A fundamental limitation to the reduction of Markov chains via aggregation

Georgios Kotsalis   Jeff S. Shamma

***Abstract*— This paper highlights a limitation of state space aggregation based model reduction of Markov chains. It is shown that within the set of Markov chains of a given dimension that admit an exact low order representation, the set of Markov chains where this exact low order representation can be extracted by means of aggregation based model reduction is a nowhere dense set.**

A. Notation

- $\mathbb{R}$ is the set of real numbers,
$$\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$$
is the set of integers.
- We further use the following restrictions:
$$\begin{aligned} \mathbb{R}_+ &= \{x \in \mathbb{R} \mid x > 0\}, \\ \mathbb{Z}_+ &= \{x \in \mathbb{Z} \mid x > 0\}, \\ \mathbb{N} &= \mathbb{Z}_+ \cup \{0\}. \end{aligned}$$
- The letters k, l, m, n will typically denote positive integers.
- $\mathbb{Z}_n$ denotes the set of positive integers less than or equal to n, i.e.
$$\mathbb{Z}_n = \{x \in \mathbb{Z}_+ \mid x \le n\} = \{1, \dots, n\}.$$
- For two sets $\mathbb{A}, \mathbb{B}$ their set difference is denoted by
$$\mathbb{A} - \mathbb{B} = \{x \in \mathbb{A} \mid x \notin \mathbb{B}\}.$$
- For a finite set $\mathbb{A}$ its cardinality is denoted by $|\mathbb{A}|$.
- For a finite ordered set $\mathbb{A} = \{a_1, \dots, a_m\}$, define the bijection $i_{\mathbb{A}} : \mathbb{A} \to \mathbb{Z}_m$ with $i_{\mathbb{A}}[a_k] = k,\ k \in \mathbb{Z}_m$.
- Given a set $\mathbb{X} \subset \mathbb{R}$, the set of n-by-m matrices with entries in $\mathbb{X}$ is denoted by $\mathbb{X}^{n \times m}$. For instance, when $\mathbb{X} = \mathbb{R}$,
$$A \in \mathbb{R}^{n \times m} \Leftrightarrow A = \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nm} \end{bmatrix},$$
where $a_{ij} \in \mathbb{R},\ i \in \mathbb{Z}_n,\ j \in \mathbb{Z}_m$. If a capital letter is used to denote a matrix, (e.g. A), then the corresponding lower case letter with subscript ij refers to the entry in the ith row and jth column, (e.g., a_{ij}).

Georgios Kotsalis and Jeff S. Shamma are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. Email: georgios.kotsalis@ece.gatech.edu, and shamma@gatech.edu.

Research was supported in part by ARO/MURI W911NF-09-1-0553, AFOSR/MURI FA9550-10-1-0573

- The tranpose of the matrix $A \in \mathbb{R}^{n \times m}$ is denoted by A^T, where $A^T \in \mathbb{R}^{m \times n}$.
$$C = A^T \Rightarrow c_{ij} = a_{ji},\ i \in \mathbb{Z}_n,\ j \in \mathbb{Z}_m.$$
- For $A \in \mathbb{R}^{n \times m}$, $A(:, k)$ designates the kth column,
$$A(:, k) = \begin{bmatrix} a_{1k} \\ \vdots \\ a_{nk} \end{bmatrix},\ k \in \mathbb{Z}_m.$$
Accordingly $A = \begin{bmatrix} A(:,1), & \dots, & A(:,m) \end{bmatrix}$.
- For $A \in \mathbb{R}^{n \times m}$, the notation $A > 0$ indicates that A consists of positive elements, i.e.,
$$A > 0 \Leftrightarrow a_{ij} \in \mathbb{R}_+,\ i \in \mathbb{Z}_n,\ j \in \mathbb{Z}_m.$$
Similarly $A \ge 0 \Leftrightarrow a_{ij} \ge 0,\ i \in \mathbb{Z}_n,\ j \in \mathbb{Z}_m$.
- For $A \in \mathbb{R}^{n \times n}$ the notation $A \succ 0$, indicates that A is a positive definite matrix, i.e.,
$$A \succ 0 \Leftrightarrow x^T A x \in \mathbb{R}_+,\ x \in \mathbb{R}^n.$$
Similarly the notation $A \succeq 0$ indicates that A is a positive semi-definite matrix.
- The identity matrix in $\mathbb{R}^{n \times n}$ is written as I_n.
- Let $(\sigma_1, \dots, \sigma_n)$ be a permutation of $(1, \dots, n)$. The associated permutation matrix $A \in \mathbb{R}^{n \times n}$ is given by
$$a_{ij} = \begin{cases} 1 & \text{if } j = \sigma_i, \\ 0 & \text{otherwise,} \end{cases}$$
where $i \in \mathbb{Z}_n,\ j \in \mathbb{Z}_n$. In each row (or column) of a permutation matrix there is exactly one entry with value one; all other entries are zero. The set of all permutation matrices in $\mathbb{R}^{n \times n}$ is denoted by $\mathbb{P}_{(n,n)}$.
- Let $\mathbb{R}^n$ denote the vector space of real n-vectors.
$$x \in \mathbb{R}^n \Leftrightarrow x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix},\ x_i \in \mathbb{R},\ i \in \mathbb{Z}_n.$$
We refer to x_i, $i \in \mathbb{Z}_n$, as the ith component of x. We are identifying $\mathbb{R}^n$ with $\mathbb{R}^{n \times 1}$ and so the members of $\mathbb{R}^n$ are column vectors. On the other hand the elements of $\mathbb{R}^{1 \times n}$ are row vectors:
$$x \in \mathbb{R}^{1 \times n} \Leftrightarrow x = (x_1, \dots, x_n),\ x_i \in \mathbb{R},\ i \in \mathbb{Z}_n.$$
- For $x \in \mathbb{R}^n$,
$$\|x\| = \sqrt{\sum_{i=1}^n x_i^2} = \sqrt{x^T x},\quad \|x\|_1 = \sum_{i=1}^n |x_i|.$$

- The simplex in $\mathbb{R}^n$ is denoted by $\mathbb{S}_{(n)}$, i.e.,
$$\mathbb{S}_{(n)} = \{x \in \mathbb{R}^n \mid \|x\|_1 = 1,\ x_i \geq 0,\ i \in \mathbb{Z}_n\}.$$
- The vector consisting of n ones is denoted by 1_n, i.e., $1_n \in \mathbb{R}^n$ and $1_n^T = (1, \ldots, 1)$.
- For $x \in \mathbb{R}^n$, $A = \mathrm{diag}[x] \in \mathbb{R}^{n \times n}$ denotes the diagonal matrix where for $i \in \mathbb{Z}_n,\ j \in \mathbb{Z}_m$
$$a_{ij} = \begin{cases} x_i & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$
- For $A \in \mathbb{R}^{n \times n}$,
$$\mathrm{trace}[A] = \sum_{i=1}^{n} a_{ii}, \quad \|A\|_F = \mathrm{trace}[A^T A].$$
- The probability of an event A is written $\mathbf{Pr}[A]$, and the expectation of the random variable X is written $\mathbf{E}[X]$.
- Given a set $\mathbb{X}$, let $S = \mathrm{rand}[\mathbb{X}]$ stand for S being sampled uniformly out of the elements of $\mathbb{X}$.

I. Discrete-time, stationary processes on finite alphabets

In this work we deal exclusively with discrete-time, stationary stochastic processes that take values on finite alphabets. By finite alphabet we mean a set with a finite number of elements that are ordered. When referring to such a stochastic process, the attributes discrete-time and stationary might be omitted, but they will be tacitly assumed.

Consider the finite alphabet $\mathbb{A}$. Finite sequences of symbols from $\mathbb{A}$ are called strings. The symbols in a string are concatenated from right to left. For example suppose $v_i \in \mathbb{A}, i \in \mathbb{Z}_n$ and consider the string $v = v_n \ldots v_1$, then the symbol v_1 preceedes v_2, v_2 preceedes v_3, and so forth. $\mathbb{A}^*$ denotes the set of all finite strings, including the empty string $\varnothing$. Given two strings, $v = v_n \ldots v_1 \in \mathbb{A}^*$, $u = u_m \ldots u_1 \in \mathbb{A}^*$, denote by $uv = u_m \ldots u_1 v_n \ldots v_1$ the string obtained by concatenation of u to v. For $v \in \mathbb{A}^*$, $|v|$ (the length of v) is the numbers of symbols in v. By convention $|\varnothing| = 0$. $\mathbb{A}_{(n)} \subset \mathbb{A}^*$ is the set of strings with length n. Consider a stochastic process $\mathcal{Z} = \{Z_t; t \in \mathbb{Z}\}$ with values on the finite alphabet $\mathbb{A}$, over some probability space $(\Omega, \mathcal{F}, \mathbf{P})$. For $u = u_k \ldots u_1 \in \mathbb{A}_{(k)}$ and $t \in \mathbb{Z}$ we shall use for the event $\{Z_{t-1} = u_1, \ldots, Z_{t-k} = u_k\}$ the shorthand notation $\{Z_t^{(-)} = u\}$ and similarly $\{Z_t^{(+)} = u\} = \{Z_{t+1} = u_1, \ldots, Z_{t+k} = u_k\}$.

Definition 1.1: Given a stochastic process $\mathcal{Z} = \{Z_t; t \in \mathbb{Z}\}$ with values on the finite alphabet $\mathbb{A}$ we define
$$p_{\mathcal{Z}}[v_n \ldots v_1] := \mathbf{Pr}[Z_t^{(+)} = v_n \ldots v_1],$$
where $v_n \ldots v_1 \in \mathbb{A}_{(n)}, t \in \mathbb{Z}$. The **probability function** of the process $\mathcal{Z}$ is the map $p_{\mathcal{Z}} : \mathbb{A}^* \to [0, 1]$.

Remarks:
- The probability function is independent of the time instant t by stationarity.
- The stochastic process is fully defined by its probability function, therefore we use the notation $\mathcal{Z} = \{Z_t; t \in \mathbb{Z}\}$ and $p_{\mathcal{Z}}$ interchangeably.
- The probability function satisfies the standard axioms of probability theory, the following property can be readily verified
$$p_{\mathcal{Z}}[v] = \sum_{u \in \mathbb{A}_{(k)}} p_{\mathcal{Z}}[uv],\ v \in \mathbb{A}^*.$$

Definition 1.2: Two stochastic processes $\mathcal{Z} = \{Z_t; t \in \mathbb{Z}\}$ and $\tilde{\mathcal{Z}} = \{\tilde{Z}_t; t \in \mathbb{Z}\}$ taking values on the same finite alphabet $\mathbb{A}$ are called (statistically) **equivalent**, and we write $\mathcal{Z} \overset{\mathrm{p}}{\sim} \tilde{\mathcal{Z}}$, if
$$p_{\mathcal{Z}}[v] = p_{\tilde{\mathcal{Z}}}[v],\ v \in \mathbb{A}^*. \tag{1}$$

Remark:
- The two stochastic processes $\mathcal{Z}$ and $\tilde{\mathcal{Z}}$ must only coincide in their probability laws in order to be equivalent. They need not be defined on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$. The only requirement is that the strings in $\mathbb{A}^*$ produced under $\mathcal{Z}$ and $\tilde{\mathcal{Z}}$ exhibit the same statistics.

In the next section we introduce a class of processes that can be described in finite terms. They are called finite rank processes and have been investigated in the context of realization theory of Hidden Markov processes, [1], [3], [4].

A. Finite rank processes and quasi-realizations

Consider a stochastic process $\mathcal{Z} = \{Z_t; t \in \mathbb{Z}\}$ taking values on the finite alphabet $\mathbb{A}$. The set $\mathbb{A}^*$ is countable since it is a countable union of finite sets. One can introduce an order on $\mathbb{A}^*$, let $\mathbb{V} = \{v_1, v_2, \ldots\}$ correspond to a first lexical ordering and $\mathbb{U} = \{u_1, u_2, \ldots\}$ correspond to a last lexical ordering of $\mathbb{A}^*$. Let $H_{(\mathcal{Z})}$ stand for the doubly infinite matrix with
$$h_{(\mathcal{Z})ij} := p_{\mathcal{Z}}[v_i u_j] \tag{2}$$
where v_i is the ith element of $\mathbb{V}$, $i \in \mathbb{Z}_n$ and u_j is the jth element of $\mathbb{U}$, $j \in \mathbb{Z}_m$.

Definition 1.3: Consider a stochastic process $\mathcal{Z} = \{Z_t; t \in \mathbb{Z}\}$ with values on a finite alphabet $\mathbb{A}$. If $\mathrm{rank}[H_{(\mathcal{Z})}]$ is finite then $\mathcal{Z}$ is a **finite rank process**.

Definition 1.4: Consider a finite rank process $\mathcal{Z} = \{Z_t; t \in \mathbb{Z}\}$ with values on a finite alphabet $\mathbb{A}$. Suppose that $\mathrm{rank}[H_{(\mathcal{Z})}] = k$ and let $n \geq k$. Let $c \in \mathbb{R}^n$, $A : \mathbb{A} \to \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$. The tuple $\mathbf{Q}_{\mathcal{Z}} = (c, A, b)$ is a size n **quasi-realization** of the probability function $p_{\mathcal{Z}}$ if the following conditions hold:
$$\begin{aligned} p_{\mathcal{Z}}[v_m \ldots v_1] &= c^T A[v_m] \ldots A[v_1] b,\ v_m \ldots v_1 \in \mathbb{A}_{(m)} \\ c^T &= c^T (\sum_{v \in \mathbb{A}} A[v]) \\ b &= (\sum_{v \in \mathbb{A}} A[v]) b \end{aligned}$$
If additionally $n = k$, then the quasi-realization is called **minimal**, (regular).

Theorem 1.1: Every finite rank process has a minimal quasi-realization.

Proof: See for instance [1],[4]. ■

Theorem 1.2: Suppose $\mathcal{Z} = \{Z_t; t \in \mathbb{Z}\}$ is a finite rank process with values on a finite alphabet $\mathbb{A}$. Let $\mathbf{Q}_{\mathcal{Z}} = (c, A, b)$

and $\tilde{\mathbf{Q}}_{\mathcal{Z}} = (\tilde{c}, \tilde{A}, \tilde{b})$ be two size n, minimal, quasi-realizations of $p_{\mathcal{Z}}$. There exists a non singular matrix $T \in \mathbb{R}^{n\times n}$ such that

$$\begin{aligned} c^T &= \tilde{c}^T T^{-1} \\ A[v] &= T\tilde{A}[v]T^{-1}, \quad v \in \mathbb{A} \\ b &= T\tilde{b}. \end{aligned}$$

Proof: See for instance [1],[4]. ■

Remarks:

- The finite rank property allows the description of a given stochastic process in finite terms by means of quasi-realizations.
- A minimal quasi-realization can be extracted by generalizing standard state-variable realization algorithms from linear system theory to the situation where the underlying transfer function is rational in more than one indeterminates.

B. Markov chains and their statistical description

Throughout this section we will make use of the finite alphabet $\mathbb{X} = \{x_1, \dots, x_n\}$.

Definition 1.5: A stochastic process $\mathcal{X} = \{X_t; t \in \mathbb{Z}\}$ taking values on $\mathbb{X}$ is a (first order) **Markov chain** if

$$\mathbf{Pr}[X_t = v_0 \mid X_t^{(-)} = u] = \mathbf{Pr}[X_t = v_0 \mid X_{t-1} = u_1], \quad (3)$$

where $v_0 \in \mathbb{X}, u = u_m \dots u_1 \in \mathbb{X}_{(m)}, t \in \mathbb{Z}$.

Definition 1.6: Consider a Markov chain $\mathcal{X} = \{X_t; t \in \mathbb{Z}\}$ with values on $\mathbb{X}$. The matrix $P \in \mathbb{R}^{n\times n}$ with

$$p_{ij} = \mathbf{Pr}[X_t = x_i | X_{t-1} = x_j], \; i \in \mathbb{Z}_n, j \in \mathbb{Z}_n, t \in \mathbb{Z},$$

is the **stochastic transition matrix** of $\mathcal{X}$. The vector $\pi \in \mathbb{R}^n$ with

$$\pi_i = \mathbf{Pr}[X_t = x_i], \; i \in \mathbb{Z}_n, t \in \mathbb{Z},$$

is a **stationary distribution** of $\mathcal{X}$.

It follows directly from Definition 1.6 that $P \in [0,1]^{n\times n}$, $P \; 1_n = 1_n$, $\pi \in \mathbb{S}_{(n)}$ and $P\;\pi = \pi$.

Theorem 1.3: Consider a Markov chain $\mathcal{X} = \{X_t; t \in \mathbb{Z}\}$ with values on $\mathbb{X}$ and stochastic transition matrix P. Suppose there exists k such that $P^k > 0$. Then the Markov chain $\mathcal{X}$ has exactly one stationary distribution π.

Proof: See for instance [2]. ■

Assumption:

- In this work we assume that for a given Markov chain $\mathcal{X} = \{X_t; t \in \mathbb{Z}\}$ with values on $\mathbb{X}$ the entries of the associated stochastic transition matrix are strictly positive, i.e. $P > 0$. The hypothesis of Theorem 1.3 is satisfied with $k = 1$ and as such the corresponding stationary distribution π is unique, moreover:

$$\pi_i > 0, \; i \in \mathbb{Z}_n.$$

Theorem 1.4: Consider a Markov chain $\mathcal{X} = \{X_t; t \in \mathbb{Z}\}$ taking values on $\mathbb{X}$ with stochastic transition matrix P and stationary distribution π. The probability of a string in $\mathbb{X}^*$ is given by

$$p_{\mathcal{X}}[v_m \dots v_1] = \sum_{s\in\mathbb{Z}_n} P_{i[v_m]i[v_{n-1}]} \dots P_{i[v_1]s}\pi_s, \qquad (4)$$

where $v_m \dots v_1 \in \mathbb{X}_{(m)}$.

Proof: See for instance [2]. ■

Theorem 1.4 shows that for a given Markov chain $\mathcal{X} = \{X_t; t \in \mathbb{Z}\}$ with values on $\mathbb{X}$, the pair $\mathbf{M}_{\mathcal{X}} = (\pi, P)$ fully specifies its probability function $p_{\mathcal{X}}$.

Definition 1.7: Consider a Markov chain $\mathcal{X} = \{X_t; t \in \mathbb{Z}\}$ with values on $\mathbb{X}$. The **statistical parameters** of the **Markov chain** $\mathcal{X}$ are given by the pair $\mathbf{M}_{\mathcal{X}} = (\pi, P)$.

An alternative way of specifying the statistical information of a given Markov chain $\mathcal{X} = \{X_t; t \in \mathbb{Z}\}$ taking values on $\mathbb{X}$ is through the matrix $G_{(\mathcal{X})} \in \mathbb{R}_+^{n\times n}$ where

$$g_{(\mathcal{X})ij} := \mathbf{Pr}[X_{t+1} = i, X_t = j], \quad i \in \mathbb{X}, j \in \mathbb{X}, t \in \mathbb{Z}. \quad (5)$$

The matrix $G_{(\mathcal{X})}$ is a submatrix of $H_{(\mathcal{X})}$, defined in Equation (2), corresponding to strings of length 2. The next lemma shows that if two Markov chains on the same alphabet are equivalent, then they have the same statistical parameters.

Lemma 1.1: Consider two Markov chains $\mathcal{X} = \{X_t; t \in \mathbb{Z}\}$ and $\tilde{\mathcal{X}} = \{\tilde{X}_t; t \in \mathbb{Z}\}$ taking values on the same alphabet $\mathbb{X}$. If $\mathcal{X} \overset{\mathrm{p}}{\sim} \tilde{\mathcal{X}}$ then $\mathbf{M}_{\mathcal{X}} = \mathbf{M}_{\tilde{\mathcal{X}}}$.

Proof: Equation (5) can be written in matrix form as $G_{(\mathcal{X})} = P \; \mathrm{diag}[\pi]$ and therefore

$$\begin{aligned} \pi &= \mathrm{diag}[\pi] \; 1_n = P \; \mathrm{diag}[\pi] \; 1_n = G_{(\mathcal{X})} \; 1_n, \qquad (6) \\ P &= G_{(\mathcal{X})} \; (\mathrm{diag}[\pi])^{-1}. \qquad (7) \end{aligned}$$

Note that since $\pi_i > 0$, $i \in \mathbb{Z}_n$ the matrix $\mathrm{diag}[\pi] \succ 0$ and therefore it is invertible. Equivalence in the sense of Equation (1) implies that $G_{(\mathcal{X})} = G_{(\tilde{\mathcal{X}})}$. Given relations (6) and (7) one can conclude that $\pi = \tilde{\pi}$ and $P = \tilde{P}$. ■

The space of Markov chains taking values on $\mathbb{X}$ is denoted by $\mathscr{M}_{\mathbb{X}}$ and the space of their statistical parameters by $\mathcal{M}_{\mathbb{X}}$, i.e.

$$\begin{aligned} \mathscr{M}_{\mathbb{X}} &= \{\mathcal{X} \mid \mathcal{X} \text{ is a Markov chain on } \mathbb{X}\}, \\ \mathcal{M}_{\mathbb{X}} &= \{\mathbf{M}_{\mathcal{X}} = (\pi, P) \mid \mathcal{X} \in \mathscr{M}_{\mathbb{X}}\}. \end{aligned}$$

C. Hidden Markov processes and their statistical description

Throughout this section we will make use of the finite alphabets $\mathbb{X} = \{x_1, \dots, x_n\}$ and $\mathbb{Y} = \{y_1, \dots, y_m\}$. There are several types of Hidden Markov processes (HMP's). They always involve a pair of stochastic processes, $(\mathcal{X} = \{X_t; t \in \mathbb{Z}\}, \mathcal{Y} = \{Y_t; t \in \mathbb{Z}\})$ that takes values on $\mathbb{X} \times \mathbb{Y}$. The hidden process $\mathcal{X}$ is a Markov chain. The observable process $\mathcal{Y}$ is a function, possibly randomized, of $\mathcal{X}$. We will make use of the basic definitions and notation introduced in the context of realization theory of HMP's. One can find them in slightly varying language, along with detailed derivations, in [1], [3], [4]. At the core of realization theory of HMP's lies the concept of a finite stochastic system (FSS).

Definition 1.8: The pair of stochastic processes $(\mathcal{X} = \{X_t; t \in \mathbb{Z}\}, \mathcal{Y} = \{Y_t; t \in \mathbb{Z}\})$ taking values on $\mathbb{X} \times \mathbb{Y}$ is called a **finite stochastic system**, if

$$\begin{aligned} &\mathbf{Pr}[X_t = \sigma_0, Y_t = v_0 \mid X_t^{(-)} = \tau, Y_t^{(-)} = u] = \\ &\mathbf{Pr}[X_t = \sigma_0, Y_t = v_0 \mid X_{t-1} = \tau_1], \qquad (8) \end{aligned}$$

where $\sigma_0 \in \mathbb{X}$, $\tau = \tau_k \dots \tau_1 \in \mathbb{X}_{(k)}$, $v_0 \in \mathbb{Y}$, $u = u_l \dots u_1 \in \mathbb{Y}_{(l)}$, $t \in \mathbb{Z}$.

Remarks:

- The sets $\mathbb{X}, \mathbb{Y}$ are called the state and output space respectively. Accordingly $\mathcal{X}$ is the state and $\mathcal{Y}$ the output process.
- From Equation (8) it follows that $(\mathcal{X}, \mathcal{Y}) \in \mathscr{M}_{\mathbb{X}\times\mathbb{Y}}$ and that $\mathcal{X} \in \mathscr{M}_{\mathbb{X}}$.
- Additionaly from Equation (8) it follows that $\{Y_t; t \in \mathbb{Z}\}$ is a probabilistic function of the Markov chain $\{X_{t-1}; t \in \mathbb{Z}\}$ in the sense that:
$$\mathbf{Pr}[Y_t = v_0 \mid X_t^{(-)} = \tau, Y_t^{(-)} = u] = \mathbf{Pr}[Y_t = v_0 \mid X_{t-1} = \tau_1],$$
where $\tau = \tau_k \dots \tau_1 \in \mathbb{X}_{(k)}$, $v_0 \in \mathbb{Y}$, $u = u_l \dots u_1 \in \mathbb{Y}_{(l)}$, $t \in \mathbb{Z}$.

Definition 1.9: Consider a FSS $(\mathcal{X} = \{X_t; t \in \mathbb{Z}\}, \mathcal{Y} = \{Y_t; t \in \mathbb{Z}\})$ taking values on $\mathbb{X} \times \mathbb{Y}$. Let $\pi \in \mathbb{S}_{(n)}$ and $M : \mathbb{Y} \to \mathbb{R}^{n\times n}$ with

$$\begin{aligned} m[y]_{ij} &= \mathbf{Pr}[X_t = x_i, Y_t = y | X_{t-1} = x_j], \\ \pi_i &= \mathbf{Pr}[X_t = x_i], \end{aligned}$$

where $y \in \mathbb{Y}, i \in \mathbb{Z}_n, j \in \mathbb{Z}_n, t \in \mathbb{Z}$. The **statistical parameters** of the **FSS** $(\mathcal{X}, \mathcal{Y})$ are given by the tuple $\mathbf{F}_{(\mathcal{X},\mathcal{Y})} = (1_n, M, \pi)$.

Remarks:

- If P denotes the stochastic transition matrix P of the state process $\mathcal{X}$, then
$$P = \sum_{y\in\mathbb{Y}} M[y], \text{ and } P\ \pi = \pi.$$

Theorem 1.5: Consider a FSS $(\mathcal{X} = \{X_t; t \in \mathbb{Z}\}, \mathcal{Y} = \{Y_t; t \in \mathbb{Z}\})$ taking values on $\mathbb{X}\times\mathbb{Y}$ with statistical parameters $\mathbf{F}_{(\mathcal{X},\mathcal{Y})} = (1_n, M, \pi)$. The probability of a string in $\mathbb{Y}^*$ is given by

$$p_{\mathcal{Y}}[v_l \dots v_1] = 1_n^T M[v_l] \dots M[v_1]\pi,$$

where $v = v_l \dots v_1 \in \mathbb{Y}_{(l)}$.

Proof: See for instance [3]. ■

The space of FSS's taking values on $\mathbb{X} \times \mathbb{Y}$ is denoted by $\mathscr{F}_{\mathbb{X}\times\mathbb{Y}}$, i.e.

$$\mathscr{F}_{\mathbb{X}\times\mathbb{Y}} = \{(\mathcal{X}, \mathcal{Y}) \mid (\mathcal{X}, \mathcal{Y}) \text{ is a FSS on } \mathbb{X} \times \mathbb{Y}\}.$$

Definition 1.10: A stochastic process $\{Y_t; t \in \mathbb{Z}\}$ taking values on $\mathbb{Y}$ is a **hidden Markov process** of size n if there exists $(\mathcal{X}, \tilde{\mathcal{Y}}) \in \mathscr{F}_{\mathbb{X}\times\mathbb{Y}}$ such that $\mathcal{Y} \overset{\mathrm{p}}{\sim} \tilde{\mathcal{Y}}$. The FSS $(\mathcal{X}, \tilde{\mathcal{Y}})$ is a **representation** of the HMP $\mathcal{Y}$.

Remark:

- Since $\mathcal{Y} \overset{\mathrm{p}}{\sim} \tilde{\mathcal{Y}}$ it follows that $p_{\mathcal{Y}}[v] = p_{\tilde{\mathcal{Y}}}[v],\ v \in \mathbb{Y}^*$. Taking theorem 1.5 into account one can see that the statistical parameters $\mathbf{F}_{(\mathcal{X},\tilde{\mathcal{Y}})}$ fully specify the probability function $p_{\mathcal{Y}}$.

Definition 1.11: Consider a size n HMP $\mathcal{Y} = \{Y_t; t \in \mathbb{Z}\}$ and let the FSS $(\mathcal{X} = \{X_t; t \in \mathbb{Z}\}, \tilde{\mathcal{Y}} = \{\tilde{Y}_t; t \in \mathbb{Z}\})$ be a representation of $\mathcal{Y}$ with statistical parameters $\mathbf{F}_{(\mathcal{X},\tilde{\mathcal{Y}})} = (1_n, M, \pi)$. The tuple $\mathbf{H}_{\mathcal{Y}} = \mathbf{F}_{(\mathcal{X},\tilde{\mathcal{Y}})}$ is a **realization** of the probability function $p_{\mathcal{Y}}$.

The space of size n HMP's $\mathcal{Y} = \{Y_t, t \in \mathbb{Z}\}$ taking values on $\mathbb{Y}$ is denoted by $\mathscr{H}_{\mathbb{Y},(n)}$, and the space of realizations of their respective probability function by $\mathcal{H}_{\mathbb{Y},(n)}$, i.e.

$$\begin{aligned} \mathscr{H}_{\mathbb{Y},(n)} &= \{\mathcal{Y} = \{Y_t, t \in \mathbb{Z}\} \mid \mathcal{Y} \text{ is a size } n \text{ HMP on } \mathbb{Y}\}, \\ \mathcal{H}_{\mathbb{Y},(n)} &= \{\mathbf{H}_{\mathcal{Y}} = (1_n, M, \pi) \mid \mathcal{Y} \in \mathscr{H}_{\mathbb{Y},(n)}\}. \end{aligned}$$

Representations of HMP's and as such realizations of their respective probability function are not unique. The following lemma can serve as a starting point for constructing two distinct realizations of the probability function of a given HMP.

Lemma 1.2: Consider $\mathcal{Y} \in \mathscr{H}_{\mathbb{Y},(n)}$ and $(\mathcal{X}^{(i)}, \mathcal{Y}^{(i)}) \in \mathscr{F}_{\mathbb{X}\times\mathbb{Y}}$, $i \in \{1, 2\}$ with statistical parameters $\mathbf{F}_{(\mathcal{X}^{(i)},\mathcal{Y}^{(i)})} = (1_n, M^{(i)}, \pi^{(i)})$, $i \in \{1, 2\}$. Suppose that the FSS $(\mathcal{X}^{(1)}, \mathcal{Y}^{(1)})$ is a representation of the HMP $\mathcal{Y}$ and suppose that there exists a non singular matrix $T \in \mathbb{R}^{n\times n}$ with $1_n^T\ T = 1_n^T$, such that

$$\begin{aligned} M^{(2)}[y] &= T^{-1}\ M^{(1)}[y]\ T, \quad y \in \mathbb{Y}, \\ \pi^{(2)} &= T^{-1}\pi^{(1)}. \end{aligned}$$

Then the FSS $(\mathcal{X}^{(2)}, \mathcal{Y}^{(2)})$ is a representation of the HMP $\mathcal{Y}$ as well.

Proof: Since the FSS $(\mathcal{X}^{(1)}, \mathcal{Y}^{(1)})$ is a representation of the HMP $\mathcal{Y}$ its statistical parameters provide a realization of the probability function $p_{\mathcal{Y}}$, i.e. $p_{\mathcal{Y}}[v_l \dots v_1] = 1_n^T M^{(1)}[v_l] \dots M^{(1)}[v_1]\pi^{(1)}$ where $v = v_l \dots v_1 \in \mathbb{Y}_{(l)}$. By direct substitution one has $p_{\mathcal{Y}}[v_l \dots v_1] = p_{\mathcal{Y}^{(2)}}[v_l \dots v_1] = 1_n^T M^{(2)}[v_l] \dots M^{(2)}[v_1]\pi^{2)}$ and therefore $\mathcal{Y} \overset{\mathrm{p}}{\sim} \mathcal{Y}^{(2)}$ showing that the FSS $(\mathcal{X}^{(2)}, \mathcal{Y}^{(2)})$ is a representation of the HMP $\mathcal{Y}$. ■

Definition 1.12: Consider a FSS $(\mathcal{X} = \{X_t; t \in \mathbb{Z}\}, \mathcal{Y} = \{Y_t; t \in \mathbb{Z}\}) \in \mathscr{F}_{\mathbb{X}\times\mathbb{Y}}$. with statistical parameters $\mathbf{F}_{(\mathcal{X},\mathcal{Y})} = (1_n, M, \pi)$. We say that the given FSS satisfies the **factorization** condition if

$$\begin{aligned} &\mathbf{Pr}[Y_t = v_0, X_t = \sigma_0 \mid X_{t-1} = \tau_1] = \\ &\mathbf{Pr}[Y_t = v_0 | X_t = \sigma_0]\mathbf{Pr}[X_t = \sigma_0 | X_{t-1} = \tau_1], \end{aligned} \tag{9}$$

where $v_0 \in \mathbb{Y}, \sigma_0 \in \mathbb{X}, \tau_1 \in \mathbb{X}, t \in \mathbb{Z}$. The matrix $O \in \mathbb{R}^{n\times m}$ with

$$o_{ij} = \mathbf{Pr}[Y_t = y_j \mid X_t = x_i],\ i \in \mathbb{Z}_n, j \in \mathbb{Z}_m, t \in \mathbb{Z}$$

is the (emission) **output matrix**.

Remarks:

- Condition (9) is written in matrix form as
$$M[y] = \mathrm{diag}[O(:, y)]\ P,\ y \in \mathbb{Y},$$
where P is the stochastic transition matrix of the underlying state process $\mathcal{X}$.
- In order to emphasize that a given FSS satisfies the factorization condition, its statistical parameters will be given by the tuple $\mathbf{F}^R_{(\mathcal{X},\mathcal{Y})} = (1_n, O, P, \pi)$.

Let $\mathscr{F}^R_{\mathbb{X}\times\mathbb{Y}} \subset \mathscr{F}_{\mathbb{X}\times\mathbb{Y}}$ denote the space of FSS's that satisfy the factorization condition (9), i.e.

$$\mathscr{F}^R_{\mathbb{X}\times\mathbb{Y}} = \{(\mathcal{X},\mathcal{Y}) \in \mathscr{F}_{\mathbb{X}\times\mathbb{Y}} \mid (\mathcal{X},\mathcal{Y}) \text{ satisfies condition (9)}\}.$$

Definition 1.13: A stochastic process $\{Y_t; t \in \mathbb{Z}\}$ taking values on $\mathbb{Y}$ is a HMP of the **random function of a Markov chain type** of size n if there exists a FSS $(\mathcal{X} = \{X_t; t \in \mathbb{Z}\}, \tilde{\mathcal{Y}} = \{\tilde{Y}_t; t \in \mathbb{Z}\}) \in \mathscr{F}^R_{\mathbb{X}\times\mathbb{Y}}$ with $\mathbf{F}^R_{(\mathcal{X},\tilde{\mathcal{Y}})} = (1_n, O, P, \pi)$ such that $\mathcal{Y} \overset{\text{p}}{\sim} \tilde{\mathcal{Y}}$.

Definition 1.14: A stochastic process $\{Y_t; t \in \mathbb{Z}\}$ taking values on $\mathbb{Y}$ is a HMP of the **deterministic function of a Markov chain type** of size n if there exists a function $f : \mathbb{X} \to \mathbb{Y}$ and a FSS $(\mathcal{X} = \{X_t; t \in \mathbb{Z}\}, \tilde{\mathcal{Y}} = \{\tilde{Y}_t; t \in \mathbb{Z}\}) \in \mathscr{F}^R_{\mathbb{X}\times\mathbb{Y}}$ with $\mathbf{F}^R_{(\mathcal{X},\tilde{\mathcal{Y}})} = (1_n, O, P, \pi)$, such that $\tilde{Y}_t = f(X_t),\ t \in \mathbb{Z}$ and $\mathcal{Y} \overset{\text{p}}{\sim} \tilde{\mathcal{Y}}$.

Remarks:

- A HMP of the random/determinstic function of a Markov chain type is abbreviated as a RF/DF HMP.
- The FSS $(\mathcal{X}, \tilde{\mathcal{Y}})$ is a representation of the RF/DF HMP $\mathcal{Y}$ and the tuple $\mathbf{H}^R_{\mathcal{Y}}/\mathbf{H}^D_{\mathcal{Y}} = (1_n, O, P, \pi)$ is a realization of the probability function $p_{\mathcal{Y}}$.
- In the case of a DF HMP $O \in \{0,1\}^{n\times m}$. In particular

$$o_{ij} = \begin{cases} 1 & \text{if } y_j = f(x_i), \\ 0 & \text{otherwise,} \end{cases}$$

  where $j \in \mathbb{Z}_m$ and $i \in \mathbb{Z}_n$.

The space of size n RF/DF HMP's $\mathcal{Y} = \{Y_t, t \in \mathbb{Z}\}$ taking values on $\mathbb{Y}$ is denoted by $\mathscr{H}^R_{\mathbb{Y},(n)}/\mathscr{H}^D_{\mathbb{Y},(n)}$, and the space of realizations of their respective probability functions by $\mathcal{H}^R_{\mathbb{Y},(n)}/\mathcal{H}^D_{\mathbb{Y},(n)}$, i.e.

$$\mathscr{H}^R_{\mathbb{Y},(n)}/\mathscr{H}^D_{\mathbb{Y},(n)} = \{\mathcal{Y} \mid \mathcal{Y} \text{ is a size } n \text{ RF/DF HMP on } \mathbb{Y}\},$$
$$\mathcal{H}^R_{\mathbb{Y},(n)}/\mathcal{H}^D_{\mathbb{Y},(n)} = \{\mathbf{H}^R_{\mathcal{Y}}/\mathbf{H}^D_{\mathcal{Y}} \mid \mathcal{Y} \in \mathscr{H}^R_{\mathbb{Y},(n)}/\mathscr{H}^D_{\mathbb{Y},(n)}\}.$$

It follows directly from the definitions that $\mathscr{M}_{\mathbb{X}} \subset \mathscr{H}^D_{\mathbb{X},(n)}$ and $\mathscr{H}^D_{\mathbb{Y},(n)} \subset \mathscr{H}^R_{\mathbb{Y},(n)} \subset \mathscr{H}_{\mathbb{Y},(n)}$. By enlarging the cardinality of the underlying state space, it is always possible to represent a HMP as a deterministic function of a Markov chain. As such the model of a HMP appearing in Definition 1.10 is the most economical in terms of state space size.

II. Aggregation Based Model Reduction

A. Markov Chains

Throughout this section we will make use of the finite alphabets $\mathbb{X} = \{x_1, \ldots, x_n\}$ and $\hat{\mathbb{X}} = \{\hat{x}_1, \ldots, \hat{x}_{\hat{n}}\}$, where $\hat{n} \le n$. The sets $\mathbb{X}$ and $\hat{\mathbb{X}}$ serve as state spaces for the high and low order Markov chain respectively. The high order Markov chain will be denoted by $\mathcal{X} \in \mathscr{M}_{\mathbb{X}}$ with $\mathbf{M}_{\mathcal{X}} = (\pi, P)$.

Aggregation based reduction applied to $\mathcal{X}$ consists of two steps: First one partitions $\mathbb{X}$ into clusters, each cluster corresponds to a unique element in $\hat{\mathbb{X}}$. The partitioning information as well as the statistical parameters of $\mathcal{X}$ enable the construction of $\hat{\mathcal{X}} \in \mathscr{M}_{\hat{\mathbb{X}}}$ frequently termed as the aggregated Markov chain. Subsequently the aggregated Markov chain $\hat{\mathcal{X}}$ is maped to a dilated process $\tilde{\mathcal{X}} \in \mathscr{M}_{\mathbb{X}}$ for the purpose of establishing a comparison with the original Markov chain $\mathcal{X}$. In order to formalize the above statements we introduce some notation.

Definition 2.1: A surjective map $\phi : \mathbb{X} \to \hat{\mathbb{X}}$ is called a **partition function**. Let $\Phi : \hat{\mathbb{X}} \to \mathcal{P}(\mathbb{X})$, $\Phi(\hat{x}) = \{x \in \mathbb{X} \mid \phi(x) = \hat{x}\}$ is the **cluster** corresponding to $\hat{x} \in \hat{\mathbb{X}}$. The set of all partition functions between $\mathbb{X}$ and $\hat{\mathbb{X}}$ is denoted by $\mathcal{F}_{\mathbb{X},\hat{\mathbb{X}}}$, i.e.

$$\mathcal{F}_{\mathbb{X},\hat{\mathbb{X}}} = \{\phi : \mathbb{X} \to \hat{\mathbb{X}} \mid \phi \text{ is a partition function}\}.$$

Remark:

- The terminology in this definition stems from the fact that the surjective map ϕ induces a partitioning of the domain into clusters that satisfy

$$\Phi(\hat{x}) \neq \varnothing, \quad \Phi(\hat{y}) \cap \Phi(\hat{z}) = \varnothing \text{ if } \hat{y} \neq \hat{z}, \qquad \bigcup_{\hat{x}\in\hat{\mathbb{X}}} \Phi(\hat{x}) = \mathbb{X}, \tag{10}$$

  where $\hat{x} \in \hat{\mathbb{X}}$, $\hat{y} \in \hat{\mathbb{X}}$, $\hat{z} \in \hat{\mathbb{X}}$.

Let $\Psi : \hat{\mathbb{X}} \to \mathcal{P}(\mathbb{Z}_n)$, $\Psi(\hat{x}) = \{s \in \mathbb{Z}_n \mid x_s \in \Phi(\hat{x})\}$ is the set of indices of the states aggregated to the cluster corresponding to $\hat{x} \in \hat{\mathbb{X}}$.

Definition 2.2: The matrix $L_{(\phi)} \in \{0,1\}^{\hat{n}\times n}$ with

$$l_{(\phi)ij} = \begin{cases} 1 & \text{if } \hat{x}_i = \phi(x_j), \\ 0 & \text{otherwise,} \end{cases}$$

where $i \in \mathbb{Z}_{\hat{n}}, j \in \mathbb{Z}_n$ is the **aggregation matrix** corresponding to ϕ. The matrix $R_{(\phi)} \in [0,1]^{n\times\hat{n}}$ with

$$r_{(\phi)ij} = \begin{cases} \frac{\pi_i}{\sum\limits_{s\in\Psi(\hat{x}_j)} \pi_s} & \text{if } \hat{x}_j = \phi(x_i), \\ 0 & \text{otherwise,} \end{cases}$$

where $i \in \mathbb{Z}_n, j \in \mathbb{Z}_{\hat{n}}$ is the **dilation matrix** corresponding to ϕ.

Remarks:

- Note that $L_{(\phi)}\ R_{(\phi)} = I_{\hat{n}}$ and that

$$\Pi_{(\phi)} = R_{(\phi)}\ L_{(\phi)}$$

  is a projection matrix in $\mathbb{R}^n$, i.e. $\Pi^2_{(\phi)} = \Pi_{(\phi)}$.
- It also holds that

$$R_{(\phi)} = \text{diag}[\pi]\ L^T_{(\phi)}\ (\text{diag}[L_{(\phi)}\ \pi])^{-1}.$$

Definition 2.3: The process $\hat{\mathcal{X}} \in \mathscr{M}_{\hat{\mathbb{X}}}$ with

$$g_{(\hat{\mathcal{X}})kl} = \mathbf{P}[\hat{X}_{t+1} = \hat{x}_k, \hat{X}_t = \hat{x}_l] := \mathbf{P}[X_{t+1} \in \Phi(\hat{x}_k), X_t \in \Phi(\hat{x}_l)],$$

where $k \in \mathbb{Z}_{\hat{n}},\ l \in \mathbb{Z}_{\hat{n}}$ is the **aggregated Markov chain** corresponding to $\mathcal{X} \in \mathscr{M}_{\mathbb{X}}$ for the partition function $\phi \in \mathcal{F}_{\mathbb{X},\hat{\mathbb{X}}}$.

Remark:

- The process $\hat{\mathcal{X}} = \{\hat{X}_t; t \in \mathbb{Z}\}$ is not to be confused with $\{\phi(X_t); t \in \mathbb{Z}\}$, a process taking values on $\hat{\mathbb{X}}$ as well, that in generic settings though is not a Markov chain.

The statistical parameters $\mathbf{M}_{\hat{\mathcal{X}}} = (\hat{\pi}, \hat{P})$ are calculated from $G_{(\hat{\mathcal{X}})}$ using relations (6) and (7) and the fact that $G_{(\hat{\mathcal{X}})} = L_{(\phi)}\ G_{(\mathcal{X})}\ L_{(\phi)}^T$, i.e.

$$\begin{aligned} \hat{\pi} &= L_{(\phi)}\ \pi, && (11) \\ \hat{P} &= L_{(\phi)}\ P\ R_{(\phi)}. && (12) \end{aligned}$$

For a given $\phi \in \mathcal{F}_{\mathbb{X},\hat{\mathbb{X}}}$ define the operator $\mathcal{A}_{(\phi)} : \mathcal{M}_{\mathbb{X}} \to \mathcal{M}_{\hat{\mathbb{X}}}$ where $\mathcal{A}_{(\phi)}[\mathcal{X}] = \hat{\mathcal{X}}$ and the statistical parameters $\mathbf{M}_{\mathcal{X}}$ and $\mathbf{M}_{\hat{\mathcal{X}}}$ are related by Equations (11) and (12).

Definition 2.4: The process $\tilde{\mathcal{X}} \in \mathscr{M}_{\mathbb{X}}$ with statistical parameters $\mathbf{M}_{\tilde{\mathcal{X}}} = (\tilde{\pi}, \tilde{P})$, where

$$\begin{aligned} \tilde{\pi} &= R_{(\phi)}\ \hat{\pi}, && (13) \\ \tilde{P} &= R_{(\phi)}\ \hat{P}\ L_{(\phi)}. && (14) \end{aligned}$$

is the **dilated Markov chain** corresponding to $\hat{\mathcal{X}} \in \mathscr{M}_{\hat{\mathbb{X}}}$ for the partition function $\phi \in \mathcal{F}_{\mathbb{X},\hat{\mathbb{X}}}$.

Remark:

- The statistical parameters of $\mathcal{X} \in \mathcal{M}_{\mathbb{X}}$ and $\tilde{\mathcal{X}} \in \mathscr{M}_{\mathbb{X}}$ are related by

$$\begin{aligned} \tilde{\pi} &= \Pi_{(\phi)}\ \pi, \\ \tilde{P} &= \Pi_{(\phi)}\ P\ \Pi_{(\phi)}. \end{aligned}$$

For a given $\phi \in \mathcal{F}_{\mathbb{X},\hat{\mathbb{X}}}$ define the operator $\mathcal{D}_{(\phi)} : \mathcal{M}_{\hat{\mathbb{X}}} \to \mathcal{M}_{\mathbb{X}}$ where $\mathcal{D}_{(\phi)}[\hat{\mathcal{X}}] = \tilde{\mathcal{X}}$ and the statistical parameters $\mathbf{M}_{\hat{\mathcal{X}}}$ and $\mathbf{M}_{\tilde{\mathcal{X}}}$ are related by Equations (13) and (14). The outcome of aggregation based model reduction of a Markov chain is error free if $\mathcal{X} \overset{\mathrm{p}}{\sim} \tilde{\mathcal{X}}$, i.e. $\mathcal{D}_{(\phi)} \circ \mathcal{A}_{(\phi)}[\mathcal{X}] \overset{\mathrm{p}}{\sim} \mathcal{X}$. Considering Lemma 1.1, the last condition translates to the fixpoint equation

$$\Pi_{(\phi)}\ P\ \Pi_{(\phi)} = P.$$

B. Hidden Markov processes

Aggregation based model reduction of HMP's will be discussed in the context of HMP's of the random function of a Markov chain type, i.e. RF HMP's. Throughout this section we will make use of the finite alphabets $\mathbb{X} = \{x_1, \dots, x_n\}$, $\mathbb{Y} = \{y_1, \dots, y_m\}$ and $\hat{\mathbb{X}} = \{\hat{x}_1, \dots, \hat{x}_{\hat{n}}\}$, where $\hat{n} \leq n$. The sets $\mathbb{X}$ and $\hat{\mathbb{X}}$ serve as state spaces for the high and low order HMP respectively. The high order HMP will be denoted by $\mathcal{Y} \in \mathscr{H}^R_{\mathbb{Y},(n)}$ with $\mathbf{H}^R_{\mathcal{Y}} = (1_n, O, P, \pi)$ being a realization of its probability function $p_{\mathcal{Y}}$. The low order HMP will be denoted by $\hat{\mathcal{Y}} \in \mathscr{H}^R_{\mathbb{Y},(\hat{n})}$ with $\mathbf{H}^R_{\hat{\mathcal{Y}}} = (1_{\hat{n}}, \hat{O}, \hat{P}, \hat{\pi})$ being a realization of its probability function $p_{\hat{\mathcal{Y}}}$.

Definition 2.5: Consider two RF HMP's $\mathcal{Y} \in \mathscr{H}^R_{\mathbb{Y},(n)}$ with $\mathbf{H}^R_{\mathcal{Y}} = (1_n, O, P, \pi)$ and $\hat{\mathcal{Y}} \in \mathscr{H}^R_{\mathbb{Y},(\hat{n})}$ with $\mathbf{H}^R_{\hat{\mathcal{Y}}} = (1_{\hat{n}}, \hat{O}, \hat{P}, \hat{\pi})$. The RF HMP $\hat{\mathcal{Y}}$ is the outcome of **aggregation based model reduction** applied to $\mathcal{Y}$ if there exists a partition function $\phi \in \mathcal{F}_{\mathbb{X},\hat{\mathbb{X}}}$ such that

$$\begin{aligned} \hat{\pi} &= L_{(\phi)}\ \pi, && (15) \\ \hat{P} &= L_{(\phi)}\ P\ R_{(\phi)}, && (16) \\ \mathrm{diag}[\hat{O}(:,y)] &= L_{(\phi)}\ \mathrm{diag}[O(:,y)]\ R_{(\phi)}, \quad y \in \mathbb{Y}. && (17) \end{aligned}$$

Remark:

- Suppose that the FSS $(\hat{\mathcal{X}}, \tilde{\hat{\mathcal{Y}}})$ is a representation of $\hat{\mathcal{Y}}$ and that the FSS $(\mathcal{X}, \tilde{\mathcal{Y}})$ is a representation of $\mathcal{Y}$. Relations (15) - (17) are derived from

$$\begin{aligned} \mathbf{P}[\hat{X}_{t+1} = \hat{x}_k, \hat{X}_t = \hat{x}_l] &= \mathbf{P}[X_{t+1} \in \Phi(\hat{x}_k), X_t \in \Phi(\hat{x}_l)], \\ \mathbf{P}[\hat{Y}_t = y_s, \hat{X}_t = \hat{x}_l] &= \mathbf{P}[\hat{Y}_t = y_s, X_t \in \Phi(\hat{x}_l)], \end{aligned}$$

where $k \in \mathbb{Z}_{\hat{n}}$, $l \in \mathbb{Z}_{\hat{n}}$, $s \in \mathbb{Z}_m$

For a given $\phi \in \mathcal{F}_{\mathbb{X},\hat{\mathbb{X}}}$ define the operator $\mathcal{A}^H_{(\phi)} : \mathscr{H}^R_{\mathbb{Y},(n)} \to \mathscr{H}^R_{\mathbb{Y},(\hat{n})}$ where $\mathcal{A}^H_{(\phi)}[\mathcal{Y}] = \hat{\mathcal{Y}}$ and the realizations $\mathbf{H}^R_{\mathcal{Y}}, \mathbf{H}^R_{\hat{\mathcal{Y}}}$ are related by equations (15) - (17). The outcome of aggregation based model reduction of a Hidden Markov process is error free if $\mathcal{Y} \overset{\mathrm{p}}{\sim} \hat{\mathcal{Y}}$, i.e. $\mathcal{A}^H_{(\phi)}[\mathcal{Y}] \overset{\mathrm{p}}{\sim} \mathcal{Y}$.

III. Projection Based Model Reduction

Throughout this section we will make use of the finite alphabets $\mathbb{X} = \{x_1, \dots, x_n\}$, $\mathbb{Y} = \{y_1, \dots, y_m\}$ and $\hat{\mathbb{X}} = \{\hat{x}_1, \dots, \hat{x}_{\hat{n}}\}$, where $\hat{n} \leq n$. The sets $\mathbb{X}$ and $\hat{\mathbb{X}}$ serve as state spaces for the high and low order HMP respectively. The high order HMP will be denoted by $\mathcal{Y} \in \mathscr{H}_{\mathbb{Y},(n)}$ with $\mathbf{H}_{\mathcal{Y}} = (1_n, M, \pi)$ being a realization of its probability function $p_{\mathcal{Y}}$. The low order HMP will be denoted by $\hat{\mathcal{Y}} \in \mathscr{H}_{\mathbb{Y},(\hat{n})}$ with $\mathbf{H}_{\hat{\mathcal{Y}}} = (1_{\hat{n}}, \hat{M}, \hat{\pi})$ being a realization of its probability function $p_{\hat{\mathcal{Y}}}$. Let $\mathcal{P}_{n,\hat{n}} = \{(U,V) \mid U \in \mathbb{R}^{\hat{n}\times n},\ V \in \mathbb{R}^{n\times\hat{n}},\ U\ V = I_{\hat{n}},\ 1_{\hat{n}}\ U = 1_{\hat{n}}\}$. Given $(U,V) \in \mathcal{P}_{n,\hat{n}}$, the matrix $\Pi_{(U,V)} = V\ U$ is an oblique projection onto the columns of V along the nullspace of U.

Definition 3.1: Consider two HMP's $\mathcal{Y} \in \mathscr{H}_{\mathbb{Y},(n)}$ with $\mathbf{H}_{\mathcal{Y}} = (1_n, M, \pi)$ and $\hat{\mathcal{Y}} \in \mathscr{H}_{\mathbb{Y},(\hat{n})}$ with $\mathbf{H}_{\hat{\mathcal{Y}}} = (1_{\hat{n}}, \hat{M}, \hat{\pi})$. The HMP $\hat{\mathcal{Y}}$ is the outcome of **projection based model reduction** applied to $\mathcal{Y}$ if there exists $(U,V) \in \mathcal{P}_{n,\hat{n}}$ such that

$$\begin{aligned} \hat{\pi} &= U\ \pi, && (18) \\ \hat{M}[y] &= U\ M[y]\ V,\ y \in \mathbb{Y}. && (19) \end{aligned}$$

For a given $(U,V) \in \mathcal{P}_{n,\hat{n}}$ define the operator $\mathcal{R}_{(U,V)} : \mathscr{H}_{\mathbb{Y},(n)} \to \mathscr{H}_{\mathbb{Y},(\hat{n})}$ where $\mathcal{R}_{(U,V)}[\mathcal{Y}] = \hat{\mathcal{Y}}$ and the realizations $\mathbf{H}_{\mathcal{Y}}, \mathbf{H}_{\hat{\mathcal{Y}}}$ are related by equations (18), (19). The outcome of projection based model reduction of a Hidden Markov process is error free if $\mathcal{Y} \overset{\mathrm{p}}{\sim} \hat{\mathcal{Y}}$, i.e. $\mathcal{R}_{(U,V)}[\mathcal{Y}] \overset{\mathrm{p}}{\sim} \mathcal{Y}$.

References

[1] B. D. O. Anderson, "The realization problem for Hidden Markov Models," *Math. Control Signals Syst.*, vol. 12, no. 1, pp. 80–122, Apr. 1999.

[2] P. Billingsley, *Probability and Measure*. Wiley, 1995.

[3] G. Picci, "On the internal structure of finite state stochastic processes," in *Recent Developments in Variable Structure Systems*, ser. Lecture Notes in Economics and Mathematical Systems. Springer, 1978, vol. 162, pp. 288–304.

[4] M. Vidyasagar, "The complete realization problem for Hidden Markov Models. a survey and some new results," *Math. Control Signals Syst.*, vol. 23, pp. 1–65, 2011.