

QUASI LUMPABILITY, LOWER-BOUNDING COUPLING MATRICES, AND NEARLY COMPLETELY DECOMPOSABLE MARKOV CHAINS*

TUĞRUL DAYAR† AND WILLIAM J. STEWART‡

Abstract. In this paper, it is shown that nearly completely decomposable (NCD) Markov chains are quasi-lumpable. The state space partition is the natural one, and the technique may be used to compute lower and upper bounds on the stationary probability of each NCD block. In doing so, a lower-bounding nonnegative coupling matrix is employed. The nature of the stationary probability bounds is closely related to the structure of this lower-bounding matrix. Irreducible lower-bounding matrices give tighter bounds compared with bounds obtained using reducible lower-bounding matrices. It is also noticed that the quasi-lumped chain of an NCD Markov chain is an ill-conditioned matrix and the bounds obtained generally will not be tight. However, under some circumstances, it is possible to compute the stationary probabilities of some NCD blocks exactly.

Key words. Markov chains, quasi lumpability, decomposability, stationary probability, aggregation–disaggregation schemes

AMS subject classifications. 60J10, 60J27, 65U05, 65F05, 65F10, 65F30

PII. S0895479895294277

1. Introduction. Markovian modeling and analysis are extensively used in many disciplines in evaluating the performance of existing systems and in analyzing and designing systems to be developed. The long-run behavior of Markovian systems is revealed through the solution of the problem

$$(1.1) \quad \boldsymbol{\pi} \mathbf{P} = \boldsymbol{\pi}, \quad \|\boldsymbol{\pi}\|_1 = 1,$$

where \mathbf{P} is the one-step stochastic transition probability matrix (i.e., discrete-time Markov chain—DTMC) and $\boldsymbol{\pi}$ is the unknown stationary probability distribution of the system under consideration. By definition, rows of \mathbf{P} and elements of $\boldsymbol{\pi}$ both sum up to 1.

In what follows, boldface capital letters denote matrices, boldface lowercase letters denote column vectors, italic lowercase and uppercase letters denote scalars, and calligraphic letters denote sets. \mathbf{e} represents a column vector of all ones and \mathbf{o} represents a row or column vector of all zeros depending on the context. The convention of representing probability distributions by row vectors is adopted.

Solving (1.1) is crucial in computing performance measures for Markovian systems. For queueing systems, these measures may be the average number of customers, the mean waiting time, or the blocking probability for a specific queue. In communication systems, they may be the total packet loss rate, the probability of an empty system, or any other relevant measure. In any case, these measures may be computed exactly if $\boldsymbol{\pi}$ is available.

* Received by the editors November 1, 1995; accepted for publication (in revised form) by D. P. O’Leary June 11, 1996. This work was initiated while T. Dayar was in the Department of Computer Science at North Carolina State University. His work is currently supported by Scientific and Technical Research Council of Turkey (TÜBİTAK) grant EEEAG-161.

<http://www.siam.org/journals/simax/18-2/29427.html>

† Department of Computer Engineering and Information Science, Bilkent University, 06533 Bilkent, Ankara, Turkey (tugrul@bilkent.edu.tr).

‡ Department of Computer Science, North Carolina State University, Raleigh, NC 27695-8206 (billy@markov.csc.ncsu.edu).

NCD Markov chains [3], [10], [16] are irreducible stochastic matrices that can be ordered so that the matrix of transition probabilities has a block structure in which the nonzero elements of the off-diagonal blocks are small compared with those of the diagonal blocks. Such matrices often arise in queueing network analysis, large-scale economic modeling, and computer systems performance evaluation, and they can be represented in the form

$$(1.2) \quad \mathbf{P}_{n \times n} = \begin{matrix} & \begin{matrix} n_1 & n_2 & \cdots & n_N \end{matrix} \\ \begin{pmatrix} \mathbf{P}_{1,1} & \mathbf{P}_{1,2} & \cdots & \mathbf{P}_{1,N} \\ \mathbf{P}_{2,1} & \mathbf{P}_{2,2} & \cdots & \mathbf{P}_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{N,1} & \mathbf{P}_{N,2} & \cdots & \mathbf{P}_{N,N} \end{pmatrix} & \begin{matrix} n_1 \\ n_2 \\ \vdots \\ n_N \end{matrix} \end{matrix}.$$

The subblocks $\mathbf{P}_{i,i}$ are square, of order n_i , with $n = \sum_{i=1}^N n_i$. Let $\boldsymbol{\pi}$ be partitioned conformally with \mathbf{P} such that $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \dots, \boldsymbol{\pi}_N)$. Each $\boldsymbol{\pi}_i$, $i = 1, 2, \dots, N$ is a row vector having n_i elements. Let $\mathbf{P} = \text{diag}(\mathbf{P}_{1,1}, \mathbf{P}_{2,2}, \dots, \mathbf{P}_{N,N}) + \mathbf{E}$. The quantity $\|\mathbf{E}\|_\infty$ is referred to as the degree of coupling, and it is taken to be a measure of the decomposability of the matrix (see [6]). If it were zero, then \mathbf{P} would be reducible.

Consider the following questions. Is it possible to obtain lower and upper bounds on the stationary probability of being in each NCD block of an NCD Markov chain in an inexpensive way? Furthermore, if the answer to the preceding question is yes, can one improve these bounds by exploiting the structure and symmetries of the chain? The motivation behind seeking answers to such questions is that in many cases performance measures of interest of systems undergoing analysis depend on the probability of being in certain groups of states. That is, probabilities need to be computed at a coarser level; each and every stationary probability is not needed. If the problem at hand is one in which the stationary probabilities of interest are those of the coupling matrix [10] corresponding to the underlying NCD Markov chain, then the technique discussed in this paper may be used to obtain answers to the above questions. Whereas if all stationary probabilities of the NCD Markov chain are to be computed, iterative aggregation–disaggregation (IAD) should be the method of choice (see [8], [2], [12], [15], [14], [16]).

In the sections to come, it is shown that NCD Markov chains are quasi-lumpable. The state space partition coincides with the NCD block partition, and the technique may be used to compute lower and upper bounds on the probability of being in each NCD block. The procedure amounts to solving linear systems of order equal to the number of NCD blocks in the chain. Thereafter, quasi lumpability is related to the polyhedra theory of Courtois and Semal for stochastic matrices [4], and it is shown that under certain circumstances the quasi-lumped chain (as defined in [5]) is a lower-bounding matrix for the coupling matrix of the NCD chain. Additionally, another substochastic matrix guaranteed to be a lower-bounding coupling matrix is given. Following this, the effects of the nonzero structure of a lower-bounding nonnegative coupling matrix on the bounds of the stationary probability of each NCD block is investigated; the results are based on the nonzero structure of a lower-bounding substochastic matrix in general, and, therefore, they may also be used in forecasting the quality of lower and upper bounds on the stationary distribution of Markov chains when Courtois and Semal’s theory is at work.

The next section provides the definitions of lumpability (see [7, section 6.3]) and quasi lumpability (see [5]), and section 3 shows how quasi lumpability applies to NCD

Markov chains. The effects of quasi lumpability on the 8×8 Courtois matrix are illustrated in section 4. The relation between the quasi-lumped chain and the coupling matrix of an NCD Markov chain is investigated in section 5. Section 6 provides information enabling one to forecast the nature of the bounds on the stationary probability of each NCD block; the idea is communicated through an illustrative example. The last section summarizes the results.

2. Lumpability vs. quasi lumpability. Lumpability is a property of some Markov chains which, if conditions are met, may be used to reduce a large state space to a smaller one. The idea is to find a partition of the original state space such that, when the states in each partition are combined to form a single state, the resulting Markov chain described by the combined states has equivalent behavior to the original chain, only at a coarser level of detail. Given that the conditions for lumpability are satisfied, it is mostly useful in systems which require the computation of performance measures dependent on the coarser analysis specified by the lumped chain (see [7, p. 123]).

DEFINITION 2.1. *A DTMC is said to be lumpable with respect to a given state space partition $\mathcal{S} = \bigcup_i \mathcal{S}_i$ with $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset \forall i \neq j$ if its transition probability matrix \mathbf{P} satisfies the lumpability condition*

$$(2.1) \quad \forall \mathcal{S}_i, \mathcal{S}_j \subset \mathcal{S} \quad \forall s \in \mathcal{S}_i : \sum_{s' \in \mathcal{S}_j} p_{s,s'} = k_{i,j} \quad \forall i, j,$$

where $k_{i,j}$ is a constant value that depends only on i and j and $p_{s,s'}$ is the one-step transition probability of going from state s to state s' . The lumped chain \mathbf{K} has $k_{i,j}$ as its i, j th entry. A similar definition applies to a continuous-time Markov chain (CTMC), where the probability matrix \mathbf{P} is substituted with the infinitesimal generator \mathbf{Q} .

To put it in another way, the lumpability condition requires the transition probability from each state in a given partition to another partition to be the same. For a given state, the probability of making a transition to a partition is the sum of the transition probabilities from the given state to each state in that partition. At this point we should stress that not all Markov chains are lumpable. In fact, only a small percentage of Markov chains arising in real-life applications is expected to be lumpable. However, in section 3 it is shown that NCD Markov chains are quasi-lumpable, that is, almost lumpable [5]. The following informative example demonstrates the concept of lumpability.

Example 2.2. Let

$$\mathbf{P} = \begin{array}{c} \begin{array}{cccc} & 1 & 2 & 3 & 4 \\ \begin{array}{l} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0.2 & 0.3 & 0.4 & 0.1 \\ 0.3 & 0.1 & 0.4 & 0.2 \\ 0.5 & 0.1 & 0.1 & 0.3 \\ 0.5 & 0.3 & 0.2 & 0 \end{pmatrix} \end{array} \end{array}.$$

We take partition $\mathcal{S} = \{1, 3\} \cup \{2, 4\}$. For this partition the lumpability condition is satisfied with $k_{1,1} = 0.6, k_{1,2} = 0.4, k_{2,1} = 0.7, k_{2,2} = 0.3$, where $\mathcal{S}_1 = \{1, 3\}, \mathcal{S}_2 = \{2, 4\}$. The lumped chain is given by

$$\mathbf{K} = \begin{array}{c} \begin{array}{cc} & \mathcal{S}_1 & \mathcal{S}_2 \\ \begin{array}{l} \mathcal{S}_1 \\ \mathcal{S}_2 \end{array} & \begin{pmatrix} 0.6 & 0.4 \\ 0.7 & 0.3 \end{pmatrix} \end{array} \end{array}.$$

DEFINITION 2.3. A DTMC is said to be ϵ quasi-lumpable with respect to a given state space partition $\mathcal{S} = \bigcup_i \mathcal{S}_i$ with $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset \forall i \neq j$ if its transition probability matrix \mathbf{P} can be written as $\mathbf{P} = \mathbf{P}^- + \mathbf{P}^\epsilon$. Here \mathbf{P}^- is a (componentwise) lower bound for \mathbf{P} that satisfies the lumpability condition

$$(2.2) \quad \forall \mathcal{S}_i, \mathcal{S}_j \subset \mathcal{S} \forall s \in \mathcal{S}_i : \sum_{s' \in \mathcal{S}_j} p_{s,s'}^- = k_{i,j} \quad \forall i \neq j$$

under the following constraints. No element in \mathbf{P}^ϵ is greater than ϵ (a small number); $\|\mathbf{P}^\epsilon\|_\infty$ assumes the minimum value among all possible alternatives (since \mathbf{P}^- and \mathbf{P}^ϵ may not be unique); $k_{i,j}$ is a constant value that depends only on i and j ; and $p_{s,s'}^-$ is the one-step transition probability of going from state s to state s' in the matrix \mathbf{P}^- (see [5, p. 224]). The computation of the quasi-lumped chain is discussed in the next section. A similar definition applies to a CTMC as in Definition 2.1.

The concept of ϵ quasi lumpability is illustrated in the following 6×6 example.

Example 2.4. Let

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \left(\begin{array}{ccc|ccc} 0.2 & 0.28 & 0.1 & 0.21 & 0.11 & 0.1 \\ 0.29 & 0.1 & 0.2 & 0.05 & 0.31 & 0.05 \\ 0.15 & 0.2 & 0.24 & 0.12 & 0.2 & 0.09 \\ \hline 0.27 & 0.18 & 0.22 & 0.18 & 0.01 & 0.14 \\ 0.18 & 0.2 & 0.3 & 0.31 & 0.01 & 0 \\ 0 & 0.25 & 0.43 & 0.07 & 0.08 & 0.17 \end{array} \right) \end{matrix}.$$

\mathbf{P}^- and \mathbf{P}^ϵ given by

$$\mathbf{P}^- = \left(\begin{array}{ccc|ccc} 0.2 & 0.28 & 0.1 & 0.2 & 0.11 & 0.1 \\ 0.29 & 0.1 & 0.2 & 0.05 & 0.31 & 0.05 \\ 0.15 & 0.2 & 0.24 & 0.12 & 0.2 & 0.09 \\ \hline 0.27 & 0.18 & 0.22 & 0.18 & 0.01 & 0.14 \\ 0.18 & 0.2 & 0.29 & 0.31 & 0.01 & 0 \\ 0 & 0.24 & 0.43 & 0.07 & 0.08 & 0.17 \end{array} \right),$$

$$\mathbf{P}^\epsilon = \left(\begin{array}{ccc|ccc} 0 & 0 & 0 & 0.01 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.01 & 0 & 0 & 0 \\ 0 & 0.01 & 0 & 0 & 0 & 0 \end{array} \right)$$

with $\epsilon = 0.01$ and state space partition $\mathcal{S} = \{1, 2, 3\} \cup \{4, 5, 6\}$ satisfy the quasi-lumpability condition in (2.2). This time $\mathcal{S}_1 = \{1, 2, 3\}$, $\mathcal{S}_2 = \{4, 5, 6\}$, and $k_{1,2} = 0.41, k_{2,1} = 0.67$. Observe that for $\epsilon = 0.01$ the given $(\mathbf{P}^-, \mathbf{P}^\epsilon)$ pair is not the only one that satisfies the quasi-lumpability condition. For instance, the following pair also satisfies (2.2):

$$\mathbf{P}^- = \left(\begin{array}{ccc|ccc} 0.2 & 0.28 & 0.1 & 0.21 & 0.1 & 0.1 \\ 0.29 & 0.1 & 0.2 & 0.05 & 0.31 & 0.05 \\ 0.15 & 0.2 & 0.24 & 0.12 & 0.2 & 0.09 \\ \hline 0.27 & 0.18 & 0.22 & 0.18 & 0.01 & 0.14 \\ 0.17 & 0.2 & 0.3 & 0.31 & 0.01 & 0 \\ 0 & 0.25 & 0.42 & 0.07 & 0.08 & 0.17 \end{array} \right),$$

$$\mathbf{P}^\epsilon = \left(\begin{array}{ccc|ccc} 0 & 0 & 0 & 0 & 0.01 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 0.01 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.01 & 0 & 0 & 0 \end{array} \right).$$

The next section provides a proof by construction for the ϵ quasi lumpability of NCD Markov chains.

3. Construction.

1. For an NCD Markov chain, let the state space be partitioned as

$$\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_N\},$$

where \mathcal{S}_i is the set of states forming the i th block and $\#(\mathcal{S}_i) = n_i$ with $n = \sum_{i=1}^N n_i$.

Form the matrix

$$(3.1) \quad \mathbf{P}^- = \begin{matrix} & \begin{matrix} n_1 & n_2 & \cdots & n_N \end{matrix} \\ \begin{pmatrix} \mathbf{P}_{1,1} & \mathbf{P}_{1,2} & \cdots & \mathbf{P}_{1,N} \\ \mathbf{P}_{2,1} & \mathbf{P}_{2,2} & \cdots & \mathbf{P}_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{N,1} & \mathbf{P}_{N,2} & \cdots & \mathbf{P}_{N,N} \end{pmatrix} & \begin{matrix} n_1 \\ n_2 \\ \vdots \\ n_N \end{matrix} \end{matrix},$$

where

$$(3.2) \quad \mathbf{P}_{i,j}^- = \begin{cases} \mathbf{P}_{i,j} & \text{if } \mathbf{P}_{i,j}\mathbf{e} = k_{i,j}\mathbf{e} \\ \mathbf{P}_{i,j} - \mathbf{P}_{i,j}^\epsilon & \text{otherwise} \end{cases} \quad \forall i \neq j.$$

Diagonal blocks of \mathbf{P}^- are the same as those of \mathbf{P} . When $\mathbf{P}_{i,j}\mathbf{e} \neq k_{i,j}\mathbf{e}$, $\mathbf{P}_{i,j}^\epsilon$ is chosen so that $(\mathbf{P}_{i,j} - \mathbf{P}_{i,j}^\epsilon)\mathbf{e} = k_{i,j}\mathbf{e}$. Here, $k_{i,j} = \min(\mathbf{P}_{i,j}\mathbf{e})$ (i.e., the minimum-valued element of the vector $\mathbf{P}_{i,j}\mathbf{e}$). As pointed out in Example 2.4, \mathbf{P}^ϵ may not be unique, and the discussion on how to choose among the alternatives available is left to after the construction. Furthermore, \mathbf{P}^ϵ has nonzero blocks (in which there is at least one nonzero element) in locations corresponding to the nonzero blocks of \mathbf{P} which do not have equal row sums. On the other hand, the number of zero blocks in \mathbf{P}^- may be more than the number of zero blocks in \mathbf{P} . In other words, there may be nonzero blocks in \mathbf{P} for which $k_{i,j} = 0$, implying $\mathbf{P}_{i,j}^- = \mathbf{0}$. Note that, if \mathbf{P}^ϵ is the null matrix, then \mathbf{P} will be exactly lumpable, and the remaining steps in the construction should be skipped.

2. Once \mathbf{P} is written as the sum of \mathbf{P}^- and \mathbf{P}^ϵ , form yet another matrix

$$(3.3) \quad \mathbf{P}^s = \begin{pmatrix} \mathbf{P}^- & \mathbf{y} \\ \mathbf{x}^T & 0 \end{pmatrix},$$

where

$$(3.4) \quad \mathbf{y} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_N \end{pmatrix} = \mathbf{P}^\epsilon \mathbf{e}$$

and $\bar{\mathbf{y}}_i$ has n_i elements. The unknown vector \mathbf{x} should be partitioned in the same way. The significance and role of \mathbf{x} in the computation of lower and upper bounds for the quasi-lumped chain is discussed in section 4. Recall the definition of an NCD Markov chain in section 1 and observe that $\|\mathbf{y}\|_\infty \leq \|\mathbf{E}\|_\infty$ (the degree of coupling of \mathbf{P}). Since $\|\mathbf{E}\|_\infty$ is a small number generally less than 0.1, one has ϵ quasi lumpability (see Definition 2.3). The small mass in the off-diagonal blocks, which prevents lumping \mathbf{P} exactly, is accumulated in an extra state.

3. Given that \mathbf{P} is not exactly lumpable (i.e., $\mathbf{y} \neq \mathbf{o}$), \mathbf{P}^s will not be lumpable. However, the lumpability condition for the i th row of blocks may be enforced by increasing some elements in $\bar{\mathbf{y}}_i$ so as to make each element equal to $\|\bar{\mathbf{y}}_i\|_\infty$ and decreasing the corresponding diagonal elements. If it is possible for any diagonal element to become negative, the diagonal of \mathbf{P}^s may be scaled by performing the transformation

$$(3.5) \quad \alpha \mathbf{P}^s + (1 - \alpha)\mathbf{I},$$

where $0 < \alpha < 1$, on \mathbf{P}^s as suggested in [5]. Denote the matrix obtained in the end $\tilde{\mathbf{P}}^s$.

4. $\tilde{\mathbf{P}}^s$ is lumpable, and it may be lumped to form the following *quasi-lumped chain* that corresponds to \mathbf{P} :

$$(3.6) \quad \mathbf{K}^s = \left(\begin{array}{cccc|c} \|\tilde{\mathbf{P}}_{1,1}^s\|_\infty & \|\tilde{\mathbf{P}}_{1,2}^s\|_\infty & \cdots & \|\tilde{\mathbf{P}}_{1,N}^s\|_\infty & \|\bar{\mathbf{y}}_1\|_\infty \\ \|\tilde{\mathbf{P}}_{2,1}^s\|_\infty & \|\tilde{\mathbf{P}}_{2,2}^s\|_\infty & \cdots & \|\tilde{\mathbf{P}}_{2,N}^s\|_\infty & \|\bar{\mathbf{y}}_2\|_\infty \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \|\tilde{\mathbf{P}}_{N,1}^s\|_\infty & \|\tilde{\mathbf{P}}_{N,2}^s\|_\infty & \cdots & \|\tilde{\mathbf{P}}_{N,N}^s\|_\infty & \|\bar{\mathbf{y}}_N\|_\infty \\ \hline \|\bar{\mathbf{x}}_1\|_1 & \|\bar{\mathbf{x}}_2\|_1 & \cdots & \|\bar{\mathbf{x}}_N\|_1 & 0 \end{array} \right).$$

5. Bounds on the stationary probability of each NCD block may be obtained using Courtois and Semal’s method [4], [13] if the $N \times N$ principal submatrix of \mathbf{K}^s is a lower-bounding coupling matrix for \mathbf{P} .

When constructing \mathbf{P}^ϵ , the nonzero elements in blocks should be arranged, if at all possible, so that there is a minimum number of nonzero columns in \mathbf{P}^ϵ . If all columns corresponding to states in \mathcal{S}_i are zero in \mathbf{P}^ϵ , then $\bar{\mathbf{x}}_i = \mathbf{o}$, and the stationary probability of the i th block may be determined exactly to working precision. An intuitive explanation for this fact is the following. The transitions in \mathbf{P}^ϵ are the transitions into and out of the extra state (in \mathbf{P}^s). Therefore, if it is not possible to make a transition to state s , say, in the matrix \mathbf{P}^ϵ (i.e., the column of \mathbf{P}^ϵ that corresponds to state s is \mathbf{o}), then it will not be possible to return to state s from the extra state. This being so, the corresponding element in \mathbf{x}^T must be zero. If all states in an NCD block possess this property, then the element in the last row of the quasi-lumped chain \mathbf{K}^s corresponding to that NCD block should be zero. A side-note is that, even though there may be multiple ways in which the nonzero entries of \mathbf{P}^ϵ can be arranged for fixed ϵ , this does not make a difference when lower and upper bounds on the stationary probability of each NCD block are computed.

The next section illustrates the construction steps on a small example and shows how to compute the corresponding quasi-lumped chain with lower and upper bounds for its stationary vector.

4. An illustrative example. Consider the 8×8 Courtois matrix [3]

$$\mathbf{P} = \left(\begin{array}{ccc|cc|cc|c} 0.85 & 0 & 0.149 & 0.0009 & 0 & 0.00005 & 0 & 0.00005 \\ 0.1 & 0.65 & 0.249 & 0 & 0.0009 & 0.00005 & 0 & 0.00005 \\ 0.1 & 0.8 & 0.0996 & 0.0003 & 0 & 0 & 0.0001 & 0 \\ \hline 0 & 0.0004 & 0 & 0.7 & 0.2995 & 0 & 0.0001 & 0 \\ 0.0005 & 0 & 0.0004 & 0.399 & 0.6 & 0.0001 & 0 & 0 \\ \hline 0 & 0.00005 & 0 & 0 & 0.00005 & 0.6 & 0.2499 & 0.15 \\ 0.00003 & 0 & 0.00003 & 0.00004 & 0 & 0.1 & 0.8 & 0.0999 \\ 0 & 0.00005 & 0 & 0 & 0.00005 & 0.1999 & 0.25 & 0.55 \end{array} \right).$$

The degree of coupling for this matrix is 0.001. From the first step of the construction, one obtains

$$\mathbf{P}^- = \left(\begin{array}{ccc|cc|cc|c} 0.85 & 0 & 0.149 & 0.0003 & 0 & 0.00005 & 0 & 0.00005 \\ 0.1 & 0.65 & 0.249 & 0 & 0.0003 & 0.00005 & 0 & 0.00005 \\ 0.1 & 0.8 & 0.0996 & 0.0003 & 0 & 0 & 0.0001 & 0 \\ \hline 0 & 0.0004 & 0 & 0.7 & 0.2995 & 0 & 0.0001 & 0 \\ 0 & 0 & 0.0004 & 0.399 & 0.6 & 0.0001 & 0 & 0 \\ \hline 0 & 0.00005 & 0 & 0 & 0.00004 & 0.6 & 0.2499 & 0.15 \\ 0.00002 & 0 & 0.00003 & 0.00004 & 0 & 0.1 & 0.8 & 0.0999 \\ 0 & 0.00005 & 0 & 0 & 0.00004 & 0.1999 & 0.25 & 0.55 \end{array} \right),$$

$$\mathbf{P}^\epsilon = \left(\begin{array}{ccc|cc|cc|c} 0 & 0 & 0 & 0.0006 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.0006 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0005 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0.00001 & 0 & 0 & 0 \\ 0.00001 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.00001 & 0 & 0 & 0 \end{array} \right).$$

$\mathbf{P} = \mathbf{P}^- + \mathbf{P}^\epsilon$ (with $\epsilon = 0.0006$), as required, and the second step of the construction gives

$$\mathbf{P}^s = \left(\begin{array}{ccc|cc|cc|c} 0.85 & 0 & 0.149 & 0.0003 & 0 & 0.00005 & 0 & 0.00005 & 0.0006 \\ 0.1 & 0.65 & 0.249 & 0 & 0.0003 & 0.00005 & 0 & 0.00005 & 0.0006 \\ 0.1 & 0.8 & 0.0996 & 0.0003 & 0 & 0 & 0.0001 & 0 & 0 \\ \hline 0 & 0.0004 & 0 & 0.7 & 0.2995 & 0 & 0.0001 & 0 & 0 \\ 0 & 0 & 0.0004 & 0.399 & 0.6 & 0.0001 & 0 & 0 & 0.0005 \\ \hline 0 & 0.00005 & 0 & 0 & 0.00004 & 0.6 & 0.2499 & 0.15 & 0.00001 \\ 0.00002 & 0 & 0.00003 & 0.00004 & 0 & 0.1 & 0.8 & 0.0999 & 0.00001 \\ 0 & 0.00005 & 0 & 0 & 0.00004 & 0.1999 & 0.25 & 0.55 & 0.00001 \\ \hline x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & 0 \end{array} \right).$$

Note that there are no transitions to states 2, 3, 6, 7, and 8 in \mathbf{P}^ϵ . Hence, x_2, x_3, x_6, x_7 , and x_8 in \mathbf{P}^s must be zero. Observe that \mathbf{P}^s is still not lumpable. For it to be lumpable, the last column should be modified. Following the third step of the construction, diagonal elements $p_{3,3}^s$ and $p_{4,4}^s$ are adjusted and one obtains

$$\tilde{\mathbf{P}}^s = \left(\begin{array}{ccc|cc|cc|c} 0.85 & 0 & 0.149 & 0.0003 & 0 & 0.00005 & 0 & 0.00005 & 0.0006 \\ 0.1 & 0.65 & 0.249 & 0 & 0.0003 & 0.00005 & 0 & 0.00005 & 0.0006 \\ 0.1 & 0.8 & 0.099 & 0.0003 & 0 & 0 & 0.0001 & 0 & 0.0006 \\ \hline 0 & 0.0004 & 0 & 0.6995 & 0.2995 & 0 & 0.0001 & 0 & 0.0005 \\ 0 & 0 & 0.0004 & 0.399 & 0.6 & 0.0001 & 0 & 0 & 0.0005 \\ \hline 0 & 0.00005 & 0 & 0 & 0.00004 & 0.6 & 0.2499 & 0.15 & 0.00001 \\ 0.00002 & 0 & 0.00003 & 0.00004 & 0 & 0.1 & 0.8 & 0.0999 & 0.00001 \\ 0 & 0.00005 & 0 & 0 & 0.00004 & 0.1999 & 0.25 & 0.55 & 0.00001 \\ \hline x_1 & 0 & x_3 & x_4 & x_5 & 0 & 0 & 0 & 0 \end{array} \right).$$

Notice that x_3 in $\tilde{\mathbf{P}}^s$ is different than zero, as opposed to what has been said before. The reason is that $p_{3,3}^s$ has been adjusted, thus making $p_{3,3}^\epsilon$ effectively a nonzero entry of value 0.0006. Therefore, the third column in \mathbf{P}^ϵ intrinsically has a nonzero entry in the diagonal position, implying a transition from the extra state to state 3. Likewise, $p_{4,4}^s$ has been adjusted, making $p_{4,4}^\epsilon$ equal to 0.0005. However, x_4 is already nonzero and need not be altered. This issue will be revisited at the end of the section. Resuming the construction, the quasi-lumped chain in step four is computed as

$$\mathbf{K}^s = \left(\begin{array}{ccc|c} 0.999 & 0.0003 & 0.0001 & 0.0006 \\ 0.0004 & 0.999 & 0.0001 & 0.0005 \\ 0.00005 & 0.00004 & 0.9999 & 0.00001 \\ \hline \|\bar{\mathbf{x}}_1\|_1 & \|\bar{\mathbf{x}}_2\|_1 & 0 & 0 \end{array} \right).$$

As suggested in the fifth step of the construction, lower and upper bounds on the stationary probability of each NCD block may be obtained by successively substituting a one for each (unknown) $\|\bar{\mathbf{x}}_i\|_1$ in the last row of \mathbf{K}^s (denote this matrix by \mathbf{K}_i^s) and solving the corresponding system

$$(4.1) \quad \mathbf{z}_i \mathbf{K}_i^s = \mathbf{z}_i, \quad \sum_{j=1}^N z_{i,j} = 1.$$

Here, \mathbf{z}_i is a probability vector of N elements. If ξ_j is the stationary probability of the j th NCD block, then lower and upper bounds on the stationary probability of block j may be computed from

$$(4.2) \quad \xi_j^{inf} = \max \left\{ \min_i(z_{i,j}); 1 - \sum_{k \neq j} \max_i(z_{i,k}) \right\},$$

$$(4.3) \quad \xi_j^{sup} = \min \left\{ \max_i(z_{i,j}); 1 - \sum_{k \neq j} \min_i(z_{i,k}) \right\}$$

(see [4, (3.26), p. 810]).

For the Courtois matrix

$$\|\bar{\mathbf{x}}_1\|_1 = 1, \|\bar{\mathbf{x}}_2\|_1 = 0 \Rightarrow \mathbf{z}_1 = [0.36923, 0.13077, 0.50000],$$

$$\|\bar{\mathbf{x}}_1\|_1 = 0, \|\bar{\mathbf{x}}_2\|_1 = 1 \Rightarrow \mathbf{z}_2 = [0.16071, 0.33929, 0.50000],$$

$$\begin{aligned} 0.16071 &\leq \xi_1 \leq 0.36923, \\ 0.13077 &\leq \xi_2 \leq 0.33929, \\ 0.50000 &\leq \xi_3 \leq 0.50000 \Rightarrow \xi_3 = 0.50000, \end{aligned}$$

and $\xi_1 + \xi_2 + \xi_3 = 1$ in five decimal digits of accuracy.

We obtained the stationary probability of each NCD block by solving for the stationary vector of the original 8×8 chain. The probabilities accurate to five decimal digits are

$$\xi_1 = 0.22253, \quad \xi_2 = 0.27747, \quad \xi_3 = 0.50000.$$

The next thing to do is to show how a distribution \mathbf{x}^T that gives the stationary probability of each NCD block may be obtained. In fact, the procedure amounts to computing x_1, x_3, x_4 , and x_5 values only, for the rest of the elements in \mathbf{x} are necessarily zero. Let $\boldsymbol{\pi}$ denote the stationary vector of \mathbf{P} (i.e., $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$, $\|\boldsymbol{\pi}\|_1 = 1$). Then

$$\begin{aligned}x_1 &= (0.0005\pi_5 + 0.00001\pi_7)/t, \\x_3 &= 0.0006\pi_3/t, \\x_4 &= (0.0006\pi_1 + 0.0005\pi_4)/t, \\x_5 &= (0.0006\pi_2 + 0.00001\pi_6 + 0.00001\pi_8)/t,\end{aligned}$$

where $t = 0.0006(\pi_1 + \pi_2 + \pi_3) + 0.0005(\pi_4 + \pi_5) + 0.00001(\pi_6 + \pi_7 + \pi_8)$. The last condition ensures that \mathbf{x}^T is a probability vector. As can be seen, the computation of \mathbf{x} requires full knowledge of $\boldsymbol{\pi}$ (which, of course, is unknown). For the Courtois matrix, the unknown entries in the last row of \mathbf{K}^s are given by

$$\begin{aligned}\|\bar{\mathbf{x}}_1\|_1 &= (0.0006\pi_3 + 0.0005\pi_5 + 0.00001\pi_7)/t, \\ \|\bar{\mathbf{x}}_2\|_1 &= (0.0006(\pi_1 + \pi_2) + 0.0005\pi_4 + 0.00001(\pi_6 + \pi_8))/t.\end{aligned}$$

Using $\boldsymbol{\pi}$, one computes $\|\bar{\mathbf{x}}_1\|_1 = 0.31213$, $\|\bar{\mathbf{x}}_2\|_1 = 0.68787$ in five decimal digits of accuracy as the combination that gives $\boldsymbol{\xi}$.

The next section relates the quasi-lumped chain to the coupling matrix of the original NCD Markov chain.

5. Quasi-lumped chain and the coupling matrix. Let \mathbf{C}^s denote the $N \times N$ principal submatrix of the quasi-lumped chain \mathbf{K}^s . For the Courtois matrix,

$$\mathbf{C}^s = \begin{pmatrix} 0.99900 & 0.00030 & 0.00010 \\ 0.00040 & 0.99900 & 0.00010 \\ 0.00005 & 0.00004 & 0.99990 \end{pmatrix}.$$

On the other hand, the entries of the coupling matrix of an NCD Markov chain are given by [11]

$$c_{i,j} = \frac{\pi_i}{\|\boldsymbol{\pi}_i\|_1} \mathbf{P}_{i,j} \mathbf{e} \quad \forall i, j.$$

For the same example, the coupling matrix in five decimal digits of accuracy is then

$$\mathbf{C} = \begin{pmatrix} 0.99911 & 0.00079 & 0.00010 \\ 0.00061 & 0.99929 & 0.00010 \\ 0.00006 & 0.00004 & 0.99990 \end{pmatrix}.$$

In this example, \mathbf{C}^s is a lower bound for the exact coupling matrix \mathbf{C} . That is, $\mathbf{C}^s \leq \mathbf{C}$. Is this always true? Before answering this question, two lemmas should be stated. In the following, $\mathbf{u} \ll \mathbf{v}$ means each element of \mathbf{u} is considerably smaller than the corresponding element of \mathbf{v} . The symbol \ll may also be used between two scalars (i.e., two vectors of one element each).

LEMMA 5.1. *Let \mathbf{P} be an NCD Markov chain with N blocks that is not exactly lumpable. Let \mathbf{C}^s be the $N \times N$ principal submatrix of the quasi-lumped chain \mathbf{K}^s corresponding to \mathbf{P} in (3.6). Then \mathbf{C}^s has entries that satisfy*

$$(5.1) \quad 0 \leq c_{i,j}^s \leq \min(\mathbf{P}_{i,j} \mathbf{e}) \ll 1 \quad \forall i \neq j,$$

$$(5.2) \quad 0 \ll \min(\mathbf{P}_{i,i}\mathbf{e}) \leq c_{i,i}^s < 1 \quad \forall i.$$

Proof. Once again introduce $k_{i,j} = \min(\mathbf{P}_{i,j}\mathbf{e})$. Now observe that

$$0 \leq k_{i,j} \ll 1 \quad \forall i \neq j,$$

$$0 \ll k_{i,i} < 1 \quad \forall i$$

are direct consequences of the following properties of NCD Markov chains [10].

- For off-diagonal blocks,

$$\mathbf{0} \leq \mathbf{P}_{i,j}\mathbf{e} \ll \mathbf{e} \quad \forall i \neq j.$$

- For diagonal blocks,

$$\mathbf{0} \ll \mathbf{P}_{i,i}\mathbf{e} \leq \mathbf{e} \quad \forall i$$

with the condition that $\mathbf{P}_{i,i}\mathbf{e} \neq \mathbf{e}$ (since \mathbf{P} is irreducible by definition).

Now inspect the off-diagonal blocks in \mathbf{P}^- (see (3.1)) given by (3.2). If $\mathbf{P}_{i,j}$ has equal row sums (i.e., $\mathbf{P}_{i,j}\mathbf{e} = k_{i,j}\mathbf{e}$), then $\mathbf{P}_{i,j}^- = \mathbf{P}_{i,j}$. Otherwise, $\mathbf{P}_{i,j}^- = \mathbf{P}_{i,j} - \mathbf{P}_{i,j}^\epsilon$, where $(\mathbf{P}_{i,j} - \mathbf{P}_{i,j}^\epsilon)\mathbf{e} = k_{i,j}\mathbf{e}$. In all cases, $\mathbf{P}_{i,j}^-\mathbf{e} = k_{i,j}\mathbf{e}$. As for the diagonal blocks in \mathbf{P}^- , each diagonal block is equal to its counterpart in \mathbf{P} . Using (3.3), a new matrix \mathbf{P}^s is formed. The only blocks (possibly) prohibiting lumpability in \mathbf{P}^s are those diagonal blocks with unequal row sums. In other words, for \mathbf{P}^s to be lumpable, each diagonal block i for which $\min(\mathbf{P}_{i,i}\mathbf{e}) \neq \max(\mathbf{P}_{i,i}\mathbf{e})$ (i.e., $\max(\bar{\mathbf{y}}_i) \neq \min(\bar{\mathbf{y}}_i)$) needs to be adjusted. The adjustment in $\mathbf{P}_{i,i}^s$ may be performed by increasing some elements in $\bar{\mathbf{y}}_i$ so as to make each element in $\bar{\mathbf{y}}_i$ equal to $\max(\bar{\mathbf{y}}_i)$ and decreasing the corresponding diagonal element in $\mathbf{P}_{i,i}^s$. The intended effect is to have $\mathbf{P}_{i,i}^s\mathbf{e} = k_{i,i}\mathbf{e}$. As a result of this diagonal adjustment, one obtains a new \mathbf{P}^s which may or may not have negative elements along the diagonal. These two cases should be analyzed in turn.

(i) There are no negative elements along the diagonal of \mathbf{P}^s . Hence, the scaling in (3.5) need not be performed. In this case, $\tilde{\mathbf{P}}^s = \mathbf{P}^s$ (i.e., $\alpha = 1$ in (3.5)) and $\tilde{\mathbf{P}}^s$ may be quasi-lumped to form \mathbf{K}^s . The effect of quasi-lumping $\tilde{\mathbf{P}}^s$ is to have

$$(5.3) \quad k_{i,j}^s = k_{i,j} \quad \forall i, j \in \{1, 2, \dots, N\}.$$

(ii) There are one or more negative elements along the diagonal of \mathbf{P}^s . The scaling in (3.5) is performed. In this case, $\tilde{\mathbf{P}}^s = \alpha\mathbf{P}^s + (1 - \alpha)\mathbf{I}$, where $0 < \alpha < 1$. The scalar α may be chosen so that the largest negative element in magnitude along the diagonal of \mathbf{P}^s becomes zero after the scaling operation and $\tilde{\mathbf{P}}_{i,i}^s \geq \mathbf{0} \quad \forall i$.

$$(5.4) \quad \tilde{\mathbf{P}}_{i,j}^s\mathbf{e} = \alpha\mathbf{P}_{i,j}^s\mathbf{e} \Rightarrow k_{i,j}^s = \alpha k_{i,j} \Rightarrow 0 \leq k_{i,j}^s \leq k_{i,j} \quad \forall i \neq j,$$

$$(5.5) \quad \tilde{\mathbf{P}}_{i,i}^s\mathbf{e} = \alpha\mathbf{P}_{i,i}^s\mathbf{e} + (1 - \alpha)\mathbf{e} \Rightarrow k_{i,i}^s = \alpha k_{i,i} + (1 - \alpha) = \begin{cases} k_{i,i} + (1 - \alpha)(1 - k_{i,i}) \\ 1 - \alpha(1 - k_{i,i}) \end{cases}$$

$$\Rightarrow k_{i,i} < k_{i,i}^s < 1 \quad \forall i.$$

Combining the above two cases with the properties of NCD chains and noticing that \mathbf{C}^s is the $N \times N$ principal submatrix of \mathbf{K}^s , one obtains the statement in the lemma. Once again it must be remarked that if (3.5) is not performed, then case (i) applies and $c_{i,j}^s = \min(\mathbf{P}_{i,j}\mathbf{e}) \forall i, j$. \square

LEMMA 5.2. *Let \mathbf{P} be an NCD Markov chain with N blocks that is not exactly lumpable. Let \mathbf{C}^s be the $N \times N$ principal submatrix of the quasi-lumped chain \mathbf{K}^s corresponding to \mathbf{P} in (3.6). Then \mathbf{C}^s has entries that satisfy*

$$(5.6) \quad \sum_j c_{i,j}^s \leq 1 \quad \forall i$$

with strict inequality for at least one i .

Proof. For the case in which scaling is not performed, the proof is straightforward and follows from (5.3):

$$\sum_j c_{i,j}^s = \sum_j k_{i,j}^s = \sum_j k_{i,j} = \sum_j \min(\mathbf{P}_{i,j}\mathbf{e}) \leq 1 \quad \forall i.$$

The fact that there is strict inequality for at least one row of blocks is a consequence of \mathbf{P} not being exactly lumpable. That is, there is at least one row of blocks in \mathbf{P} in which one of the blocks has unequal row sums; otherwise, \mathbf{P} would be exactly lumpable. When scaling is performed, one obtains

$$\begin{aligned} \sum_j c_{i,j}^s &= k_{i,i}^s + \sum_{j \neq i} k_{i,j}^s = 1 - \alpha(1 - k_{i,i}) + \alpha \sum_{j \neq i} k_{i,j} = 1 - \alpha + \alpha \sum_j k_{i,j} \\ &= 1 - \alpha + \alpha \sum_j \min(\mathbf{P}_{i,j}\mathbf{e}) \leq 1 \quad \forall i \end{aligned}$$

from (5.4) and (5.5). The strict inequality for at least one i stems from the same reason. \square

The following theorem summarizes the properties of \mathbf{C}^s .

THEOREM 5.3. *Let \mathbf{P} be an NCD Markov chain with N blocks and coupling matrix \mathbf{C} . Assume that \mathbf{P} is not exactly lumpable. Let \mathbf{C}^s be the $N \times N$ principal submatrix of the quasi-lumped chain \mathbf{K}^s corresponding to \mathbf{P} in (3.6). Then*

- (i) \mathbf{C}^s is nonnegative;
- (ii) \mathbf{C}^s is row diagonally dominant;
- (iii) \mathbf{C}^s may be reducible (although \mathbf{C} is irreducible);
- (iv) if \mathbf{C}^s is irreducible or each row of blocks in \mathbf{P} is not exactly lumpable,

then $\mathbf{I} - \mathbf{C}^s$ is a nonsingular M-matrix;

- (v) if the scaling in (3.5) is not performed, $\mathbf{C}^s \leq \mathbf{C}$;

(vi) if the scaling in (3.5) is not performed and for some i, j $\mathbf{P}_{i,j}$ has equal row sums, then $c_{i,j} = c_{i,j}^s$.

Proof. Parts (i) and (ii) follow directly from Lemma 5.1. Although the coupling matrix of an NCD Markov chain is irreducible, \mathbf{C}^s may very well be a reducible matrix. The reason for this is implicit in equation (5.1). For a given diagonal element of \mathbf{C}^s , all off-diagonal elements in the same row may be zero. This is a sufficient condition and happens, for instance, if $\min(\mathbf{P}_{i,j}\mathbf{e}) = 0 \forall i \neq j$ for a given i , and part (iii) follows. Note that it is also possible for \mathbf{C}^s to be an irreducible matrix. For part (iv), let $\mathbf{A} = \mathbf{I} - \mathbf{C}^s$. To prove that \mathbf{A} is a nonsingular M-matrix [1], the following properties need to be shown (see [9, pp. 531–532]):

1. $a_{i,i} > 0 \forall i$ and $a_{i,j} \leq 0 \forall i \neq j$.

2. A is irreducible and $a_{i,i} \geq \sum_{j \neq i} |a_{i,j}| \forall i$ with strict inequality for at least one i , or $a_{i,i} > \sum_{j \neq i} |a_{i,j}| \forall i$.

Now,

$$0 < a_{i,i} \ll 1 \forall i \quad \text{and} \quad -1 \ll a_{i,j} \leq 0 \forall i \neq j$$

follow directly from Lemma 5.1; hence, the first property is verified. The second property amounts to showing that $\sum_j c_{i,j}^s < 1 \forall i$. As indicated in Lemma 5.2, this is not true in general. However, if \mathbf{C}^s is irreducible, then so is \mathbf{A} , and the second property is also satisfied due to Lemma 5.2. On the other hand, if strict inequality holds for each row of \mathbf{C}^s in Lemma 5.2 (i.e., \mathbf{A} is strictly row diagonally dominant), the irreducibility assumption for \mathbf{C}^s may be relaxed and the second property is once again satisfied. Note that this is the case if each row of blocks in \mathbf{P} possesses at least one block with unequal row sums, and therefore it is quite likely to happen. Finally, the nonsingularity is a direct consequence of condition (I₂₉) on p. 136 of [1]. Part (v) follows from Lemma 5.1. A sufficient condition for $\mathbf{C}^s \leq \mathbf{C}$ to be true is for \mathbf{P} to be diagonally dominant or for \mathbf{P} to have diagonal elements larger than the degree of coupling. Part (vi) may be shown by noticing that $c_{i,j}^s = \min(\mathbf{P}_{i,j}\mathbf{e})$ if $\mathbf{P}_{i,j}$ has equal row sums and scaling is not performed. Hence,

$$c_{i,j} = \frac{\pi_i}{\|\pi_i\|_1} \mathbf{P}_{i,j}\mathbf{e} = \frac{\pi_i}{\|\pi_i\|_1} c_{i,j}^s \mathbf{e} = c_{i,j}^s \frac{\pi_i \mathbf{e}}{\|\pi_i\|_1} = c_{i,j}^s \quad \forall i, j. \quad \square$$

COROLLARY 5.4. *Let \mathbf{P} be an NCD Markov chain with N blocks and coupling matrix \mathbf{C} . Then \mathbf{C}^l with entries*

$$(5.7) \quad c_{i,j}^l = \min(\mathbf{P}_{i,j}\mathbf{e}) \quad \forall i, j$$

is a nonnegative, lower-bounding matrix for \mathbf{C} and \mathbf{C}^u with entries

$$(5.8) \quad c_{i,j}^u = \max(\mathbf{P}_{i,j}\mathbf{e}) = \|\mathbf{P}_{i,j}\|_\infty \quad \forall i, j$$

is a nonnegative, upper-bounding matrix for \mathbf{C} .

That $\mathbf{C} \leq \mathbf{C}^u$ follows from

$$c_{i,j} = \frac{\pi_i}{\|\pi_i\|_1} \mathbf{P}_{i,j}\mathbf{e} \leq \max(\mathbf{P}_{i,j}\mathbf{e}) \quad \forall i, j,$$

where $\pi_i/\|\pi_i\|_1$ is a probability vector. Also note that \mathbf{C}^u is irreducible because \mathbf{P} is irreducible, whereas an analogous statement is not valid for \mathbf{C}^l .

Returning to the question posed at the beginning of this section, the answer is no, \mathbf{C}^s is not necessarily a lower-bounding matrix for \mathbf{C} , but \mathbf{C}^l is. Nevertheless, for the Courtois example $\mathbf{C}^s = \mathbf{C}^l$, and \mathbf{C}^s turns out to be a lower-bounding matrix for \mathbf{C} . Note that it is possible to subtract a slack probability mass from some other element (rather than the diagonal element) in the diagonal block and avoid the scaling in equation (3.5) (see the third step of construction in section 3) to have $\mathbf{C}^s = \mathbf{C}^l$. We use the definition of quasi lumpability in [5] to be consistent in terminology. The next section investigates the relation between the nonzero structure of a substochastic lower-bounding matrix for a given Markov chain and the nature of lower and upper bounds obtained on the chain's stationary probabilities.

6. Significance of the structure of lower-bounding matrices. Given an irreducible Markov chain \mathbf{P} and a substochastic lower-bounding matrix \mathbf{P}^* (i.e., $\mathbf{0} \leq \mathbf{P}^* \leq \mathbf{P}$, $\mathbf{P}^* \neq \mathbf{0}$), one can use Courtois and Semal's technique and compute lower and upper bounds on the stationary probabilities of \mathbf{P} . The question of interest is the following. What, if any, is the relation between the nonzero structure of \mathbf{P}^* and the bounds obtained? Analogously, the same question may be posed for the coupling matrix of an NCD Markov chain that is not exactly lumpable and a nonnegative lower-bounding coupling matrix \mathbf{C}^* (such as \mathbf{C}^l of (5.7)) (i.e., $\mathbf{0} \leq \mathbf{C}^* \leq \mathbf{C}$, $\mathbf{C}^* \neq \mathbf{0}$). In order to avoid introducing new symbols and complicating the terminology further, the equivalent second question is considered. That \mathbf{C}^l and the like have weighty diagonals is immaterial in the theory developed.

Observe that $\mathbf{C}^* \geq \mathbf{0}$, $c_{i,i}^* \neq 0 \forall i$, and $\mathbf{C}^* \mathbf{e} \neq \mathbf{e}$ for the matrices of interest by definition. The principles that govern the solution of the systems

$$(6.1) \quad \mathbf{z}_i \mathbf{K}_i^* = \mathbf{z}_i \quad \sum_{j=1}^N z_{i,j} = 1 \quad \forall s_i \in \mathcal{S}^* = \{s_1, s_2, \dots, s_N\},$$

where

$$(6.2) \quad \mathbf{K}_i^* = \begin{pmatrix} \mathbf{C}^* & \mathbf{e} - \mathbf{C}^* \mathbf{e} \\ \mathbf{e}_i^T & 0 \end{pmatrix},$$

are established next. Here \mathbf{K}_i^* is a stochastic matrix (i.e., $\mathbf{K}_i^* \mathbf{e} = \mathbf{e}$), \mathbf{z}_i is a probability vector (i.e., the i th row of the stochastic matrix \mathbf{Z}), \mathcal{S}^* represents the states of the lower-bounding nonnegative (coupling) matrix, and \mathbf{e}_i denotes the i th column of the identity matrix.

The discussion that follows refers to essential and nonessential (i.e., transient) states and to the concept of reducibility in nonnegative square matrices, as presented in pages 25–26 of [16]. Furthermore, for simplicity it is assumed that \mathbf{C}^* is already in the normal form of a reducible (i.e., decomposable) nonnegative matrix. However, that \mathbf{C}^* is in reducible normal form should not be understood to mean \mathbf{C}^* is reducible.

Following the terminology in [16], let K denote the number of mutually disjoint irreducible subsets of states in \mathbf{C}^* . Let these subsets be represented by $\mathcal{S}_1^{ir}, \mathcal{S}_2^{ir}, \dots, \mathcal{S}_K^{ir}$. Note that $\mathcal{S}_i^{ir} \cap \mathcal{S}_j^{ir} = \emptyset \forall i \neq j$. In any case, the states in $\mathcal{S}^{ir} (= \bigcup_i \mathcal{S}_i^{ir})$ are referred to as essential states. If $\mathcal{S}^{ir} = \mathcal{S}^*$, there would be no transient states in \mathbf{C}^* . Moreover, if $K = 1$, \mathbf{C}^* would be irreducible; else it could be decomposed into K mutually disjoint irreducible subsets of states. Hereafter, the possibility of having a stochastic transition probability submatrix (as part of \mathbf{C}^*) corresponding to any irreducible subset of states is overruled. That is, for each irreducible subset of states, the extra column in \mathbf{K}_i^* has at least one nonzero element. If not, the irreducible subset of states for which this property does not hold may be extracted from \mathbf{C}^* and analyzed separately. On the other hand, if $\mathcal{S}^{ir} \neq \mathcal{S}^*$, there would be transient states in \mathbf{C}^* . Similarly, let $\mathcal{S}_1^{tr}, \mathcal{S}_2^{tr}, \dots, \mathcal{S}_M^{tr}$ represent the transient subsets of states, where M is the number of transient subsets of states in \mathbf{C}^* subject to the constraints $\mathcal{S}_i^{tr} \cap \mathcal{S}_j^{tr} = \emptyset \forall i \neq j$. Moreover, the mutually disjoint transient subsets of states should be ordered so that there are no transitions from \mathcal{S}_i^{tr} to \mathcal{S}_j^{tr} in $\mathcal{S}^{tr} (= \bigcup_i \mathcal{S}_i^{tr}) \forall i < j$. However, there must be a transition from a given \mathcal{S}_i^{tr} to at least one \mathcal{S}_k^{tr} for $1 \leq k < i \leq M$ or to at least one \mathcal{S}_l^{ir} for $1 \leq l \leq K$.

The following 9×9 lower-bounding nonnegative (coupling) matrix for an (NCD) Markov chain demonstrates the concepts introduced in this section.

Example 6.1. Let

$$\mathbf{C}^* = \begin{pmatrix}
 \begin{array}{c|ccc|ccc|ccc}
 0.999 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 \hline
 0 & 0.995 & 0.005 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0.002 & 0.997 & 0 & 0 & 0 & 0 & 0 & 0 \\
 \hline
 0 & 0 & 0 & 0.998 & 0.001 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0.997 & 0.003 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0.002 & 0 & 0.998 & 0 & 0 & 0 \\
 \hline
 0 & 0.001 & 0 & 0 & 0 & 0 & 0.997 & 0.002 & 0 \\
 0 & 0.002 & 0.002 & 0 & 0 & 0 & 0.001 & 0.995 & 0 \\
 \hline
 0.001 & 0 & 0 & 0 & 0 & 0.001 & 0 & 0.001 & 0.996
 \end{array}
 \end{pmatrix}.$$

For this matrix, $\mathcal{S}^{ir} = \{s_1, s_2, \dots, s_6\}$ and $\mathcal{S}^{tr} = \{s_7, s_8, s_9\}$ with $K = 3$, $M = 2$, $\mathcal{S}_1^{ir} = \{s_1\}$, $\mathcal{S}_2^{ir} = \{s_2, s_3\}$, $\mathcal{S}_3^{ir} = \{s_4, s_5, s_6\}$, $\mathcal{S}_1^{tr} = \{s_7, s_8\}$, $\mathcal{S}_2^{tr} = \{s_9\}$. Since \mathbf{C}^* is in reducible normal form, each diagonal block in \mathbf{C}^* is (and should be) irreducible. By the same token, the first transient subset of states, \mathcal{S}_1^{tr} , always has a transition to an irreducible subset of states from which the extra state is accessible. Therefore, by induction all transient subsets of states can access the extra state. For this example, the nonzero structure of \mathbf{Z} in (6.1), (6.2) is given by the following matrix in which an X represents a nonzero entry:

$$\begin{pmatrix}
 \begin{array}{c|ccc|ccc|ccc}
 X & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 \hline
 0 & X & X & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & X & X & 0 & 0 & 0 & 0 & 0 & 0 \\
 \hline
 0 & 0 & 0 & X & X & X & 0 & 0 & 0 \\
 0 & 0 & 0 & X & X & X & 0 & 0 & 0 \\
 0 & 0 & 0 & X & X & X & 0 & 0 & 0 \\
 \hline
 0 & X & X & 0 & 0 & 0 & X & X & 0 \\
 0 & X & X & 0 & 0 & 0 & X & X & 0 \\
 \hline
 X & X & X & X & X & X & X & X & X
 \end{array}
 \end{pmatrix}.$$

The following theorems summarize these observations, enabling one to forecast the nonzero structure of \mathbf{Z} for a given \mathbf{C}^* . It should be emphasized once more that each irreducible subset of states in the lower-bounding nonnegative matrices of interest should have a transition to the extra state and that the original Markov chain should not be exactly lumpable. Under these conditions, one may state the following theorems, which are valid a fortiori for an NCD Markov chain with coupling matrix \mathbf{C} such that $\mathbf{C}^* \leq \mathbf{C}$ and \mathbf{C}^* is substochastic.

THEOREM 6.2. *Let \mathbf{C}^* be a substochastic matrix. If \mathbf{C}^* is irreducible, then \mathbf{Z} given by (6.1), (6.2) is positive.*

Proof. Since $\mathbf{C}^* \mathbf{e} \neq \mathbf{e}$, there is at least one row in \mathbf{C}^* , say k , for which $(\mathbf{e} - \mathbf{C}^* \mathbf{e})_k > 0$. All states in \mathbf{C}^* form a single communicating class and the extra state in \mathbf{K}_i^* (see (6.2)) is accessible from at least one of the states in \mathbf{C}^* . Hence, \mathbf{K}_i^* is irreducible for each i , and the theorem follows. \square

Note that when \mathbf{C}^* is irreducible, $\mathcal{S}^{ir} = \mathcal{S}^*$, $K = 1$, and there are no transient states in \mathcal{S}^* . Furthermore, under the stated conditions $\mathbf{I} - \mathbf{C}^*$ is a nonsingular M-matrix.

In the statement of the following theorem, a substochastic state means a state for which the corresponding row sum is less than one.

THEOREM 6.3. *Let \mathbf{C}^* be a substochastic matrix, and let $\mathcal{S}^* = \mathcal{S}^{ir} \cup \mathcal{S}^{tr}$, $\mathcal{S}^{ir} = \bigcup_{i=1}^K \mathcal{S}_i^{ir}$, $\mathcal{S}^{tr} = \bigcup_{i=1}^M \mathcal{S}_i^{tr}$ be the state space partition of \mathbf{C}^* , where K is the number of*

disjoint irreducible subsets of states and M is the number of disjoint transient subsets of states. If \mathbf{C}^* is reducible and each irreducible subset of states in \mathbf{C}^* has at least one substochastic state, then

- (i) if s_i is an essential state and $s_i \in \mathcal{S}_k^{ir}$ for some k , then $z_{i,j} > 0$ for all $s_j \in \mathcal{S}_k^{ir}$ and $z_{i,j} = 0$ for all $s_j \notin \mathcal{S}_k^{ir}$;
- (ii) if s_i is a transient state and $s_i \in \mathcal{S}_k^{tr}$ for some k , then $z_{i,j} > 0$ for all $s_j \in (\mathcal{S}_k^{tr}$ and states accessible from $\mathcal{S}_k^{tr})$; otherwise $z_{i,j} = 0$.

Proof. Part (i) follows from the fact that the extra state is accessible from \mathcal{S}_k^{ir} , of which s_i is a member, and the last row of \mathbf{K}_i^* has a one at the i th column position in (6.2), thereby making \mathcal{S}_k^{ir} with the extra state an irreducible stochastic submatrix in \mathbf{K}_i^* . Hence, \mathbf{z}_i in (6.1) has nonzero entries only in locations corresponding to the members of \mathcal{S}_k^{ir} . Part (ii) follows from the fact that the extra state is accessible from \mathcal{S}_k^{tr} (of which s_i is a member) and all other subsets of states accessible from \mathcal{S}_k^{tr} . Hence, states in \mathcal{S}_k^{tr} and states accessible from \mathcal{S}_k^{tr} together with the extra state form an irreducible stochastic submatrix in \mathbf{K}_i^* . Again, \mathbf{z}_i has nonzero entries only in locations corresponding to the members of \mathcal{S}_k^{tr} and other states they access. \square

COROLLARY 6.4. *If the substochastic matrix \mathbf{C}^* is reducible and the k th irreducible subset of states \mathcal{S}_k^{ir} is a singleton with a substochastic state (i.e., $\mathcal{S}_k^{ir} = \{s_i\}$, $s_i \in \mathcal{S}^{ir}$), then $z_{i,j} = \delta_{i,j}$.*

Corollary 6.4 helps to identify those states for which the lower and upper bounds obtained by Courtois and Semal's technique will be 0 and 1, respectively. Such states do not contribute to the tightening of the bounds of other states. Hence, if these states are identified in advance, they may be extracted from the lower-bounding matrix, thereby reducing the size of the systems to be solved in (6.1) and (6.2).

Before stating the next corollary, we recall the definition of a reachability (or accessibility) matrix. The reachability matrix of a square matrix is constructed as follows. First, the given square matrix is represented as a directed graph. The graph must have a directed arc for each nonzero entry in the original matrix. Then a new matrix is formed whose i, j th entry is a one (zero) if and only if state j is accessible (inaccessible) from state i on the directed graph. The newly formed matrix is the reachability matrix corresponding to the original square matrix.

COROLLARY 6.5. *If each irreducible subset of states in the substochastic matrix \mathbf{C}^* has at least one substochastic state, then the nonzero structure of \mathbf{Z} in (6.1), (6.2) is identical to the nonzero structure of the reachability matrix of \mathbf{C}^* .*

Corollary 6.5 helps one to forecast the nonzero structure of \mathbf{Z} by inspecting the nonzero structure of the lower-bounding matrix; that is, one does not need to solve N systems to find out what the nonzero structure of \mathbf{Z} looks like.

A result of Theorem 6.3 and Corollaries 6.4 and 6.5 (with (4.2) and (4.3)) is that a reducible lower-bounding nonnegative matrix gives lower (upper) bounds of zero (one) for various stationary probabilities of the coupling matrix and therefore indirectly causes other stationary probabilities to be loosely bounded. In conclusion, reducible lower-bounding nonnegative matrices should be avoided whenever possible.

7. Conclusion. This paper shows that NCD Markov chains are quasi-lumpable (if not lumpable). In most cases, \mathbf{C}^s , the $N \times N$ principal submatrix of the quasi-lumped chain turns out to be a lower-bounding coupling matrix for an NCD chain with N NCD blocks. When \mathbf{C}^s is a lower-bounding coupling matrix, it may be used to compute lower and upper bounds for the stationary probabilities of the NCD blocks. If \mathbf{C}^s is not a lower-bounding coupling matrix, \mathbf{C}^l , which is guaranteed to be a lower-bounding coupling matrix, may be used instead. Bounding the station-

ary probabilities of NCD blocks from below and from above amounts to solving at most N , $(N + 1) \times (N + 1)$, systems. These linear systems differ only in the last row. Therefore, only one **LU** decomposition needs to be performed. Assuming that the transposed systems of equations are solved, the upper-triangular matrices will be different in the last columns only. Hence, the last column in each of these systems needs to be treated separately during the triangularization phase. Thereafter, all back substitutions may be performed in parallel. Consequently, a solution method such as Gaussian elimination has a time complexity of $O(N^3)$ in the computation of the bounds.

If the NCD Markov chain is sparse with symmetries in its nonzero structure, it is quite likely that some elements of the unknown vector \mathbf{x} in the quasi-lumped chain will turn out to be zero, thus tightening the bounds further as in the Courtois matrix. The more information one has regarding the distribution of the probability mass in \mathbf{x}^T , the tighter the lower and upper bounds become. In fact, there is a distribution \mathbf{x}^T which gives the stationary probability of being in each NCD block exactly to working precision. However, although ϵ is always less than or equal to the degree of coupling of the NCD Markov chain, the lower-bounding nonnegative coupling matrix will have diagonal elements close to one, and it seems that the bounds obtained by the procedure generally will not be tight. The ill-conditioned nature of NCD Markov chains is once again noticed, but this time from a different perspective.

Furthermore, when choosing lower-bounding nonnegative matrices for Markov chains, one should be on the lookout for irreducible matrices. Reducible matrices should be avoided whenever possible because they provide lower (upper) bounds of zero (one) for various stationary probabilities, thereby indirectly causing other stationary probabilities to be loosely bounded.

Acknowledgments. The authors wish to thank the referees for their remarks which led to improvements in the manuscript.

REFERENCES

- [1] A. BERMAN AND R. J. PLEMMONS, *Nonnegative Matrices in the Mathematical Sciences*, SIAM, Philadelphia, PA, 1994.
- [2] W. L. CAO AND W. J. STEWART, *Iterative aggregation/disaggregation techniques for nearly uncoupled Markov chains*, *J. Assoc. Comput. Mach.*, 32 (1985), pp. 702–719.
- [3] P.-J. COURTOIS, *Decomposability: Queueing and Computer System Applications*, Academic Press, New York, 1977.
- [4] P.-J. COURTOIS AND P. SEMAL, *Bounds for the positive eigenvectors of nonnegative matrices and for their approximations by decomposition*, *J. Assoc. Comput. Mach.*, 31 (1984), pp. 804–825.
- [5] G. FRANCESCHINIS AND R. R. MUNTZ, *Bounds for quasi-lumpable Markov chains*, *Performance Evaluation*, 20 (1994), pp. 223–243.
- [6] W. J. HARROD AND R. J. PLEMMONS, *Comparison of some direct methods for computing the stationary distributions of Markov chains*, *SIAM J. Sci. Comput.*, 5 (1984), pp. 453–469.
- [7] J. R. KEMENY AND J. L. SNELL, *Finite Markov Chains*, Van Nostrand, New York, 1960.
- [8] J. R. KOURY, D. F. MCALLISTER, AND W. J. STEWART, *Iterative methods for computing stationary distributions of nearly completely decomposable Markov chains*, *SIAM J. Alg. Disc. Meth.*, 5 (1984), pp. 164–186.
- [9] P. LANCASTER AND M. TISMENETSKY, *The Theory of Matrices*, Academic Press, New York, 1985.
- [10] C. D. MEYER, *Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems*, *SIAM Rev.*, 31 (1989), pp. 240–272.
- [11] C. D. MEYER, *Sensitivity of the stationary distribution of a Markov chain*, *SIAM J. Matrix Anal. Appl.*, 15 (1994), pp. 715–728.

- [12] P. J. SCHWEITZER, *A survey of aggregation–disaggregation in large Markov chains*, in Numerical Solution of Markov Chains, W. J. Stewart, ed., Marcel Dekker, New York, 1991, pp. 63–88.
- [13] P. SEMAL, *Analysis of Large Markov Models, Bounding Techniques and Applications*, Doctoral Thesis, Université Catholique de Louvain, Belgium, 1992.
- [14] G. W. STEWART, W. J. STEWART AND D. F. MCALLISTER, *A two–stage iteration for solving nearly completely decomposable Markov chains*, in IMA Volumes in Mathematics and its Applications 60: Recent Advances in Iterative Methods, G. H. Golub, A. Greenbaum, and M. Luskin, eds. Springer–Verlag, New York, 1994, pp. 201–216.
- [15] W. J. STEWART AND W. WU, *Numerical experiments with iteration and aggregation for Markov chains*, ORSA J. Comput., 4 (1992), pp. 336–350.
- [16] W. J. STEWART, *Introduction to the Numerical Solution of Markov Chains*, Princeton University Press, Princeton, NJ, 1994.