



HAL
open science

Weak lumpability and pseudo-stationarity of finite Markov chains

James Ledoux, Patrice Leguesdron

► **To cite this version:**

James Ledoux, Patrice Leguesdron. Weak lumpability and pseudo-stationarity of finite Markov chains. Stochastic Models, 2000, 16 (1), pp.49-67. hal-00852730

HAL Id: hal-00852730

<https://hal.science/hal-00852730>

Submitted on 21 Aug 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

WEAK LUMPABILITY AND PSEUDO-STATIONARITY OF FINITE MARKOV CHAINS

James LEDOUX, Patrice LEGUESDRON

Institut National des Sciences Appliquées
20 Avenue des Buttes de Cosmes
35043 Rennes Cedex, FRANCE
jledoux,pleguesdr@insa-rennes.fr

13 April 1999

ABSTRACT

We consider the question of whether a function of a finite-state Markov chain is also Markovian, that is whether the chain is lumpable with respect to the partition determined by the function. We explore how an initial distribution with respect to which the chain is lumpable may differ from a pseudo-stationary initial distribution. Our results give insight into Peng's condition under which the Chapman-Kolmogorov equation implies that the lumped chain is Markovian. We illustrate these ideas by treating the question of whether the absorption time of a finite-state absorbing Markov chain is geometric.

Keywords: State aggregation, Cone, Invariance of cones, Stochastic equivalence.

1 Introduction

Let us consider a homogeneous Markov chain $X = (X_n)_{n \in \mathbb{N}}$, over the state space $S = \{1, \dots, N\}$. Such a process is fully specified by its transition proba-

bility matrix (t.p.m.) P and its initial distribution α . Let $\mathcal{P} = \{C(1), \dots, C(\hat{N})\}$ be a fixed partition of S in $\hat{N} < N$ classes. We associate with the given chain X the *aggregated process* $\hat{X} = (\hat{X}_n)_{n \in \mathbb{N}}$ defined by

$$\forall l \in \hat{S} = \{1, \dots, \hat{N}\} : \quad \hat{X}_n = l \iff X_n \in C(l), \quad \text{for any integer } n. \quad (1.1)$$

It is well known that \hat{X} is not Markovian in the general case (e.g. Kemeny and Snell [6]). In particular the Markov property of \hat{X} may depend on the selected initial distribution α for X . In this paper, we consider the set of distributions α such that the Markov chain characterized by (α, P) leads to an aggregated homogeneous Markov chain \hat{X} on the state space \hat{S} . This set will be denoted by $\mathcal{A}_{\mathcal{M}}$. If $\mathcal{A}_{\mathcal{M}}$ is not empty, we say that the family of Markov chains sharing the same t.p.m. P is *weakly lumpable*. A particular case occurs when every distribution leads to an aggregated Markov chain: the family is said to be *strongly lumpable* or *lumpable* (a complete characterization is given by Theorem 2.4 and Remark 4 in Ledoux [8]). The work presented in this paper is mainly motivated by the results reported in Peng [11] where matrix P was assumed to be irreducible. The stationary distribution π associated with such a matrix P allows to define, for each state class $C(l) \in \mathcal{P}$, the conditional stationary distribution given $C(l)$, that is the N -vector $\pi^{C(l)} = (\pi^{C(l)}(i))_{i \in S}$ with $\pi^{C(l)}(i) = \pi(i) / \sum_{j \in C(l)} \pi(j)$ if $i \in C(l)$ and 0 if $i \notin C(l)$. Peng shows that, under certain conditions, the set $\mathcal{A}_{\mathcal{M}}$ reduces to the bounded convex set

$$\mathcal{A}_{\pi} = \left\{ \sum_{l \in \hat{S}} \lambda_l \pi^{C(l)} \mid \forall l \in \hat{S}, \lambda_l \in [0, +\infty[\text{ and } \sum_{l \in \hat{S}} \lambda_l = 1 \right\}$$

In the case of a reducible chain, the stationary distribution is not available but it can be replaced with a positive vector, called the pseudo-stationary distribution, and also denoted by π . From this point of view, our paper is a natural continuation of the work carried out in [8] on the weak lumpability of reducible Markov chains. As shown in [8], the weak lumpability property of X implies that a Markovian aggregation according to the partition \mathcal{P} may be achieved for each initial distribution in the convex set \mathcal{A}_{π} . The aim of our paper is to explore how the set $\mathcal{A}_{\mathcal{M}}$ may differ from \mathcal{A}_{π} .

We adopt in Section 3 a somewhat different approach from Rubino and Sericola's [12] to characterize the set $\mathcal{A}_{\mathcal{M}}$. Indeed we directly use the necessary stochastic equivalence between the aggregated process and a homogeneous Markov chain with state space \hat{S} and t.p.m. \hat{P} where \hat{P} will be defined below (see relation (3.1), page 5). This allows us to deduce the fact that any probability distribution in $\mathcal{A}_{\mathcal{M}}$ differs from the pseudo-stationary distribution in a vector which contains no information which could help us to decide on the Markovian property of the aggregated process. In particular, we characterize the situation when the set $\mathcal{A}_{\mathcal{M}}$ is reduced to the convex set \mathcal{A}_{π} . In

Section 4, we discuss in our context Peng's condition to have $\mathcal{A}_{\mathcal{M}} = \mathcal{A}_{\pi}$ which is only a sufficient condition. Finally, in Section 5, we illustrate the stochastic equivalence and the cone invariance approaches by characterizing the set of all absorption time distributions of a finite-state absorbing Markov chain which are geometric.

The next section introduces notation and definitions used throughout this paper.

2 Notation and definitions

- The set of all probability distributions over S will be denoted by \mathcal{A} .
- By convention, vectors are row vectors. Entry i of a vector α is denoted by $\alpha(i)$. Column vectors are indicated by means of the transpose operator $(\cdot)^{\top}$. The vector with all its components equal to 1 (resp. 0) is denoted by $\mathbf{1}$ (resp. $\mathbf{0}$).
- Entry (i, j) of a matrix P is denoted by $P(i, j)$. For any $C_1, C_2 \subseteq S$, let $P_{C_1 C_2}$ be the sub-matrix of P given by $(P(i, j))_{i \in C_1, j \in C_2}$.
- The cardinality $|C(l)|$ of the class $C(l)$ is denoted by $n(l)$. We assume the states of S ordered such that $C(l) = \{n(1) + \dots + n(l-1) + 1, \dots, n(1) + \dots + n(l)\}$ for $1 \leq l \leq \hat{N}$ (with $n(0) = 0$).
- For any subset C of S and $\alpha \in \mathcal{A}$, if $\sum_{i \in C} \alpha(i) \neq 0$, α^C is the stochastic N -vector $(\alpha^C(i))_{i \in S}$ of conditional probabilities given C , that is $\alpha^C(i) = \alpha(i) / \sum_{j \in C} \alpha(j)$ if $i \in C$ and 0 if $i \notin C$.
- For any $l \in \hat{S}$, let I_l be the $N \times N$ matrix defined by $I_l(i, i) = 1$ if $i \in C(l)$ and $I_l(i, j) = 0$ otherwise.
- Let \hat{S}^* be the set of all finite sequences of elements from \hat{S} . Let \hat{S}_l^* be, for any $l \in \hat{S}$, its subset constituted of sequences with first element l .
- Let $s = l_0 l_1 \dots l_k$ be any sequence of elements of \hat{S} . The integer k will be denoted by $\lg(s)$. It has to be interpreted as the number of transitions in the lumped states $C(l_0), \dots, C(l_k)$.

We associate with the t.p.m. P over the state space S , the following $N \times N$ matrix

$$P^s = \begin{cases} I_{l_0} & \text{if } \lg(s) = 0, \\ I_{l_0} P I_{l_1} \dots I_{l_{k-1}} P I_{l_k} & \text{if } \lg(s) > 0. \end{cases}$$

For the t.p.m. \hat{P} over the state space \hat{S} , \hat{P}^s will denote the scalar

$$\prod_{i=1}^{\lg(s)} \hat{P}(l_{i-1}, l_i),$$

which is the transition probability from l_0 to l_k along the path $s = l_0 l_1 \dots l_k$ computed from \hat{P} . We adopt the convention that $\hat{P}^s = 1$ if $\lg(s) = 0$.

Let us consider the following example to clarify the notation. The set $\mathcal{P} = \{C(1) = \{1\}, C(2) = \{2, 3\}\}$ is a partition of the state space $S = \{1, 2, 3\}$. We thus have $\hat{N} = 2$ and $\hat{S} = \{1, 2\}$. Matrix P is

$$P = \left(\begin{array}{c|cc} 1/4 & 1/4 & 1/2 \\ \hline 0 & 1/6 & 5/6 \\ 7/8 & 1/8 & 0 \end{array} \right).$$

For example, if $s_1 = 12$, $s_2 = 121$, $s_3 = 122$ ($s_1, s_2, s_3 \in \hat{S}_1^*$) then we obtain

$$P^{s_1} = \left(\begin{array}{c|cc} 0 & 1/4 & 1/2 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right), P^{s_2} = \left(\begin{array}{c|cc} 7/16 & 0 & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right), P^{s_3} = \left(\begin{array}{c|cc} 0 & 5/48 & 5/24 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right).$$

Cones

The basic definitions of the cones are from Berman and Plemmons [2]. Throughout this subsection \mathcal{C} denotes a subset of \mathbb{R}^n . For any \mathcal{C} , $\text{Span}(\mathcal{C})$ refers to the linear hull of \mathcal{C} . $\text{Cone}(\mathcal{C})$ denotes the set of all finite nonnegative linear combinations of elements of \mathcal{C} . The elements of \mathcal{C} are called the generators of $\text{Cone}(\mathcal{C})$. If $\text{Cone}(\mathcal{C}) = \mathcal{C}$ then \mathcal{C} is called a *cone*. A finitely generated cone is called a *polyhedral cone*. A polyhedral cone is a closed convex set. The *dimension* $\dim(\mathcal{C})$ of a subset \mathcal{C} is the dimension of the affine hull of \mathcal{C} .

Definition 2.1 Let \mathcal{C}_1 and \mathcal{C}_2 be cones of \mathbb{R}^n . The cone \mathcal{C} is said to be the direct sum of \mathcal{C}_1 and \mathcal{C}_2 , which is denoted by $\mathcal{C} = \mathcal{C}_1 \oplus \mathcal{C}_2$, if $\text{Span}(\mathcal{C}_1) \cap \text{Span}(\mathcal{C}_2) = \{\mathbf{0}\}$ and $\mathcal{C} = \mathcal{C}_1 + \mathcal{C}_2$.

3 Weak lumpability and pseudo-stationarity

Let $(X_n)_{n \in \mathbb{N}}$ be a homogeneous Markov chain over the state space S , given by its t.p.m. P and its initial distribution α ; when necessary we denote it by (α, P) . We also denote the aggregated process $(\hat{X}_n)_{n \in \mathbb{N}}$ associated with (α, P) (see relation (1.1)) by $\text{agg}(\alpha, P, \mathcal{P})$.

We are concerned with reducible Markov chains. We assume that the states of S are ordered such that P is the lower block-triangular matrix

$$P = \left(\begin{array}{cccc} P_{I(1)I(1)} & 0 & \cdots & 0 \\ * & P_{I(2)I(2)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ * & \cdots & * & P_{I(K)I(K)} \end{array} \right)$$

where $\mathcal{I} = (I(k))_{k=1}^K$ is the usual partition of S in communicating classes or irreducibility classes. Henceforth, partition \mathcal{P} will be a refinement of the above partition \mathcal{I} : each class $C(l)$ ($l \in \hat{S}$) corresponds to a unique k such that $C(l) \subseteq I(k)$. Since the nonnegative matrix $P_{I(k)I(k)}$ is irreducible for each $k \in \{1, \dots, K\}$, there exists a unique stochastic left eigenvector v_k corresponding to the spectral radius of $P_{I(k)I(k)}$. We can form a unique N -vector \tilde{v}_k from the $|I(k)|$ -vector v_k by: the restriction $(\tilde{v}_k(i))_{i \in I(k)}$ of \tilde{v}_k to $I(k)$ is equal to v_k and its components outside $I(k)$ are 0. We define the following family of N -vectors $\{\pi_l, l \in \hat{S}\}$ by

$$\pi_l = \tilde{v}_k I_l \quad \text{if } k \text{ is such that } C(l) \subseteq I(k).$$

Consequently, the N -vector π defined by

$$\pi = \frac{1}{K} \sum_{l \in \hat{S}} \pi_l$$

is positive and stochastic. We have $\pi^{C(l)} = \pi_l^{C(l)}$ for every $l \in \hat{S}$. Vector π will be referred to as the *pseudo-stationary distribution* associated with the reducible matrix P . We can consider the following convex subsets of \mathbb{R}^N

$$\mathcal{C}_\pi = \bigoplus_{l \in \hat{S}} \text{Cone}(\pi_l), \quad \mathcal{A}_\pi = \mathcal{C}_\pi \cap \mathcal{A}.$$

Finally, we define the $\hat{N} \times \hat{N}$ stochastic matrix \hat{P} by

$$\forall l, m \in \hat{S}, \quad \hat{P}(l, m) = \pi_l^{C(l)} I_l P I_m \mathbf{1}^\top. \quad (3.1)$$

We are going to illustrate the above material with the following example.

Example

Let us consider the following partition $\mathcal{P} = \{C(1) = \{1\}, C(2) = \{2, 3\}, C(3) = \{4\}, C(4) = \{5, 6, 7\}\}$ of the state space $S = \{1, 2, 3, 4, 5, 6, 7\}$. The reducible t.p.m. P is

$$P = \left(\begin{array}{ccc|ccc} 1/4 & 1/4 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/6 & 5/6 & 0 & 0 & 0 & 0 \\ 7/8 & 1/8 & 0 & 0 & 0 & 0 & 0 \\ \hline 1/7 & 0 & 0 & 3/14 & 3/14 & 3/14 & 3/14 \\ \hline 1/8 & 1/24 & 0 & 1/6 & 1/6 & 1/6 & 1/3 \\ 1/12 & 0 & 0 & 1/8 & 3/8 & 1/4 & 1/6 \\ 0 & 0 & 1/12 & 3/8 & 1/8 & 1/4 & 1/6 \end{array} \right).$$

The partition \mathcal{I} is defined by the two communicating classes $I(1) = \{1, 2, 3\}$ and $I(2) = \{4, 5, 6, 7\}$. The stochastic left eigenvectors corresponding to the

spectral radius of respective matrices $P_{I(1)I(1)}$ and $P_{I(2)I(2)}$ are $v_1 = (7/16, 3/16, 6/16)$, $v_2 = (1/4, 1/4, 1/4, 1/4)$. The 7-vectors \tilde{v}_1 and \tilde{v}_2 associated with v_1 and v_2 are $\tilde{v}_1 = (7/16, 3/16, 6/16, 0, 0, 0, 0)$, $\tilde{v}_2 = (0, 0, 0, 1/4, 1/4, 1/4, 1/4)$. The vectors π_l ($l \in \hat{S} = \{1, 2, 3, 4\}$) are

$$\begin{aligned}\pi_1 &= (7/16, 0, 0, 0, 0, 0, 0) & \pi_2 &= (0, 3/16, 6/16, 0, 0, 0, 0), \\ \pi_3 &= (0, 0, 0, 1/4, 0, 0, 0) & \pi_4 &= (0, 0, 0, 0, 1/4, 1/4, 1/4)\end{aligned}$$

and allow us to consider as pseudo-stationary distribution

$$\pi = \frac{1}{2}(7/16, 3/16, 6/16, 1/4, 1/4, 1/4, 1/4).$$

Therefore, we obtain as polyhedral cone \mathcal{C}_π

$$\mathcal{C}_\pi = \left\{ \sum_{l=1}^4 \lambda_l \pi_l \mid \forall l \in \{1, 2, 3, 4\}, \lambda_l \in [0, +\infty[\right\}.$$

It follows that the bounded convex set \mathcal{A}_π is given by

$$\mathcal{A}_\pi = \left\{ \sum_{l=1}^4 \lambda_l \pi^{C(l)} \mid \forall l \in \{1, 2, 3, 4\}, \lambda_l \in [0, +\infty[\text{ and } \sum_{l=1}^4 \lambda_l = 1 \right\}$$

with

$$\begin{aligned}\pi^{C(1)} &= (1, 0, 0, 0, 0, 0, 0), & \pi^{C(2)} &= (0, 1/3, 2/3, 0, 0, 0, 0), \\ \pi^{C(3)} &= (0, 0, 0, 1, 0, 0, 0) & \pi^{C(4)} &= (0, 0, 0, 0, 1/3, 1/3, 1/3).\end{aligned}$$

The 4×4 t.p.m. \hat{P} over the state space \hat{S} is from relation (3.1)

$$\hat{P} = \begin{pmatrix} 1/4 & 3/4 & 0 & 0 \\ 7/12 & 5/12 & 0 & 0 \\ 1/7 & 0 & 3/14 & 9/14 \\ 5/72 & 1/24 & 2/9 & 2/3 \end{pmatrix}.$$

Remark 3.1 If we consider partitions of S which are not a refinement of \mathcal{I} , then new facts have to be taken into consideration. Indeed, when a class $C(l)$ lumps states of different irreducibility classes of a reducible chain, the aggregated process may be an irreducible Markov chain. In the same way, the t.p.m. of an aggregated Markov chain may depend on the initial distribution of the original one. Consequently, a complete description of weak lumpability becomes of limited interest (for more details, see the technical report associated with reference [8]).

The approach developed in [6], [12] consists of identifying the conditional probability $\mathbb{P}_\alpha(X_{n+1} \in C(m) | X_n \in C(l), \dots, X_0 \in C(l_0))$ defined for Markov chain (α, P) , as the state probability $\mathbb{P}_\beta(X_1 \in C(m))$ associated with Markov chain (β, P) . Vector β is the conditional distribution $(\mathbb{P}_\alpha(X_n = i | X_n \in C(l), \dots, X_0 \in C(l_0)))_{i \in S}$ which may depend on n, l_0, \dots, l . Consequently, $\text{agg}(\alpha, P, \mathcal{P})$ is a homogeneous Markov chain if such a probability $\mathbb{P}_\beta(X_1 \in C(m))$ depends only on l and m . If so, it is the transition probability from state l to m for the aggregated Markov chain. This approach has been extended to the present context in [8] and leads to the computation of solutions to linear systems.

We adopt a somewhat different approach which has the advantage of giving insight into the relationship between the pseudo-stationary distribution π and any distribution α such that $\text{agg}(\alpha, P, \mathcal{P})$ is a Markov chain. Let $\hat{\alpha}$ be the stochastic vector on \hat{S} defined by $\hat{\alpha}(l) = \sum_{i \in C(l)} \alpha(i) = \mathbb{P}_\alpha(X_0 \in C(l))$ for each $l \in \hat{S}$. The process $\text{agg}(\alpha, P, \mathcal{P})$ is a homogeneous Markov chain with t.p.m. \hat{P} if $\text{agg}(\alpha, P, \mathcal{P})$ and the Markov chain $(\hat{\alpha}, \hat{P})$ have the same finite dimensional joint distributions, that is for every $n \geq 0$ and for any $(l_0, \dots, l_n) \in \hat{S}^{n+1}$

$$\mathbb{P}_\alpha(X_0 \in C(l_0), \dots, X_n \in C(l_n)) = \hat{\alpha}(l_0) \hat{P}(l_0, l_1) \cdots \hat{P}(l_{n-1}, l_n).$$

The two processes are said to be stochastically equivalent (e.g. Gray [4]). This relation can be rewritten with the notation presented in Section 2 as

$$\forall l \in \hat{S}, \forall s \in \hat{S}_l^*, \quad \alpha P^s \mathbf{1}^\top = \hat{\alpha}(l) \hat{P}^s. \quad (3.2)$$

The concept of equivalence is heavily used in the finite automata theory and in the topic of hidden Markov chains (see Amari et al. [1] for a closely related problem). Some of our notation comes from this research area.

We consider the set

$$\mathcal{C}_M \stackrel{\text{def}}{=} \{ \alpha \geq \mathbf{0} \mid \forall l \in \hat{S}, \forall s \in \hat{S}_l^*, \quad \alpha P^s \mathbf{1}^\top = \hat{\alpha}(l) \hat{P}^s \}.$$

Note that $\mathcal{A}_M = \mathcal{C}_M \cap \mathcal{A}$ is the set of all probability distributions α such that $\text{agg}(\alpha, P, \mathcal{P})$ is a Markov chain with t.p.m. \hat{P} . The aim of this subsection is to analyze properties of set \mathcal{C}_M when it is not reduced to $\{\mathbf{0}\}$, or equivalently when $\mathcal{A}_M \neq \emptyset$. We recall a result from [8] which will be useful in the sequel.

Result 3.1 *If $\mathcal{C}_M I_l \neq \{\mathbf{0}\}$ for $l \in \hat{S}$, then $\mathcal{C}_M I_l$ is a polyhedral cone such that*

$$\pi_l \in \mathcal{C}_M I_l \subseteq \mathcal{C}_M \quad \text{and} \quad \forall m \in \hat{S}, \quad \mathcal{C}_M I_l P I_m \subseteq \mathcal{C}_M I_m.$$

The set \mathcal{C}_M is the polyhedral cone

$$\mathcal{C}_M = \bigoplus_{l \in \hat{S}} [\mathcal{C}_M I_l].$$

The following statement gives a sufficient condition to have weak lumpability

$$\mathcal{C}_\pi P \subseteq \mathcal{C}_\pi \implies \mathcal{C}_\pi \subseteq \mathcal{C}_M.$$

From Result 3.1, it is clear that, for any $l \in \hat{S}$, a non-trivial set $\mathcal{C}_{\mathcal{M}}I_l$ always includes vector π_l . In such a case, π_l satisfies relation (3.2) and it follows that

$$\forall s \in \hat{S}_l^*, \quad \hat{P}^s = \frac{\pi_l}{\pi_l \mathbf{1}^\top} P^s \mathbf{1}^\top = \pi_l^{C(l)} P^s \mathbf{1}^\top.$$

Therefore, relation (3.2) may be rewritten as: for each $l \in \hat{S}$

$$\begin{aligned} \alpha \in \mathcal{C}_{\mathcal{M}}I_l &\iff \forall s \in \hat{S}_l^*, \alpha P^s \mathbf{1}^\top = \hat{\alpha}(l) \pi_l^{C(l)} P^s \mathbf{1}^\top \text{ and } \alpha = \alpha I_l \\ &\iff \forall s \in \hat{S}_l^*, [\alpha - \hat{\alpha}(l) \pi_l^{C(l)}] P^s \mathbf{1}^\top = 0 \text{ and } \alpha = \alpha I_l. \end{aligned} \quad (3.3)$$

Let us now define the following linear subspace of \mathbb{R}^N

$$\mathcal{N} = \{ u \in \mathbb{R}^N \mid \forall s \in \hat{S}^*, u P^s \mathbf{1}^\top = 0 \}.$$

The set \mathcal{N} is the direct sum

$$\mathcal{N} = \bigoplus_{l \in \hat{S}} [\mathcal{N}I_l]$$

where $\mathcal{N}I_l = \{ u \in \mathbb{R}^N I_l \mid \forall s \in \hat{S}_l^*, u P^s \mathbf{1}^\top = 0 \}$. Note that the linear subspace $\mathcal{N}I_l$ is orthogonal to $\text{Span}(\mathbf{1})$ and $\mathcal{N}I_l \cap \mathbb{R}_+^N = \{\mathbf{0}\}$.

It is clear from (3.3) that $\alpha \in \mathcal{C}_{\mathcal{M}}I_l$ is equivalent to requiring $\alpha \in (\mathcal{N}I_l \oplus \text{Cone}(\pi_l)) \cap \mathbb{R}_+^N$. So the following theorem holds and states that $\mathcal{C}_{\mathcal{M}}I_l$ (for any $l \in \hat{S}$) differs from $\text{Cone}(\pi_l)$ in a set which contains no information as to the Markovian nature of the aggregated process.

Theorem 3.2 *For each $l \in \hat{S}$, let us set $k(l) = \dim(\mathcal{N}I_l)$. If $\mathcal{C}_{\mathcal{M}}I_l \neq \{\mathbf{0}\}$ then*

$$\mathcal{C}_{\mathcal{M}}I_l = (\mathcal{N}I_l \oplus \text{Cone}(\pi_l)) \cap \mathbb{R}_+^N \text{ and } \dim(\mathcal{C}_{\mathcal{M}}I_l) = k(l) + 1.$$

proof. It remains to show that $\dim(\mathcal{C}_{\mathcal{M}}I_l) = k(l) + 1$. We have $\mathcal{N}I_l \cap \mathbb{R}_+^N = \{\mathbf{0}\}$; $\mathcal{C}_{\mathcal{M}}I_l$ is also $(\mathcal{N}I_l \oplus \text{Span}(\pi_l)) \cap \mathbb{R}_+^N$, which allows us to write $\dim(\mathcal{C}_{\mathcal{M}}I_l) = \dim(\mathcal{N}I_l \oplus \text{Span}(\pi_l)) = k(l) + 1$. ■

Corollary 3.3 *For any $l \in \hat{S}$ such that $\mathcal{C}_{\mathcal{M}}I_l \neq \{\mathbf{0}\}$, we have*

$$\mathcal{C}_{\mathcal{M}}I_l = \text{Cone}(\pi_l) \iff \mathcal{N}I_l = \{\mathbf{0}\}.$$

In particular, we have

$$\mathcal{C}_{\mathcal{M}} = \mathcal{C}_\pi \iff [\mathcal{N} = \{\mathbf{0}\} \text{ and } \mathcal{C}_\pi P \subseteq \mathcal{C}_\pi].$$

proof. For any $l \in \hat{S}$ such that $\mathcal{C}_{\mathcal{M}}I_l \neq \{\mathbf{0}\}$, we have $\dim(\mathcal{C}_{\mathcal{M}}I_l) = k(l) + 1$ from Theorem 3.2. Now, if $\mathcal{C}_{\mathcal{M}}I_l = \text{Cone}(\pi_l)$ then $k(l) = \dim(\mathcal{N}I_l) = 0$. In such a case, the linear space $\mathcal{N}I_l$ reduces to $\{\mathbf{0}\}$. The converse implication is obvious.

The equality $\mathcal{C}_{\mathcal{M}} = \mathcal{C}_{\pi}$ says that $\mathcal{C}_{\mathcal{M}}I_l = \text{Cone}(\pi_l)$ for every $l \in \hat{S}$. Therefore the first part gives $\mathcal{N} = \bigoplus_{l \in \hat{S}} \mathcal{N}I_l = \{\mathbf{0}\}$ and $\mathcal{C}_{\pi}P \subseteq \mathcal{C}_{\pi}$ follows from the invariance of $\mathcal{C}_{\mathcal{M}}$ under matrix P given by Result 3.1. Conversely, if $\mathcal{C}_{\pi}P \subseteq \mathcal{C}_{\pi}$, it follows from Result 3.1 that $\mathcal{C}_{\pi} \subseteq \mathcal{C}_{\mathcal{M}}$ and $\mathcal{C}_{\mathcal{M}}I_l \neq \{\mathbf{0}\}$ for all l . Finally $\mathcal{N} = \{\mathbf{0}\}$ or $\mathcal{N}I_l = \{\mathbf{0}\}$ for all $l \in \hat{S}$ gives $\mathcal{C}_{\mathcal{M}}I_l = \text{Cone}(\pi_l)$ for all $l \in \hat{S}$ with the first part of the proof. ■

Remark 3.2 Let us say that any class $I(k) \in \mathcal{I}$ corresponds to $\uplus_{l \in L_k} C(l)$ for some subset L_k of \hat{S} . It should be pointed out that if $\mathcal{C}_{\mathcal{M}}I_l \neq \{\mathbf{0}\}$ for some $l \in L_k$, then $\mathcal{C}_{\mathcal{M}}I_m \neq \{\mathbf{0}\}$ for all $m \in L_k$. Indeed, for any states $i \in C(l)$ and $j \in C(m)$, there exists an integer n such that $P^n(i, j) > 0$ since matrix $P_{I(k)I(k)}$ is irreducible. Consequently, the vector $\pi_l P^n I_m$ is nonzero (because $(\pi_l(i))_{i \in C(l)} > \mathbf{0}$) and it belongs to $\mathcal{C}_{\mathcal{M}}I_m$ with the decomposition and invariance properties given by Result 3.1 ($\pi_l \in \mathcal{C}_{\mathcal{M}}$, $\pi_l P^n \in \mathcal{C}_{\mathcal{M}}$ and $\pi_l P^n I_m \in \mathcal{C}_{\mathcal{M}}I_m$).

Example continued

We compute the set \mathcal{N} associated with the 7×7 matrix P from the numerical example. If the vector u is in \mathcal{N} , it must satisfy the conditions $uI_1\mathbf{1}^T = uI_2\mathbf{1}^T = uI_3\mathbf{1}^T = uI_4\mathbf{1}^T = 0$ which are equivalent to

$$u(1) = 0, \quad (3.4)$$

$$u(2) + u(3) = 0, \quad (3.5)$$

$$u(4) = 0, \quad (3.6)$$

$$u(5) + u(6) + u(7) = 0. \quad (3.7)$$

It follows from equations (3.4) and (3.6) that $\mathcal{N}I_1 = \mathcal{N}I_3 = \{\mathbf{0}\}$. Now, vector u must also verify $uI_2PI_2\mathbf{1}^T = 0$ or $8u(2) + u(3) = 0$. It becomes clear that the only solution to the previous equation associated with (3.5) is $u(2) = u(3) = 0$, that is $\mathcal{N}I_2 = \{\mathbf{0}\}$. Note that $uI_4PI_1\mathbf{1}^T = uI_4PI_2\mathbf{1}^T = 0$ is equivalent to $3u(5) + 2u(6) = u(5) + 2u(7) = 0$, which, once combined with equation (3.7), gives $u(5) = u(6) = u(7) = 0$ or $\mathcal{N}I_4 = \{\mathbf{0}\}$. Finally the set \mathcal{N} is reduced to $\{\mathbf{0}\}$. Corollary 3.3 states that if $\mathcal{C}_{\mathcal{M}}I_l$ ($l = 1, 2, 3, 4$) are different from $\{\mathbf{0}\}$ then they are respectively $\text{Cone}(\pi_l)$ ($l = 1, 2, 3, 4$). We can easily check that

$$\mathcal{C}_{\pi}P \subseteq \mathcal{C}_{\pi}.$$

We conclude that $\mathcal{C}_{\mathcal{M}} = \mathcal{C}_{\pi}$ according to Corollary 3.3.

Remark 3.3 The following example shows that condition $\mathcal{N} = \{\mathbf{0}\}$ on the right hand side of the last equivalence of the Corollary 3.3 cannot be dropped. Indeed, for the partition $\mathcal{P} = \{C(1) = \{1\}, C(2) = \{2, 3\}\}$ and matrix P given by

$$P = \left(\begin{array}{c|cc} 1/2 & 1/4 & 1/4 \\ \hline 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \end{array} \right),$$

we have $\mathcal{C}_\pi P \subseteq \mathcal{C}_\pi = \text{Cone}((1, 0, 0); (0, 1, 1))$; but any initial distribution $\alpha \in \mathcal{A}$ is such that $\text{agg}(\alpha, P, \mathcal{P})$ is a Markov chain, that is $\mathcal{C}_\pi \subset \mathcal{C}_\mathcal{M} = \mathbb{R}_+^3$ (see characterization of strong lumpability in [6] or in [8]).

For any fixed $l \in \hat{S}$, let us define the following condition (R_l) that will be related, in the next section, to Peng's condition for an irreducible matrix P :

$$(R_l) : \text{Span}(I_l P^k I_m \mathbf{1}^\top, m \in \hat{S}, k \geq 0) = \mathbb{R}^N I_l.$$

We just prove that condition (R_l) is sufficient to have $\mathcal{N}I_l = \{\mathbf{0}\}$ and therefore, in such a case, the set $\mathcal{C}_\mathcal{M}I_l$ will be $\text{Cone}(\pi_l)$ if it does not reduce to $\{\mathbf{0}\}$.

Lemma 3.4 *If condition (R_l) holds then we have $\mathcal{N}I_l = \{\mathbf{0}\}$.*

proof. Let u be any vector of $\mathcal{N}I_l \subseteq \mathcal{N}$. We have by definition $uI_l = u$ and $uP^s \mathbf{1}^\top = 0$ for all $s \in \hat{S}^*$. Note that $I_j I_m = \mathbf{0}$ for $j \neq m$ and I_m otherwise. Consequently, for any $m \in \hat{S}$,

$$\forall s \in \hat{S}^*, uI_l P^s I_m \mathbf{1}^\top = 0.$$

Finally, for every $m \in \hat{S}$, we can write for any $k \in \mathbb{N}$

$$\begin{aligned} \sum_{s \in \hat{S}^*, \text{lg}(s)=k} uI_l P^s I_m \mathbf{1}^\top = 0 &\iff uI_l \left[\sum_{s \in \hat{S}^*, \text{lg}(s)=k} P^s \right] I_m \mathbf{1}^\top = 0 \\ &\iff uI_l P^k I_m \mathbf{1}^\top = 0 \\ &\iff u[I_l P^k I_m \mathbf{1}^\top] = 0. \end{aligned}$$

We deduce from the equality $\text{Span}(I_l P^k I_m \mathbf{1}^\top, k \geq 0, m \in \hat{S}) = \mathbb{R}^N I_l$ that the $n(l)$ -vector $(u(i))_{i \in C(l)}$ is in the orthogonal complement of $\mathbb{R}^{n(l)}$, hence $u = uI_l = \mathbf{0}$. So, the set $\mathcal{N}I_l$ reduces to $\{\mathbf{0}\}$. \blacksquare

If the condition (R_l) is fulfilled for $l = 1, \dots, Q$ then Lemma 3.4 states that $\bigoplus_{l=1}^Q \mathcal{N}I_l = \{\mathbf{0}\}$. Consequently, by Corollary 3.3, for any $l = 1, \dots, Q$ each "component" $\mathcal{C}_\mathcal{M}I_l$ of $\mathcal{C}_\mathcal{M}$ not reduced to $\{\mathbf{0}\}$ is necessarily equal to $\text{Cone}(\pi_l)$.

4 Connection with Peng's work

In this section, we confine ourselves to the irreducible case to make the correspondence with Peng's work easier. However, it is not difficult to see that all forthcoming results may be stated for any reducible t.p.m. P as long as the partition \mathcal{P} is a refinement of the usual partition \mathcal{I} (see Remark 3.2 and [8]). For an irreducible matrix P , vector π is the usual stationary distribution of the Markov chain, that is the unique stochastic vector solution to $\pi P = \pi$.

Peng studies in [11] the connection between the homogeneous Markov property and the Chapman-Kolmogorov equation. That was also the purpose of Burke and Rosenblatt [3] and Hachigan [5] under various assumptions on the t.p.m. P and the initial distribution. The transition probabilities of process $\text{agg}(\pi, P, \mathcal{P})$ satisfy the Chapman-Kolmogorov equation if

$$\forall k \geq 1, \quad \left[\mathbb{P}_\pi(X_k \in C(m) \mid X_0 \in C(l)) = \widehat{P}^k(l, m) \quad \forall l, m \in \widehat{S} \right]. \quad (4.1)$$

Peng gives a condition under which relation (4.1) implies that the process $\text{agg}(\pi, P, \mathcal{P})$ is Markovian and $\mathcal{C}_M = \mathcal{C}_\pi$. We briefly restate the main results from [11] in terms of cones. First of all, we recall the condition used by Peng.

(Γ_Q): Let $Q \leq \widehat{N}$ be an integer and $q = n(1) + \dots + n(Q) \leq N$. The first q elements of the column vectors of matrices $P^k V$ ($k \geq 0$) span \mathbb{R}^q

where V is the $N \times \widehat{N}$ matrix defined by $V(i, l) = 1$ if $i \in C(l)$ and 0 otherwise. The first result states that, under condition (Γ_Q), the first Q projections $\{\mathcal{C}_M I_l \mid l = 1, \dots, Q\}$ of set \mathcal{C}_M depend only on π and \mathcal{P} .

Result 4.1 (Th 3.2, [11]) *Under condition (Γ_Q),*

$$\mathcal{C}_M \neq \{0\} \implies \mathcal{C}_M = \bigoplus_{l=1}^Q \text{Cone}(\pi_l) \oplus_{l=Q+1}^{\widehat{N}} \mathcal{C}_l$$

where unknown cone \mathcal{C}_l is such that $\mathcal{C}_l I_l = \mathcal{C}_l$.

In the borderline case where $Q = \widehat{N}$, we have the following equivalent statements.

Result 4.2 (Th 3.1, [11]) *Under condition ($\Gamma_{\widehat{N}}$), the following are equivalent:*

- (a) *the transition probabilities of $\text{agg}(\pi, P, \mathcal{P})$ satisfy the Chapman-Kolmogorov equation;*
- (b) $\mathcal{C}_M = \mathcal{C}_\pi$.

First, we rely on condition (Γ_Q) to conditions (R_l) $l = 1, \dots, Q$ (defined before Lemma 3.4). Indeed, the m th column vector ($1 \leq m \leq \widehat{N}$) of the $N \times \widehat{N}$ matrix $P^k V$ ($k \geq 0$) corresponds to $P^k I_m \mathbf{1}^\top$. Next, the first $q = n(1) + \dots + n(Q)$ entries of the m th column vector of $P^k V$ are also the q first entries of vector $\sum_{l=1}^Q I_l P^k I_m \mathbf{1}^\top$. Consequently, condition (Γ_Q) is the same as (R_l) with $l = 1, \dots, Q$ or

$$\bigoplus_{l=1}^Q \text{Span}(I_l P^k I_m \mathbf{1}^\top, m \in \widehat{S}, k \geq 0) = \bigoplus_{l=1}^Q \mathbb{R}^N I_l.$$

Now, Result 4.1 follows from Result 3.1, Lemma 3.4 and Corollary 3.3.

Let us state the connection between the Markov property and the Chapman-Kolmogorov equation under condition $(\Gamma_{\widehat{N}})$ given by Result 4.2. We have the following equivalent formulation of Result 4.2 (a) from relation (4.1)

$$\forall k \geq 0, (U_\pi P - U_\pi P V U_\pi) P^k V = \mathbf{0}$$

where the l th row of the $\widehat{N} \times N$ matrix U_π is vector $\pi^{C(l)}$. It is now clear that the equality $U_\pi P = U_\pi P V U_\pi$ will follow from the condition $\text{Span}(P^k V, k \geq 0) = \mathbb{R}^N$. Equality $U_\pi P = U_\pi P V U_\pi$ gives $\mathcal{C}_\pi P \subseteq \mathcal{C}_\pi$ and we obtain from Result 3.1 that $\mathcal{C}_\pi \subseteq \mathcal{C}_\mathcal{M}$. Result 4.2 (b), or $\mathcal{C}_\mathcal{M} = \mathcal{C}_\pi$, follows from Lemma 3.4.

Finally, note that condition $(\Gamma_{\widehat{N}})$ is not necessary to derive Result 4.2 (b) from $\mathcal{C}_\pi P \subseteq \mathcal{C}_\pi$. For instance, let us consider the following irreducible matrix

$$P = \left(\begin{array}{cc|cc} 1/6 & 1/6 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/6 & 1/6 \\ \hline 1/6 & 1/6 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/6 & 1/6 \end{array} \right)$$

with partition $\mathcal{P} = \{C(1) = \{1, 2\}, C(2) = \{3, 4\}\}$. The stationary distribution of this doubly stochastic matrix is $\pi = (1/4, 1/4, 1/4, 1/4)$. We have $\pi^{C(1)} P = \pi^{C(2)} P = \pi$ and hence $\mathcal{C}_\pi P \subseteq \mathcal{C}_\pi$. Next, it is easy to check that $\mathcal{N} = \{\mathbf{0}\}$ and we deduce that $\mathcal{C}_\mathcal{M} = \mathcal{C}_\pi$ from Corollary 3.3. However, condition $(\Gamma_{\widehat{N}})$ is not fulfilled since $\dim(\text{Span}(P^k I_1 \mathbf{1}^\top, P^k I_2 \mathbf{1}^\top, k \geq 0)) = 3$.

5 On the PH-representation of a geometric distribution

Throughout this section, C denotes a finite set with $|C|$ elements. Let α be a sub-stochastic vector on C and A be a $|C| \times |C|$ sub-stochastic matrix. We assume that matrix $I - A$ is invertible (I denotes the identity matrix). A random variable which takes values in \mathbb{N} is said to have a PH-distribution with

parameters (α, A) , which is denoted by $\text{PH}(\alpha, A)$, if its distribution function F is given by

$$\forall k \geq 1, F(k) = 1 - \alpha A^k \mathbf{1}^\top \quad \text{and} \quad F(0) = 1 - \alpha \mathbf{1}^\top.$$

A distribution $\text{PH}(\alpha, A)$ can be viewed as the distribution of the time to absorption in the absorbing Markov chain over the state space $C \cup \{|C| + 1\}$, with initial distribution $(\alpha, 1 - \alpha \mathbf{1}^\top)$ and t.p.m.

$$P = \left(\begin{array}{c|c} A & \mathbf{1}^\top - A\mathbf{1}^\top \\ \hline \mathbf{0} & 1 \end{array} \right).$$

A question addressed by Norberg [9] is characterizing the probability distributions on C such that $\text{PH}(\alpha, A)$ is actually a geometric distribution with some fixed nonnegative scalar λ as parameter. Such a geometric distribution is denoted by $\text{Geo}(\lambda)$. In this section, \mathcal{A} is the set of all probability distributions on C . We consider the following set

$$\mathcal{G}_A(\lambda) \stackrel{\text{def}}{=} \{\alpha \in \mathcal{A} \mid \text{PH}(\alpha, A) = \text{Geo}(\lambda)\}.$$

A stochastic vector α is in $\mathcal{G}_A(\lambda)$ if and only if

$$\forall k \geq 1, \alpha A^k \mathbf{1}^\top = \lambda^k. \tag{5.1}$$

The case when matrix A is primitive is tackled in [9], essentially with the help of the classic asymptotical expression of the power of such a matrix

$$A^k = \lambda^k w^\top v + o(\lambda^k), \tag{5.2}$$

where λ is the spectral radius of A , v and w are respectively the left and right eigenvector corresponding to λ . Norberg shows that

1. the spectral radius of A must be the parameter λ of the geometric distribution;
2. $\mathcal{G}_A(\lambda)$ is a closed convex set;
3. $\mathcal{G}_A(\lambda) = \mathcal{A}$ if and only if $\mathbf{1}$ is the right eigenvector of matrix A .

We prove the following general result by using the invariance of cones.

Theorem 5.1 *Let C be a finite set and A be a sub-stochastic matrix on the set C . If there exists a probability distribution α on C such that $\text{PH}(\alpha, A) = \text{Geo}(\lambda)$ then λ is an eigenvalue of A . Moreover, if matrix A is irreducible then λ is the spectral radius of A and the stochastic left eigenvector corresponding to λ is in $\mathcal{G}_A(\lambda)$.*

proof. Let us define the following set

$$\mathcal{C} = \{\alpha \geq \mathbf{0} \mid \forall k \geq 1, \alpha(A^k - \lambda^k I)\mathbf{1}^\top = 0\}.$$

\mathcal{C} is a polyhedral cone from the theorem of Cayley-Hamilton and [2, Th 2.5]. It is easy to check that cone \mathcal{C} is invariant under matrix A . Therefore, a slight modification of Theorem 3.2 in [2] (see [8, Lemma 3.3]) states that \mathcal{C} contains a nonnegative left eigenvector v of matrix A . If so, λ is an eigenvalue of matrix A . Moreover, when A is irreducible, matrix A admits only one nonnegative left eigenvector which corresponds to the spectral radius of A . ■

Now a similar approach to that of Section 3 enables us to compute the set $\mathcal{G}_A(\lambda)$. We may restate relation (5.1) as follows: the process $agg((\alpha, 0), P, \mathcal{P})$, with the partition $\mathcal{P} = \{C, \{|C| + 1\}\}$, is a homogeneous Markov chain with t.p.m.

$$\hat{P} = \begin{pmatrix} \lambda & 1 - \lambda \\ 0 & 1 \end{pmatrix}.$$

Let us introduce the set

$$\mathcal{N}_A = \{u \in \mathbb{R}^{|C|} \mid uA^k\mathbf{1}^\top = 0 \text{ for } k \geq 0\}$$

which is the orthogonal complement of $\text{Span}(A^k\mathbf{1}^\top, k \geq 0)$. The proofs of Theorem 5.1 and Theorem 3.2 give the following result in the present context.

Corollary 5.2 *If $\mathcal{G}_A(\lambda) \neq \emptyset$ then there exists a stochastic left eigenvector v corresponding to the eigenvalue λ of matrix A such that $\mathcal{G}_A(\lambda) = \mathcal{C}_A(\lambda) \cap \mathcal{A}$ with*

$$\mathcal{C}_A(\lambda) = (\text{Cone}(v) \oplus \mathcal{N}_A) \cap \mathbb{R}_+^{|C|}.$$

In such a case, we have $(\mathcal{G}_A(\lambda) = \{v\}) \iff \mathcal{N}_A = \{\mathbf{0}\}$ and $(\mathcal{G}_A(\lambda) = \mathcal{A}) \iff \mathbf{1}$ is a right eigenvector corresponding to λ .

Remark 5.1 Matrix A is said to be PH-simple according to O’Cinneide [10] if, for all $\alpha_1, \alpha_2 \in \mathcal{A}$, the equality $\text{PH}(\alpha_1, A) = \text{PH}(\alpha_2, A)$ implies that $\alpha_1 = \alpha_2$. From a discrete-time counterpart of Theorem 1 of [10], condition $\mathcal{N}_A = \{\mathbf{0}\}$ ensures that matrix A is PH-simple. In such a case, if geometric distribution $\text{Geo}(\lambda)$ admits the PH-representation $\text{PH}(v, A)$ then this is the only one in the form $\text{PH}(\alpha, A)$, or equivalently $\mathcal{G}_A(\lambda) = \{v\}$.

The first part of Norberg’s result was implicitly shown in Ledoux [7] in a general framework which does not allow the use of formula (5.2). As noted in [9], the weak lumpability is related to the current problem. Indeed, the Markov property for $agg(\alpha, P, \mathcal{P})$ requires distributions of sojourn times in each class $C(l)$ of \mathcal{P} for the Markov chain (α, P) which are $\text{Geo}(\hat{P}(l, l))$. It is known that the distribution of the n th sojourn time in a subset of states

$C(l)$ for a Markov chain with irreducible t.p.m. P is $\text{PH}(v_n, P_{C(l)C(l)})$, where v_n ($n \geq 1$) is a probability distribution over the set $C(l)$. Consequently, for every $n \geq 1$, the distribution of the n th sojourn time in $C(l)$ is $\text{Geo}(\widehat{P}(l, l))$ if $v_n \in \mathcal{G}_{P_{C(l)C(l)}}(\widehat{P}(l, l))$. Starting with the PH-representation of distributions of sojourn times in each class $C(l)$ and using the canonical Jordan form of the matrix $P_{C(l)C(l)}$, it was shown in [7] that $\widehat{P}(l, l)$ (for all $l \in \widehat{S}$) is an eigenvalue of matrix $P_{C(l)C(l)}$. A similar argument should lead to a proof of Theorem 5.1. However we have preferred to give a proof illustrating how the invariance of cones can be used to derive spectral properties. Finally, we note that, for a given matrix P and a partition \mathcal{P} , we have for every $l \in \widehat{S}$

$$u \in \mathcal{N}I_l \implies (u(i))_{i \in C(l)} \in \mathcal{N}_{P_{C(l)C(l)}}(\widehat{P}(l, l)).$$

So, the set $\mathcal{N}I_l$ reduces to $\{\mathbf{0}\}$ if $\mathcal{N}_{P_{C(l)C(l)}}(\widehat{P}(l, l)) = \{\mathbf{0}\}$. In such a case, the set $\mathcal{C}_{\mathcal{M}}I_l$, if it is different from $\{\mathbf{0}\}$, is $\text{Cone}(\pi_l)$. If so, we have $\pi_{C(l)}P_{C(l)C(l)} = \widehat{P}(l, l)\pi_{C(l)}$ where $\pi_{C(l)}$ is the $n(l)$ -vector $(\pi_l(i))_{C(l)}$. Thus, vector $\pi_{C(l)}$ is a positive left eigenvector of matrix $P_{C(l)C(l)}$ corresponding to the eigenvalue $\widehat{P}(l, l)$.

Acknowledgments

The authors are grateful to the anonymous referees and the Editor for their careful reading of the manuscript and for their helpful comments.

References

- [1] Amari, S.-I., Ito, H. and Kobayashi, K. Identifiability of hidden Markov information sources and their minimum degrees of freedom. *IEEE Trans. Inform. Theory* **31** (1992), 324–333.
- [2] Berman, A. and Plemmons, R.J. *Nonnegative Matrices in the Mathematical Sciences*. Academic Press, 1979.
- [3] Burke, C.J. and Rosenblatt, M. A Markov function of a Markov chain. *Ann. Math. Statist.* **29** (1958), 1112–1122.
- [4] Gray, R.M. *Probability, Random processes, and Ergodic Properties*. Springer-Verlag, 1988.
- [5] Hachigan, J. Collapsed Markov chains and the Chapman-Kolmogorov equation. *Ann. Math. Statist.* **34** (1963), 233–237.
- [6] Kemeny, J.G. and Snell, J.L. *Finite Markov chains*. Springer-Verlag, 1976.
- [7] Ledoux, J. A necessary condition for weak lumpability. *Oper. Res. Letters* **13** (1993), 165–168.
- [8] Ledoux, J. A geometric invariant in weak lumpability of finite Markov chains. *J. Appl. Prob.* **34** (1997), 847–858. Detailed version available at <ftp.inria.fr> as `INRIA/publication/publi-ps-gz/RR/RR-2801.ps.gz`
- [9] Norberg, T. On the time a Markov chain spends in a lumped state. *J. Appl. Prob.* **34** (1997), 340–345.
- [10] O’Cinneide, C.A. On non-uniqueness of representations of phase-type distributions. *Commun. Statist.-Stochastic Models* **5** (1989), 247–259.
- [11] Peng, N. On weak lumpability of a finite Markov chain. *Statist. and Probab. Letters* **27** (1996), 313–318.
- [12] Rubino, G. and Sericola, B. A finite characterization of weak lumpable Markov processes. Part I: The discrete time case. *Stoch. Proc. and Appl.* **38** (1991), 195–204.