

ON THE RELATIONSHIPS BETWEEN LUMPABILITY AND FILTERING OF FINITE STOCHASTIC SYSTEMS

JAMES LEDOUX,* *Université de Poitiers and CNRS UMR 6086*

LANGFORD B. WHITE,** *University of Adelaide*

GARY D. BRUSHE,*** *Defence Science and Technology Organisation*

Abstract

The aim of this paper is to provide the conditions necessary to reduce the complexity of state filtering for finite stochastic systems (FSSs). A concept of lumpability for FSSs is introduced. In this paper we assert that the unnormalised filter for a lumped FSS has linear dynamics. Two sufficient conditions for such a lumpability property to hold are discussed. We show that the first condition is also necessary for the lumped FSS to have linear dynamics. Next, we prove that the second condition allows the filter of the original FSS to be obtained directly from the filter for the lumped FSS. Finally, we generalise an earlier published result for the approximation of a general FSS by a lumpable FSS.

Keywords: Optimal filtering; model reduction; hidden Markov chain; discrete Markovian arrival process

2000 Mathematics Subject Classification: Primary 60J10; 93E11

1. Introduction

In this paper we deal with a class of homogeneous Markov chains (MCs) $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ with finite state space $\mathcal{Y} \times \mathcal{X}$ and whose transition probabilities satisfy

$$\begin{aligned} P\{(Y_{t+1}, X_{t+1}) = (y, x) \mid (X_t, Y_t)\} &= P\{(Y_{t+1}, X_{t+1}) = (y, x) \mid X_t\} \\ &= D_y(X_t, x). \end{aligned} \tag{1}$$

Such a Markov model is called a *finite stochastic system (FSS)* in [1], [15], and [18], and is parameterised by the family of matrices $(D_y, y \in \mathcal{Y})$. A direct consequence of (1) is that $\{X_t\}_{t \in \mathbb{N}}$ is a Markov chain with transition probabilities

$$P(X_t, x) := P\{X_{t+1} = x \mid X_t\} = \sum_{y \in \mathcal{Y}} D_y(X_t, x). \tag{2}$$

As a result, $\sum_{y \in \mathcal{Y}} D_y$ is a stochastic matrix.

Received 26 June 2007; revision received 8 August 2008.

* Postal address: Département de Mathématiques, Université de Poitiers, BP 30179, Téléport 2, Boulevard Marie et Pierre Curie, F-86962 Futuroscope-Chasseneuil cedex, France.

** Postal address: School of Electrical and Electronic Engineering, University of Adelaide, SA 5005, Australia.

*** Postal address: Signals Analysis Discipline, C3I Division, Defence Science and Technology Organisation, PO Box 1500, Edinburgh, SA 5111, Australia.

Note that any discrete-time Markovian arrival process $\{(N_t, X_t)\}_{t \in \mathbb{N}}$ (see [3]) defines an FSS setting $Y_t := N_t - N_{t-1}$ for $t \geq 1$. Using Bayes' formula, we have from (1) the following general factorisation property:

$$D_y(X_t, x) = P\{X_{t+1} = x \mid X_t\} P\{Y_{t+1} = y \mid X_{t+1} = x, X_t\}. \tag{3}$$

Thus, the 'output process' $\{Y_t\}_{t \in \mathbb{N}}$ may be thought of as generated by transitions of the Markov chain $\{X_t\}_{t \in \mathbb{N}}$. Here $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is called a *hidden Markov chain (HMC)* when the probabilities $P\{Y_{t+1} = y \mid X_{t+1} = x, X_t\}$ in (3) do not depend on X_t and the time t . If, for any t , the conditional distribution of Y_{t+1} given $X_{t+1} = x$ is denoted by $G(x, \cdot)$, this means that the transition probabilities (1) have the following special factorisation property:

$$D_y(X_t, x) = P\{X_{t+1} = x \mid X_t\} P\{Y_{t+1} = y \mid X_{t+1} = x\} = P(X_t, x)G(x, y). \tag{4}$$

Note that, in general, the distribution of Y_0 given X_0 is assumed to be $G(X_0, \cdot)$. It is well known that, for an FSS $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$, the process $\{(Y_t, (X_t, X_{t-1}))\}_{t \in \mathbb{N} - \{0\}}$ forms an HMC with state space $\mathcal{Y} \times \{(i, j) \in \mathcal{X} \times \mathcal{X} \mid P(i, j) > 0\}$. In this sense, assumptions (1) and (4) on $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ are equivalent. However, only FSSs are considered in this paper for the following reasons. First, the factorisation property (3) or (4) introduces no simplification in addressing the lumping and filtering problems. Moreover, the transformation of results on HMCs into results on FSSs involves some notational and technical difficulties which are unhelpful to the reader. In contrast, any result on an FSS can be applied to an HMC by replacing, everywhere, $D_y(X_t, x)$ by $P(X_t, x)G(x, y)$. Second, conceptual difficulties arise when you try to discuss some Markovian models widely used in stochastic modelling in the framework of HMCs [6].

For the rest of this paper, $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ will denote an FSS. In this context, $\{Y_t\}_{t \in \mathbb{N}}$ and $\{X_t\}_{t \in \mathbb{N}}$ are called the *observed process* and the *state process*, respectively. The aim of this paper is to provide a complexity reduction of the filtering for FSSs. More precisely, consider the a posteriori probabilities

$$\pi_t(x) := P\{X_t = x \mid Y_t, \dots, Y_0\} \quad \text{for all } x \in \mathcal{X} \text{ and } t \in \mathbb{N}. \tag{5}$$

Let π_t denote the probability distribution $(\pi_t(x))_{x \in \mathcal{X}}$ on \mathcal{X} . Here $\{\pi_t\}_{t \in \mathbb{N}}$ is called the (state) *filter process* associated with $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$. We know that the filter process is the solution of the nonlinear recursive equation (see, e.g. [13])

$$\pi_{t+1}(x) = \frac{\sum_{x_0 \in \mathcal{X}} \pi_t(x_0) D_{Y_{t+1}}(x_0, x)}{\sum_{x \in \mathcal{X}} \sum_{x_0 \in \mathcal{X}} \pi_t(x_0) D_{Y_{t+1}}(x_0, x)}, \tag{6}$$

where $D_{Y_{t+1}}$ is defined in a natural way. Another standard way of filtering is to use the unnormalised filter $\{\rho_t\}_{t \in \mathbb{N}} = \{(\rho_t(x, Y_t, \dots, Y_0))_{x \in \mathcal{X}}\}_{t \in \mathbb{N}}$, where, for every $t \in \mathbb{N}$ and $x \in \mathcal{X}$, $\rho_t(x, \cdot)$ is the positive measure on \mathcal{Y}^{t+1} defined by

$$\rho_t(x, y_t, \dots, y_0) := P\{X_t = x, Y_t = y_t, \dots, Y_0 = y_0\} \quad \text{for all } (y_t, \dots, y_0) \in \mathcal{Y}^{t+1}. \tag{7}$$

We denote, for short, $\rho_t(x, Y_t, \dots, Y_0)$ by $\rho_t(x)$. The conditional probability (5) is obtained from

$$\pi_t(x) = \frac{\rho_t(x)}{\sum_{z \in \mathcal{X}} \rho_t(z)}. \tag{8}$$

The unnormalised filter has the main advantage of being the solution of the linear recursive equation

$$\rho_{t+1}(x) = \sum_{x_0 \in \mathcal{X}} \rho_t(x_0) D_{Y_{t+1}}(x_0, x). \tag{9}$$

We are interested in computing the filter associated with the function $\{g(X_t)\}_{t \in \mathbb{N}}$ of the Markov chain $\{X_t\}_{t \in \mathbb{N}}$, where $\text{card}(g(\mathcal{X})) < \text{card}(\mathcal{X})$, that is,

$$\hat{\pi}_t(w) := \mathbb{P}\{g(X_t) = w \mid Y_t, \dots, Y_0\} \quad \text{for all } w \in g(\mathcal{X}).$$

We introduce the unnormalised filter $\{\hat{\rho}_t\}_{t \in \mathbb{N}} = \{(\hat{\rho}_t(w, Y_t, \dots, Y_0))_{w \in g(\mathcal{X})}\}_{t \in \mathbb{N}}$ associated with the lumped process $\{g(X_t)\}_{t \in \mathbb{N}}$, that is, $\hat{\rho}_t(w, y_t, \dots, y_0)$ is defined, as in (7), by

$$\hat{\rho}_t(w, y_t, \dots, y_0) := \mathbb{P}\{g(X_t) = w, Y_t = y_t, \dots, Y_0 = y_0\}. \tag{10}$$

Using the same convention as in (8), we have

$$\hat{\pi}_t(w) = \frac{\hat{\rho}_t(w)}{\sum_{z \in g(\mathcal{X})} \hat{\rho}_t(z)}.$$

The problem here is that $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is not an FSS in general, so we cannot use a linear recursive equation as in (9) for the computation of the unnormalised filter $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ of $\{g(X_t)\}_{t \in \mathbb{N}}$. The purpose of this paper is to propose conditions under which $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ could be derived using $\text{card}(g(\mathcal{X}))$ -dimensional matrix computations. A direct way is to look for conditions under which $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an FSS, so that the filter $\{\hat{\pi}_t\}_{t \in \mathbb{N}}$ and the unnormalised filter $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ satisfy a recursive equation as in (6) and, respectively, (9). A related problem was discussed in [21] for HMCs. In [21], a concept of strong lumpability for HMCs was defined. In that paper a general procedure for testing lumpability and deriving the associated lumped states was described. The present paper briefly discusses lumpability for FSSs using a somewhat more explicit relationship between lumpability of MCs and FSSs. Recalling that the focus is on the dynamics of the filter of lumped FSSs, the main contributions of the paper are as follows.

1. The lumped filter $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ has linear dynamics irrespective of the probability distribution of (Y_0, X_0) if and only if the FSS is strongly lumpable with respect to the function g .
2. A new condition is introduced for the unnormalised filter to have linear dynamics for some specific probability distributions of (Y_0, X_0) . Furthermore, this condition asserts that the filter $\{\pi_t\}_{t \in \mathbb{N}}$ can be computed directly from the lumped filter $\{\hat{\pi}_t\}_{t \in \mathbb{N}}$.

In Section 2 we revisit the basic results on lumpability of Markov chains in order to discuss the lumpability of FSSs in Section 3. The fact that the lumped filter $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ has linear dynamics if and only if the FSS is strongly lumpable is proven in Theorem 2. This new condition requiring $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ to have linear dynamics for a specific probability distribution of (Y_0, X_0) is introduced in Subsection 3.2, where we also show that, when this property is true, the filter $\{\pi_t\}_{t \in \mathbb{N}}$ for the original FSS can be computed directly from the lumped filter $\{\hat{\pi}_t\}_{t \in \mathbb{N}}$. In Section 4 we discuss the problem of approximating an MC or an FSS by a strongly lumpable one, and propose algorithms for computing such approximations.

2. Lumpable Markov chains

Let $\{Z_t\}_{t \in \mathbb{N}}$ be a homogeneous Markov chain with state space $\mathcal{Z} = \{1, \dots, N\}$. Consider a function f from \mathcal{Z} into $f(\mathcal{Z}) = \{1, \dots, n\}$ with $n < N$. Such a map is called a *lumping map*. For notational convenience, f is assumed to be nondecreasing. This function defines a partition $\mathcal{Z}_i, i = 1, \dots, n$, of \mathcal{Z} , where $\mathcal{Z}_i := f^{-1}(\{i\})$. The number of states in the subset \mathcal{Z}_i is denoted by N_i . We define the *lumping matrix* associated with this partition as the

$N \times n$ matrix L , where $L(j, i) = 1$ when $j \in \mathcal{Z}_i$ and 0 elsewhere. Next, we introduce the $n \times N$ matrix U ,

$$U := (L^\top L)^{-1} L^\top = \text{diag}\left(\frac{1}{N_i}\right)^{-1} L^\top. \tag{11}$$

The i th row of U is an N -dimensional probability vector which has precisely N_i nonzero elements, each with identical value $1/N_i$. Let Q be the $N \times N$ transition matrix of $\{Z_t\}_{t \in \mathbb{N}}$. Let $\mathbf{0}$ and $\mathbf{1}$ denote a matrix or vector in which each entry is equal to 0 or 1, respectively, with their dimensions being defined by the context. Any vector is a row vector. In particular, the linear space $\ker(L)$ is defined by the set of vectors $v \in \mathbb{R}^N$ such that $vL = \mathbf{0}$. We will denote by I_k the $k \times k$ identity matrix.

The Markov chain $\{Z_t\}_{t \in \mathbb{N}}$ is said to be *lumpable* with respect to the function f and a specified initial distribution of Z_0 if $\{f(Z_t)\}_{t \in \mathbb{N}}$ is a homogeneous Markov chain. The Markov chain is said to be *strongly lumpable* with respect to f if it is lumpable with respect to f for every probability distribution of Z_0 .

Lemma 1. *The following statements are equivalent for $\{Z_t\}_{t \in \mathbb{N}}$ to be strongly lumpable.*

(a) For all $w_1, w_2 \in f(\mathcal{Z})$,

$$P\{f(Z_{t+1}) = w_2 \mid Z_t = z_1\} = P\{Z_{t+1} \in f^{-1}(w_2) \mid Z_t = z_1\} = \sum_{z_2 \in f^{-1}(w_2)} Q(z_1, z_2)$$

is independent of $z_1 \in f^{-1}(w_1)$; this conditional probability defines the transition probability from w_1 to w_2 for the Markov chain $\{f(Z_t)\}_{t \in \mathbb{N}}$.

(b) The transition matrix Q has the following block structure:

$$Q = \begin{pmatrix} Q_{11} & \cdots & Q_{1n} \\ \vdots & & \vdots \\ Q_{n1} & \cdots & Q_{nn} \end{pmatrix},$$

where Q_{ij} is an $N_i \times N_j$ matrix which satisfies $Q_{ij} \mathbf{1}^\top = q_{ij} \mathbf{1}^\top$ for some nonnegative constant q_{ij} . The matrix $\hat{Q} := (q_{ij})_{i,j=1,\dots,n}$ is stochastic.

(c) $QL = LUQL$.

(d) $\ker(L)Q \subset \ker(L)$.

(e) $\ker(L) \subset \ker(QL)$.

(f) Let $V^\top = \{v_1^\top, \dots, v_N^\top\}$, where the v_i are the right singular vectors of L^\top . We have

$$VQV^\top = \begin{pmatrix} B_{11} & B_{12} \\ \mathbf{0} & B_{22} \end{pmatrix}, \tag{12}$$

where $B_{11} \in \mathbb{R}_+^{n \times n}$.

In this case, $\{f(Z_t)\}_{t \in \mathbb{N}}$ is a Markov chain with transition matrix $\hat{Q} := UQL$.

Conditions (d) and (f) were proved in [21] for $n = 2$. Condition (d) for any $n \geq 2$ was derived in [10] from a very general criterion for a function of a Markov chain to be a Markov

chain. That condition (f) for strong lumpability is valid for any $n \geq 2$ is proved in Appendix A. The other criteria are reviewed from [10, Theorem 10].

Now we introduce another criterion for $\{f(Z_t)\}_{t \in \mathbb{N}}$ to be Markovian for some probability distribution of Z_0 . It was stated in [12] for finite state spaces and generalised in [16] for general state spaces.

Definition 1. ([10].) Let L be the lumping matrix associated with lumping map f . A stochastic matrix Q is called an R-P matrix if there exist an $n \times N$ stochastic matrix Λ such that

$$\Lambda L = I_n \quad \text{and} \quad \Lambda Q = \Lambda Q \Lambda.$$

The following lemma states a known result on the lumpability of a Markov chain whose transition matrix is an R-P matrix. In this case, note that the Markov property of $\{f(Z_t)\}_{t \in \mathbb{N}}$ depends on the probability distribution of Z_0 . This is a so-called *weak lumpability* condition [12].

Lemma 2. Let $\{Z_t\}_{t \in \mathbb{N}}$ be a Markov chain with an R-P transition matrix Q . Then the process $\{f(Z_t)\}_{t \in \mathbb{N}}$ is a Markov chain when the probability of Z_0 is any convex combination of the n rows of matrix Λ . Moreover, its transition matrix is given by $\hat{Q} = \Lambda Q \Lambda$.

When an MC has an R-P transition matrix with Λ defined as in (11), it is said to be exactly lumpable [4]. See [10] for further properties of Markov chains with an R-P transition matrix.

Lumpability of MCs has been found to be relevant in various areas (see, e.g. [2], [9], [14], and [20]). This is especially true in performance evaluation, where various modelling formalisms (e.g. stochastic automata networks, Petri nets, and algebra processes) have been developed for model simplification. Every model specified by these formalisms has an underlying (continuous-time) MC, but, in general, with a very large state space. Thus, one objective is to avoid having to generate such a Markov graph. Therefore, the focus is on equivalence relations between basic objects of the formalism and on the development of efficient algorithms to aggregate equivalence classes. It is well known that some concepts of equivalence relations are directly connected to lumpability of the underlying MC (see, e.g. [7] and [11] for strong lumpability and [19] for the R-P condition, and the references therein). But, in some sense, the lumpability of an MC is only used through the fundamental results given in [12] (statements (a)–(c) of Lemma 1 or Lemma 2). In this way, the main contribution to the theory of lumpability of MCs is an efficient algorithm to find the optimal (strong) lumping map associated with an MC [5]. It will be clear from the next section that lumpability concepts for FSSs are directly related to lumpability for the bivariate MC. Thus, we do not contribute here to the theory of lumpability through new criteria, but we recall some equivalent forms which are not well known. The aim of the paper is to investigate the connection between lumpability and the dynamics of filters for FSSs.

3. Lumpable finite stochastic systems and filtering

In the special case of an HMC, Spreij [17] discussed the conditions under which the observed process $\{Y_t\}_{t \in \mathbb{N}}$ is a Markov chain. The state process was assumed to be irreducible. This problem was solved in [10] under no particular assumption. The basic idea was to interpret the process $\{Y_t\}_{t \in \mathbb{N}}$ as the function $f(Y_t, X_t) = Y_t$ of the Markov chain $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ and to use the criteria for lumpability of Markov chains. A similar idea, though not explicitly stated, was used in [21] to discuss the problem that we are interested in. Indeed, we are concerned here

with the function $f(Y_t, X_t) = (Y_t, g(X_t))$ of the Markov chain $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$, where g is some lumping map for the Markov chain $\{X_t\}_{t \in \mathbb{N}}$.

Definition 2. An FSS $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is said to be lumpable with respect to the lumping map g from \mathcal{X} into $g(\mathcal{X})$ if $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an FSS for some probability distribution of (Y_0, X_0) . An FSS $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is said to be strongly lumpable with respect to g if it is lumpable with respect to g for every probability distribution of (Y_0, X_0) .

When an FSS is lumpable, note that both of the following conditions are satisfied: (i) $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is lumpable with respect to the function $f(y, x) = (y, g(x))$ from $\mathcal{Y} \times \mathcal{X}$ into $\mathcal{Y} \times g(\mathcal{X})$ and (ii) the Markov chain $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is lumpable with respect to the function $f(y, x) = x$ from $\mathcal{Y} \times g(\mathcal{X})$ into $g(\mathcal{X})$ (the lumpability property above is relative to the same probability distribution of (Y_0, X_0)). However, the converse is not correct because the second condition only asserts that the conditional probabilities

$$P\{g(X_{t+1}) = w_{t+1} \mid Y_t = y_t, g(X_t) = w_t\}$$

are independent of y_t .

For the rest of this paper, $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ denotes an FSS with finite state space $\mathcal{Y} \times \mathcal{X} = \{1, \dots, M\} \times \{1, \dots, N\}$ and g denotes a lumping map from \mathcal{X} into $g(\mathcal{X}) = \{1, \dots, n\}$. Recall that the transition matrix P of the Markov chain $\{X_t\}_{t \in \mathbb{N}}$ is given by $\sum_{y \in \mathcal{Y}} D_y$ (see (2)).

3.1. Strong lumpability for FSSs

Let us introduce our main result on the strong lumpability of FSSs.

Theorem 1. Let L be the lumping $N \times n$ matrix associated with g . The FSS $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is strongly lumpable with respect to g if and only if any of the following conditions are satisfied.

(a) For every $y \in \mathcal{Y}$ and all $w_1, w_2 \in g(\mathcal{X})$,

$$P\{Y_{t+1} = y, g(X_{t+1}) = w_2 \mid X_t = x_1\} = \sum_{x_2 \in g^{-1}(w_2)} D_y(x_1, x_2) \tag{13}$$

is independent of $x_1 \in g^{-1}(w_1)$.

(b) For each $y \in \mathcal{Y}$, the $N \times N$ matrix D_y has the following block structure:

$$D_y = \begin{pmatrix} E_{11}(y) & \cdots & E_{1n}(y) \\ \vdots & & \vdots \\ E_{n1}(y) & \cdots & E_{nn}(y) \end{pmatrix}, \tag{14}$$

where $E_{ij}(y)$ is an $N_i \times N_j$ matrix which satisfies

$$E_{ij}(y)\mathbf{1}^\top = d_{ij}(y)\mathbf{1}^\top \tag{15}$$

for some nonnegative constant $d_{ij}(y)$. The matrices $\hat{D}_y := (d_{ij}(y))_{i,j=1,\dots,n}$, $y \in \mathcal{Y}$, are such that $\sum_y \hat{D}_y$ is a stochastic matrix.

(c) For every $y \in \mathcal{Y}$,

$$D_y L = L U D_y L, \tag{16}$$

where $U := (L^\top L)^{-1} L^\top$.

- (d) For every $y \in \mathcal{Y}$, $\ker(\mathbf{L})\mathbf{D}_y \subset \ker(\mathbf{L})$.
- (e) For every $y \in \mathcal{Y}$, $\ker(\mathbf{L}) \subset \ker(\mathbf{D}_y\mathbf{L})$.
- (f) Set $\mathbf{V}^\top := \{\mathbf{v}_1^\top, \dots, \mathbf{v}_N^\top\}$, where the \mathbf{v}_i are the right singular vectors of \mathbf{L}^\top . Then we have, for any $y \in \mathcal{Y}$,

$$\mathbf{V}\mathbf{D}_y\mathbf{V}^\top = \begin{pmatrix} \mathbf{B}_y & \mathbf{R}_y \\ \mathbf{0} & \mathbf{Z}_y \end{pmatrix},$$

where the \mathbf{B}_y are nonnegative.

Moreover, $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an FSS with characteristic matrices $(\hat{\mathbf{D}}_y, y \in \mathcal{Y})$ given by

$$\hat{\mathbf{D}}_y := \mathbf{U}\mathbf{D}_y\mathbf{L} \quad \text{with} \quad \mathbf{U} := (\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top. \tag{17}$$

Proof. The main step is to verify that the FSS $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is strongly lumpable if and only if it is strongly lumpable as a Markov chain. The direct statement is obvious from the definition of an FSS. Now, assume that the Markov chain $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is strongly lumpable. From the first statement of Lemma 1, this is equivalent to the statement that, for every $y \in \mathcal{Y}$, and all $w_1, w_2 \in g(\mathcal{X})$,

$$\mathbb{P}\{Y_{t+1} = y_1, g(X_{t+1}) = w_2 \mid Y_t = y_0, X_t = x_1\} \tag{18}$$

is independent of $x_1 \in g^{-1}(w_1)$. Since $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is an FSS, the probabilities above are independent of y_0 . Therefore, $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an FSS for any probability distribution of (Y_0, X_0) and $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is strongly lumpable as an FSS.

Next, the equivalence of statements (a)–(f) is deduced from the six equivalent conditions of Lemma 1 for the Markov chain $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ to be strongly lumpable (see Appendix B).

Comment 1. In this paper we do not discuss the continuous-time counterpart of FSSs. But, time discretisation by the standard uniformisation technique will provide the corresponding lumpability results for continuous-time FSSs. In this context, the entries of matrix \mathbf{D}_y for an FSS must be interpreted as transition rates between states of the bivariate MC defining an FSS. Then, entries of matrix \mathbf{D}_y are very similar to the labels of the derivation graph for a component in a performance evaluation process algebra (PEPA) model, this graph being the basis of generating the underlying continuous-time MC. In this framework, relation (13) may be thought of as the basic equality to define the concept of strong equivalence for PEPA. Note that such a relation is considered as a definition and it is shown that strong equivalence for PEPA implies strong lumpability of the underlying MC. In this way, this corresponds to the fact that Theorem 1(a) (or, basically, Theorem 1(b)) implies that the FSS is strongly lumpable with respect to g . We refer the reader to [7] and [11] for details and connections to other stochastic process algebras.

Comment 2. When $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is an HMC, we know from (4) that

$$\mathbf{D}_y = \mathbf{P} \operatorname{diag}(\mathbf{G}(\cdot, y))$$

and the probability distribution of (Y_0, X_0) is given by

$$(\beta \operatorname{diag}(\mathbf{G}(\cdot, 1)), \dots, \beta \operatorname{diag}(\mathbf{G}(\cdot, M))),$$

where β is the probability distribution of X_0 . In general, the requirement that $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is a Markov chain for every distribution of X_0 is weaker than the Markov property of $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ for every probability distribution of (Y_0, X_0) .

Comment 3. In contrast to the Markov chain case, the ‘off-diagonal’ condition,

- (C) for every $y \in \mathcal{Y}$ and all $w_1 \neq w_2 \in g(\mathcal{X})$, the probability $\sum_{x_2 \in g^{-1}(w_2)} D_y(x_1, x_2)$ is independent of $x_1 \in g^{-1}(w_1)$,

does not assert that the FSS is strongly lumpable (see Example 1, below). The definition in [21, p. 2301] for an HMC to be (strongly) lumpable must be replaced by Theorem 1(a).

The main result of this subsection states that the strong lumpability condition provides the only way of ensuring that the unnormalised lumped filter $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ has linear dynamics.

Theorem 2. *Let $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ be an FSS with characteristic matrices $(D_y, y \in \mathcal{Y})$. Consider the lumping map g with $N \times n$ matrix L , and let $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ be the associated unnormalised lumped filter. Then the following statements are equivalent.*

- (a) *For any probability distribution of (Y_0, X_0) , $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ has the linear dynamics*

$$\hat{\rho}_{t+1} = \hat{\rho}_t \hat{D}_{Y_{t+1}} \quad \text{for all } t \geq 0 \tag{19}$$

for some family of $n \times n$ nonnegative matrices $\hat{D}_y, y \in \mathcal{Y}$, such that $\sum_{y \in \mathcal{Y}} \hat{D}_y$ is a stochastic matrix.

- (b) *$\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is a strongly lumpable FSS with respect to g .*
- (c) *$\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is a strongly lumpable Markov chain with respect to g .*

As a result, under any of the equivalent conditions of Theorem 1, the filter $\{\hat{\pi}_t\}_{t \in \mathbb{N}}$ associated with the lumped process $\{g(X_t)\}_{t \in \mathbb{N}}$ is given by

$$\hat{\pi}_t = \frac{\hat{\rho}_t}{\hat{\rho}_t \mathbf{1}^\top},$$

where the unnormalised filter $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ satisfies (19) with $\hat{D}_y = U D_y L, y \in \mathcal{Y}$.

Proof. The equivalence of statements (b) and (c) has been checked in the proof of Theorem 1. It is immediate from Theorem 1 that statement (b) implies statement (a). Indeed, if any of the four conditions in Theorem 1 is satisfied then $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an FSS with characteristic matrices $(\hat{D}_y, y \in \mathcal{Y})$ whatever the probability distribution of (Y_0, X_0) . Thus, the unnormalised filter satisfies a linear equation of the type (9), that is (19). A direct calculation may give some insight into this fact. Let $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ be the unnormalised filter associated with the Markov chain $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$. It follows, from (7) and (10), that $\{\rho_t\}_{t \in \mathbb{N}}$ and $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ are always related by

$$\hat{\rho}_t = \rho_t L \quad \text{for all } t \geq 0. \tag{20}$$

Then we can write, for every $t \geq 0$,

$$\begin{aligned} \hat{\rho}_{t+1} &= \rho_{t+1} L \\ &= \rho_t D_{Y_{t+1}} L && \text{from (9)} \\ &= \rho_t L U D_{Y_{t+1}} L && \text{from (16)} \\ &= \hat{\rho}_t U D_{Y_{t+1}} L && \text{from (20)} \\ &= \hat{\rho}_t \hat{D}_{Y_{t+1}} && \text{from (17)}. \end{aligned}$$

It remains to prove that statement (a) implies statement (b). This involves checking that statement (a) implies that, for any probability distribution of (Y_0, X_0) , the unnormalised filter $\{\rho_t\}_{t \in \mathbb{N}}$ satisfies the set of linear equations

$$\rho_t(D_{Y_{t+1}}L - L\hat{D}_{Y_{t+1}}) = \mathbf{0} \quad \text{for all } t \geq 0. \tag{21}$$

Indeed, we know from the algebraic manipulation above that, for any $t \geq 0$,

$$\hat{\rho}_{t+1} = \rho_t D_{Y_{t+1}} L,$$

and it follows from condition (19) that

$$\hat{\rho}_{t+1} = \hat{\rho}_t \hat{D}_{Y_{t+1}} = \rho_t L \hat{D}_{Y_{t+1}}.$$

Combining the two representations of $\hat{\rho}_{t+1}$, we obtain (21).

Now we can write

$$\rho_0 = \sum_{y_0 \in \mathcal{Y}} 1_{\{Y_0=y_0\}} \alpha_{y_0},$$

where $\alpha_{y_0} = (P\{Y_0 = y_0, X_0 = x_0\})_{x_0 \in \mathcal{X}}$ and $1_{\{\cdot\}}$ is the indicator function. If we consider any probability distribution for (Y_0, X_0) , it is clear that

$$\text{span}(\alpha_{y_0}, y_0 \in \mathcal{Y}) = \mathbb{R}^N.$$

It follows from (21) that

$$\alpha(D_{Y_1}L - L\hat{D}_{Y_1}) = \mathbf{0} \quad \text{for all } \alpha \in \mathbb{R}^N$$

or, for all $y_1 \in \mathcal{Y}$,

$$D_{y_1}L - L\hat{D}_{y_1} = \mathbf{0}.$$

Also, for any $y_1 \in \mathcal{Y}$, the equality $D_{y_1}L = L\hat{D}_{y_1}$ implies that $UD_{y_1}L = UL\hat{D}_{y_1} = \hat{D}_{y_1}$. Then $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is a strongly lumpable FSS with respect to g from Theorem 1.

Example 1. Let us consider the following HMC (see (4)) with $N = 3, M = 2$, the lumping map $g(1) = 1, g(2) = g(3) = 2$, and the matrices

$$P = \left(\begin{array}{c|cc} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \hline \frac{1}{4} & \frac{1}{3} & \frac{5}{12} \end{array} \right), \quad G = \left(\begin{array}{cc|c} \frac{1}{2} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{2} & \frac{4}{5} & \frac{4}{5} \end{array} \right), \quad L = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

This gives an FSS with characteristic matrices

$$D_1 = P \text{diag}(G(\cdot, 1)) \quad \text{and} \quad D_2 = P \text{diag}(G(\cdot, 2))$$

as follows:

$$D_1 = \left(\begin{array}{c|cc} \frac{1}{4} & \frac{1}{15} & \frac{1}{30} \\ \frac{1}{8} & \frac{1}{10} & \frac{1}{20} \\ \hline \frac{1}{8} & \frac{1}{15} & \frac{1}{12} \end{array} \right), \quad D_2 = \left(\begin{array}{c|cc} \frac{1}{4} & \frac{4}{15} & \frac{2}{15} \\ \frac{1}{8} & \frac{2}{5} & \frac{1}{5} \\ \hline \frac{1}{8} & \frac{4}{15} & \frac{1}{3} \end{array} \right),$$

and we have

$$\mathbf{P}\mathbf{L} = \left(\begin{array}{c|c} \frac{1}{2} & \frac{1}{2} \\ \hline \frac{1}{4} & \frac{3}{4} \\ \hline \frac{1}{4} & \frac{3}{4} \end{array} \right), \quad \mathbf{D}_1\mathbf{L} = \left(\begin{array}{c|c} \frac{1}{4} & \frac{1}{10} \\ \hline \frac{1}{8} & \frac{3}{20} \\ \hline \frac{1}{8} & \frac{3}{20} \end{array} \right), \quad \mathbf{D}_2\mathbf{L} = \left(\begin{array}{c|c} \frac{1}{4} & \frac{2}{5} \\ \hline \frac{1}{8} & \frac{3}{5} \\ \hline \frac{1}{8} & \frac{3}{5} \end{array} \right).$$

Then statement (a) is met and $\{(Y_t, g(X_t))\}_{t \geq 0}$ is an FSS with characteristic matrices

$$\hat{\mathbf{D}}_1 = \left(\begin{array}{c|c} \frac{1}{4} & \frac{1}{10} \\ \hline \frac{1}{8} & \frac{3}{20} \end{array} \right), \quad \hat{\mathbf{D}}_2 = \left(\begin{array}{c|c} \frac{1}{4} & \frac{2}{5} \\ \hline \frac{1}{8} & \frac{3}{5} \end{array} \right).$$

Now, we deduce from Theorem 2 that the filters can be computed from observations (Y_0, \dots, Y_t) as follows (see also (16)): for any probability vector α on \mathcal{X} ,

$$\begin{aligned}
 \rho_t &= \alpha \operatorname{diag}(\mathbf{G}(\cdot, Y_0))\mathbf{D}_{Y_1} \cdots \mathbf{D}_{Y_t}, \\
 \hat{\rho}_t &= \rho_t \mathbf{L} = \alpha \operatorname{diag}(\mathbf{G}(\cdot, Y_0))\mathbf{L}\hat{\mathbf{D}}_{Y_1} \cdots \hat{\mathbf{D}}_{Y_t} = \hat{\rho}_0 \hat{\mathbf{D}}_{Y_1} \cdots \hat{\mathbf{D}}_{Y_t}, \\
 \pi_t &= \frac{\alpha \operatorname{diag}(\mathbf{G}(\cdot, Y_0))\mathbf{D}_{Y_1} \cdots \mathbf{D}_{Y_t}}{\alpha \operatorname{diag}(\mathbf{G}(\cdot, Y_0))\mathbf{D}_{Y_1} \cdots \mathbf{D}_{Y_t} \mathbf{1}^\top}, \quad \text{and} \quad \hat{\pi}_t = \frac{\hat{\rho}_0 \hat{\mathbf{D}}_{Y_1} \cdots \hat{\mathbf{D}}_{Y_t}}{\hat{\rho}_0 \hat{\mathbf{D}}_{Y_1} \cdots \hat{\mathbf{D}}_{Y_t} \mathbf{1}^\top}.
 \end{aligned}$$

In fact, Theorem 2 asserts that the previous equalities hold for the FSS with characteristic matrices $(\mathbf{D}_1, \mathbf{D}_2)$ deduced from this HMC, and are valid for every probability distribution of (Y_0, X_0) (see related Comment 2).

Now replace matrix \mathbf{G} in the previous model by

$$\mathbf{G}_1 = \left(\begin{array}{c|c|c} \frac{1}{2} & \frac{1}{4} & \frac{1}{3} \\ \hline \frac{1}{2} & \frac{3}{4} & \frac{2}{3} \end{array} \right).$$

A direct calculation gives

$$\begin{aligned}
 &P\{Y_2 = 1, g(X_2) = 2 \mid Y_1 = 1, g(X_2) = 2, Y_0 = 2, X_0 = 1\} = \frac{77}{360} \\
 \text{and} \quad &P\{Y_2 = 1, g(X_2) = 2 \mid Y_1 = 2, g(X_2) = 2, Y_0 = 2, X_0 = 1\} = \frac{77}{936},
 \end{aligned}$$

so that $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is not an FSS (and not an HMC as well). Note that $\{g(X_t)\}_{t \in \mathbb{N}}$ is an MC for every probability distribution of X_0 from Lemma 1 (see the form of matrix $\mathbf{P}\mathbf{L}$ above). Finally, it is easily checked that

$$\mathbf{D}_1\mathbf{L} = \left(\begin{array}{c|c} \frac{1}{4} & \frac{5}{36} \\ \hline \frac{1}{8} & \frac{5}{24} \\ \hline \frac{1}{8} & \frac{2}{9} \end{array} \right), \quad \mathbf{D}_2\mathbf{L} = \left(\begin{array}{c|c} \frac{1}{4} & \frac{13}{36} \\ \hline \frac{1}{8} & \frac{13}{24} \\ \hline \frac{1}{8} & \frac{19}{36} \end{array} \right).$$

Therefore, condition (C) in Comment 3 is satisfied, although the FSS is not lumpable.

3.2. Rogers–Pitman’s condition for FSSs

In this subsection we deal with a criterion which will be deduced from Rogers–Pitman’s criterion for the lumpability of MCs. In order to ease the exposition, $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ will be

thought of as a univariate Markov chain $\{Z_t\}_{t \in \mathbb{N}}$ with state space $\mathcal{Z} := \{1, \dots, NM\}$ using a lexicographic ordering of the elements of $\mathcal{Y} \times \mathcal{X}$:

$$Z_t = (y - 1)N + x \iff (Y_t, X_t) = (y, x), \quad x \in \mathcal{X}, y \in \mathcal{Y}.$$

When $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is an FSS, its $NM \times NM$ transition matrix has the form

$$Q = \begin{pmatrix} D_1 & D_2 & \cdots & D_M \\ \vdots & \vdots & & \vdots \\ D_1 & D_2 & \cdots & D_M \end{pmatrix}, \tag{22}$$

or in a compact form

$$Q = (\mathbf{1}_M^\top \otimes I_N)(D_1 \cdots D_M), \tag{23}$$

using the Kronecker product ‘ \otimes ’ of matrices. The probability distribution of Z_0 is the NM vector α obtained by listing the components of the $M \times N$ matrix $(P\{Y_0 = y_0, X_0 = x_0\})_{y_0 \in \mathcal{Y}, x_0 \in \mathcal{X}}$ using the lexicographic ordering. Next, the process $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is associated with the function $\{f(Z_t)\}_{t \in \mathbb{N}}$ of $\{Z_t\}_{t \in \mathbb{N}}$, where f is defined by

$$f((y - 1)N + x) := (y - 1)n + g(x) \quad \text{for all } y \in \mathcal{Y} \text{ and } x \in \mathcal{X}.$$

Then the corresponding lumping $MN \times Mn$ matrix is

$$L_f = \begin{pmatrix} L_g & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & L_g \end{pmatrix}, \tag{24}$$

where $\mathbf{0} \in \mathbb{R}^{N \times n}$ and L_g is the lumping $N \times n$ matrix associated with g . In a compact form, $L_f = I_M \otimes L_g$.

It is clear that $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is a Markov chain (an FSS) for

$$(P\{Y_0 = y_0, X_0 = x_0\})_{y_0 \in \mathcal{Y}, x_0 \in \mathcal{X}}$$

if and only if $\{f(Z_t)\}_{t \in \mathbb{N}}$ is a Markov chain (an FSS) with α as the probability distribution of Z_0 . Now, we state a condition for $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ to be an FSS for specific probability distributions of (Y_0, X_0) .

Theorem 3. *Let Λ_g be a stochastic $n \times N$ matrix such that*

$$\Lambda_g L_g = I_n. \tag{25}$$

Let us introduce the subset \mathcal{P} of probability distributions ν on $\mathcal{Y} \times \mathcal{X}$ such that

$$\nu(y, x) = \begin{cases} \Lambda_g(w, x) & \text{if } g(x) = w, \\ 0 & \text{otherwise.} \end{cases}$$

We denote the set of convex combinations of elements of \mathcal{P} by $\text{conv}(\mathcal{P})$.

If we have

$$\Lambda_g D_y = \Lambda_g D_y L_g \Lambda_g \quad \text{for every } y \in \mathcal{Y} \tag{26}$$

then $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an FSS for every probability distribution of (Y_0, X_0) in $\text{conv}(\mathcal{P})$. Its characteristic matrices $(\hat{D}_y, y \in \mathcal{Y})$ are given by

$$\hat{D}_y := \Lambda_g D_y L_g \quad \text{for all } y \in \mathcal{Y}. \tag{27}$$

Proof. The first step consists in checking that condition (26) asserts that the process $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an MC for every probability distribution of (Y_0, X_0) in $\text{conv}(\mathcal{P})$. Using the univariate framework proposed in the beginning of this subsection, we must check that the process $\{f(Z_t)\}_{t \in \mathbb{N}}$ satisfies Definition 1. Let us introduce the stochastic $Mn \times MN$ block diagonal matrix

$$\Lambda_f = I_M \otimes \Lambda_g. \tag{28}$$

Observe that $\Lambda_f L_f = I_{Mn}$ from relation (25). Each row of the matrix Λ_f corresponds to a unique probability distribution in the set \mathcal{P} . Next, using a block decomposition of each of the matrices, it is easily seen, from (23), (24), and (28), that

$$\Lambda_f Q = \begin{pmatrix} \Lambda_g D_1 & \Lambda_g D_2 & \cdots & \Lambda_g D_M \\ \vdots & \vdots & & \vdots \\ \Lambda_g D_1 & \Lambda_g D_2 & \cdots & \Lambda_g D_M \end{pmatrix}$$

and

$$\Lambda_f Q L_f = \begin{pmatrix} \Lambda_g D_1 L_g & \Lambda_g D_2 L_g & \cdots & \Lambda_g D_M L_g \\ \vdots & \vdots & & \vdots \\ \Lambda_g D_1 L_g & \Lambda_g D_2 L_g & \cdots & \Lambda_g D_M L_g \end{pmatrix}.$$

Therefore, $\Lambda_f Q = \Lambda_f Q L_f \Lambda_f$ if and only if we have, for every $y \in \mathcal{Y}$,

$$\Lambda_g D_y = \Lambda_g D_y L_g \Lambda_g.$$

This last equality is just (26). From Lemma 2 we deduce that $\{f(Z_t)\}_{t \in \mathbb{N}}$ is an MC for every probability distribution of Z_0 which is a convex combination of the rows of Λ_f . Its transition matrix \hat{Q} has the form

$$\hat{Q} := \Lambda_f Q L_f = \begin{pmatrix} \hat{D}_1 & \hat{D}_2 & \cdots & \hat{D}_M \\ \vdots & \vdots & & \vdots \\ \hat{D}_1 & \hat{D}_2 & \cdots & \hat{D}_M \end{pmatrix},$$

where $\hat{D}_y := \Lambda_g D_y L_g$. Thus, from the structure of the matrix \hat{Q} , it is clear that $\{f(Z_t)\}_{t \in \mathbb{N}}$ is an FSS (see, e.g. (22)).

Comment 4. When $\Lambda_g = (L_g^\top L_g)^{-1} L_g^\top$, condition (26) in Theorem 3 reads as follows: for every $y \in \mathcal{Y}$ and all $w_1, w_2 \in g(\mathcal{X})$,

$$\sum_{x_1 \in g^{-1}(w_1)} D_y(x_1, x_2)$$

is independent of $x_2 \in g^{-1}(w_2)$. This relation is the exact lumpability condition discussed in [4] and corresponds to backward simulation (see [19] and the references therein). This appears

as a ‘dual condition’ of that reported in Theorem 1 (see (13)). The duality between strong lumpability and Rogers–Pitman’s condition is made explicit in the context of Markov chains in [10, Paragraph 2.2.4].

The next result shows the interest in Rogers–Pitman’s condition for filtering an FSS.

Theorem 4. *Using the same notation as in Theorem 3, let $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ be an FSS with characteristic matrices $(D_y, y \in \mathcal{Y})$. Then the following two statements are equivalent.*

(a) *For any $y \in \mathcal{Y}$,*

$$\Lambda_g D_y = \Lambda_g D_y L_g \Lambda_g. \tag{29}$$

(b) *For any probability distribution of (Y_0, X_0) in $\text{conv}(\mathcal{P})$, for any $t \geq 0$,*

$$\hat{\rho}_{t+1} = \hat{\rho}_t \hat{D}_{Y_{t+1}} \quad \text{and} \quad \rho_t = \hat{\rho}_t \Lambda_g \tag{30}$$

for some family of $n \times n$ matrices $(\hat{D}_y, y \in \mathcal{Y})$ such that $\sum_y \hat{D}_y$ is stochastic.

Proof. First we prove that statement (a) implies statement (b). The expressions for ρ_0 and $\hat{\rho}_0$ in connection with our choice of probability distribution of (Y_0, X_0) in $\text{conv}(\mathcal{P})$ can be written from (7), (20), and (25) as

$$\rho_0 = \sum_{y_0 \in \mathcal{Y}} 1_{\{Y_0=y_0\}} \alpha_{y_0} \Lambda_g \quad \text{and} \quad \hat{\rho}_0 = \rho_0 L_g = \sum_{y_0 \in \mathcal{Y}} 1_{\{Y_0=y_0\}} \alpha_{y_0} \tag{31}$$

with

$$\alpha_{y_0} \in \mathbb{R}_+^n \quad \text{and} \quad \sum_{y_0 \in \mathcal{Y}} \alpha_{y_0} \mathbf{1}^\top = 1 \quad \text{for all } y_0 \in \mathcal{Y}.$$

The leftmost equality in (30) follows from the fact that $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an FSS with characteristic matrices $(\hat{D}_y, y \in \mathcal{Y})$ for every distribution of (Y_0, X_0) in $\text{conv}(\mathcal{P})$. Proof of the rightmost equality is by induction on t . For $t = 0$, the relation is deduced from (31):

$$\rho_0 = \sum_{y_0 \in \mathcal{Y}} 1_{\{Y_0=y_0\}} \alpha_{y_0} \Lambda_g = \left(\sum_{y_0 \in \mathcal{Y}} 1_{\{Y_0=y_0\}} \alpha_{y_0} \right) \Lambda_g = \hat{\rho}_0 \Lambda_g.$$

Assume that $\rho_t = \hat{\rho}_t \Lambda_g$. Then,

$$\begin{aligned} \rho_{t+1} &= \rho_t D_{Y_{t+1}} \\ &= \hat{\rho}_t \Lambda_g D_{Y_{t+1}} \\ &= \hat{\rho}_t \hat{D}_{Y_{t+1}} \Lambda_g \quad \text{from (27) and (29)} \\ &= \hat{\rho}_{t+1} \Lambda_g, \end{aligned}$$

from the leftmost equality in (30).

Conversely, assume that statement (b) holds. Then, from relations (30), for every $t \in \mathbb{N}$,

$$\rho_{t+1} = \rho_t D_{Y_{t+1}} = \hat{\rho}_t \Lambda_g D_{Y_{t+1}} \quad \text{and} \quad \rho_{t+1} = \hat{\rho}_{t+1} \Lambda_g = \hat{\rho}_t \hat{D}_{Y_{t+1}} \Lambda_g.$$

In combining the two representations of ρ_{t+1} for any $t \in \mathbb{N}$ we obtain

$$\hat{\rho}_t (\Lambda_g D_{Y_{t+1}} - \hat{D}_{Y_{t+1}} \Lambda_g) = \mathbf{0}$$

or, for all $y \in \mathcal{Y}$,

$$\hat{\rho}_t(\Lambda_g D_y - \hat{D}_y \Lambda_g) = \mathbf{0}.$$

In particular, we have, for $t = 0$,

$$\hat{\rho}_0(\Lambda_g D_y - \hat{D}_y \Lambda_g) = \mathbf{0} \quad \text{for all } y \in \mathcal{Y}. \tag{32}$$

Now we write, from (31),

$$\hat{\rho}_0 = \sum_{y_0 \in \mathcal{Y}} 1_{\{Y_0=y_0\}} \alpha_{y_0},$$

where $\alpha_{y_0} \in \mathbb{R}_+^n$ and $\sum_{y_0} \alpha_{y_0} \mathbf{1}^\top = 1$. Now, it is clear that $\text{span}(\alpha_{y_0}, y_0 \in \mathcal{Y}) = \mathbb{R}^n$. Then we deduce from (32) that

$$\alpha(\Lambda_g D_y L_g - \hat{D}_y \Lambda_g) = \mathbf{0} \quad \text{for all } y \in \mathcal{Y} \text{ and all } \alpha \in \mathbb{R}^n,$$

so that

$$\Lambda_g D_y L_g = \hat{D}_y \Lambda_g \quad \text{for all } y \in \mathcal{Y}.$$

Next, for any $y \in \mathcal{Y}$, the equality $\Lambda_g D_y = \hat{D}_y \Lambda_g$ implies that $\Lambda_g D_y L_g = \hat{D}_y \Lambda_g L_g = \hat{D}_y$. Then statement (a) is met.

Example 2. Let us illustrate the results by an example. We consider the same lumping map as in Example 1 for the FSS with characteristic matrices (D_1, D_2) :

$$D_1 := \left(\begin{array}{c|cc} \frac{1}{4} & \frac{1}{20} & \frac{1}{20} \\ \hline \frac{5}{24} & \frac{1}{5} & \frac{2}{15} \\ \frac{1}{8} & \frac{2}{15} & \frac{1}{5} \end{array} \right), \quad D_2 := \left(\begin{array}{c|cc} \frac{1}{4} & \frac{1}{5} & \frac{1}{5} \\ \hline \frac{1}{8} & \frac{2}{15} & \frac{1}{5} \\ \frac{5}{24} & \frac{1}{5} & \frac{2}{15} \end{array} \right).$$

Since $D_1(2, 1) \neq D_1(3, 1)$, the FSS is not strongly lumpable from Theorem 1. Note that the transition matrix of the Markov chain $\{X_t\}_{t \in \mathbb{N}}$,

$$P = D_1 + D_2 = \left(\begin{array}{c|cc} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \hline \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \right),$$

is strongly lumpable from Lemma 1. We introduce the following matrix Λ_g :

$$\Lambda_g := \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Note that P is also an R-P matrix for matrix Λ_g from Lemma 2.

Next we obtain

$$\hat{D}_1 := \Lambda_g D_1 L_g = \begin{pmatrix} \frac{1}{4} & \frac{1}{10} \\ \frac{1}{6} & \frac{1}{3} \end{pmatrix}, \quad \hat{D}_2 := \Lambda_g D_2 L_g = \begin{pmatrix} \frac{1}{4} & \frac{2}{5} \\ \frac{1}{6} & \frac{1}{3} \end{pmatrix}.$$

It is easily checked that $\Lambda_g D_i = \hat{D}_i \Lambda_g$ for $i = 1, 2$, so that $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an FSS with characteristic matrices (\hat{D}_1, \hat{D}_2) for every probability distribution of (Y_0, X_0) of the form $\alpha \Lambda_g = \alpha(1)(1, 0, 0) + (1 - \alpha(1))(0, \frac{1}{2}, \frac{1}{2})$, where α is any stochastic vector on $g(\mathcal{X})$. For any sequence of observations (Y_0, \dots, Y_T) , the filters are given from Theorem 3 as follows. For every stochastic two-dimensional vector α ,

$$\hat{\rho}_t = \alpha \hat{D}_{Y_0} \cdots \hat{D}_{Y_t} \quad \text{and} \quad \rho_t = \hat{\rho}_t \Lambda_g = \alpha \hat{D}_{Y_0} \cdots \hat{D}_{Y_t} \Lambda_g = \alpha \Lambda_g D_{Y_0} \cdots D_{Y_t}.$$

4. Approximation by strongly lumpable models

In [21] a procedure was derived that yielded a two-lumpable approximation to a given MC, that is, an approximation of a given MC by a strongly lumpable MC with two lumps ($n = 2$). By extension, two-lumpable approximations to HMCs were obtained. If this procedure is recursively repeated on each of the two lumpings until no more lumpings are possible then the approximate n -lumpable HMC will be determined and the filters for the approximate states computed. However, as the number of lumpings, n , approaches the number of states, N , in the original HMC, the question arises as to the benefits of determining any more lumpings. Also, there is no obvious procedure to obtain an optimal approximation when the lumping mapping is unknown. Note that, for large state spaces, a crucial question is how to search for a lumping map. But this point is beyond the scope of the present paper. We mention in passing that there is an algorithm with $O(k \log N)$ time complexity (where k is the number of nonzero transition probabilities) that computes the optimal lumping map for any MC [5] (which is optimal in terms of the reduction of the state space complexity). However, it remains unknown as to how to find a best lumping in some approximate sense when the MC is not lumpable.

In Subsection 4.1 we generalise the problem presented in [21] for MCs for $n > 2$, i.e. we assume that we have knowledge of the lumping map g for the process, and wish to determine the closest approximation (in the Frobenius norm sense) of the transition probability matrix on the assumed lumping. The results are generalised to FSSs using a process outlined in Subsection 4.2.

4.1. Approximation of MCs

Let

$$\mathcal{T}_N(L) = \{\Psi \in \mathbb{R}^{N \times N} : \ker(L)\Psi \subset \ker(L)\}.$$

Then $\Psi \in \mathcal{T}_N(L)$ if and only if

$$\Psi = V^T \begin{pmatrix} B_{11} & B_{12} \\ \mathbf{0} & B_{22} \end{pmatrix} V, \tag{33}$$

where $B_{11} \in \mathbb{R}^{n \times n}$, $B_{12} \in \mathbb{R}^{n \times (N-n)}$, $B_{22} \in \mathbb{R}^{(N-n) \times (N-n)}$, $V^T = [v_1^T, \dots, v_N^T]$, and the v_i s are the right singular vectors of L^T , arranged such that $\ker(L) = \text{span}\{v_{n+1}, \dots, v_N\}$. It can be seen directly from (33) that $\mathcal{T}_N(L)$ has dimension $N^2 - n(N - n)$. Note that we do not restrict elements of $\mathcal{T}_N(L)$ to be stochastic or even nonnegative matrices. The proof of this result follows in the same way as in the proof of Lemma 1 given in Appendix A.

Lemma 3. *The orthogonal projection onto the subspace $\mathcal{T}_N(L)$ of $\mathbb{R}^{N \times N}$ is given by*

$$\Pi(Q) = Q - (I_N - \Phi)Q\Phi, \quad \text{where } \Phi = V^T \begin{pmatrix} I_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} V. \tag{34}$$

Proof. Let $Q \in \mathbb{R}^{N \times N}$. Then we seek a matrix $\Psi \in \mathcal{T}_N(L)$ which minimises $\|Q - \Psi\|$. Let $B = V\Psi V^T$ and $C = VQV^T$. We know that B has the form of the block matrix in (33). Partition C commensurately with B , then, clearly, $\|Q - \Psi\| = \|C - B\|$ is minimised by taking

$$\Psi = V^T \begin{pmatrix} C_{11} & C_{12} \\ \mathbf{0} & C_{22} \end{pmatrix} V.$$

It is simple to verify that indeed Ψ has the form (34).

If Q is a stochastic matrix then, in general, $\Pi(Q)$ will not be. In order to obtain a stochastic matrix corresponding to the transition probability matrix of a strongly lumpable MC, we employ the method of alternating convex projections [23]. Apart from the above subspace constraint, we also need to force nonnegativity of the elements of the approximant and the property that the row sums are unity. We do this by successively applying the following two projection operators to $\Pi(Q)$. Let us consider the subset of $\mathbb{R}^{N \times N}$ consisting of nonnegative matrices. It can be shown that the (convex) projection $\Pi_+(M)$ of any $M \in \mathbb{R}^{N \times N}$ onto this subset is

$$\Pi_+(M)(i, j) = \begin{cases} M(i, j) & \text{if } M(i, j) \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Now consider the affine subspace of $\mathbb{R}^{N \times N}$ consisting of matrices with all row sums unity. It can be shown that the (affine) projection of $M \in \mathbb{R}^{N \times N}$ onto this subspace is given by

$$\Pi_1(M) = M \left(I_N - \frac{\mathbf{1}^T \mathbf{1}}{N} \right) + \frac{\mathbf{1}^T \mathbf{1}}{N}. \tag{35}$$

The algorithm proposed in [21] consists of repeatedly applying each of the three projection operators above to the given matrix Q until convergence is noted. It is known that the algorithm will converge to a stochastic matrix being the transition probability matrix of a strongly lumpable Markov chain, although the optimality of the approximation has not been demonstrated.

Let us illustrate how the procedure works using two examples. The first is didactic, whilst the second illustrates the convergence behaviour of the algorithm in a more realistic case. Finally, we compare the filtering performance for the aggregated states of the optimal filter, and the filter obtained from a lumpable approximation, to the model.

Example 3. Consider an MC with $N = 3$ states, which we wish to lump into $n = 2$ groups under the lumping matrix

$$L = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

We start with a matrix Q_0 which is constructed by randomly perturbing the transition probability matrix Q of an MC lumpable under L (with appropriate renormalisation),

$$Q = \left(\begin{array}{c|cc} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \hline \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \right).$$

The initial iterate is determined by adding a random matrix with independent and identically distributed entries uniform on $[-0.1, 0.1]$ and renormalising, giving

$$Q_0 = \begin{pmatrix} 0.5471 & 0.2292 & 0.2237 \\ 0.3012 & 0.4434 & 0.2554 \\ 0.3117 & 0.3390 & 0.349 \end{pmatrix}.$$

The algorithm terminates after one iteration, giving the approximate transition matrix

$$Q_1 = \begin{pmatrix} 0.5471 & 0.2292 & 0.2237 \\ 0.3064 & 0.4408 & 0.2528 \\ 0.3064 & 0.3416 & 0.3520 \end{pmatrix},$$

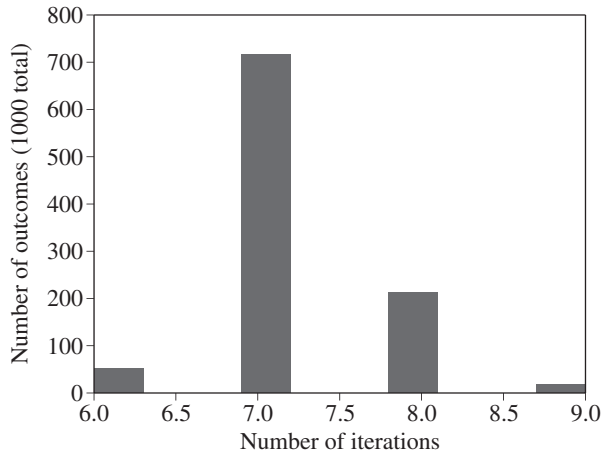


FIGURE 1: Histogram of the number of iterations of the convex projection algorithm for $N = 100$ and $n = 2$.

corresponding to the transition matrix of an MC lumpable under L . Note, however, that the algorithm does not recover Q . Note that $\|Q_1 - Q\| = 0.1525$ and $\|Q_1 - Q_0\| = 0.091$, so indeed the solution is closer to the initial iterate than to Q .

The next example has $N = 100$ and $n = 2$. The sizes of the groups were $N_1 = 40$ and $N_2 = 60$. We ran 1000 independent realisations with random initial transition matrices. The number of iterations until the relative error between successive iterates dropped to 10^{-8} was determined, and a histogram of the results is shown in Figure 1. These results show, and are supported by other experiments, that the algorithm converges in only a few steps, even for relatively large problems.

4.2. Approximation of FSSs

An FSS is characterised by a finite set $(D_y, y \in \mathcal{Y})$ of substochastic matrices. The above approximation procedure can also be applied to FSSs with one modification. The lumpability and nonnegativity constraints can be applied to each D_y independently, however, the stochasticity constraint needs to be applied to the $N \times NM$ matrix $(D_1 D_2 \cdots D_M)$. An appropriate modification of (35) is easily deduced.

Whilst a full study of the performance of the approximate filtering scheme deduced from the above procedure in specific applications is beyond the scope of this paper, we provide the following simple example for the purpose of illustrating the potential utility of the technique.

Example 4. Let the state space be $\mathcal{X} = \{1, \dots, N\}$ for some integer $N > 1$, and let the observation space be $\mathcal{Y} = \{1, \dots, N^2\}$. We define the observations according to

$$Y_t = [N(X_{t-1} - 1) + X_t + V_t],$$

where V_t is an independent and identically distributed sequence of zero-mean normal random variables with variance σ^2 and $[\cdot]$ denotes taking the integer part, with appropriate consideration for the boundaries. Thus, Y_t is a lexicographic ordering of (X_{t-1}, X_t) with a disturbance increasing with increasing σ^2 owing to the V_t . Therefore, the conditional probability distribution of Y_t given (X_{t-1}, X_t) is specified. The state transition probabilities were randomly

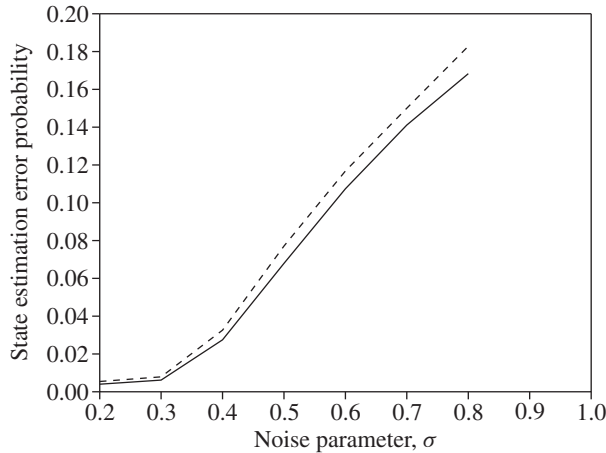


FIGURE 2: Performance of the optimal (solid line) and approximate (dashed line) filters.

chosen and the model remained fixed over all experiments. The state aggregation was the two subsets $\{1, \dots, N_1\}$ and $\{N_1 + 1, \dots, N\}$ for some $1 \leq N_1 < N$.

To compare the performance of the optimal filter for the aggregated states to the filter derived from the approximate two-lumpable approximation, we generated 10 000 independent realisations of the above FSS, each of length 100 samples. We chose $N = 5$ and $N_1 = 2$. Figure 2 shows the state estimation errors for each case, as σ^2 varies. As expected, the performance of both filters degrades with increasing σ^2 , the approximate filter performs well and remains surprisingly robust.

5. Conclusions

In this paper we have generalised the strong lumpability results for the HMCs of [21] in a number of significant ways. We have established the theory for the strong lumpability of FSSs, which include HMCs as a special case. A (weak) lumpability condition based on [12] and [16] was introduced and discussed for FSSs. The main results concerned the linear dynamics of the lumped FSS. First, a necessary and sufficient condition for the dynamics of the lumped process to be linear was established. Second, using the (weak) lumpability condition, the filter for the FSS was derived directly from the filter for the lumped FSS (and the lumpability condition was shown to be necessary for this specific derivation to hold). Finally, an algorithm for the approximation of an arbitrary FSS by a strongly lumpable FSS was proposed.

Appendix A. Proof that Lemma 1(d)–(f) are equivalent for any $n > 2$

We assume that the right singular vectors of L^\top are ordered so that

$$\ker(L) = \text{span}\{v_{n+1}, \dots, v_N\}.$$

The (i, j) th element of the matrix VQV^\top is $v_i Q v_j^\top$. Let $i \in \{n+1, \dots, N\}$ and $j \in \{1, \dots, n\}$, and consider $v_i Q \in \ker(L)Q \subset \ker(L)$ (by Lemma 1(d)). Thus, $v_i Q \perp v_j$, since the right singular vectors are orthogonal [8, Theorem 2.5.2]. Hence, the lower left-hand block of VQV^\top , as shown in (12), is 0 as claimed. Now the range of L^\top is spanned by the rows of L , and these are precisely the usual unit orthonormal basis. Thus, there is an ordering of the v_i , consistent

with the above ordering such that the matrix of left singular vectors of L^\top is the identity matrix of size n . Thus, we can write the singular value decomposition [8] $L^\top = SV$, where $S = [S_1 \mathbf{0}]$ and S_1 is an $n \times n$ diagonal matrix with strictly positive entries. Let $V^\top = [V_1^\top \ V_2^\top]$, where V_1^\top is of size $N \times n$. Consider

$$\begin{aligned} L(L^\top L)^{-1/2} &= V^\top S^\top (SS^\top)^{-1/2} \\ &= V_1^\top S_1 (S_1 S_1)^{-1/2} \\ &= V_1^\top. \end{aligned}$$

The matrix on the left is nonnegative, so is V_1^\top . Therefore, in (12), the matrix $B_{11} = V_1 Q V_1^\top$ is nonnegative (since Q has nonnegative entries).

Now let $x \in \ker(L)Q$. Then there are scalars $\alpha_{n+1}, \dots, \alpha_N$ such that

$$x = \sum_{i=n+1}^N \alpha_i v_i Q.$$

Let $j \in \{1, \dots, n\}$, and consider

$$x v_j^\top = \sum_{i=n+1}^N \alpha_i v_i Q v_j^\top = 0,$$

by Lemma 1(f). Thus, $x \perp \text{span}\{v_1, \dots, v_n\} = \ker(L)^\perp$, so $x \in \ker(L)$, establishing Lemma 1(d).

Appendix B. Direct proof that Theorem 1(a)–(f) are equivalent

First, we know that the conditional probability in (18) is independent of y_0 . It follows from (1) that they are equal to

$$P\{Y_{t+1} = y_1, g(X_{t+1}) = w_2 \mid Y_t = y_0, X_t = x_1\} = \sum_{x_2 \in g^{-1}(w_2)} D_{y_1}(x_1, x_2).$$

This is exactly what is required in (a). Statement (a) can be reformulated in matrix form as follows: for every $y_1 \in \mathcal{Y}$,

$$D_{y_1} L = L \hat{D}_{y_1}, \tag{36}$$

where \hat{D}_{y_1} is a nonnegative $n \times n$ matrix and $D_{y_1}(w_1, w_2)$ is the common value of the probabilities in (13). Then, using $UL = I_n$, we obtain (16). By expressing the matrix equality (36) element by element and using (1), statement (a) is obtained. Therefore, statements (a) and (c) are equivalent.

Next assume that (c) holds. Let $x \in \ker(L)$. Then it follows from (16) that

$$x D_y L = x L U D_y L = \mathbf{0}.$$

Therefore, $x D_y \in \ker(L)$ and statement (d) is satisfied.

If (d) holds then, for any $x \in \ker(L)$, we have $x D_y \in \ker(L)$, that is, $x D_y L = \mathbf{0}$. Therefore, $x \in D_y \ker(L)$ and (e) is valid. Next, statement (e), i.e. $\ker(L) \subseteq \ker(D_y L)$, implies that there exists an $n \times n$ matrix \hat{D} such that $D_y L = L \hat{D}$ (see, e.g. [22]). If we multiply from the left

the previous relation by U , we obtain (16) and statement (c) is satisfied. Thus, statements (c), (d), and (e) are equivalent.

Note that (14) and (15) in statement (b) are just a rewriting of (13), which gives insight into the structure of the matrix D_y required for having strong lumpability. That $\sum_{y \in \mathcal{Y}} \sum_{j=1}^n c_{k,j}(y) = 1$ for each $k = 1, \dots, n$ follows from the fact that $\sum_y D_y$ is a stochastic matrix.

Finally, the equivalence of statements (d) and (f) can be deduced as in Appendix A for the MC case.

Acknowledgement

The authors thank the anonymous referee for his/her comments, which led to a substantial improvement in the presentation of the paper.

References

- [1] BARAS, J. S. AND FINESSO, L. (1992). Consistent estimation of the order of hidden Markov chains. In *Stochastic Theory and Adaptive Control* (Lecture Notes Control Inf. Sci. **184**), eds T. E. Duncan and B. Pasik-Duncan, Springer, Berlin, pp. 26–39.
- [2] BENOIT, A., BRENNER, L., FERNADES, P. AND PLATEAU, B. (2004). Aggregation of stochastic automata networks with replicas. *Linear Algebra Appl.* **386**, 111–136.
- [3] BREUER, L. AND BAUM, D. (2005). *An Introduction to Queuing Theory*. Springer, Berlin.
- [4] BUCHHOLZ, P. (1994). Exact and ordinary lumpability in finite Markov chains. *J. Appl. Prob.* **31**, 59–75.
- [5] DERISAVI, S., HERMANN, H. AND SANDERS, W. H. (2003). Optimal state-space lumping in Markov chains. *Inf. Process. Lett.* **87**, 309–315.
- [6] EPHRAIM, Y. AND MERHAV, N. (2002). Hidden Markov processes. *IEEE Trans. Inf. Theory* **48**, 1518–1569.
- [7] GILMORE, S., HILLSTON, J. AND RIBAUDO, M. (2001). An efficient algorithm for aggregating PEPA models. *IEEE Trans. Software Eng.* **27**, 449–464.
- [8] GOLUB, G. H. AND VAN LOAN, C. F. (1996). *Matrix Computations*, 3rd edn. Johns Hopkins University Press, Baltimore, MD.
- [9] GRINFELD, M. AND KNIGHT, P. A. (2000). Weak lumpability in the k -sat problem. *Appl. Math. Lett.* **13**, 49–53.
- [10] GURVITS, L. N. AND LEDOUX, J. (2005). Markov property for a function of a Markov chain: a linear algebra approach. *Linear Algebra Appl.* **404**, 85–117.
- [11] HERMANN, H., HERZOG, U. AND KATOEN, J.-P. (2002). Process algebra for performance evaluation. *Theoret. Comput. Sci.* **274**, 43–87.
- [12] KEMENY, J. G. AND SNELL, J. L. (1976). *Finite Markov Chains*. Springer, New York.
- [13] LEDOUX, J. (2005). Recursive filters for partially observable finite Markov chains. *J. Appl. Prob.* **42**, 687–697.
- [14] PAIGE, R. L. (2001). Classification and lumpability in the stochastic Hopfield model. *Adv. Appl. Prob.* **33**, 930–943.
- [15] PICCI, G. (1978). On the internal structure of finite state stochastic processes. In *Recent Developments in Variable Structure Systems* (Lecture Notes Econom. Math. Systems **162**), eds R. R. Mohler and A. Ruberti, Springer, Berlin, pp. 288–304.
- [16] ROGERS, L. C. G. AND PITMAN, J. W. (1981). Markov functions. *Ann. Prob.* **9**, 573–582.
- [17] SPREIJ, P. J. C. (2001). On the Markov property of a finite hidden Markov chain. *Statist. Prob. Lett.* **52**, 279–288.
- [18] SPREIJ, P. J. C. (2003). On hidden Markov chains and finite stochastic systems. *Statist. Prob. Lett.* **62**, 189–201.
- [19] SPROSTON, J. AND DONATELLI, S. (2006). Backward bisimulation in Markov chain model checking. *IEEE Trans. Software Eng.* **32**, 531–546.
- [20] TIAN, J. AND LIN, X.-S. (2005). Colored coalescent theory. *Discrete Continuous Dynamical Systems* **2005**, 833–845.
- [21] WHITE, L. B., MAHONY, R. AND BRUSHE, G. D. (2000). Lumpable hidden Markov models – model reduction and reduced complexity filtering. *IEEE Trans. Automatic Control* **45**, 2297–2306.
- [22] WONHAM, W. M. (1979). *Linear Multivariable Control: A Geometric Approach*. Springer, New York.
- [23] YOULA, D. C. (1987). Mathematical theory of image restoration by the method of convex projections. In *Image Recovery: Theory and Applications*, ed. H. Stark, Academic Press, New York, pp. 29–77.