Weak Lumpability of Finite Markov Chains and Positive Invariance of Cones

James Ledoux

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Weak lumpability of finite Markov chains and positive invariance of cones

James Ledoux

N° 2801

Fevrier 1996

PROGRAMME 1

Rapport de recherche

ISSN 0249-6399

INRIA
RENNES

Weak lumpability of finite Markov chains and positive invariance of cones

James Ledoux*

Programme 1 — Architectures parallèles, bases de données, réseaux et systèmes distribués
Projet Model

Rapport de recherche n° 2801 — Fevrier 1996 — 16 pages

Abstract: We consider weak lumpability of general finite homogeneous Markov chains evolving in discrete time, that is when a lumped Markov chain with respect to a partition of the initial state space is also a homogeneous Markov chain. We show that weak lumpability is equivalent to the existence of a decomposable polyhedral cone which is positively invariant by the transition probability matrix of the original chain. It allows us, in a unified way, to derive new results on lumpability of reducible Markov chains and to obtain spectral properties associated with lumpability.

Key-words: Markov chain, States aggregation, Weak lumpability, Polyhedral cone, Positive invariance.

(Résumé : tsvp)

*INSA, Campus de Beaulieu 35043 Rennes Cedex, E-mail ledoux@{univ-rennes1.fr},{irisa.fr}

Agrégation faible de chaînes de Markov finies et invariance positive de cônes

Résumé : Nous nous intéressons à la propriété d'agrégation faible de chaînes de Markov finies évoluant en temps discret, c'est à dire quand une chaîne de Markov, agrégée selon une partition de l'espace d'état initial, est encore markovienne homogène. Nous montrons que cette propriété est équivalente à l'existence d'un cône polyhédrique décomposable qui est invariant par la matrice des probabilités de transition de la chaîne originale. Cela nous permet, d'une manière unifée, de dériver de nouveaux résultats sur l'agrégation de chaînes de Markov réductibles et d'obtenir des propriétés spectrales associées à l'agrégation.

Mots-clé : Chaîne de Markov, Agrégation d'états, Cône polyhédrique, Invariance positive

1 Introduction

Let us consider a homogeneous Markov chain X, in discrete or continuous time, on a finite state space, denoted by S, which is assumed to be $S = \{1, \ldots, N\}$. Let $\mathcal{P} = \{C(1), \ldots, C(M)\}$ be a fixed partition of S in $M < N$ classes. We associate with the given chain X the aggregated chain Y, over the state space $\hat{S} = \{1, \ldots, M\}$, defined by:

$$Y_t = l \iff X_t \in C(l), \quad \text{for any } t.$$

We are interested in the initial distributions of X which give an aggregated homogeneous Markov chain Y. If such a distribution exists, we say that the family of Markov chains sharing the same transition probability matrix (t.p.m.) is *weakly lumpable*. The former work on lumpability was concerned with the *strong lumpability* situation, that is, when any initial distribution leads to an aggregated homogeneous Markov chain. It has been characterized for irreducible Markov chains in Kemeny and Snell (1976) and for a class of reducible chains in Abdel-Moneim and Leysieffer (1984). The weak lumpability problem has been addressed in Kemeny and Snell (1976), Abdel-Moneim and Leysieffer (1982), Rubino and Sericola (1991) for irreducible Markov chains. Under an additional condition, the (homogeneous) Markovian property for the aggregated chain Y has recently been proved in Peng (1995) to be equivalent to require Y to satisfying the Chapman-Kolmogorov equations and then the set of all initial distributions leading to an aggregated Markov chain is explicitly given. As shown in Ledoux et al. (1994), weak lumpability of absorbing Markov chains with an only one irreducible transient state class, is closely related with the previous works using a resetting from absorbing states on the transient ones according to the quasi-stationary distribution. Aggregation of any finite continuous time Markov chain can be replaced in the discrete time context (see Ledoux (1995)) and, therefore, will not be discussed in the sequel. If the state space S is countably infinite we refer to Ball and Yeo (1993) and Ledoux (1995).

A large amount of work on lumpability Markov chains is concerned in proving the unicity of the transition probability matrix of the Markovian aggregated chains. The linear system approach, initialized by Abdel-Moneim and Leysieffer (1982) and fully developed in Rubino and Sericola (1991), to compute the set of all initial distributions leading to an aggregated Markov chain is centered on this property. However, this unicity property may not hold for a general finite Markov chain. The purpose of this paper is to place emphasis on geometrical properties associated with the weak lumpability condition when we are interested in aggregated Markov chains sharing the same t.p.m. In particular, it can be used to give an unified view of the previous works and to derive new results for general finite Markov chains. After reviewing some preliminaries on polyhedral cones, we analyze in Section 2, for a general finite Markov chain with transition probability matrix P, the set of all initial distributions which give aggregated Markov chains sharing the same t.p.m. Pointing out the relation between lumpability and positive invariance of cones in Section 3, we show that this set is non empty if there exists a family of M polyhedral cones which are "invariant" under sub-matrices of matrix P. This result allows us to state in Section 4 that if the partition $\mathcal{P}$ is a refinement of the partition of the state space S induced by the usual "communication" equivalence relation, then we obtain an explicit formula for the transition probability matrix of any Y, which depends only on $\mathcal{P}$ and P. Previous works can be viewed as direct applications of this result. Throughout Section 3 and Section 4, various properties reported in Ledoux (1993), Abdel-Moneim and Leysieffer (1984) and Peng (1995) are extended to general finite Markov chains and new spectral results are also derived.

Notation

The set of all probability distributions on S will be denoted by $\mathcal{A}$. The support D of a probability distribution α is defined as the greatest subset of S apart from the distribution is zero, i.e. $\alpha(i) = 0$ for all $i \in E \setminus D$ and $\alpha(i) \neq 0$ for $i \in D$.

By convention, vectors are row vectors. Column vectors are indicated by means of the transpose operator $(.)^{\mathrm{T}}$. The vector with all its components equal to 1 (resp. 0) is denoted merely by 1 (resp. 0). The vector

e_i denotes the ith vector of the canonical basis of $\mathbb{R}^N$. We denote by I the identity matrix and by $\mathrm{diag}(v)$ (by $\mathrm{diag}(H_i)$) the (block) diagonal matrix with generic diagonal (block) entry $v(i)$ (the matrix H_i), the dimensions being defined by the context.

The cardinality of the class $C(l)$ is denoted by $n(l)$. We assume the states of S ordered such that $C(l) = \{n(1) + \cdots + n(l-1) + 1, \ldots, n(1) + \cdots + n(l)\}$ for $1 \leq l \leq M$ (with $n(0) = 0$.)

For any subset C of S (whose cardinality is n) and $\alpha \in \mathcal{A}$, the restriction of α to C, i.e. the vector $(\alpha(i), i \in C)$, is denoted by α_C or $R_C\alpha$. On the other hand, a vector β on $[0,1]^{n(l)}$ can be viewed as the vector on $[0,1]^N$ defined by: $[R_l^{-1}\beta](i) = 0$ if $i \notin C(l)$ and $[R_l^{-1}\beta](i) = \beta(i - n(1) - \cdots - n(l-1))$ if $i \in C(l)$. If $\mathcal{C}$ is a subset of $\mathcal{A}$ (resp. of $[0,1]^{n(l)}$) then $R_l\mathcal{C}$ (resp. $R_l^{-1}\mathcal{C}$) denotes the set $\{\alpha_{C(l)} \ / \ \alpha \in \mathcal{C}\} \subseteq [0,1]^{n(l)}$ (resp. $\{R_l^{-1}\beta \ / \ \beta \in \mathcal{C}\} \subseteq \mathbb{R}^N$.)

If $C \subseteq S$ and $\alpha_C 1^{\mathrm{T}} \neq 0$, α^C is the vector of $\mathcal{A}$ defined by $\alpha^C(i) = \alpha(i)/\sum_{j \in C} \alpha(j)$ if $i \in C$ and by 0 if $i \notin C$.

2 Preliminaries on cones and weak lumpability

2.1 Cone, polyhedral cone of $\mathbb{R}^n$

The basic definitions on the cones are reviewed from Berman and Plemmons (1979). Throughout this subsection $\mathcal{C}$ denotes a subset of $\mathbb{R}^n$. For any $\mathcal{C}$, $\mathrm{Span}(\mathcal{C})$ (resp. $\mathrm{Aff}(\mathcal{C})$) refers to the linear (resp. affine) hull of $\mathcal{C}$. The set $\mathrm{Cone}(\mathcal{C})$ denotes the conical hull of C that is the set of all finite nonnegative linear combinations of the elements of $\mathcal{C}$. The elements of $\mathcal{C}$ are called the generators of $\mathrm{Cone}(\mathcal{C})$. If $\mathrm{Cone}(\mathcal{C}) = \mathcal{C}$ then $\mathcal{C}$ is called a *cone*. $\mathrm{Conv}(\mathcal{C})$ is the set of all finite convex linear combinations of the elements of $\mathcal{C}$. The *dimension* of a subset $\mathcal{C}$ is defined by $\dim(\mathcal{C}) = \dim \mathrm{Aff}(\mathcal{C})$. The interior of $\mathcal{C}$ relative to the affine space A is denoted by $\mathrm{int}_{\mathrm{A}}(\mathcal{C})$.

Definition 2.1 *A polyhedral cone $\mathcal{C}$ of $\mathbb{R}^n$ is the solution set of a system of linear homogeneous inequalities, i.e. $\mathcal{C} = \{ x \in \mathbb{R}^n \ / \ xH \geq 0 \}$ where $H \in \mathbb{R}^{n \times m}$. Such a cone is a closed convex subset of $\mathbb{R}^n$.*

We recall that a bounded solution set of a system of linear inequalities is called a *polytope* of $\mathbb{R}^n$.

A convex cone $\mathcal{C}$ is *pointed* if $\mathcal{C} \cap (-\mathcal{C}) = \{0\}$ and *solid* if $\mathrm{int}_{\mathbb{R}^n}(\mathcal{C}) \neq \emptyset$. Note that a convex subset $\mathcal{C}$ is such that $\mathrm{int}_{\ \mathrm{Aff}(\mathcal{C})}(\mathcal{C}) \neq \emptyset$. Finally, a closed, pointed, solid convex cone is called a *proper* cone.

An *extremal* of a pointed polyhedral cone $\mathcal{C}$ is an element which can never be written as a nonnegative linear combinations of others elements of $\mathcal{C}$. A pointed polyhedral cone is finitely generated by their extremals.

Definition 2.2 *Let $\mathcal{C}$ be a cone of $\mathbb{R}^n$, $\mathcal{C}_1$ and $\mathcal{C}_2$ be two sub-cones of $\mathcal{C}$. The cone $\mathcal{C}$ is the direct sum of $\mathcal{C}_1$ and $\mathcal{C}_2$, that is denoted by $\mathcal{C} = \mathcal{C}_1 \bigoplus \mathcal{C}_2$, if $\mathrm{Span}(\mathcal{C}_1) \cap \mathrm{Span}(\mathcal{C}_2) = \{0\}$ and $\mathcal{C} = \mathcal{C}_1 + \mathcal{C}_2$. If such cones $\mathcal{C}_1$ and $\mathcal{C}_2$ exist and are distinct from $\{0\}$, then the cone $\mathcal{C}$ is said to be decomposable.*

2.2 Weak lumpability of a finite Markov chain

Let $X = (X_n)_{n \geq 0}$ be a homogeneous Markov chain over state space S, given by its transition probability matrix $P = (P(i,j))_{i,j \in E}$ and its initial distribution α; when necessary we denote it by (α, P). Let $P(i, C)$ denote the transition probability of moving in one step from state i to the subset C of S, that is $P(i,C) = \sum_{j \in C} P(i,j)$. Let $P_{C(l)C(m)}$ be the $n(l) \times n(m)$ sub-matrix of P given by $(P(i,j))_{i \in C(l), j \in C(m)}$. We denote the aggregated chain constructed from (α, P) with respect to the partition $\mathcal{P}$ by $agg(\alpha, P, \mathcal{P})$.

Definition 2.3 *A sequence $(C_0, C_1, \ldots, C_j)$ of subsets of S is called possible for the initial distribution α if $\mathbb{P}_\alpha(X_0 \in C_0, X_1 \in C_1, \ldots, X_j \in C_j) > 0$. Given any distribution $\alpha \in \mathcal{A}$ and a possible sequence*

$(C_0, C_1, \ldots, C_j)$ *for* α*, we can define the vector* $f(\alpha, C_0, C_1, \ldots, C_j) \in \mathcal{A}$ *recursively by:*

$$\begin{aligned} f(\alpha, C_0) &= \alpha^{C_0} \\ f(\alpha, C_0, C_1, \ldots, C_k) &= (f(\alpha, C_0, C_1, \ldots, C_{k-1})P)^{C_k} \qquad k \geq 1. \end{aligned}$$

For any $C \in \mathcal{C}$*,* $\mathcal{A}(\alpha, C)$ *denotes the subset of all distributions of the form* $f(\alpha, C_0, \ldots, C_k, C)$*.*

The approach developed in Kemeny and Snell (1976) and in Rubino and Sericola (1989) consists in rewriting the conditional expression $\mathbb{P}_\alpha(X_{n+1} \in C(m) \mid X_n \in C(l), X_{n-1} \in C_{n-1}, \ldots, X_0 \in C_0)$ (defined for any $(C_0, C_1, \ldots, C_{n-1}, C(l))$ possible for α) as $\mathbb{P}_\beta(X_1 \in C(m))$ with $\beta = f(\alpha, C_0, \ldots, C_{n-1}, C(l))$, that is, in including the past into the initial distribution. A necessary and sufficient condition for Y to be a homogeneous Markov chain can be exhibited without any particular assumption on X.

Result 2.4 *The chain* $Y = agg(\alpha, P, \mathcal{C})$ *is a homogeneous Markov chain if and only if* $\forall l, m \in \hat{S}$*, the probability* $\mathbb{P}_\beta(X_1 \in C(m))$ *is the same for every* $\beta \in \mathcal{A}(\alpha, C(l))$*. This common value is the transition probability for the chain* Y *to move from state* l *to state* m*.*

Remark 1 For a chain (α, P), the set $\mathcal{A}(\alpha, C)$ may be empty. That implies that we can never access to the states of the class C with α as initial distribution. Consequently, C can be eliminated of $\mathcal{P}$ with any repercussions on the analysis of the aggregated process.

$\triangle$

Result 2.4 determines the transition probability matrix of the aggregated process which may depend on the initial distribution α as shown by the following example.

Example 1 Let us consider the two irreducible Markov chains with respective t.p.m. P_1 and P_2 :

$$P_1 = \left(\begin{array}{c|cc} 1/4 & 1/4 & 1/2 \\ \hline 0 & 1/6 & 5/6 \\ 7/8 & 1/8 & 0 \end{array}\right) \quad P_2 = \left(\begin{array}{cc|cc} 1/6 & 1/6 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/6 & 1/6 \\ \hline 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{array}\right).$$

It has been shown in Kemeny and Snell (1976) that, for P_1, the set of all initial distribution leading to an aggregated Markov chain according to the partition $\mathcal{P}_1 = \{\{1\}, \{2, 3\}\}$ is $A_1 = \{\lambda e_1 + (1-\lambda)(0, 1/3, 2/3) \,/\, 0 \leq \lambda \leq 1\}$. Following the same way as in Rubino and Sericola (1991), the corresponding set for P_2 and $\mathcal{P}_2 = \{\{1, 2\}, \{3, 4\}\}$ is $A_2 = \{\lambda(1/2, 1/2, 0, 0) + \mu e_3 + (1 - \lambda - \mu)e_4 \,/\, 0 \leq \lambda, \mu \leq 1\}$. Let P be the matrix $\mathrm{diag}(P_i)$ and $\mathcal{P} = \{\{1, 6, 7\}, \{2, 3, 4, 5\}\}$. It is easy to convince ourself that the chain $agg(e_1, P, \mathcal{P})$ is a homogeneous Markov chain. Its transition probability matrix is

$$\widehat{P}^{(e_1)} = \begin{pmatrix} 1/4 & 3/4 \\ 7/12 & 5/12 \end{pmatrix}.$$

In the same way, the chain $agg(e_7, P, \mathcal{P})$ is a homogeneous Markov chain with t.p.m.

$$\widehat{P}^{(e_7)} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

We note that $\widehat{P}_{e_1}(1, 1) = 1/4$ is an eigenvalue of matrix $P_{C(1)C(1)}$ which is distinct from the spectral radius $1/2$ of the nonnegative matrix.

$\triangleleft$

Let us consider a probability distribution α and eliminate all the classes of the partition $\mathcal{P}$ which can never be accessed by the chain (α, P), that is the classes $C(l)$ $(l \in \hat{S})$ such that $\mathcal{A}(\alpha, C(l)) = \emptyset$. We obtain a new state space and a new partition of the subset of S which are again denoted by S and $\mathcal{P}$ for not making the notation heavier. With that new partition $\mathcal{P}$, all the sets $\mathcal{A}(\alpha, C(l))$ are not empty and we can define matrix $\widehat{P}^{(\alpha)}$ in the following manner: for each $l \in \hat{S}$, let β be a vector in $\mathcal{A}(\alpha, C(l))$ and set

$$\widehat{P}^{(\alpha)}(l,m) \stackrel{\text{def}}{=} \mathbb{P}_\beta(X_1 \in C(m)) \quad \forall m \in \hat{S}.$$

If $agg(\alpha, P, \mathcal{P})$ is a homogeneous Markov process then it follows from Result 2.4 that $\widehat{P}^{(\alpha)}$ is the transition probability matrix of this aggregated chain. Let us now define the set, denoted by $\mathcal{A}_{\mathcal{M}}(\widehat{P}^{(\alpha)})$, of all initial distributions β leading to an aggregated homogeneous Markov process $agg(\beta, P, \mathcal{P})$ with transition probability matrix $\widehat{P}^{(\alpha)}$:

$$\mathcal{A}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) = \{ \ \beta \in \mathcal{A} \ / \ agg(\beta, P, \mathcal{C}) \text{ is a homogeneous Markov chain with t.p.m. } \widehat{P}^{(\alpha)} \ \}$$

To lighten the presentation, $\mathcal{A}_{\mathcal{M}}$ will also refer to $\mathcal{A}_{\mathcal{M}}(\widehat{P}^{(\alpha)})$ if there is no ambiguity. The aim of this subsection is to analyze properties of this set when it is not empty.

Let us define the following matrices.

- For any $l \in \hat{S}$, P_l denotes the $n(l) \times N$ sub-matrix of P: $(P(i,j))_{i \in C(l), j \in S}$.
- $\widetilde{P}$ denotes the $N \times M$ matrix defined by: $\forall i \in S$, $\forall m \in \hat{S}$, $\widetilde{P}(i,m) = P(i, C(m))$. For any $l \in \hat{S}$, we denote by $\widetilde{P}_l$ the $n(l) \times M$ sub-matrix of $\widetilde{P}$: $(\widetilde{P}(i,m))_{i \in C(l), m \in \hat{S}}$.
- The lth row of the stochastic matrix $\widehat{P}^{(\alpha)}$ is denoted by $\widehat{P}_l^{(\alpha)}$.
- For all $l \in \hat{S}$, we set $H_l = \widetilde{P}_l - 1^{\mathrm{T}} \widehat{P}_l^{(\alpha)}$ $(n(l) \times M)$ and for any $j \geq 1$, we define the following $N \times M^{j+1}$ block diagonal matrices

$$H^{[1]} = \operatorname{diag}(H_l), \qquad H^{[j+1]} = \operatorname{diag}(P_l H^{[j]}). \tag{1}$$

We are in position to adopt the linear system approach from Rubino and Sericola (1991) and in the same manner, we have:

$$\mathcal{A}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) = \bigcap_{j \geq 1} \mathcal{A}^j \qquad \text{where } \ \mathcal{A}^j = \{ \ \beta \in \mathcal{A} \ / \ \beta H^{[k]} = 0, \quad \text{for } k \geq 1 \ \}.$$

Now, each polytope $\mathcal{A}^j$ can be seen as the trace on the set $\mathcal{A}$ of the following polyhedral cone:

$$\mathcal{C}^j \ \stackrel{\text{def}}{=} \ \{\beta \geq 0 \ / \ \beta H^{[k]} = 0, \ \text{for } 1 \leq k \leq j\}, \tag{2}$$

that is $\mathcal{A}^j = \mathcal{C}^j \cap \mathcal{A}$ for $j \geq 0$ (with the convention $\mathcal{C}^0 = \mathbb{R}_+^N$ and $\mathcal{A}^0 = \mathcal{A}$). Consequently, we note that $\mathcal{A}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) = \mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) \cap \mathcal{A}$ where $\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) \stackrel{\text{def}}{=} \bigcap_{j \geq 1} \mathcal{C}^j$ and we have

$$\mathcal{A}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) \neq \emptyset \Longleftrightarrow \mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) \neq \{0\}.$$

Now, if we note that $\mathcal{C}^{j+1}$ is deduced from $\mathcal{C}^j$ by attaching the (eventually) additional constraints ($\beta H^{[j+1]} = 0$) and that $\dim(\mathcal{C}^1) \leq N$ then the following extension of Theorem 3.4 from Rubino and Sericola (1991) is intuitively clear:

$$\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) = \mathcal{C}^N \tag{3}$$

where N is the number of states of the original chain.

We note from the diagonal structure of the matrices $H^{[j]}$, that, for $\beta \geq 0$, $\beta \in \mathcal{C}^j \Longleftrightarrow \forall l \in \hat{S}, \ R_l^{-1} \beta_{C(l)} \in \mathcal{C}^j$. It allows us to derive part of the following lemma.

Lemma 2.5 *Let us set* $\mathcal{C}_l^j = R_l \mathcal{C}^j$ *for every* $l \in \hat{S}$. *We have, for all* $j \geq 1$, $\mathcal{C}^j = \bigoplus_{l \in \hat{S}} R_l^{-1} \mathcal{C}_l^j$ *where* $R_l^{-1} \mathcal{C}_l^j \subseteq \mathcal{C}^j$ *is a polyhedral cone of* $\mathbb{R}^N$ *(*$\mathcal{C}_l^j$ *is a polyhedral cone of* $\mathbb{R}^{n(l)}$*.)*

proof. We can check from the definition of the sets $\mathcal{C}^j$ (see (1), (2)) that for $j \geq 1$,

$$\mathcal{C}_l^j = \left\{ \beta \in \mathbb{R}_+^{n(l)} \ / \ \beta H_l = 0 \ \text{ and } \ \beta P_l H^{[k]} = 0, \ 1 \leq k \leq j-1 \right\}. \tag{4}$$

Consequently, $\mathcal{C}_l^j$ (resp. $R_l^{-1} \mathcal{C}_l^j$) is a polyhedral cone of $\mathbb{R}^{n(l)}$ (resp. $\mathbb{R}^N$). ■

The well-known necessary and sufficient condition reported in Kemeny and Snell (1976) for having strong lumpability of $(., P)$ with an irreducible matrix P can be extended to a general stochastic matrix. The only requirement is that all the aggregated chains share the same t.p.m. $\widehat{P}$. In that case, by definition, the family $(., P)$ of Markov chains is strongly lumpable if $\mathcal{A}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) = \mathcal{A}$ or $\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) = \mathbb{R}_+^N$ for any $\alpha \in \mathcal{A}$. In fact, it is equivalent to require that $\mathcal{A}^1 = \mathcal{A}^0$ or $\mathcal{C}^1 = \mathcal{C}^0$. Now, $\mathcal{C}^1 = \mathbb{R}_+^N$ is equivalent to $H^{[1]} = 0$ or to $(H_l = 0, \ \forall l \in \hat{S})$ which are precisely the conditions given by the following theorem.

Theorem 2.6 *If we require that all the aggregated chains share the same transition probability matrix, then the family* $(., P)$ *of Markov chains is strongly lumpable if and only if for each pair of classes* $C(l)$ *and* $C(m)$, $P(i, C(m))$ *does not depend on* $i \in C(l)$.

In particular, this result is necessary to derive some results in Abdel-Moneim and Leysieffer (1984) though the characterization explicitly used is the Kemeny and Snell's one with the irreducibility assumption.

3 Lumpability and positive invariance

Definition 3.1 *A matrix* A *leaves a cone* $\mathcal{C}$ *of* $\mathbb{R}^N$ *invariant or matrix* A *is nonnegative on the cone* $\mathcal{C}$, *that will be denoted by* $A \overset{\mathcal{C}}{\geq} 0$, *if for every* $x \in \mathcal{C}$ *the vector* $xA \in \mathcal{C}$ *(i.e.* $\mathcal{C}A \subseteq \mathcal{C}$*). The cone* $\mathcal{C}$ *is said to be positively invariant by matrix* A.

Some spectral properties of matrices leaving a proper cone invariant are reviewed from Berman and Plemmons (1979).

Result 3.2 *If matrix* A *leaves a proper cone* $\mathcal{C}$ *invariant then the spectral radius* $\rho(A)$ *is an eigenvalue of* A *and* $\mathcal{C}$ *contains a left eigenvector of* A *corresponding to* $\rho(A)$.

Note that a nonnegative matrix is a matrix which leaves the proper cone $\mathbb{R}_+^N$ of $\mathbb{R}^N$ invariant. We will deal with cones which are not solid. Consequently, we have to derive a weaker result than the previous one.

Lemma 3.3 *If matrix* A *leaves a closed, pointed convex cone* $\mathcal{C}$ *invariant then there exists an nonnegative eigenvalue* λ *of* A *such that* $\mathcal{C}$ *contains a left eigenvector of* A *associated with* λ.

If a nonnegative matrix A *is irreducible and leaves a closed, convex cone* $\mathcal{C} \subseteq \mathbb{R}_+^N$ *invariant then* $\mathcal{C}$ *contains the positive left eigenvector corresponding to the spectral radius* $\rho(A)$.

proof. Matrix A represents the matrix of a linear operator f on $\mathbb{R}^N$ with respect to the canonical basis (with the convention that $f(e_i)$, for every $i \in N$, is the ith row of matrix A, that is $f(x) = xA$ for all $x \in \mathbb{R}^N$.) Matrix A is nonnegative on $\mathcal{C}$ means that $f(\mathcal{C}) \subseteq \mathcal{C}$. Consequently, f leaves the linear subspace $L = \text{Span}(\mathcal{C}) \subseteq \mathbb{R}^N$ invariant and it implies that the restriction of f to the subspace L, denoted by $f_{|L}$, is a linear operator from L to L. The cone $\mathcal{C}$ is also invariant by $f_{|L}$ and is solid with respect to L. Thus, the proper cone $\mathcal{C}$ is positively invariant by the matrix $A_{|L}$ of the operator $f_{|L}$. The first part of Result 3.2 can be applied to $A_{|L}$ and conclusions are associated with the spectral radius of that matrix. However, the eigenvectors and the spectral radius of $f_{|L}$ are eigenvectors and a nonnegative eigenvalue of the initial linear operator f on domain $\mathbb{R}^N$, that gives the first part of the lemma.

If the nonnegative matrix A is irreducible, then there exists an unique (up to a constant multiple) left eigenvector of A in $\mathbb{R}_+^N$ (in fact in int$\mathbb{R}_+^N$ i.e. it is a positive left eigenvector) which corresponds to the spectral radius of A. Now, for any closed, (pointed) convex cone $\mathcal{C} \subseteq \mathbb{R}_+^N$, if $A \overset{\mathcal{C}}{\geq} 0$ then we deduce from the first part of the proof that there exists a nonnegative left eigenvector of matrix A in $\mathcal{C} \subseteq \mathbb{R}_+^N$. Since there is only one left eigenvector of A in $\mathbb{R}_+^N$, it is positive and associated with the spectral radius of the matrix A. The second part of the lemma holds. ∎

We want emphasize that the positive invariance of polytope, used in Lemma 3.5 from Rubino and Sericola (1991) as a simple stop test in their incremental computation of $\mathcal{A}_{\mathcal{M}}$ from the $\mathcal{A}^j$ ones, is a central geometric invariant of the weak lumpability property as soon as we are interested in aggregated Markov chains sharing the same transition probability matrix.

Theorem 3.4 *The set $\mathcal{A}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) \neq \emptyset$ or $\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) \neq \{0\}$ if and only if there exists a polyhedral cone $\mathcal{C} \subseteq \mathcal{C}^1$, different from $\{0\}$, such that $P \overset{\mathcal{C}}{\geq} 0$ and $\mathcal{C}$ is the direct sum $\bigoplus_{l \in \hat{S}} R_l^{-1}\mathcal{C}_l$ where $\mathcal{C}_l \overset{\text{def}}{=} R_l C$ for all $l \in \hat{S}$.*

proof. Suppose that $\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) \neq \{0\}$. Let us verify that $\mathcal{C}_{\mathcal{M}}$ fulfills the required conditions. We have $\mathcal{C}_{\mathcal{M}} = \mathcal{C}^N$ from relation (3). Since $\mathcal{C}^N = \mathcal{C}^{N+1}$, we have for any vector $\beta = \sum_{l \in \hat{S}} R_l^{-1}\beta_{C(l)} \in \mathcal{C}^N$ and for any j such that $1 \leq j \leq N$

$$\begin{aligned} \beta H^{[j+1]} = 0 \quad &\Longleftrightarrow \quad \forall l \in \hat{S}, \quad \beta_{C(l)} P_l H^{[j]} = 0 \text{ (by definition of system } H^{[j+1]}) \\ &\Longrightarrow \quad \beta P H^{[j]} = \sum_{l \in \hat{S}} \beta_{C(l)} P_l H^{[j]} = 0, \end{aligned}$$

that is $\beta P \in \mathcal{C}^N$. The set $\mathcal{C}_{\mathcal{M}}$ is the direct sum of its M "projections" from relation (3) and Lemma 2.5. To obtain the decomposability of $\mathcal{C}_{\mathcal{M}}$, it remains to establish that these sets are distinct from $\{0\}$. If there exists $l \in \hat{S}$ such that $\mathcal{C}_l = R_l \mathcal{C}_{\mathcal{M}} = \{0\}$, that implies that the original chain (α, P) can never accessed to the state class $C(l)$. Indeed, if $\mathcal{A}(\alpha, C(l)) \neq \emptyset$ then there exists $n \geq 0$ such that $R_l[\alpha P^n] \neq 0$ and $R_l[\alpha P^n] \in \mathcal{C}_l$ with the positive invariance of $\mathcal{C}_{\mathcal{M}}$ by matrix P. To conclude, recall that the considered partition $\mathcal{P}$ contains only the state classes of the original state space S which are accessed by (α, P).

Conversely, if there exists a polyhedral cone $\mathcal{C} \subseteq \mathcal{C}^1$, which is distinct from $\{0\}$ and is positively invariant by P, such that $\mathcal{C} = \bigoplus_{l \in \hat{S}} R_l^{-1}\mathcal{C}_l$ then we show by induction that

$$\mathcal{C} \subseteq \mathcal{C}^j \quad \forall j \geq 1.$$

The first step is obvious. Let us assume that $\mathcal{C} \subseteq \mathcal{C}^j$ with $j > 1$. For every $\beta \in \mathcal{C}$, we have $R_l^{-1}\beta_{C(l)} \in \mathcal{C} \subseteq \mathcal{C}^j$ for all $l \in \hat{S}$ (since $\mathcal{C}$ is decomposable), next $[R_l^{-1}\beta_{C(l)}]P = \beta_{C(l)}P_l \in \mathcal{C} \subseteq \mathcal{C}^j$ for all $l \in \hat{S}$ (because $P \overset{\mathcal{C}}{\geq} 0$). We conclude that $[\forall l \in \hat{S}, \beta_{C(l)} P_l H^{[j]} = 0]$ or $\beta H^{[j+1]} = 0$. Thus, we have $\mathcal{C} \subseteq \mathcal{C}^{j+1}$. Finally, we obtain $\mathcal{C} \subseteq \bigcap_{j \geq 1} \mathcal{C}^j = \mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)})$. ∎

The above proof gives a sufficient condition for finding initial distributions which lead to an aggregated homogeneous Markov chain (with fixed t.p.m.)

Corollary 3.5 *Let $\widehat{P}$ be any stochastic matrix. We define the polyhedral cone $\mathcal{C}^1$ associated with $\widehat{P}$ as in* (2) *and suppose that it is distinct from $\{0\}$. If there exists a polyhedral cone $\mathcal{C} \subseteq \mathcal{C}^1$ which is decomposable in $\bigoplus_{l \in \hat{S}} \mathcal{C}_l$ and such that $P \overset{\mathcal{C}}{\geq} 0$ then $\mathcal{C} \subseteq \mathcal{C}_{\mathcal{M}}(\widehat{P})$.*

Using decomposability property, Theorem 3.4 can be reformulated with "local" characteristics. That gives the main result of this section.

Theorem 3.6 *The set* $\mathcal{A}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) \neq \emptyset$ *or* $\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) \neq \{0\}$ *if and only if there exists a family of* M *polyhedral cones* $(\mathcal{C}_l)_{l \in \hat{S}}$*, distinct from* $\{0\}$*, such that*

$$\begin{cases} \mathcal{C}_l \subseteq \mathcal{C}_l^1 \subseteq \mathbb{R}_+^{n(l)} & \forall l \in \hat{S}, \\ \mathcal{C}_l P_{C(l)C(m)} \subseteq \mathcal{C}_m & \forall l, m \in \hat{S}. \end{cases}$$

Remark 2 The polyhedral cone $\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)})$, when it is distinct from $\{0\}$, satisfies the conditions of Theorem 3.6. From Corollary 3.5, it follows that $\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)})$ is the largest polyhedral sub-cone of $\mathcal{C}^1$ which is positively invariant by P and decomposable in M polyhedral cones. However, it may exist a smaller polyhedral sub-cone of $\mathcal{C}^1$ than $\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)})$ which is only positively invariant. Indeed, let us return to the Example 1. We consider the transition probability matrix P_1 and the partition $\mathcal{P} = \{C(1) = \{1\}, C(2) = \{2,3\}\}$ of $\{1,2,3\}$. Recall that $\mathcal{A}_{\mathcal{M}}(\widehat{P}^{(e_1)}) = \mathcal{A}^1 = \mathcal{C}_{\mathcal{M}} \cap \mathcal{A}$ with $\mathcal{C}_{\mathcal{M}} = \mathcal{C}^1 = \{\lambda(1,0,0) + \mu(0,1,2) \ / \ \lambda, \mu \geq 0\}$. Let us define $\mathcal{C} = \mathcal{C}_{\mathcal{M}} P \stackrel{\text{def}}{=} \{\beta P \ / \ \beta \in \mathcal{C}_{\mathcal{M}}\}$. The two extremals of the polyhedral cone $\mathcal{C}$ are $r_1 = (1,1,2)$ and $r_2 = (21,5,10)$ that is $\mathcal{C} = \{\lambda r_1 + \mu r_2 \ / \ \lambda, \mu \geq 0\}$. We can check that $\mathcal{C} \subset \mathcal{C}_{\mathcal{M}} = \mathcal{C}^1$ and that $\mathcal{C}$ is positively invariant by P (since $\mathcal{C}P \subseteq \mathcal{C}_{\mathcal{M}} P = \mathcal{C}$).

$\triangle$

Theorems 3.4 and 3.6 can be associated with the Lemma 3.3 to give the following corollary.

Corollary 3.7 *If* $\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) \neq \{0\}$ *then it contains a nonnegative left eigenvector corresponding to a nonnegative eigenvalue of* P.

For each $l \in \hat{S}$*, the cone* $R_l \mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)})$ *of* $\mathbb{R}_+^{n(l)}$ *contains a nonnegative left eigenvector corresponding to the nonnegative eigenvalue* $\widehat{P}^{(\alpha)}(l,l)$ *of* $P_{C(l)C(l)}$.

proof. The first assertion is a direct consequence from positive invariance of $\mathcal{C}_{\mathcal{M}}$ and from Lemma 3.3. Since $[R_l \mathcal{C}_{\mathcal{M}}] P_{C(l)C(l)} \subseteq [R_l \mathcal{C}_{\mathcal{M}}]$ (Theorem 3.6), there exists a nonzero left eigenvector v_l in $R_l \mathcal{C}_{\mathcal{M}}$ associated with an eigenvalue ρ_l of $P_{C(l)C(l)}$ with Lemma 3.3; and if so, we have with vector $f(R_l^{-1} v_l, C(l))$ as initial distribution for the original chain $\widehat{P}^{(\alpha)}(l,l) = \mathbb{P}_{f(R_l^{-1} v_l, C(l))}(X_1 \in C(l)) = v_l P_{C(l)C(l)} 1^{\mathrm{T}} / v_l 1^{\mathrm{T}} = \rho_l$. ∎

Remark 3 The fact that $\widehat{P}^{(\alpha)}(l,l)$ is an eigenvalue of $P_{C(l)C(l)}$ completely generalizes the result given in Ledoux (1993) for an irreducible original chain. It was based on the fact that Markovian property induces geometric sojourn times in each classe $C(l)$ and on the Jordan's canonical form of a matrix. We recall that $\widehat{P}^{(\alpha)}(l,l)$ may not be $\rho(P_{C(l)C(l)})$ (see Example 1.)

$\triangle$

From Corollary 3.7, a cone which may fulfill the sufficient condition for weak lumpability given in Corollary 3.5 is the one which can be formed from a family $\{v_l, \ l \in \hat{S}\}$ of nonnegative left eigenvectors (and nonzero vectors) associated with the family of sub-matrices $\{P_{C(l)C(l)}, \ l \in \hat{S} \ \}$. Let us set

$$\mathcal{C}_v \stackrel{\text{def}}{=} \operatorname{Cone}(\{R_l^{-1} v_l, \ l \in \hat{S}\}) = \bigoplus_{l \in \hat{S}} \operatorname{Cone}(R_l^{-1} v_l). \tag{5}$$

Since $v_l \neq 0$ for every $l \in \hat{S}$, the following stochastic matrix, denoted by $\widehat{P}$, can be defined by

$$\widehat{P}_l = f(R_l^{-1} v_l, C(l)) \widetilde{P} \quad \forall l \in \hat{S}.$$

Thus, we deduce from Corollary 3.5 that

$$P \overset{\mathcal{C}_v}{\geq} 0 \Longrightarrow \mathcal{C}_v \subseteq \mathcal{C}_{\mathcal{M}}(\widehat{P}).$$

Such a situation raises with the *exact lumpability* property described in Schweitzer (1984). Indeed, it corresponds to assume that for all $l \in \hat{S}$, $\sum_{i\in C(l)} P(i,j)$ depends only on l and m for every $j \in C(m)$. Consequently, for every $l \in \hat{S}$, the vector $v_l = 1_{C(l)}$ is a left eigenvector corresponding to the eigenvalue $\sum_{i\in C(l)} P(i,j)$ of nonzero matrix $P_{C(l)C(l)}$ such that

$$1_{C(l)}P_{C(l)C(m)} = [\sum_{i\in C(l)} P(i,j)]\ 1_{C(m)} \qquad \forall m \in \hat{S},$$

and we have $\mathcal{C}_v \subseteq \mathcal{C}_{\mathcal{M}}(\widehat{P})$ according to the previous discussion. The fact that exact lumpability implies weak lumpability is well known.

The following corollary takes advantage of the identification of the sub-cone $\mathcal{C}_v$ of $\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) = \mathcal{C}^N$ defined in (5) and of the affine independence of the M vectors $R_l^{-1}v_l$ (i.e. dim $\mathcal{C}_v = M$).

Corollary 3.8 *We have $\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) \neq \{0\}$ if and only if $\mathcal{C}_l^{N-M} \neq \{0\}$ for all $l \in \hat{S}$. In that case, we have*

$$\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) = \bigoplus_{l\in\hat{S}} R_l^{-1}\mathcal{C}_l^{N-M}.$$

When $C(l)$ is is an irreducible state class of S then Corollary 3.7 and Lemma 3.3 give the following additional assertions. The final part also uses the positive invariance properties of cones $R_l\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)})$ $(l \in \hat{S})$ given in Theorem 3.6.

Corollary 3.9 *Let us assume that $\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)}) \neq \{0\}$. If $P_{C(l)C(l)}$ is an irreducible matrix then $R_l\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)})$ contains only one left eigenvector v_l of $P_{C(l)C(l)}$ and this vector is positive. Moreover, $\widehat{P}^{(\alpha)}(l,l)$ is the spectral radius of $P_{C(l)C(l)}$. Thus, we necessarily have*

$$\widehat{P}_l^{(\alpha)} = f(R_l^{-1}v_l, C(l))\widetilde{P}$$

with any initial distribution in $\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)})$ whose support D is such that $D \cap C(l) \neq \emptyset$; moreover, for any state class $C(m)$ which can be accessed from a state of $C(l)$ (i.e. there exists a possible sequence $(C(l), C(i_1), \ldots, C(i_k), C(m))$ for some e_i with $i \in C(l)$), we have $v_lP_{C(l)C(i_1)}\cdots P_{C(i_k)C(m)} \neq 0$ is in $R_m\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)})$ and

$$\widehat{P}_m^{(\alpha)} = f(R_m^{-1}[v_lP_{C(l)C(i_1)}\cdots P_{C(i_k)C(m)}], C(m))\widetilde{P}.$$

Remark 4 We have shown that $\widehat{P}^{(\alpha)}(l,l)$ is necessarily the spectral radius of $P_{C(l)C(l)}$ when $C(l)$ is irreducible. That improves the results given in Section 4 from Ledoux (1993).

△

4 Lumpability of reducible Markov chains

The previous results can be applied to the aggregation of Markov chains with respect to a partition $\mathcal{P}$ which is a refinement of the partition of S corresponding to the usual communication equivalence relation. This partition is denoted by $\mathcal{I} = (I_k)_{k\in J}$ throughout this section. The elements of $\mathcal{I}$ are called the communication classes or the irreducibility classes and $|J|$ denotes the cardinality of $\mathcal{I}$. Such a state class I_k induces an irreducible sub-matrix $P_{I_kI_k}$ of P. Consequently, we can associated with each state class I_k, the unique stochastic left eigenvector v_k of $P_{I_kI_k}$ corresponding to the spectral radius of $P_{I_kI_k}$. Throughout this section, we assume that the states of S are ordered such that P is a lower block-triangular matrix

$$P = \begin{pmatrix} P_{I_1I_1} & 0 & \cdots & 0 \\ * & P_{I_2I_2} & \ddots & \vdots \\ * & * & \ddots & 0 \\ * & * & * & P_{I_{|J|}I_{|J|}} \end{pmatrix}.$$

Partition $\mathcal{P}$ is a *refinement* of the partition $\mathcal{I}$ if $\forall l \in \hat{S}$, $\exists! k \in J$ such that $C(l) \subseteq I_k$. For each $k \in J$, there exists $L_k \subseteq \hat{S}$ such that $I_k = \uplus_{l \in L_k} C(l)$. Any nonnegative vector β on I_k can be seen as an element of $\otimes_{l \in L_k} \mathbb{R}_+^{n(l)}$. Consequently, we denote the vector on S, $\sum_{l \in L_k} R_l^{-1} \beta_{C(l)}$, by $R_{I_k}^{-1}\beta$.

Definition 4.1 *A family of communication classes $(I_{i_0}, \cdots, I_{i_n})$ is called a path if each class $I_{i_{k-1}}$ has an access to the class I_{i_k} for $k = 1, \ldots, n$ (that is there exists a state in $I_{i_{k-1}}$ which communicates with a state of I_{i_k}.) We call I_{i_0} the starting point and I_{i_n} the end point of the path.*

Theorem 4.2 *Let us assume that partition $\mathcal{P}$ is a refinement of the partition $\mathcal{I} = (I_k)_{k \in J}$ of S. We have the family of vectors $(v_k)_{k \in J}$, v_k being the stochastic left eigenvector associated with the spectral radius of matrix $P_{I_kI_k}$. If $\alpha \in \mathcal{A}$ is such that $\alpha_{I_k} \neq 0$ and $agg(\alpha, P, \mathcal{P})$ is a homogeneous Markov chain, then, for any m such that I_m belongs to a path with starting point I_k, we have $agg(R_{I_m}^{-1} v_m, P, \mathcal{P})$ is a homogeneous Markov chain and for all $l \in \hat{S}$ such that $C(l) \subseteq I_m$:*

$$\widehat{P}_l^{(\alpha)} = f(R_{I_m}^{-1} v_m, C(l))\widetilde{P}; \tag{6}$$

moreover the family $\mathcal{F}_m$ composed of vectors $R_l^{-1}(v_m)_{C(l)}$ is such that $\mathrm{Cone}(\mathcal{F}_m) \subseteq \mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)})$.

Remark 5 The previous theorem can be interpreted as follows: if a state of a class I_k is allowed to be an initial state of our Markovian model then all the rows of matrix $\widehat{P}$ corresponding to the state classes of the $\mathcal{P}$ included in I_k or in the element of a path with starting point I_k, are necessarily given by formula (6) and depend only on $\mathcal{I}$ and P.

$\triangle$

proof. We have $I_k = \uplus_{l \in L_k} C(l)$ for some $L_k \subseteq \hat{S}$. We deduce from Theorem 3.6 that if $agg(\alpha, P, \mathcal{P})$ is a homogeneous Markov chain then there exists a pointed polyhedral cone, defined by $\mathcal{C}_{I_k} = \bigoplus_{l \in L_k} R_l^{-1} \mathcal{C}_l$, such that cone $R_{I_k}\mathcal{C}_{I_k}$ is positively invariant by the irreducible matrix $P_{I_kI_k}$. Lemma 3.3 states that this last cone contains the stochastic left eigenvector v_k corresponding to the spectral radius of $P_{I_kI_k}$. Since all the distributions of cone $\mathcal{C}_{I_k}$ lead to an aggregated Markov chain with the same t.p.m. $\widehat{P}^{(\alpha)}$, we derive that $agg(R_{I_k}^{-1} v_k, P, \mathcal{P})$ is a homogeneous Markov chain and that $\widehat{P}_l^{(\alpha)} = f(R_{I_k}^{-1} v_k, C(l))\widetilde{P}$ for every $l \in \hat{S}$ such that $C(l) \subseteq I_k$.

Let us now consider a path with starting point I_k and assume that there exists a distribution α such that $\alpha_{I_k} \neq 0$ and $agg(\alpha, P, \mathcal{P})$ is Markov. The chain $agg(R_{I_k}^{-1} v_k, P, \mathcal{P})$ is also a homogeneous Markov chain from the first part of the proof. Since $\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)})$ is positively invariant by P, we have that for any $n \geq 0$, $R_{I_k}^{-1} v_k P^n \in \mathcal{C}_{\mathcal{M}}$. The class I_k communicate with any element I_i of the path. Consequently, let i be fixed, there exists $n_i > 0$ such that for $w_i = R_{I_k}^{-1} v_k P^{n_i}$, $R_{I_i} w_i \neq 0$ and $agg(w_i, P, \mathcal{P})$ is a homogeneous Markov chain. The rows of matrix $\widehat{P}$ corresponding to the classes of $\mathcal{P}$ included in I_i are necessarily given by $\widehat{P}_l^{(\alpha)} = f(R_{I_i}^{-1} v_i, C(l))\widetilde{P}$, from the first part of the proof.

The last part of the theorem follows from the fact that each $R_l v_m$ is in $R_l \mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)})$ (since $R_{I_m}^{-1} v_m \in \mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)})$) and from the conical property of $\mathcal{C}_{\mathcal{M}}(\widehat{P}^{(\alpha)})$. ∎

Remark 6 If we wish that all initial distributions on $\mathcal{A}$ lead to an aggregated homogeneous Markov chain (strong lumpability property) then, for all $k \in \hat{S}$, there must exist such a distribution whose support contains states from class I_k. Thus, Theorem 4.2 allows us to conclude that all the aggregated chains share the same transition probability. Consequently, the unicity condition on this matrix required in Theorem 2.6 can be dropped.

$\triangle$

Algorithm

We propose to briefly specify an algorithm which can be used in the present context to compute probability distributions leading to an aggregated homogeneous Markov chain. However, an analogous algorithm also works for a general finite Markov chain from Section 2. The algorithm below follows the same lines as the Rubino and Sericola's one for an irreducible stochastic matrix P. We just want to point out here that all computation may be performed "locally", that is in working with characteristics corresponding to the state classes of $\mathcal{P}$. Note that the set $\mathcal{C}_{\mathcal{M}}(\widehat{P})$ resulting from the algorithm contains all the initial distributions leading to an aggregated homogeneous Markov chain with respect to $\mathcal{P}$. In the process of computation, a set $\mathcal{C}_l^j$ may be found to be $\{0\}$. In such a case, we have $\mathcal{C}_{I_k} = \bigoplus_{m \in L_k} R_m^{-1} \mathcal{C}_m^j = \{0\}$ if $C(l) \subseteq I_k = \uplus_{m \in L_k} C(m)$. Consequently, the states of I_k are ignored in the sequel.

Preliminary step:

Compute the stochastic left eigenvectors $\{v_k, k \in J\}$ of matrices $\{P_{I_k I_k},\ k \in J\}$.
Compute matrix $\widehat{P}$ with for any $l \in \hat{S}$, $\widehat{P}_l = f(R_{I_m}^{-1} v_m, C(l))\widetilde{P}$ if $C(l) \subseteq I_m$.

Loop $j = 1, \ldots, N-M$
First step:

Form the polyhedral cones $\mathcal{C}_l^j$ $(l \in \hat{S})$ from the recursive definition (4).
Compute the extremals for each $\mathcal{C}_l^j$, that is the minimal finite family of nonzero vectors r_i such that $\mathcal{C}_l^j = \text{Cone}(r_1, \ldots, r_n)$.

Second step:

If for each $l \in \hat{S}$ such that $\mathcal{C}_l^j \neq \{0\}$, $(i = 1, \ldots, n$: $r_i P_{C(l)C(l)} \in \mathcal{C}_m^j)$ for all $m \in \hat{S}$
then $\mathcal{C}_{\mathcal{M}}(\widehat{P}) = \bigoplus_{l \in \hat{S}} R_l^{-1} \mathcal{C}_l^j$ **exit**

Endloop

Example 2 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 transition probability matrix P is given by:

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

The partition in communication classes is $\mathcal{I} = \{I_1 = \{1,2,3\}\}, I_2 = \{4,5,6,7\}\}$. The stochastic left eigenvectors corresponding to spectral radius of respective matrices $P_{I_1 I_1}$ and $P_{I_2 I_2}$ are $v_1 = (7/16, 3/16, 6/16)$,

$v_2 = (1/4, 1/4, 1/4, 1/4)$. Matrix $\widehat{P}$ is given by

$$\widehat{P}_1 = (1/4, 3/4, 0, 0), \qquad \widehat{P}_2 = (7/12, 5/12, 0, 0),$$
$$\widehat{P}_3 = (1/7, 0, 3/14, 9/14), \qquad \widehat{P}_4 = (5/12, 1/24, 2/9, 2/3).$$

Let us form the matrices H_1, H_2, H_3, H_4:

$$H_1 = H_3 = 0, \quad H_2 = \begin{pmatrix} -7/12 & 7/12 & 0 & 0 \\ 7/24 & -7/24 & 0 & 0 \end{pmatrix}, \quad H_4 = \begin{pmatrix} 1/18 & 0 & -1/18 & 0 \\ 1/72 & -1/24 & -7/72 & 1/8 \\ -5/72 & 1/24 & 11/72 & -1/8 \end{pmatrix}.$$

The nonnegative solutions to the homogeneous system associated with each previous matrix define the four following polyhedral cones $\mathcal{C}_1^1, \mathcal{C}_2^1, \mathcal{C}_3^1, \mathcal{C}_4^1$ (see formula (4)):

$$\mathcal{C}_1^1 = \mathcal{C}_3^1 = \mathbb{R}_+; \quad \mathcal{C}_2^1 = \mathrm{Cone}(v_2'); \quad \mathcal{C}_4^1 = \mathrm{Cone}(v_4');$$

with $v_2' = (1,2) = R_2 v_1 / R_2 v_1 1^{\mathrm{T}}$ and $v_4' = (1,1,1) = R_4 v_2 / R_4 v_2 1^{\mathrm{T}}$. Note that v_2' (resp. v_4') is the positive left eigenvector (up to a constant multiplicative) corresponding to the spectral radius $\widehat{P}(2,2) = 5/12$ (resp. $\widehat{P}(4,4) = 2/3$) of the irreducible matrix $P_{C(2)C(2)}$ (resp. $P_{C(4)C(4)}$.) It is easy to check that the conditions of the Theorem 3.6 are met and thus $\mathcal{C}_{\mathcal{M}}(\widehat{P}) \neq \{0\}$. If we construct the cone $\mathcal{C}_{v'} = \mathrm{Cone}(\{e_1, (R_2^{-1} v_2'), e_3, (R_4^{-1} v_4')\})$, then we observe that $\mathcal{C}^1 = \mathcal{C}_{v'}$. It follows that $\mathcal{C}_{\mathcal{M}}(\widehat{P}) = \mathcal{C}^1 = \mathcal{C}_{v'}$.

◁

Let us define the following positive vector on S, $v \stackrel{\text{def}}{=} \sum_{k \in J} R_{I_k}^{-1} v_k$, the convex subsets of $\mathbb{R}^N$

$$\mathcal{C}_v = \bigoplus_{l \in \hat{S}} \mathrm{Cone}(R_l^{-1} v_{C(l)}), \qquad \mathcal{A}_v = \mathcal{C}_v \cap \mathcal{A}$$

and matrix $\widehat{P}$ by $\widehat{P}_l = f(v, C(l))\widetilde{P}$ for all $l \in \hat{S}$. In the previous example, we found that $\mathcal{C}_{\mathcal{M}}(\widehat{P}) = \mathcal{C}_v$ or $\mathcal{A}_{\mathcal{M}}(\widehat{P}) = \mathcal{A}_v$. We can verify (with Theorem 3.6) that

$$\mathcal{A}_{\mathcal{M}}(\widehat{P}) = \mathcal{A}_v \Longrightarrow f(v, C(l), C(m)) = f(v, C(m)) \ \forall l, m \in \hat{S}. \tag{7}$$

On the other hand, property in the right hand side implies that $\mathcal{A}_v \subseteq \mathcal{A}_{\mathcal{M}}(\widehat{P})$ with Corollary 3.5. Thus, it is a sufficient condition for weak lumpability with matrix $\widehat{P}$ as noted in Kemeny and Snell (1976) for irreducible matrix P. We also note that the right hand side in (7) gives for all $l \in \hat{S}$, $v_{C(l)} P_{C(l)C(l)} = \widehat{P}(l,l) v_{C(l)}$. Thus, for all $l \in \hat{S}$, $v_{C(l)}$ is a positive left eigenvector of matrix $P_{C(l)C(l)}$ corresponding to the eigenvalue $\widehat{P}(l,l)$. It can be useful to know when the converse implication of (7) holds. It is shown to be valid in Peng (1995) under the irreducible assumption for the initial matrix P and the additional condition (Γ):

$$(\Gamma)\text{: the column vectors of matrices } P^k V \ (k \geq 0) \text{ span } \mathbb{R}^N$$

where V is the $N \times M$ matrix defined by $V(i,l) = 1$ if $i \in C(l)$ and 0 otherwise. The previous comments precise some relation between the various equivalent conditions given in Theorem 3.1 from Peng (1995). Since this theorem is based only on the condition (Γ) and the unicity of the t.p.m. associated with any aggregated chain from $\mathcal{A}_{\mathcal{M}}(\widehat{P})$, it can be directly extended to our context. Note that all Peng's results hold in the context of Section 2.

Theorem 4.3 *Let us assume that partition $\mathcal{P}$ is a refinement of the partition $\mathcal{I}$ of S. Under the condition (Γ), the following are equivalent:*

1. *$agg(v/v1^{\mathrm{T}}, P, \mathcal{P})$ satisfies to the Chapman-Kolmogorov equations;*

2. $f(v, C(l), C(m)) = f(v, C(m))$ *for all* $l, m \in \hat{S}$*;*

3. $\mathcal{A}_{\mathcal{M}}(\widehat{P}) = \mathcal{A}_v$.

proof. Let α be a probability distribution of $\mathcal{A}$. Let us define the lth row, for all $l \in \hat{S}$, of the $M \times N$ matrix U_α by $f(\alpha, C(l))$. Note that $U_v P V U_v = U_v P$ is equivalent to the property reported in statement 2.

First, we show that statement 1 implies the second one. If $agg(v/v1^{\mathrm{T}}, P, \mathcal{P})$ satisfies to the Chapman-Kolmogorov equations:

$$\widehat{P}^k = U_v P^k V \quad \forall k \geq 1.$$

Consequently, we have for all $k \geq 0$

$$\begin{aligned} U_v P^{k+1} V &= \widehat{P}^{k+1} \\ &= \widehat{P}\widehat{P}^k \\ &= \widehat{P} U_v P^k V. \end{aligned}$$

We get $\forall k \geq 0$, $[U_v P - \widehat{P} U_v] P^k V = 0$. Under the condition (Γ), we conclude to

$$U_v P - \widehat{P} U_v = 0 \Longleftrightarrow U_v P = U_v P V U_v.$$

Now, let us assume that $U_v P V U_v = U_v P$. As previously noted, it implies that $\mathcal{C}_v$ is positively invariant by matrix P and $\mathcal{C}_v \subseteq \mathcal{C}_{\mathcal{M}}(\widehat{P})$. Consequently, X is weak lumpable and $\mathcal{A}_v \subseteq \mathcal{A}_{\mathcal{M}}(\widehat{P})$. It remains to state that under the condition (Γ), $\mathcal{C}_v = \mathcal{C}_{\mathcal{M}}(\widehat{P})$. Let α be a vector in $\mathcal{A}_{\mathcal{M}}(\widehat{P})$. Since $agg(\alpha, P, \mathcal{P})$ is a homogeneous Markov chain, $agg(\alpha, P, \mathcal{P})$ satisfies to the Chapman-Kolmogorov equations:

$$\widehat{P}^k = U_\alpha P^k V \quad \forall k \geq 1.$$

Consequently, under (Γ), we obtain in the same manner as in the first part of the proof

$$U_\alpha P = U_\alpha P V U_\alpha.$$

In a second part, we show that $U_\alpha P = U_v P$ follows from the the unicity of the transition matrix for any aggregated chain from an initial distribution $\alpha \in \mathcal{A}_{\mathcal{M}}(\widehat{P})$. We necessarily have for all $k \geq 0$

$$\begin{aligned} (\widehat{P}^{(\alpha)})^{k+1} = \widehat{P}^{k+1} &\Longleftrightarrow U_\alpha P^{k+1} V = U_v P^{k+1} V \\ &\Longleftrightarrow [U_\alpha P - U_v P] P^k V = 0. \end{aligned}$$

We deduce from the condition (Γ) that $U_\alpha P = U_v P$, which implies that $f(\alpha^{C(l)} P, C(m)) = f(v^{C(l)} P, C(m))$ for all $l, m \in \hat{S}$.

Combining the both parts, we have for all $\alpha \in \mathcal{A}_{\mathcal{M}}(\widehat{P})$, $\forall l, m \in \hat{S}$

$$\left\{ \begin{array}{l} f(\alpha, C(l), C(m)) = f(\alpha, C(m)), \\ f(\alpha, C(l), C(m)) = f(v, C(l), C(m)). \end{array} \right.$$

Consequently, we have obtained $[\forall l \in \hat{S}\ f(\alpha, C(l)) = f(v, C(l))]$ for all $\alpha \in \mathcal{A}_{\mathcal{M}}(\widehat{P})$ that is $\mathcal{A}_{\mathcal{M}}(\widehat{P}) = \mathcal{A}_v$.

This proof holds also in the context of Section 2 because, apart from the condition (Γ), we only need that all aggregated chains concerned with share the same transition probability matrix. ■

Theorem 4.2 of this section can be applied to derive the two main published results on weak lumpability. The first one deals with irreducible matrix P, that is $\mathcal{I}$ reduces to only one class.

Corollary 4.4 (Rubino and Sericola 1991) *The transition probability matrix P of the original chain is assumed to be irreducible. If $agg(\alpha, P, \mathcal{P})$ is a homogeneous Markov chain then $agg(\pi, P, \mathcal{P})$ is also a homogeneous Markov chain where π is the stochastic vector solution to $\pi P = \pi$. The t.p.m. $\widehat{P}$ is the same for any aggregated homogeneous Markov chain and is given by $\widehat{P}_l = f(\pi, C(l))\widetilde{P}, \quad l \in \hat{S}$.*

Remark 7 Let us define the following polyhedral cone $\mathcal{C}_\pi = \bigoplus_{l\in\hat{S}} \text{Cone}(\{R_l^{-1}\pi_{C(l)})$. As noted previously, the positive invariance of the set $\mathcal{C}_\pi$ by P implies that $\mathcal{C}_\pi \subseteq \mathcal{C}_{\mathcal{M}}(\widehat{P})$ and thus we have $\mathcal{A}_\pi = \mathcal{C}_\pi \cap \mathcal{A} \subseteq \mathcal{A}_{\mathcal{M}}(\widehat{P})$. The interest of this condition comes from the fact that it can be checked from the "data" of the problem.

$\triangle$

A second family of Markov chains can also be treated with Theorem 4.2.

Corollary 4.5 (Ledoux et al. 1994) *Let us consider a family of Markov chain with transition probability matrix P such that the partition of S induced by the communication equivalence relation is $\mathcal{I} = \{I_1, I_2\}$: where I_1 contains one absorbing state and I_2 all the transient ones. If there exists $\alpha \in \mathcal{A}$ such that $\alpha_{I_2} \neq 0$ and $agg(\alpha, P, \mathcal{P})$ is a homogeneous Markov chain then $agg((0, v), P, \mathcal{P})$ is also a homogeneous Markov chain with v is the stochastic vector solution to $vP_{I_2I_2} = \rho v$, where ρ is the spectral radius of the matrix $P_{I_2I_2}$. We recall that v is called the quasi-stationary distribution associated with the family $(., P)$. The t.p.m. $\widehat{P}$ is the same for any homogeneous Markov chain $agg(\alpha, P, \mathcal{P})$ with an initial distribution α whose support contains transient states. It is given by $\widehat{P}_1 = e_1$ and $\widehat{P}_l = f((0, v), C(l))\widetilde{P}, \quad l \in \hat{S} \setminus \{1\}$.*

Finally, as noted in Remark 7, if we have the opportunity to make a nontrivial aggregation with respect to the partition $\mathcal{P}$ then that must hold for any probability distribution of the polytope defined by $\mathcal{A}_v = \text{Conv}(\{(0, v)^{C(l)}, \ / \ l \in \hat{S}\})$.

Conclusion

This paper extends to general finite Markov chains the linear system approach used in Abdel-Moneim and Leysieffer (1982), Rubino and Sericola (1991) for weak lumpability problem. In adopting here the viewpoint of positive invariance of polyhedral cones, we propose new results on weak/strong lumpability of a finite Markov chain. Most of our results are expressed with "local" characteristics of the chain, that is to the level of the state classes of the partition. This allows us to derive (or extend) spectral properties associated with exact aggregation. In a general manner, our work specifies some (geometrical) invariant corresponding to the lumpability requirement which are promising for study related problems: investigate formally the weak lumpability of strongly structured Markovian models and analyze sensitivity to the "data" of the exact aggregation feasibility. We do not go into further details here.

References

A.M. Abdel-Moneim and F.W. Leysieffer, Weak lumpability in finite Markov chains, J. Appl. Probab. 19 (1982) 685–691.

A.M. Abdel-Moneim and F.W. Leysieffer, Lumpability for non-irreducible finite Markov chains, J. Appl. Probab. 21 (1984) 567–574.

F. Ball and G. Yeo, Lumpability and marginalisability for continuous-time Markov chains, J. Appl. Probab. 29 (1993) 518–528.

A. Berman and R. Plemmons, Nonnegative Matrices in the Mathematical Sciences (Academic Press, 1979)

J.G. Kemeny and J.L. Snell, Finite Markov chains (Springer-Verlag, 1976).

J. Ledoux, A necessary condition for weak lumpability, Oper. Res. Letters 13 (1993) 165–168.

J. Ledoux, G. Rubino and B. Sericola, Exact aggregation of absorbing Markov processes using quasi-stationary distribution, J. Appl. Probab. 31 (1994) 626–634.

J. Ledoux, On weak lumpability of denumerable Markov chains, Stat. and Probab. Letters 25 (1995) 329–339.

N.F. Peng, On weak lumpability of a finite Markov chain, To appear in Stat. and Probab. Letters (1995)

G. Rubino and B. Sericola, On weak lumpability in Markov chains, J. Appl. Probab. 26 (1989) 446–457.

G. Rubino and B. Sericola, A finite caracterization of weak lumpable Markov processes. Part I: The discrete time case, Stochastic Process. Appl. 38 (1991) 195–204.

P. Schweitzer, Aggregation methods for large Markov chains, in: Iazeolla et al, eds, Mathematical Computer Performance and Reliability (Elsevier-North Holland, 1984).

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399