguments are difficult to appreciate, and they do not fully illuminate the basic underlying mechanisms. This is corroborated by the fact that although the conclusions of the Simon–Ando theory are fundamental to the development of material in his text, Courtois [3] chooses to omit the Simon–Ando proofs on the grounds that they have "little relevance" to subsequent developments.

The purpose of the latter portion of this exposition is to apply the concept of stochastic complementation in order to develop the theory for nearly completely reducible systems in a unified, clear, and simple manner while simultaneously sharpening some results and generalizing others. The discussion is divided into three parts—the short-run dynamics, the middle-run dynamics, and the long-run dynamics.

8. Short-run dynamics. The object of the short-run analysis is to examine the behavior of the distribution

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_0\mathbf{P}^n \quad \text{where } \boldsymbol{\pi}_0 \text{ is arbitrary,}$$

for smaller values of n. The following theorem provides the basis for doing this.

THEOREM 8.1. *For an $m \times m$ irreducible stochastic matrix with a k-level partition*

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

let $\mathbf{S}$ be the completely reducible stochastic matrix

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{S}_{kk} \end{pmatrix},$$

and assume that each stochastic complement $\mathbf{S}_{ii}$ is primitive.[3] *Let the eigenvalues of $\mathbf{S}_{m\times m}$ be ordered as*

$$\lambda_1 = \lambda_2 = \cdots = \lambda_k = 1 > |\lambda_{k+1}| \geq |\lambda_{k+2}| \geq \cdots \geq |\lambda_m|,$$

and assume that $\mathbf{S}$ is similar to a diagonal matrix [4]

$$\mathbf{Z}^{-1}\mathbf{S}\mathbf{Z} = \begin{pmatrix} \mathbf{I}_{k\times k} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} \quad \text{where } \mathbf{D} = \begin{pmatrix} \lambda_{k+1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m \end{pmatrix}.$$

If $\mathbf{S}^\infty$ denotes the limit

$$\mathbf{S}^\infty = \lim_{n\to\infty} \mathbf{S}^n = \begin{pmatrix} \mathbf{e}\mathbf{s}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{e}\mathbf{s}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{e}\mathbf{s}_k \end{pmatrix}$$

where $\mathbf{s}_i$ is the stationary distribution vector for $\mathbf{S}_{ii}$, then

$$\|\mathbf{P}^n - \mathbf{S}^\infty\|_\infty \leq n\delta + \kappa|\lambda_{k+1}|^n \tag{8.1}$$

where δ is the deviation from complete reducibility as defined in Definition 6.1*, and where κ is the constant*

$$\kappa = \|\mathbf{Z}\|_\infty \|\mathbf{Z}^{-1}\|_\infty.$$

Moreover, if the (row) vector norm defined by $\|\mathbf{r}\|_1 = \sum_i |r_i|$ is used, then for every $n = 1, 2, \cdots$, the difference between $\boldsymbol{\pi}_n$ and the limit

$$\mathbf{s} = \lim_{n\to\infty} \boldsymbol{\pi}_0 \mathbf{S}^n$$

can be measured by

$$\|\boldsymbol{\pi}_n - \mathbf{s}\|_1 \leq n\delta + \kappa|\lambda_{k+1}|^n. \tag{8.2}$$

[3] The primitivity assumption is included here for clarity of exposition because it insures that limits exist in the strong sense. Although Theorems 5.1 and 5.2 show that almost all stochastic complements are primitive, it will later be argued that primitivity is not needed to reach the same general conclusions of this theorem.

[4] This assumption is also for clarity of exposition—it will later be indicated how the results of this theorem can be preserved without this diagonalizability assumption.

Proof. Begin by observing that the identity

$$\mathbf{P}^n - \mathbf{S}^n = \mathbf{S}^{n-1}(\mathbf{P}-\mathbf{S}) + \mathbf{S}^{n-2}(\mathbf{P}-\mathbf{S})\mathbf{P} + \cdots + \mathbf{S}(\mathbf{P}-\mathbf{S})\mathbf{P}^{n-2} + (\mathbf{P}-\mathbf{S})\mathbf{P}^{n-1}$$

is valid for all $n = 1, 2, \cdots$. Use this together with (6.4) and the fact that

$$\|\mathbf{P}^j\|_\infty = \|\mathbf{S}^j\|_\infty = 1 \quad \text{for every } j$$

in order to conclude that

$$\|\mathbf{P}^n - \mathbf{S}^n\|_\infty \leq n\delta. \tag{8.3}$$

By writing

$$\mathbf{S}^n = \mathbf{Z}\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^n \end{pmatrix}\mathbf{Z}^{-1} \quad \text{and} \quad \lim_{n\to\infty}\mathbf{S}^n = \mathbf{Z}\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{Z}^{-1},$$

it is clear that

$$\mathbf{S}^n - \mathbf{S}^\infty = \mathbf{Z}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^n \end{pmatrix}\mathbf{Z}^{-1}.$$

Taking norms produces

$$\|\mathbf{S}^n - \mathbf{S}^\infty\|_\infty \leq \|\mathbf{Z}\|_\infty |\lambda_{k+1}|^n \|\mathbf{Z}^{-1}\|_\infty = \kappa|\lambda_{k+1}|^n. \tag{8.4}$$

Now use (8.4) together with (8.3) to write

$$\|\mathbf{P}^n - \mathbf{S}^\infty\|_\infty = \|\mathbf{P}^n - \mathbf{S}^n + \mathbf{S}^n - \mathbf{S}^\infty\|_\infty \leq \|\mathbf{P}^n - \mathbf{S}^n\|_\infty + \|\mathbf{S}^n - \mathbf{S}^\infty\|_\infty \leq n\delta + \kappa|\lambda_{k+1}|^n,$$

which is the desired conclusion (8.1). The second conclusion (8.2) follows by noting that the inequality $\|\mathbf{rA}\|_1 \leq \|\mathbf{r}\|_1 \|\mathbf{A}\|_\infty$ holds for row vectors $\mathbf{r}$ and square matrices $\mathbf{A}$ so that

$$\|\boldsymbol{\pi}_n - \mathbf{s}\|_1 = \|\boldsymbol{\pi}_0(\mathbf{P}^n - \mathbf{S}^\infty)\|_1 \leq \|\mathbf{P}^n - \mathbf{S}^\infty\|_\infty \leq n\delta + \kappa|\lambda_{k+1}|^n. \qquad \square$$

The results of Theorem 8.1 motivate the following definition.

DEFINITION 8.1. *For each $\epsilon > 0$, there is an associated* ***short-run stabilization interval*** *$\mathcal{I}(\epsilon)$ which is defined to be the set*

$$\mathcal{I}(\epsilon) = \left\{ n \;\middle|\; \|\mathbf{P}^n - \mathbf{S}^\infty\|_\infty < \epsilon \right\}.$$

The set

$$E(\epsilon) = \left\{ n \;\middle|\; \delta n + \kappa|\lambda_{k+1}|^n < \epsilon \right\}$$

is referred to as the ***estimated short-run stabilization interval,*** *and the function defined by*

$$f(x) = \delta x + \kappa|\lambda_{k+1}|^x$$

is called the ***estimating function.*** *Notice that Theorem* 8.1 *insures $E(\epsilon) \subseteq \mathcal{I}(\epsilon)$.*

Examining the characteristics of the graph of the estimating function $f(x)$ provides an indication of the nature of the short-run stabilization interval $\mathcal{I}(\epsilon)$. Refer to Fig. 2 and notice that as long as ϵ is greater than the minimum value of f, which occurs at

$$x_{\min} = \frac{\ln\left(\frac{-\delta}{\kappa \ln|\lambda_{k+1}|}\right)}{\ln|\lambda_{k+1}|},$$

the estimated short-run stabilization interval $E(\epsilon)$ is nonempty. Because $f(x)$ is asymptotic to the line $y = \delta x$, the length of the interval $E(\epsilon)$ increases as $\delta \to 0$, i.e., $E(\epsilon)$ grows as the $\mathbf{P}_{ij}$'s$\to \mathbf{0}$ in the underlying partitioned matrix $\mathbf{P}$.

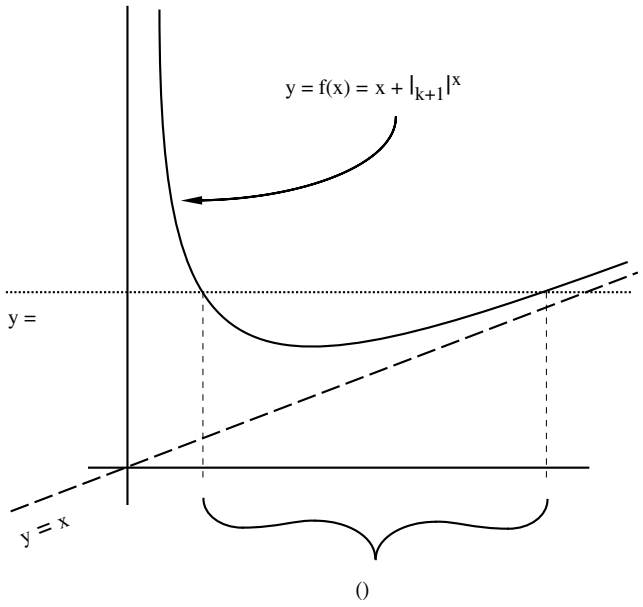

FIG. 2

To appreciate the relationship between the short-run stabilization interval $\mathcal{I}(\epsilon)$ and its estimate $E(\epsilon)$, consider the following example in which $\mathbf{P}$ is the nearly completely reducible matrix

$$\mathbf{P} = \left(\begin{array}{cc|cc} .6999 & .3000 & .0001 & 0 \\ .2000 & .7996 & 0 & .0004 \\ \hline .0003 & 0 & .7997 & .2000 \\ 0 & .0007 & .1000 & .8993 \end{array}\right). \tag{8.5}$$

Let $g(n)$ be the function defined by

$$g(n) = \left\|\mathbf{P}^n - \mathbf{S}^{\infty}\right\|_{\infty},$$

and let $f(n)$ be the estimating function

$$f(n) = \delta n + \kappa|\lambda_{k+1}|^n.$$

For the matrix (8.5), $\delta = .0014$, $\kappa = 8.657$, and $\lambda_{k+1} = .6996$ (rounded to four significant digits). The graph of $g(n)$, shown in Fig. 3, illustrates how rapidly powers of $\mathbf{P}$ initially approach the matrix $\mathbf{S}^\infty$ and then how $\mathbf{P}^n$ gradually pulls away from $\mathbf{S}^\infty$.

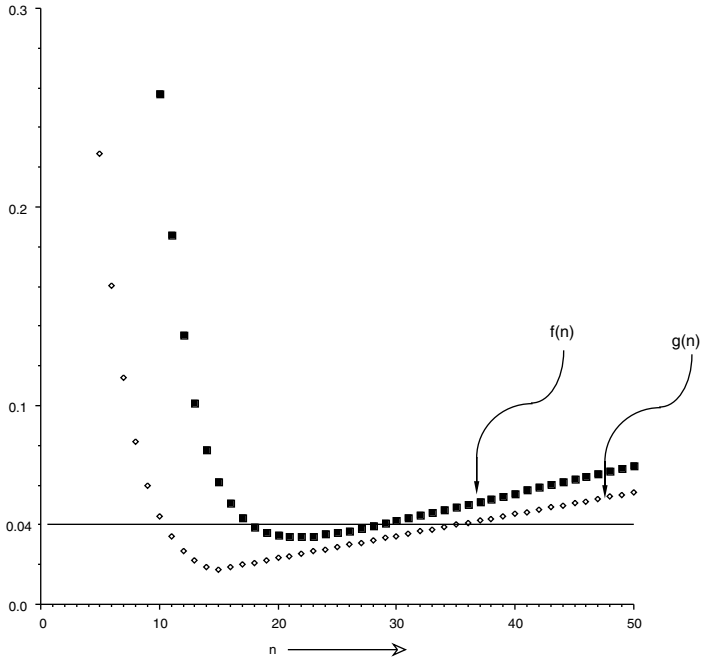

FIG. 3

For this example (which is not pathological), the graphs in Fig. 3 show that the estimating function $f(n)$ gives a good indication of the evolution of the process in the sense that $f(n)$ is essentially parallel with $g(n)$. That is, $f(n)$ decreases, stabilizes, and then increases more or less in concert with $g(n)$. Although necessarily conservative, the estimating function $f(n)$ is not overly pessimistic in estimating the length of the short-run stabilization period. For example, if ϵ is taken to be $\epsilon = .04$, then—as depicted in Fig. 3—the associated short-run stabilization interval is

$$\mathcal{I}(.04) = [11, 35]$$

while the estimated short-run stabilization interval is

$$E(.04) = [18, 28].$$

COROLLARY 8.1. *If, for a given $\epsilon > 0$, n lies in the interval defined by*

$$\frac{\ln\ \epsilon/2\kappa}{\ln|\lambda_{k+1}|} < n < \frac{\epsilon}{2\delta}, \tag{8.6}$$

then

$$\|\mathbf{P}^n - \mathbf{S}^\infty\|_\infty < \epsilon,$$

and for every $\boldsymbol{\pi}_0$,

$$\|\boldsymbol{\pi}_n - \mathbf{s}\|_1 < \epsilon.$$

More precisely,

$$\left\{ n \;\middle|\; \frac{\ln\ \epsilon/2\kappa}{\ln|\lambda_{k+1}|} < n < \frac{\epsilon}{2\delta} \right\} \subseteq E(\epsilon) \subseteq \mathcal{I}(\epsilon).$$

Proof. The fact that $|\lambda_{k+1}| < 1$ means $\ln|\lambda_{k+1}| < 0$ so that the left-hand inequality in (8.6) implies

$$\ln|\lambda_{k+1}|^n < \ln\ \epsilon/2\kappa,$$

and hence $\kappa|\lambda_{k+1}|^n < \epsilon/2$. Similarly, the right-hand inequality in (8.6) says that $n\delta < \epsilon/2$, and the desired conclusions follow from Theorem 8.1. ∎

To understand the significance of a short-run stabilization interval, suppose there are m_i states in the ith subclass $\mathcal{Q}_i$ (i.e., $\mathbf{P}_{ii}$ is $m_i \times m_i$), and observe that if $\mathbf{s}_i$ is the stationary distribution vector for $\mathbf{S}_{ii}$, then for an initial distribution $\boldsymbol{\pi}_0$ partitioned as

$$\boldsymbol{\pi}_0 = \begin{pmatrix} \boldsymbol{\pi}_0^{(1)} & \boldsymbol{\pi}_0^{(2)} & \cdots & \boldsymbol{\pi}_0^{(k)} \end{pmatrix}$$

in which $\boldsymbol{\pi}_0^{(i)}$ contains m_i components, the limiting distribution $\mathbf{s}$ is given by

$$\mathbf{s} = \boldsymbol{\pi}_0\mathbf{S}^\infty = \boldsymbol{\pi}_0 \begin{pmatrix} \mathbf{e}\mathbf{s}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{e}\mathbf{s}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{e}\mathbf{s}_k \end{pmatrix} = \begin{pmatrix} \nu_1\mathbf{s}_1 & \nu_2\mathbf{s}_2 & \cdots & \nu_k\mathbf{s}_k \end{pmatrix}$$

where the coefficient ν_i associated with the ith subclass is given by

$$\nu_i = \boldsymbol{\pi}_0^{(i)}\mathbf{e} = \sum_{h=1}^{m_i} \left(\boldsymbol{\pi}_0^{(i)}\right)_h.$$

Suppose that n belongs to a short-run stabilization interval $\mathcal{I}(\epsilon)$. Roughly speaking, this means that n simultaneously satisfies the two conditions

$$|\lambda_{k+1}|^n << 1 \quad \text{and} \quad n << \frac{1}{\delta}, \tag{8.7}$$

and for such values of n, Theorem 8.1 guarantees that

$$\boldsymbol{\pi}_n \approx \mathbf{s} = \begin{pmatrix} \nu_1\mathbf{s}_1 & \nu_2\mathbf{s}_2 & \cdots & \nu_k\mathbf{s}_k \end{pmatrix}$$

where the ν_i's are constant—they depend only on $\boldsymbol{\pi}_0$. Therefore, each closely coupled subclass $\mathcal{Q}_i$ will approximately be in a local equilibrium which is defined by the stationary distribution vector $\mathbf{s}_i$ as long as n satisfies (8.7).

In other words, the chain initially evolves in such a way that each closely coupled subclass begins to approach a short-run equilibrium—defined by the $\mathbf{s}_i$'s—completely independent of the evolution in all other subclasses. As the chain continues to evolve, there will exist a period (the short-run stabilization interval $\mathcal{I}$) in which each subclass—and hence the entire chain—is approximately stable. As time proceeds beyond $\mathcal{I}$, the chain will move away from $\mathbf{s}$, the approximate point of short-run stability, and it will evolve along a path en route toward global stability.

The existence of an approximate short-run local equilibrium is a major conclusion of the classical Simon–Ando theory, but the description of the short-run stabilization interval and the estimate given in Definition 8.1 or (8.7) is new. The fact that the approximate short-run local equilibrium can be described by the stationary distribution vectors of the individual stochastic complements also appears to be new. The preceding remarks are formally summarized in the following theorem.

THEOREM 8.2 (Short-run dynamics). *For an irreducible stochastic matrix with a k-level partition*

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

in which the stochastic complements $(\mathbf{S}_{ii})_{m_i \times m_i}$ are each primitive and diagonalizable, let

$$\boldsymbol{\pi}_0 = \begin{pmatrix} \boldsymbol{\pi}_0^{(1)} & \boldsymbol{\pi}_0^{(2)} & \cdots & \boldsymbol{\pi}_0^{(k)} \end{pmatrix}$$

be an arbitrary initial distribution, and let

$$|\lambda_{k+1}| = \max_{\lambda \neq 1} \left\{ |\lambda| \;\middle|\; \lambda \in \bigcup_{i=1}^{k} \sigma(\mathbf{S}_{ii}) \right\}.$$

If the magnitudes of the off-diagonal blocks $\mathbf{P}_{ij}$ are small enough to guarantee the existence of values of n such that

$$|\lambda_{k+1}|^n << 1 \quad \text{and} \quad n << \frac{1}{\delta}, \tag{8.8}$$

then as long as n satisfies (8.8)*, it must be the case that*

$$\mathbf{P}^n \approx \begin{pmatrix} \mathbf{S}_{11}^{\infty} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22}^{\infty} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{S}_{kk}^{\infty} \end{pmatrix} = \begin{pmatrix} \mathbf{e}\mathbf{s}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{e}\mathbf{s}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{e}\mathbf{s}_k \end{pmatrix}$$

and

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_0 \mathbf{P}^n \approx \begin{pmatrix} \nu_1 \mathbf{s}_1 & \nu_2 \mathbf{s}_2 & \cdots & \nu_k \mathbf{s}_k \end{pmatrix}$$

where $\mathbf{s}_i$ is the stationary distribution vector for $\mathbf{S}_{ii}$, and where

$$\nu_i = \boldsymbol{\pi}_0^{(i)} \mathbf{e} = \sum_{h=1}^{m_i} \left(\boldsymbol{\pi}_0^{(i)} \right)_h .$$

Before turning to the middle-run and long-run dynamics, observe that the assumption of Theorems 8.1 and 8.2 that $\mathbf{S}$ is diagonalizable is not necessary—it merely makes the conclusions easier to grasp. In the more general case, there exists a nonsingular matrix $\mathbf{Z}$ such that

$$\mathbf{Z}^{-1}\mathbf{S}\mathbf{Z} = \begin{pmatrix} \mathbf{I}_{k \times k} & \mathbf{0} \\ \mathbf{0} & \mathbf{J} \end{pmatrix}$$

where $\mathbf{J}$ is in Jordan canonical form, and the spectral radius of $\mathbf{J}$ is $\rho(\mathbf{J}) = |\lambda_{k+1}|$. It is well known (see [19] or [11]) that for every square matrix $\mathbf{A}$ and each $\epsilon > 0$, there exists a matrix norm $\|\!|\!| \bullet |\!|\!|$ such that

$$\rho(\mathbf{A}) \leq |\!|\!|\mathbf{A}|\!|\!| \leq \rho(\mathbf{A}) + \epsilon.$$

Clearly, this norm can be made to be meaningful for row vectors $\mathbf{r}$ simply by concatenating enough zero rows to $\mathbf{r}$ so as to fill out a square matrix. This means that for every n and for each $\epsilon > 0$, there is a norm $\|\!|\!| \bullet |\!|\!|$ such that

$$|\lambda_{k+1}|^n \leq |\!|\!|\mathbf{J}^n|\!|\!| \leq |\lambda_{k+1}|^n + \epsilon,$$

and (8.2) can be replaced by

$$|\!|\!|\boldsymbol{\pi}_n - \mathbf{s}|\!|\!| \leq n\delta + \kappa\left(|\lambda_{k+1}|^n + \epsilon\right) \tag{8.9}$$

where $\kappa = |\!|\!|\boldsymbol{\pi}_0\mathbf{Z}|\!|\!|\ |\!|\!|\mathbf{Z}^{-1}|\!|\!|$. Although the analysis is more involved, (8.9) used in place of (8.2) leads to the same general conclusions given in Theorems 8.1 and 8.2.

9. Middle-run dynamics. According to the short-run dynamics,

$$\mathbf{P}^n \approx \mathbf{S}^\infty \quad \text{and} \quad \boldsymbol{\pi}_n \approx \mathbf{s} = \begin{pmatrix} \nu_1\mathbf{s}_1 & \nu_2\mathbf{s}_2 & \cdots & \nu_k\mathbf{s}_k \end{pmatrix}$$

as long as n belongs to a short-run stabilization interval characterized by (8.8). As the chain continues to evolve, and as n advances beyond the short-run stabilization interval, $\mathbf{P}^n$ moves away from $\mathbf{S}^\infty$ and tends toward $\mathbf{P}^\infty$ (assuming $\mathbf{P}$ is primitive), while the distribution $\boldsymbol{\pi}_n$ diverges away from $\mathbf{s}$ and moves along a path en route to the global stationary distribution vector $\boldsymbol{\pi}$. However, after the short-run stabilization period has ended, components of $\boldsymbol{\pi}_n$ belonging to the same subclass $\mathcal{Q}_i$ will nevertheless continue to remain in approximate equilibrium *relative to each other*. The following theorems precisely articulate the sense in which this occurs.

THEOREM 9.1. *For an $m \times m$ irreducible stochastic matrix*

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

with stochastic complements $\mathbf{S}_{ii}$ which are each primitive, let $\mathbf{s}_i$ be the stationary distribution vector for $\mathbf{S}_{ii}$, and let $\boldsymbol{\pi}_n = \boldsymbol{\pi}_0\mathbf{P}^n$ where $\boldsymbol{\pi}_0$ is an arbitrary initial distribution. If $\epsilon > 0$ is a number such that the associated short-run stabilization interval $\mathcal{I}(\epsilon)$ is nonempty, then for each integer n beyond $\mathcal{I}(\epsilon)$, there exist scalars β_i (which vary with n) such that the vector

$$\mathbf{v}_n = \begin{pmatrix} \beta_1\mathbf{s}_1 & \beta_2\mathbf{s}_2 & \cdots & \beta_k\mathbf{s}_k \end{pmatrix}$$

satisfies the inequality

$$\|\boldsymbol{\pi}_n - \mathbf{v}_n\|_1 < \epsilon.$$

Proof. If n is an integer beyond the short-run stabilization interval $\mathcal{I}(\epsilon)$, then there exist positive integers q and r such that $n = q + r$, where

$$q \in \mathcal{I}(\epsilon).$$

Write $\boldsymbol{\pi}_n$ as

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_{q+r} = \boldsymbol{\pi}_r \mathbf{P}^q$$

where $\boldsymbol{\pi}_r$ is partitioned as

$$\boldsymbol{\pi}_r = \begin{pmatrix} \boldsymbol{\pi}_r^{(1)} & \boldsymbol{\pi}_r^{(2)} & \cdots & \boldsymbol{\pi}_r^{(k)} \end{pmatrix}.$$

Define $\mathbf{v}_n$ to be the vector

$$\begin{aligned} \mathbf{v}_n = \boldsymbol{\pi}_r \mathbf{S}^\infty &= \boldsymbol{\pi}_r \begin{pmatrix} \mathbf{S}_{11}^\infty & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22}^\infty & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{S}_{kk}^\infty \end{pmatrix} = \boldsymbol{\pi}_r \begin{pmatrix} \mathbf{e}\mathbf{s}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{e}\mathbf{s}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{e}\mathbf{s}_k \end{pmatrix} \\ &= \begin{pmatrix} \beta_1 \mathbf{s}_1 & \beta_2 \mathbf{s}_2 & \cdots & \beta_k \mathbf{s}_k \end{pmatrix} \end{aligned}$$

where

$$\beta_i = \boldsymbol{\pi}_r^{(i)} \mathbf{e} = \sum_{h=1}^{m_i} \left(\boldsymbol{\pi}_r^{(i)} \right)_h. \tag{9.1}$$

Because $q \in \mathcal{I}(\epsilon)$, we may conclude that

$$\|\boldsymbol{\pi}_n - \mathbf{v}_n\|_1 = \|\boldsymbol{\pi}_r \mathbf{P}^q - \boldsymbol{\pi}_r \mathbf{S}^\infty\|_1 \leq \|\boldsymbol{\pi}_r\|_1 \|\mathbf{P}^q - \mathbf{S}^\infty\|_\infty < \epsilon. \qquad \square$$

The short-run dynamics established the fact that as long as $n \in \mathcal{I}(\epsilon)$,

$$\boldsymbol{\pi}_n \approx \begin{pmatrix} \nu_1 \mathbf{s}_1 & \nu_2 \mathbf{s}_2 & \cdots & \nu_k \mathbf{s}_k \end{pmatrix} \tag{9.2}$$

where the ν_i's are constant—they depend only on the initial distribution $\boldsymbol{\pi}_0$. The middle-run dynamics guarantees that for all n beyond $\mathcal{I}(\epsilon)$,

$$\boldsymbol{\pi}_n \approx \begin{pmatrix} \beta_1 \mathbf{s}_1 & \beta_2 \mathbf{s}_2 & \cdots & \beta_k \mathbf{s}_k \end{pmatrix} \tag{9.3}$$

where the β_i's vary with n. Therefore, if states $\mathcal{S}_i$ and $\mathcal{S}_j$ each belong to the same subclass—say $\mathcal{Q}_r$—and if $\mathcal{S}_i$ and $\mathcal{S}_j$ represent the r_ith and r_jth states within $\mathcal{Q}_r$, then it is clear from (9.2) and (9.3) that the ratio between the ith and jth components in $\boldsymbol{\pi}_n$ during the short run is

$$\frac{(\boldsymbol{\pi}_n)_i}{(\boldsymbol{\pi}_n)_j} \approx \frac{(\nu_r \mathbf{s}_r)_{r_i}}{(\nu_r \mathbf{s}_r)_{r_j}} = \frac{(\mathbf{s}_r)_{r_i}}{(\mathbf{s}_r)_{r_j}},$$

and this ratio is approximately maintained beyond the short run because during the middle run,

$$\frac{(\boldsymbol{\pi}_n)_i}{(\boldsymbol{\pi}_n)_j} \approx \frac{(\beta_r \mathbf{s}_r)_{r_i}}{(\beta_r \mathbf{s}_r)_{r_j}} = \frac{(\mathbf{s}_r)_{r_i}}{(\mathbf{s}_r)_{r_j}}.$$

This realization is the second fundamental feature of the Simon–Ando theory, which is formally stated below.

THEOREM 9.2 (Middle-run dynamics). *Consider an irreducible stochastic matrix*

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

for which each stochastic complement is primitive, and suppose that the off-diagonal blocks $\mathbf{P}_{ij}$ *are sufficiently small in magnitude so as to guarantee the existence of a nonempty short-run stabilization interval* $\mathcal{I}$*. After* n *enters* $\mathcal{I}$*, the ratio between any two components in* $\boldsymbol{\pi}_n$*, within the same closely coupled subclass* $\mathcal{Q}_r$*, remains approximately constant for all time thereafter. This ratio is approximately the same as the ratio between the corresponding components in the stationary distribution vector* $\mathbf{s}_r$ *of the associated stochastic complement* $\mathbf{S}_{rr}$*. In other words, after* n *enters* $\mathcal{I}$*, the relations*

$$\frac{(\boldsymbol{\pi}_n)_i}{(\boldsymbol{\pi}_n)_j} \approx \frac{(\mathbf{s}_r)_{r_i}}{(\mathbf{s}_r)_{r_j}} \tag{9.4}$$

are maintained throughout the entire evolution of the chain, and the ratios in (9.4) *are independent of the initial distribution* $\boldsymbol{\pi}_0$*.*

10. Long-run dynamics. The only observation which needs to be made concerning long-run behavior is that the short-run and middle-run relative stability expressed by (9.4) is maintained in the limit. For the time being, assume that $\mathbf{P}$ is primitive so that its stationary distribution vector is given as the limit

$$\boldsymbol{\pi} = \lim_{n\to\infty} \boldsymbol{\pi}_n = \lim_{n\to\infty} \boldsymbol{\pi}_0 \mathbf{P}^n,$$

and recall from Theorem 4.1 there exist constants ξ_i (the coupling factors) such that

$$\boldsymbol{\pi} = \begin{pmatrix} \xi_1\mathbf{s}_1 & \xi_2\mathbf{s}_2 & \cdots & \xi_k\mathbf{s}_k \end{pmatrix}.$$

Theorem 9.1 guarantees that during the middle run,

$$\boldsymbol{\pi}_n \approx \begin{pmatrix} \beta_1\mathbf{s}_1 & \beta_2\mathbf{s}_2 & \cdots & \beta_k\mathbf{s}_k \end{pmatrix}$$

where the β_i's vary with n, and hence it appears that each β_i should eventually settle down to the constant coupling factor ξ_i. To rigorously demonstrate this, recall from (9.1) that each β_i is of the form

$$\beta_i = \boldsymbol{\pi}_r^{(i)}\mathbf{e} = \sum_{h=1}^{m_i} \left(\boldsymbol{\pi}_r^{(i)}\right)_h$$

where $\boldsymbol{\pi}_r$ is a distribution partitioned as

$$\boldsymbol{\pi}_r = \begin{pmatrix} \boldsymbol{\pi}_r^{(1)} & \boldsymbol{\pi}_r^{(2)} & \cdots & \boldsymbol{\pi}_r^{(k)} \end{pmatrix},$$

and where $r = n - q$ for some fixed positive integer q. Use the fact that

$$\boldsymbol{\pi}_r \to \boldsymbol{\pi} = \begin{pmatrix} \xi_1\mathbf{s}_1 & \xi_2\mathbf{s}_2 & \cdots & \xi_k\mathbf{s}_k \end{pmatrix}$$

in conjunction with (4.1) to conclude

$$\lim_{n\to\infty} \beta_i = \lim_{r\to\infty} \boldsymbol{\pi}_r^{(i)}\mathbf{e} = \xi_i \mathbf{s}_i\mathbf{e} = \xi_i.$$

The entire evolution of a nearly completely reducible chain can now be formally summarized by the following theorem.

THEOREM 10.1 (Summary). *Consider a nearly completely reducible chain whose transition matrix*

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

as well as each stochastic complement $\mathbf{S}_{ii}$ *is primitive. For an arbitrary initial distribution* $\boldsymbol{\pi}_0$*, the distribution* $\boldsymbol{\pi}_n = \boldsymbol{\pi}_{n-1}\mathbf{P} = \boldsymbol{\pi}_0\mathbf{P}^n$ *evolves as follows:*

Short-run evolution—*The components in* $\boldsymbol{\pi}_n$ *begin to evolve in such a way that groups of components corresponding to closely coupled subclasses independently tend toward their own approximate local equilibrium. The approximate local equilibrium values within each closely coupled subclass are proportional to the values in the stationary distribution* $\mathbf{s}_i$ *of the associated stochastic complement* $\mathbf{S}_{ii}$.

Short-run stabilization—*As the chain continues to evolve, there is an interval of time, the short-run stabilization interval* $\mathcal{I}$*, during which the chain is approximately stable in the sense that*

$$\boldsymbol{\pi}_n \approx (\, \alpha_1\mathbf{s}_1 \quad \alpha_2\mathbf{s}_2 \quad \cdots \quad \alpha_k\mathbf{s}_k \,) \tag{10.1}$$

where the α_i*'s are constants which depend only on* $\boldsymbol{\pi}_0$.

Middle-run evolution—*As* n *passes beyond* $\mathcal{I}$*, the distribution* $\boldsymbol{\pi}_n$ *moves away from the short-run stabilization vector given in* (10.1)*, but* $\boldsymbol{\pi}_n$ *nevertheless maintains an approximate relative stability between components corresponding to the same closely coupled subclass in the sense that*

$$\boldsymbol{\pi}_n \approx (\, \beta_1\mathbf{s}_1 \quad \beta_2\mathbf{s}_2 \quad \cdots \quad \beta_k\mathbf{s}_k \,) \tag{10.2}$$

where the β_i*'s vary with* n.

Long-run evolution—*The* β_i*'s in* (10.2) *begin to settle down to limiting values which define the stationary distribution vector for the entire chain.*

Long-run stabilization—*As* $n \to \infty$, $\boldsymbol{\pi}_n \to \boldsymbol{\pi}$, *and* $\beta_i \to \xi_i$ *where the* ξ_i*'s are the constant coupling factors given in Theorem* 4.1*. So for practical purposes, there is some point in time past which*

$$\boldsymbol{\pi}_n \approx (\, \xi_1\mathbf{s}_1 \quad \xi_2\mathbf{s}_2 \quad \cdots \quad \xi_k\mathbf{s}_k \,).$$

11. Relaxing primitivity. An explicit hypothesis of Theorem 10.1 is that $\mathbf{P}$ and each stochastic complement $\mathbf{S}_{ii}$ are primitive. Although it is implicit in the Simon–Ando development [18] and in treatment by Courtois [3], the primitivity assumption is nevertheless present in these places as well. Of course, primitivity is not an overly restrictive hypothesis because it is frequently present in practical problems. However, the primitivity assumption is not necessary. The purpose of this section is

to give a brief indication of how and why the conclusions of Theorem 10.1 remain valid in a more general sense without the assumption of primitivity.

Suppose that

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

as well as its associated stochastic complements $\mathbf{S}_{ii}$ are not necessarily primitive, and assume only that $\mathbf{P}$ is irreducible. According to Theorem 2.3, all stochastic complements $\mathbf{S}_{ii}$ are therefore irreducible, and hence $\mathbf{P}$ as well as the $\mathbf{S}_{ii}$'s each possess a unique stationary distribution vector—denote them by $\boldsymbol{\pi}$ and $\mathbf{s}_i$, respectively. If $\mathbf{S}$ is the matrix

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{S}_{kk} \end{pmatrix},$$

and if $\mathbf{P}^{(n)}$ and $\mathbf{S}^{(n)}$ denote the averages

$$\mathbf{P}^{(n)} = \frac{\mathbf{I} + \mathbf{P} + \mathbf{P}^2 + \cdots + \mathbf{P}^{n-1}}{n} \quad \text{and} \quad \mathbf{S}^{(n)} = \frac{\mathbf{I} + \mathbf{S} + \mathbf{S}^2 + \cdots + \mathbf{S}^{n-1}}{n},$$

then it is well known (see [11]) that

$$\mathbf{P}^{(n)} \to \mathbf{P}^{\infty} = \mathbf{e}\boldsymbol{\pi} \quad \text{and} \quad \mathbf{S}^{(n)} \to \mathbf{S}^{\infty} = \begin{pmatrix} \mathbf{e}\mathbf{s}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{e}\mathbf{s}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{e}\mathbf{s}_k \end{pmatrix}.$$

If the off-diagonal blocks $\mathbf{P}_{ij}$ in $\mathbf{P}$ are quite small in magnitude, then Theorem 6.1 guarantees $\mathbf{P} \approx \mathbf{S}$ so that $\mathbf{P}^n \approx \mathbf{S}^n$, and hence $\mathbf{P}^{(n)} \approx \mathbf{S}^{(n)}$, as long as n is not too large. Moreover, if the magnitudes of the $\mathbf{P}_{ij}$'s are sufficiently small, then there are some values of n such that each $\mathbf{S}_{ii}^{(n)}$ is near its limiting value $\mathbf{S}_{ii}^{\infty}$ before $\mathbf{P}^{(n)}$ and $\mathbf{S}^{(n)}$ have moved too far apart. That is, if $\mathbf{P}$ is nearly completely reducible, then there is a short-run stabilization interval $\mathcal{I}$ such that

$$\mathbf{P}^{(n)} \approx \mathbf{S}^{(n)} \approx \mathbf{S}^{\infty} = \begin{pmatrix} \mathbf{e}\mathbf{s}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{e}\mathbf{s}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{e}\mathbf{s}_k \end{pmatrix}$$

as long as $n \in \mathcal{I}$, and during this period,

$$\boldsymbol{\pi}_{(n)} = \boldsymbol{\pi}_0\mathbf{P}^{(n)} \approx \mathbf{s} = \boldsymbol{\pi}_0\mathbf{S}^{\infty} = \begin{pmatrix} \nu_1\mathbf{s}_1 & \nu_2\mathbf{s}_2 & \cdots & \nu_k\mathbf{s}_k \end{pmatrix}$$

where the ν_i's are the constants given by

$$\nu_i = \boldsymbol{\pi}_0^{(i)}\mathbf{e} = \sum_{h=1} \left(\boldsymbol{\pi}_0^{(i)}\right)_h.$$

In such a manner, the analogue of short-run stabilization can be established for the more general case, and essentially the same arguments used in §8 can be used to provide rigorous proofs. Similarly, the middle-run and long-run dynamics of a non-primitive nearly completely reducible chain are derived from essentially the same arguments used in §9 and §10 thereby providing a direct extension of Theorem 10.1 to the more general case by simply replacing $\mathbf{P}^n$, $\mathbf{S}_{ii}^n$, and $\boldsymbol{\pi}_n$ by $\mathbf{P}^{(n)}$, $\mathbf{S}_{ii}^{(n)}$, and $\boldsymbol{\pi}_{(n)}$, respectively.

12. Acknowledgments. The author is indebted to Paul Schweitzer, Moshe Haviv, Bruce Mattingly, and W. J. Stewart for their useful comments which helped to improve this exposition.

REFERENCES

[1] S. Balsamo and B. Pandolfi, *Bounded aggregation in Markovian networks,* in Computer Performance and Reliability, G. Iazeolla, P. J. Courtois, and O. J. Boxma, eds., Elsevier North–Holland, Amsterdam, New York, 1988, pp. 73–91.

[2] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences,* Academic Press, New York, 1979.

[3] P. J. Courtois, *Decomposability: Queueing And Computer System Applications,* Academic Press, New York, 1977

[4] ———, *Analysis of large markovian models by parts—Applications to queueing network models,* in Messung, Modellierung und Bewertung von Rechensystemen, GI/NTH-Fachtagung, Springer-Verlag, Dortmund, 1985, pp. 1–10.

[5] P. J. Courtois and P. Semal, *Block iterative algorithms for stochastic matrices,* Linear Algebra Appl., 76(1986), pp. 59–70.

[6] ———, *Bounds on conditional steady-state distributions in large Markovian queueing models,* in Teletraffic Analysis and Computer Performance Evaluation, O. J. Boxma, J. W. Cohen, and H. C. Tijms, eds., Elsevier North–Holland, Amsterdam, New York, 1986, pp. 499–520.

[7] F. R. Gantmacher, *Matrix Theory*—Vol. II, Chelsea, New York, 1960.

[8] W. K. Grassmann, M. I. Taksar, and D. P. Heyman, *Regenerative analysis and steady state distributions for Markov chains,* Oper. Res., 33(1985), pp. 1107–1116.

[9] M. Haviv, *Aggregation/disaggregation methods for computing the stationary distribution of a Markov chain,* SIAM J. Numer. Anal., 22(1987), pp. 952–966.

[10] M. Haviv and L. Van Der Heyden, *Perturbation bounds for the stationary probabilities of a finite Markov chain,* Adv. in Appl. Probab., 16(1984), pp. 804–818.

[11] R. A. Horn and C. A. Johnson, *Matrix Analysis,* Cambridge University Press, New York, 1985.

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

[13] R. Lal and U. N. Bhat, *Reduced systems in Markov chains and their applications in queueing theory,* Queueing Systems, 2(1987), pp. 147–172.

[14] R. B. Mattingly, *Vector and parallel algorithms for computing the stationary distribution vector of an irreducible Markov chain,* Ph.D. thesis, North Carolina State University, Raleigh, NC, 1988.

[15] R. B. Mattingly and C. D. Meyer, *Computing the stationary distribution vector of an irreducible Markov chain on a shared memory multiprocessor,* Tech. Report 11138801, Center for Research in Scientific Computation, North Carolina State University, Raleigh, NC, 1988.

[16] ———, *Stochastic complementation and the GTH algorithm for Markov chains,* Tech. Report 11228801, Center for Research in Scientific Computation, North Carolina State University, Raleigh, NC, 1988.

[17] C. D. Meyer, *Uncoupling the Perron eigenvalue problem,* Tech. Report 12038701, Center for Research in Scientific Computation, North Carolina State University, Raleigh, NC, 1988; Linear Algebra Appl., March 1989.

[18] H. A. Simon and A. Ando, *Aggregation of variables in dynamic systems,* Econometrica, 29(1961), pp. 111–138.

[19] G. W. Stewart, *Introduction To Matrix Computations,* Academic Press, New York, 1973.

[20] H. Vantilborgh, *Aggregation with an error of $O(\epsilon^2)$,* J. Assoc. Comput. Mach, 32(1985), pp. 162–190.
[21] R. L. Zarling, *Numerical solutions of nearly completely decomposable queueing networks,* Ph.D. thesis, University of North Carolina, Chapel Hill, NC, 1976.