Stochastic Processes and their Applications 38 (1991) 195-204
North-Holland

A finite characterization of weak lumpable Markov processes. Part I: The discrete time case

Gerardo Rubino and Bruno Sericola
IRISA-INRIA, Campus de Beaulieu, 35042 Rennes, France

Received 10 August 1989
Revised 19 June 1990 and 10 September 1990

We consider an irreducible and homogeneous Markov chain (discrete time) with finite state space. Given a partition of the state space, it is of interest to know if the aggregated process constructed from the first one with respect to the partition is also Markov homogeneous. We give a characterization of this situation by means of a finite algorithm. This algorithm computes the set of all initial probability distributions of the starting homogeneous Markov chain leading to an aggregated homogeneous Markov chain.

Markov chains * aggregation * weak lumpability

1. Introduction

This paper is the natural continuation of the work performed in Rubino and Sericola (1989). To remain compatible with this previous paper, we conserve the same notation. Let us first recall the problem of weak lumping in Markov chains.

Let $X=(X_n)_{n\geq 0}$ be a homogeneous irreducible Markov chain evolving in discrete time. The state space is assumed to be finite and denoted by $E=\{1,2,\ldots,N\}$. The stationary distribution of X is denoted by π. Let us denote by $\mathcal{B}=\{B(1),B(2),\ldots,B(M)\}$ a partition of the state space and by $n(m)$ the cardinal of $B(m)$. We assume the states of E ordered such that

$$\begin{aligned}
&B(1)=\{1,\ldots,n(1)\},\\
&\quad\vdots\\
&B(m)=\{n(1)+\cdots+n(m-1)+1,\ldots,n(1)+\cdots+n(m)\},\\
&\quad\vdots\\
&B(M)=\{n(1)+\cdots+n(M-1)+1,\ldots,N\}.
\end{aligned}$$

With the given process X we associate the aggregated stochastic process Y with values in the set $F=\{1,2,\ldots,M\}$, defined by

$$Y_n=m \overset{\text{def}}{\iff} X_n\in B(m)\quad \text{for all } n\geq 0.$$

It is easily checked from this definition and the irreducibility of X that the process Y obtained is also irreducible in the following sense: for all $m \in F$, for all $l \in F$ such that $\mathbb{P}(Y_0 = l) > 0$, there exists $n > 0$ such that $\mathbb{P}(Y_n = m \mid Y_0 = l) > 0$.

The Markov property of X means that given the state in which X is at the present time, the future and the past of the process are independent. Clearly, this is no longer true for the aggregated process Y in the general case (see the example below). This paper deals with the conditions under which the process Y is also a homogeneous Markov chain.

The homogeneous Markov chain X is given by its transition probability matrix P and its initial distribution α; we shall denote it by (α, P) when necessary. The (i, j) entry of matrix P is denoted by $P(i, j)$. We shall denote by $\mathrm{agg}(\alpha, P, \mathscr{B})$ the aggregated process constructed from (α, P) over the partition $\mathscr{B}$. Let us denote by $\mathscr{A}$ the set of all probability vectors with N entries.

Of course, it is possible to have the situation in which $\mathrm{agg}(\alpha, P, \mathscr{B})$ is a Markov homogeneous chain for any $\alpha \in \mathscr{A}$. This happens, for instance, in the following example where $E = \{1, 2, 3\}$, $B(1) = \{1\}$, $B(2) = \{2, 3\}$ and

$$P = \left(\begin{array}{c|cc} 1-p & p & 0 \\ \hline q & 0 & 1-q \\ q & 1-q & 0 \end{array}\right)$$

with $0 < p \leqslant 1$ and $0 < q < 1$. In this case, X is said to be *strongly lumpable* with respect to the partition $\mathscr{B}$ and this is a well known property with a simple characterization. For every $i \in E$ and $m \in F$, let us denote by $P(i, B(m))$ the transition probability of passing in one step from state i to the subset $B(m)$ of E, that is $P(i, B(m)) \stackrel{\text{def}}{=} \sum_{j \in B(m)} P(i, j)$. Then, X is strongly lumpable with respect to the partition $\mathscr{B}$ if and only if for every pair of elements l, $m \in F$, the probability $P(i, B(m))$ has the same value for every $i \in B(l)$. This common value is the transition probability from state l to state m for the aggregated homogeneous Markov chain Y.

The opposite situation in which for any initial distribution α, the process $\mathrm{agg}(\alpha, P, \mathscr{B})$ is not Markov is illustrated by the following example. Consider again a three state chain with $B(1) = \{1\}$, $B(2) = \{2, 3\}$ and

$$P = \left(\begin{array}{c|cc} 1-p & p & 0 \\ \hline 0 & 0 & 1 \\ 1 & 0 & 0 \end{array}\right).$$

For any value of p with $0 < p \leqslant 1$ and for any initial probability distribution, we have

$$\mathbb{P}(X_{n+1} \in B(1) \mid X_n \in B(2), X_{n-1} \in B(1)) = 0$$

and

$$\mathbb{P}(X_{n+1} \in B(1) \mid X_n \in B(2), X_{n-1} \in B(2)) = 1.$$

In Kemeny and Snell (1976), the authors show that a third situation is possible, in which there exists a proper subset of $\mathscr{A}$ denoted here by $\mathscr{A}_{\mathscr{M}}$ such that for any

$\alpha \in \mathscr{A}_{\mathscr{M}}$, the process agg$(\alpha, P, \mathscr{B})$ is Markov homogeneous while for any $\alpha \notin \mathscr{A}_{\mathscr{M}}$ the aggregated process is not Markov. In this more general situation, X is said to be *weakly lumpable* with respect to the given partition $\mathscr{B}$ (see Kemeny and Snell, 1976, for some examples). Moreover, in the same reference a sufficient condition to weak lumpability is given.

In Abdel-Moneim and Leysieffer (1982), the authors analyse the problem of the characterization of weak lumpability. An interesting technique is introduced but we have shown in Rubino and Sericola (1989) that their characterization is wrong.

The work performed here is an extension of Rubino and Sericola (1989). We obtain a finite characterization of weak lumpability by means of an algorithm which computes the set $\mathscr{A}_{\mathscr{M}}$. In Rubino and Sericola (1989), $\mathscr{A}_{\mathscr{M}}$ is given as an infinite intersection of sets. The main result obtained in this paper is the reduction of this infinite intersection to a finite one.

The paper is organized as follows. The next section gives the background material and notation and recalls the main results obtained in Rubino and Sericola (1989). In Section 3, we show how the searched set $\mathscr{A}_{\mathscr{M}}$ can be obtained by means of a finite algorithm. Section 4 presents an example in which the set $\mathscr{A}_{\mathscr{M}}$ is computed and Section 5 concludes the paper.

2. Notation and main results of Rubino and Sericola (1989)

By convention, the vectors will be row vectors. Column vectors will be indicated by means of the transposition operator $(\cdot)^{\mathrm{T}}$. A vector with all its entries equal to 1 will be denoted simply by 1; its dimension will be defined by the context.

For each $l \in F$ and $\alpha \in \mathscr{A}$ we denote by $T_l \cdot \alpha$ the 'restriction' of α to the subset $B(l)$. That is, $T_l \cdot \alpha$ is a vector with $n(l)$ entries and its ith entry is $(T_l \cdot \alpha)(i) = \alpha(n(1)+\cdots+n(l-1)+i)$, $i = 1, 2, \ldots, n(l)$. When $T_l \cdot \alpha \neq 0$, we denote by $\alpha^{B(l)}$ the vector of $\mathscr{A}$ defined by $\alpha^{B(l)}(i) = \alpha(i)/\|T_l \cdot \alpha\|$ if i belongs to $B(l)$, 0 otherwise, where $\|\gamma\|$ denotes the sum of the components of the nonnegative real vector γ. Each time when in the sequel we shall write a vector of the form $\gamma^{B(m)}$ with $\gamma \in \mathscr{A}$, we implicitly mean that this vector is defined, i.e. $T_m \cdot \gamma \neq 0$.

To clarify this notation, let us give some examples. Suppose that $N = 5$ and $\mathscr{B} = \{B(1), B(2)\}$ with $B(1) = \{1, 2, 3\}$ and $B(2) = \{4, 5\}$. Then for $\alpha = (\frac{1}{10}, \frac{1}{10}, \frac{1}{5}, \frac{2}{5}, \frac{1}{5})$:

$$\alpha^{B(1)} = (\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{2}, 0, 0), \qquad \alpha^{B(2)} = (0, 0, 0, \tfrac{2}{3}, \tfrac{1}{3}),$$

$$T_1 \cdot \alpha = (\tfrac{1}{10}, \tfrac{1}{10}, \tfrac{1}{5}), \qquad T_2 \cdot \alpha = (\tfrac{2}{5}, \tfrac{1}{5}),$$

$$T_1 \cdot \alpha^{B(1)} = (\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{2}), \qquad T_2 \cdot \alpha^{B(2)} = (\tfrac{2}{3}, \tfrac{1}{3});$$

for $\alpha = (\frac{1}{2}, \frac{1}{2}, 0, 0, 0)$:

$$\alpha^{B(1)} = (\tfrac{1}{2}, \tfrac{1}{2}, 0, 0, 0) \quad \text{and} \quad \alpha^{B(2)} \text{ is not defined since } T_2 \cdot \alpha = (0, 0).$$

Recall the definition of the main object of this analysis.

$$\mathcal{A}_{\mathcal{M}} \stackrel{\text{def}}{=} \{\alpha \in \mathcal{A} \mid Y = \operatorname{agg}(\alpha, P, \mathcal{B}) \text{ is a homogeneous Markov chain}\}.$$

For $l \in F$, we shall denote by $\tilde{P}_l$ the $n(l) \times M$ matrix defined by

$$\tilde{P}_l(j, m) \stackrel{\text{def}}{=} P(n(1) + \cdots + n(l-1) + j, B(m)), \quad 1 \leqslant j \leqslant n(l), \quad m \in F.$$

In other words, $\tilde{P}_l(j, m)$ is the probability of passing in one step from the jth state in $B(l)$ to the set $B(m)$.

When the set $\mathcal{A}_{\mathcal{M}}$ is not empty, the transition probability matrix of the Markov chain Y, denoted by $\hat{P}$, is the same for every $\alpha \in \mathcal{A}_{\mathcal{M}}$. In this case, its lth row $\hat{P}_l$ can be computed by $\hat{P}_l = (T_l \cdot \pi^{B(l)})\tilde{P}_l$ (see Rubino and Sericola, 1989).

To illustrate these definitions, consider the following matrix:

$$P = \left(\begin{array}{ccc|c} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{8} & \frac{3}{8} & \frac{1}{4} & \frac{1}{4} \\ \frac{3}{8} & \frac{1}{8} & \frac{1}{4} & \frac{1}{4} \\ \hline \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array}\right)$$

with $\mathcal{B} = \{B(1), B(2)\}$, $B(1) = \{1, 2, 3\}$ and $B(2) = \{4\}$.

We have

$$\tilde{P}_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{4} \\ \frac{3}{4} & \frac{1}{4} \end{pmatrix}, \qquad \tilde{P}_2 = (\tfrac{3}{4} \ \tfrac{1}{4}),$$

$$\pi = (\tfrac{3}{13}, \tfrac{3}{13}, \tfrac{3}{13}, \tfrac{4}{13}) \quad \text{and} \quad \hat{P} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{3}{4} & \frac{1}{4} \end{pmatrix}.$$

Let α be an element of $\mathcal{A}$, k a positive integer and $i_1, i_2, \ldots, i_k$ a sequence of k elements of F. We define recursively the vector $f(\alpha, B(i_1), \ldots, B(i_k)) \in \mathcal{A}$ by

$$f(\alpha, B(i_1)) = \alpha^{B(i_1)},$$

$$f(\alpha, B(i_1), \ldots, B(i_h)) = (f(\alpha, B(i_1), \ldots, B(i_{h-1}))P)^{B(i_h)}, \quad 2 \leqslant h \leqslant k,$$

whenever these operations make sense. The reader can check the following probabilistic interpretation of f. If $\beta = f(\alpha, B(i_0), \ldots, B(i_n))$ then

$$\beta(j) = \mathbb{P}(X_n = j \mid X_n \in B(i_n), \ldots, X_0 \in B(i_0)).$$

That is, β is the probability distribution of X_n given $(X_0 \in B(i_0), \ldots, X_n \in B(i_n))$.

Using this definition, we define the basic sequence of sets

$$\mathcal{A}^1 \stackrel{\text{def}}{=} \{\alpha \in \mathcal{A} \mid (T_l \cdot \alpha^{B(l)})\tilde{P}_l = \hat{P}_l \text{ for every } l \in F\}$$

and for $j \geqslant 2$,

$$\mathcal{A}^j \stackrel{\text{def}}{=} \{\alpha \in \mathcal{A} \mid \forall \beta = f(\alpha, B(i_1), \ldots, B(i_k)) \text{ with } k \leqslant j, \text{ we have } \beta \in \mathcal{A}^1\}.$$

As an example, the reader can check that for the previous 4-dimensional transition probability matrix, the set $\mathcal{A}^1$ can be written as

$$\mathcal{A}^1 = \{\alpha \in \mathcal{A} \mid \alpha = \lambda(\tfrac{1}{3}, t, \tfrac{2}{3} - t) + (1 - \lambda)(0, 0, 0, 1), 0 \leqslant t \leqslant \tfrac{2}{3}, 0 \leqslant \lambda \leqslant 1\}.$$

We have then the basic result (proved in Rubino and Sericola, 1989),

$$\mathcal{A}_{\mathcal{M}} = \bigcap_{j \geqslant 1} \mathcal{A}^j.$$

In the same paper, some properties of $\mathcal{A}_{\mathcal{M}}$ are given. Let us say that a subset $\mathcal{U}$ of $\mathcal{A}$ is *stable by right product by* P iff for all $x \in \mathcal{U}$ the vector $xP \in \mathcal{U}$. The principal properties of the set $\mathcal{A}_{\mathcal{M}}$ that we need here are

$$\alpha \in \mathcal{A}_{\mathcal{M}} \;\Rightarrow\; \alpha^{B(l)} \in \mathcal{A}_{\mathcal{M}} \quad \forall l \in F, \tag{1}$$

$$\alpha \in \mathcal{A}_{\mathcal{M}} \;\Rightarrow\; \frac{1}{n} \sum_{k=1}^{n} \alpha P^k \in \mathcal{A}_{\mathcal{M}} \quad \forall n \geqslant 1, \tag{2}$$

$$\mathcal{A}_{\mathcal{M}} \neq \emptyset \;\Rightarrow\; \pi \in \mathcal{A}_{\mathcal{M}}, \tag{3}$$

$$\mathcal{A}_{\mathcal{M}} \neq \emptyset \;\Rightarrow\; \mathcal{A}_{\mathcal{M}} \text{ is a convex closed set}, \tag{4}$$

$$\mathcal{A}^1 \text{ is stable by right product by } P \Leftrightarrow \mathcal{A}_{\mathcal{M}} = \mathcal{A}^1, \tag{5}$$

$$\mathcal{A}^{j+1} \subseteq \mathcal{A}^j \quad \forall j \geqslant 1, \tag{6}$$

$$\text{If } \exists j \geqslant 1 \text{ such that } \mathcal{A}^{j+1} = \mathcal{A}^j \quad \text{then } \mathcal{A}_{\mathcal{M}} = \mathcal{A}^{j+k} \quad \forall k \geqslant 1, \tag{7}$$

$$\mathcal{A}_{\mathcal{M}} = \bigcap_{j \geqslant 1} \mathcal{A}^j. \tag{8}$$

See Rubino and Sericola (1989) for the proofs of these properties. In the next section, we show that the sequence of sets $(\mathcal{A}^j)_{j \geqslant 1}$ is a stationary sequence whose limit is equal to $\mathcal{A}^N$. That is, $\mathcal{A}_{\mathcal{M}} = \mathcal{A}^N$, where N denotes the cardinal of the state space E of the process X. Moreover, a finite algorithm to compute this set is given.

3. An algorithm to compute $\mathcal{A}_{\mathcal{M}}$

To obtain the announced result, we need some preliminary lemmas. The first one gives a way to compute the set $\mathcal{A}^{j+1}$ as a function of $\mathcal{A}^j$.

Lemma 3.1. $\forall j \geqslant 1$,

$$\mathcal{A}^{j+1} = \{\alpha \in \mathcal{A}^j \mid \alpha^{B(l)} P \in \mathcal{A}^j \textit{ for every } l \in F\}.$$

Proof. The proof is a direct consequence of the definition of f. It only uses the property

$$f(\alpha, B(i_1), B(i_2), \ldots, B(i_k)) = f(\alpha^{B(i_1)}P, B(i_2), \ldots, B(i_k))$$

which is immediate. □

For every $l \in F$, we denote by P_l the $n(l) \times N$ sub-matrix of P corresponding to the transition probabilities of X from the states of $B(l)$ to the states of E. Let us define for every $l \in F$, the $n(l) \times M$ matrix H_l as $H_l = \tilde{P}_l - 1^{\mathrm{T}} \hat{P}_l$ and recall that a polytope of $\mathbb{R}^N$ is a bounded convex polyhedral subset of $\mathbb{R}^N$ (Rockafellar, 1970).

Lemma 3.2. *$\forall j \geqslant 1$, $\mathscr{A}^j$ is a polytope of $\mathbb{R}^N$.*

Proof. Verify first that $\mathscr{A}^1$ can be written as

$$\mathscr{A}^1 = \{\alpha \in \mathscr{A} \mid \alpha H = 0\}$$

where H is the $N \times M^2$ block diagonal matrix whose lth block is H_l. When necessary, we shall also write $H = \mathrm{Diag}(H_l)$. So, $\mathscr{A}^1$ is a polytope (see Rockafellar, 1970).

Now, for every $j \geqslant 1$, we define the $N \times M^{j+1}$ matrix $H^{[j]}$ as

$$H^{[1]} = H, \qquad H^{[j+1]} = \mathrm{Diag}(P_l H^{[j]}) \quad \text{for } j \geqslant 1.$$

The previous lemma allows us to write

$$\mathscr{A}^{j+1} = \{\alpha \in \mathscr{A}^j \mid \alpha H^{[j+1]} = 0\}.$$

The result follows by definition (see Rockafellar, 1970). $\square$

For α and $\beta \in \mathbb{R}^N$, we denote by $D(\alpha, \beta)$ the line containing the two points α and β. That is $D(\alpha, \beta) = \{\lambda\alpha + (1-\lambda)\beta, \lambda \in \mathbb{R}\}$. Recall that the dimension of a convex set $\mathscr{C}$ of $\mathbb{R}^N$ is the dimension of the smallest affin subset containing $\mathscr{C}$; we shall denote it by $\dim(\mathscr{C})$.

Lemma 3.3. *$\forall j \geqslant 1$ such that $\mathscr{A}^{j+1} \neq \emptyset$, we have*

$$\dim(\mathscr{A}^{j+1}) = \dim(\mathscr{A}^j) \;\Rightarrow\; \mathscr{A}^{j+1} = \mathscr{A}^j.$$

Proof. We have seen that for all $j \geqslant 1$, $\mathscr{A}^{j+1} \subseteq \mathscr{A}^j$. If the dimension is equal to 0, the result follows immediately. Suppose now that the common dimension is at least 1. Let α and β be two points belonging to $\mathscr{A}^{j+1}$ and let γ be a point of $D(\alpha, \beta) \cap \mathscr{A}^j$. Such a point exists since the two sets have the same dimension. Now, there exists $\lambda \in \mathbb{R}$ such that $\gamma = \lambda\alpha + (1-\lambda)\beta$. That is, using the previous lemma, $\gamma H^{[j+1]} = \lambda\alpha H^{[j+1]} + (1-\lambda)\beta H^{[j+1]} = 0$. This proves that $\gamma \in \mathscr{A}^{j+1}$ and then that $\mathscr{A}^{j+1} = \mathscr{A}^j$. $\square$

We are now ready to prove the main result of this paper. Recall that N denotes the cardinal of the state space of the starting homogeneous Markov chain X.

Theorem 3.4. $\mathscr{A}_{\mathscr{M}} = \mathscr{A}^N$.

Proof. Consider the sequence $\mathscr{A}^1, \ldots, \mathscr{A}^N$. From property (6) of the previous section, this sequence is decreasing. If two consecutive elements of $\mathscr{A}^1, \ldots, \mathscr{A}^N$ are equal, property (7) of the previous section leads to $\mathscr{A}_{\mathscr{M}} = \mathscr{A}^N$. If all its elements are different, the previous lemma concludes that the sequence of their dimensions is strictly decreasing and so, since $\dim(\mathscr{A}) = N-1$, we have $\dim(\mathscr{A}^j) < N-k$ for $j = 1, 2, \ldots, N-1$, and then $\mathscr{A}^N = \emptyset$. Property (8) of the previous section leads to the result. $\square$

Before giving the algorithm which computes the set $\mathscr{A}_{\mathcal{M}}$, we need another interesting property given in the following lemma and which generalizes property (5) of the previous section.

Lemma 3.5. *$\mathscr{A}^j$ is stable by right product by $P \Leftrightarrow \mathscr{A}_{\mathcal{M}} = \mathscr{A}^j$.*

Proof. $\mathscr{A}_{\mathcal{M}}$ being stable by right product by P (property (2) of the previous section), the implication from the right to the left is immediate. Conversely, suppose that $\mathscr{A}^j$ is stable by right product by P. Verify then that

$$\forall j \geq 1, \quad \alpha \in \mathscr{A}^j \Leftrightarrow \alpha^{B(l)} \in \mathscr{A}^j \text{ for every } l \in F.$$

Then Lemma 3.1 leads to $\mathscr{A}^{j+1} = \mathscr{A}^j$ and property (7) implies $\mathscr{A}_{\mathcal{M}} = \mathscr{A}^j$. $\square$

To determine the set $\mathscr{A}_{\mathcal{M}}$, we can proceed as follows. Assume the chain X not strongly lumpable with respect to the partition $\mathscr{B}$. The first step consists of calculating the set $\mathscr{A}^1$ and determining if it is stable by right product by P. If $\mathscr{A}^1$ is stable by right product by P, Lemma 3.5 says that $\mathscr{A}_{\mathcal{M}} = \mathscr{A}^1$. If $\mathscr{A}^1$ is not stable by right product by P, we first verify whether for all $l \in F$ the vector $\pi^{B(l)}P$ is in $\mathscr{A}^1$. If there exists $l \in F$ such that $\pi^{B(l)}P \notin \mathscr{A}^1$, then $\pi \notin \mathscr{A}^2$ and $\mathscr{A}_{\mathcal{M}} = \emptyset$ (consequence of properties (3) and (7) of the previous section). If $\pi \in \mathscr{A}^2$ then we calculate this set and we repeat the previous operations.

Denote by Ψ the following procedure, where $\mathscr{P}(\mathscr{A})$ is the set of all subsets of $\mathscr{A}$:

$$\Psi : \mathscr{P}(\mathscr{A}) \to \mathscr{P}(\mathscr{A}), \quad \mathscr{U} \mapsto \{\alpha \in \mathscr{U} \mid \alpha^{B(l)}P \in \mathscr{U} \text{ for every } l \in F\}.$$

With this notation, $\mathscr{A}^{j+1} = \Psi(\mathscr{A}^j)$ (Lemma 3.1). The algorithm has then the following form.

if X is strongly lumpable **then** $\mathscr{A}_{\mathcal{M}} := \mathscr{A}$; **stop**
else $\mathscr{U} := \mathscr{A}^1$
  loop
    if $\mathscr{U}$ is stable by P **break by** *found* **endif**
    if $\exists l \in F : \pi^{B(l)}P \notin \mathscr{U}$ **break by** *empty* **endif**
    $\mathscr{U} := \Psi(\mathscr{U})$
  endloop
  exits
    exit found: $\mathscr{A}_{\mathcal{M}} := \mathscr{U}$
    exit empty: $\mathscr{A}_{\mathcal{M}} := \emptyset$
  endexits
endif

Remark that all the sets considered in this algorithm are polytopes and so they can be determined in a unique way by giving their vertices.

4. Example

To illustrate our algorithm, let us take as an example the process $X=(\cdot, P)$ on the state space $E=\{1, 2, 3, 4, 5, 6\}$ with

$$P=\left(\begin{array}{cccc|cc} \frac{3}{14} & \frac{3}{14} & \frac{3}{14} & \frac{3}{14} & \frac{1}{14} & \frac{1}{14} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{12} & \frac{1}{12} \\ \frac{1}{8} & \frac{3}{8} & \frac{1}{4} & \frac{1}{6} & \frac{1}{24} & \frac{1}{24} \\ \frac{3}{8} & \frac{1}{8} & \frac{1}{4} & \frac{1}{6} & \frac{1}{24} & \frac{1}{24} \\ \hline \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{10} & \frac{1}{10} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{10} & \frac{1}{10} \end{array}\right)$$

and consider the partition $\mathscr{B}=\{B(1), B(2)\}$ where $B(1)=\{1,2,3,4\}$ and $B(2)=\{5,6\}$. We have $N=6$, $M=2$, $n(1)=4$, $n(2)=2$. We see that X is not strongly lumpable with respect to the partition $\mathscr{B}$. The stationary distribution π of X is

$$\pi=\left(\tfrac{42}{193}\ \tfrac{42}{193}\ \tfrac{42}{193}\ \tfrac{42}{193}\ \tfrac{25}{386}\ \tfrac{25}{386}\right)$$

and we have

$$\pi^{B(1)}=\left(\tfrac{1}{4}\ \tfrac{1}{4}\ \tfrac{1}{4}\ \tfrac{1}{4}\ 0\ 0\right), \qquad \pi^{B(2)}=\left(0\ 0\ 0\ 0\ \tfrac{1}{2}\ \tfrac{1}{2}\right).$$

The matrices $\tilde{P}_1$ and $\tilde{P}_2$ are given by

$$\tilde{P}_1=\begin{pmatrix} \frac{6}{7} & \frac{1}{7} \\ \frac{5}{6} & \frac{1}{6} \\ \frac{11}{12} & \frac{1}{12} \\ \frac{11}{12} & \frac{1}{12} \end{pmatrix}, \qquad \tilde{P}_2=\begin{pmatrix} \frac{4}{5} & \frac{1}{5} \\ \frac{4}{5} & \frac{1}{5} \end{pmatrix}.$$

This leads to

$$\hat{P}_1=\left(\tfrac{37}{42}\ \tfrac{5}{42}\right), \qquad \hat{P}_2=\left(\tfrac{4}{5}\ \tfrac{1}{5}\right).$$

The polytope $\mathscr{A}^1$ can be determined by the matrix $H^{[1]}$ (see the proof of Lemma 3.2) given by

$$H^{[1]}=\begin{pmatrix} -\frac{1}{42} & \frac{1}{42} & 0 & 0 \\ -\frac{1}{24} & \frac{1}{24} & 0 & 0 \\ \frac{1}{28} & -\frac{1}{28} & 0 & 0 \\ \frac{1}{28} & -\frac{1}{28} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Therefore, the vertices of the polytope $\mathscr{A}^1$ are

$$\alpha_1=\left(\tfrac{3}{5}\ 0\ 0\ \tfrac{2}{5}\ 0\ 0\right), \quad \alpha_2=\left(\tfrac{3}{5}\ 0\ \tfrac{2}{5}\ 0\ 0\ 0\right), \quad \alpha_3=\left(0\ \tfrac{3}{7}\ 0\ \tfrac{4}{7}\ 0\ 0\right),$$

$$\alpha_4=\left(0\ \tfrac{3}{7}\ \tfrac{4}{7}\ 0\ 0\ 0\right), \quad \alpha_5=(0\ 0\ 0\ 0\ 0\ 1), \quad \alpha_6=(0\ 0\ 0\ 0\ 1\ 0).$$

The set $\mathscr{A}^1$ is not stable by right product by P since $\alpha_2 P \notin \mathscr{A}^1$ ($\alpha_2 PH^{[1]} \neq 0$). Furthermore, $\pi^{B(1)}P \in \mathscr{A}^1$ and $\pi^{B(2)}P \in \mathscr{A}^1$ since $\pi^{B(1)}PH^{[1]} = 0$ and $\pi^{B(2)}PH^{[1]} = 0$. This means that the set $\mathscr{A}^2 \neq \emptyset$ ($\pi \in \mathscr{A}^2$). The polytope $\mathscr{A}^2$ can be determined by the matrix $H^{[2]}$ given by

$$H^{[2]} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{168} & -\frac{1}{168} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{168} & \frac{1}{168} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The vertices of the polytope $\mathscr{A}^2$ are α_1, α_5, α_6 and α_7 where

$$\alpha_7 = (0\ \tfrac{3}{7}\ \tfrac{3}{7}\ \tfrac{1}{7}\ 0\ 0).$$

The set $\mathscr{A}^2$ is not stable by right product by P since $\alpha_1 P \notin \mathscr{A}^2$ ($\alpha_1 PH^{[2]} \neq 0$). Furthermore, $\pi^{B(1)}P \in \mathscr{A}^2$ and $\pi^{B(2)}P \in \mathscr{A}^2$ since $\pi^{B(1)}PH^{[2]} = 0$ and $\pi^{B(2)}PH^{[2]} = 0$. This means that $\mathscr{A}^3 \neq \emptyset$ ($\pi \in \mathscr{A}^3$). The polytope $\mathscr{A}^3$ can be determined by the matrix $H^{[3]}$ given by

$$H^{[3]} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{1344} & -\frac{1}{1344} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{1344} & \frac{1}{1344} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The vertices of the polytope $\mathscr{A}^3$ are α_5, α_6 and $\pi^{B(1)}$. Finally, the polytope $\mathscr{A}^3$ is stable by right product by P since $\alpha_5 PH^{[3]} = 0$, $\alpha_6 PH^{[3]} = 0$ and $\pi^{B(1)}PH^{[3]} = 0$. So, we have $\mathscr{A}_{\mathscr{M}} = \mathscr{A}^3$, that is

$$\mathscr{A}_{\mathscr{M}} = \{\lambda(0, 0, 0, 0, 0, 1) + \mu(0, 0, 0, 0, 1, 0) + \eta(\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, 0, 0),$$
$$\lambda, \mu, \eta \in [0, 1], \lambda + \mu + \eta = 1\}.$$

5. Conclusions

In this paper, we have analysed the set of all initial probability distributions of an irreducible and homogeneous Markov chain which lead to a homogeneous aggregated Markov chain given the transition probability matrix and a partition of the state space. We have obtained a constructive characterization of this set by means of a finite algorithm. In a subsequent paper we analyse the continuous time case. Basically, it is shown that it is always possible to come down to the discrete time case using the uniformization technique. The case of homogeneous Markov processes with absorbing states seems to be the first possible direction to extend these results.

References

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

J.G. Kemeny and J.L. Snell, Finite Markov Chains (Springer, New York–Heidelberg–Berlin, 1976).

R.T. Rockafellar, Convex Analysis (Princeton Univ. Press, Princeton, NJ, 1970).

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