Bulletin of the *Transilvania* University of Braşov • Series III: Mathematics, Informatics, Physics,
Vol 5(54) 2012, Special Issue: *Proceedings of the Seventh Congress of Romanian Mathematicians*, 23-32,
published by Transilvania University Press, Braşov and Publishing House of the Romanian Academy

APPLICATION OF DEPENDENCE WITH COMPLETE CONNECTIONS TO HIDDEN MARKOV MODELS

Cristina ANTON POPESCU [1]

Abstract

We consider a hidden Markov model with mutually independent observations. To this model we associate a random system with complete connections. Using the random system, we study the ergodic properties of the prediction filter corresponding to the hidden Markov model.

2000 *Mathematics Subject Classification:* 60G35, 93E11, 60J05
Key words: hidden Markov model, random system with complete connections, prediction filter, geometric ergodicity

1 Indroduction

Hidden Markov models (HMMs) have been extensively applied to problems in econometrics, computational biology, speech recognition and fault detection.

A HMM is formed by a hidden Markov chain $\{S_n, n \geq 1\}$ and a stochastic process $\{Y_n, n \geq 1\}$, with distributions depending on $\{S_n, n \geq 1\}$. Usually, the hidden sequence $\{S_n, n \geq 1\}$ is a finite homogeneous Markov chain, and the observations $\{Y_n, n \geq 1\}$ are mutually independent given the sequence $\{S_n, n \geq 1\}$. The prediction filter, $\{\mathbf{P}_\bullet(S_n|Y_{n-1}, \ldots, Y_1)\}$, where $(\Omega, F, \mathbf{P}_\bullet)$ is the probability space, is very important in the inference algorithms commonly used for HMMs.

The dependence with complete connections concerns the couple formed with a Markov chain, and a stochastic process with distributions depending on the states of the Markov chain. Hence, it is natural to associate a random system with complete connections (RSCC) ([6]) with a HMM. This was noticed in 1975 by Kaijser ([7]) who studied the ergodicity of the filter under the assumption of subrectangularity. More recently, in 1990, Arapostathis and Marcus [1] have also associated a RSCC with a HMM with binary observations, but they did not further develop this idea.

In this paper, we illustrate how techniques and results regarding the dependence with complete connections can be used to study the ergodic properties of the HMMs. When

[1]Department of Mathematics and Statistics, *University of Alberta*, Canada, e-mail: popescuc@macewan.ca

some of the parameters of the HMM are unknown, the ergodicity has statistical applications in estimation problems.

Different assumptions can be made about the state space and the distributions of $\{S_n, n \geq 1\}$ and $\{Y_n, n \geq 1\}$. Besides Kaijser's approach, under compactness or locally compactness hypotheses, Kunita ([9]) and Stettner ([10]) have proved the existence of an invariant probability distribution for the prediction filter. In this paper, we recover this result using a different approach. We consider a HMM similar to the one presented by Le Gland and Mevel in [4]. With this model, we associate a RSCC, and as a consequence of the properties of the RSCC, we obtain the ergodicity of the Markov chain formed by the prediction filter.

2 Geometric ergodicity of the prediction filter

We consider the same HMM as in [4], which is formed by the unobserved random sequence $\{S_n, n \geq 1\}$, and the observations sequence $\{Y_n, n \geq 1\}$, defined on the probability space $(\Omega, F, \mathbf{P}_\bullet)$ with values in the finite set $S = \{1, \ldots, M\}$ and in $\mathbb{R}^d$, respectively. We suppose that the sequence $\{S_n, n \geq 1\}$ is a homogeneous Markov chain with the initial probability distribution $\mathbf{p}_\bullet = (\mathbf{p}_\bullet{}^i)$, and the transition probability matrix $Q = (q^{i,j})$, $i, j \in S$. Hence, for any $i, j \in S$ and any integer $n \geq 1$, we have

$$\mathbf{P}_\bullet(S_1 = i) = \mathbf{p}_\bullet{}^i, \quad \mathbf{P}_\bullet(S_{n+1} = j | S_n = i) = q^{i,j}.$$

We also suppose that the conditional probability distribution of the observation Y_n given that $\{S_n = i\}$ is absolutely continuous with respect to a non-negative and σ- finite measure λ, and it has a positive density

$$\mathbf{P}_\bullet(Y_n \in A | S_n = i) = \int_A b^i(y) \lambda(dy),$$

for any integer $n \geq 1$, $i \in S$ and $A \in B\left(\mathbb{R}^d\right)$, where for any integer $n \geq 1$ we denote by $B(\mathbb{R}^n)$ the collection of Borel sets on $\mathbb{R}^n$. Moreover, the observations $\{Y_n, n \geq 1\}$ are mutually independent given the sequence $\{S_n, n \geq 1\}$:

$$\mathbf{P}_\bullet(Y_n \in A_n, \ldots, Y_1 \in A_1 | S_n = i_n, \ldots, S_1 = i_1) = \prod_{k=1}^{n} \mathbf{P}_\bullet(Y_k \in A_k | S_k = i_k),$$

for any integer $n \geq 1$, any $i_1, \ldots i_n \in S$ and any sets $A_n, \ldots A_1 \in B\left(\mathbb{R}^d\right)$.

Let consider the prediction filter $p_n = (p_n^i)$, where $p_1 = \mathbf{p}_\bullet$, and for any $n > 1$ and all $i \in S$,

$$p_n^i = \mathbf{P}_\bullet(S_n = i | Y_{n-1}, \ldots, Y_1).$$

We now have the following forward Baum equation [4]

$$p_{n+1} = \frac{Q^t B_\bullet(Y_n) p_n}{b_\bullet^t(Y_n) p_n}, \quad n \geq 1$$

where ^{t} denotes the transpose of a matrix or a vector, and for any $y \in \mathbb{R}^d$, $b_\bullet(y) = (b^i(y))$, $B_\bullet(y) = \mathrm{diag}(b^i(y))$. To emphasize the dependency with respect to the initial condition and the observations, we use the same notation as in [4], and we put

$$p_{n+1} = f[Y_n, \ldots, Y_1, p_1] = \frac{M_{n,1}p_1}{e^t M_{n,1}p_1} = M_{n,1} \cdot p_1, \tag{2.1}$$

where $e = (1, \ldots, 1)^t$ is a M-dimensional vector, $\cdot$ denotes the projective product [3], and for any integers $n \geq l$

$$M_{n,l} = Q^t B_\bullet(Y_n) \cdots Q^t B_\bullet(Y_l).$$

Example 2.1. *As noted in [4], an example of a HMM which satisfies all the previous assumptions is a model with observations $\{Y_n, n \geq 1\}$ of the following form*

$$Y_n = h(S_n) + V_n,$$

where $\{V_n, n \geq 1\}$ is a Gaussian white noise sequence independent of $\{S_n, n \geq 1\}$, with symmetric and positive definite covariance matrix D, (i.e. $V_n \sim N(0, D)$), and h is a mapping from S to $\mathbb{R}^d$.

Let define

$$W = \{w \in \mathbb{R}^M : w_i \geq 0, i \in S, \sum_{i=1}^{M} w_i = 1\}, \tag{2.2}$$

and consider on W the topology induced by the Euclidian space $\mathbb{R}^M$. Let W denote the collection of the Borel sets on W and $\|\cdot\|_1$ the L_1- norm on $\mathbb{R}^M$. Hence, for any $u = (u_i) \in \mathbb{R}^M$ and for any $M \times M$ matrix $Z = (Z_{i,j})$, we have

$$\|u\|_1 = \sum_{i \in S} |u_i|, \quad \|Z\|_1 = \max_{j \in S} \sum_{i \in S} |Z_{i,j}|.$$

Notice that $\{p_n, n \geq 1\}$ is a Markov chain with the transition probability

$$\Pi(p_{n+1} \in E | p_n = p) = \sum_{j \in S} p^j \int_{\mathbb{R}^d} b^j(y) 1_E(f[y, p]) \lambda(dy),$$

for any $n \geq 1$, $E \in W$ and $p \in W$. Here and throughout the paper, 1_E denotes the indicator function of the set E.

For any real-valued, bounded and $(W, B(R))$ measurable function g defined on W, and any $p \in W$, we have

$$Ug(p) = \mathbf{E}_\bullet[g(p_{n+1}) | p_n = p] = \sum_{j \in S} p^j \int_{\mathbb{R}^d} b^j(y) g(f[y, p]) \lambda(dy), \tag{2.3}$$

and, for any $n \geq 1$,

$$\begin{aligned} U^n g(p) = \mathbf{E}_\bullet[g(p_{n+1}) | p_1 = p] = &\sum_{i_1, \ldots, i_n \in S} p^{i_1} q^{i_1, i_2} \cdots q^{i_{n-1}, i_n} \\ &\times \int_{\mathbb{R}^{nd}} b^{i_1}(y_1) \cdots b^{i_n}(y_n) g(f[y_n, \ldots, y_1, p]) \lambda^{(n)}(dy^{(n)}), \end{aligned} \tag{2.4}$$

where $(\mathbb{R}^{nd}, B(\mathbb{R}^{nd}), \lambda^{(n)})$ is the product measurable space, $\lambda^{(n)} = \lambda \otimes \ldots \otimes \lambda$, and $y^{(n)} = (y_1, \ldots, y_n)$.

It is easy to verify that the quadruple $\{(W, W), (\mathbb{R}^d, B(\mathbb{R}^d)), u, P\}$ is a RSCC ([6], definition 1.1.1, page 5), where $u : W \times \mathbb{R}^d \to W$,

$$u(p, y) = f[y, p] = \frac{Q^t B_\bullet(y) p}{b_\bullet^t(y) p}, y \in \mathbb{R}^d, p \in W,$$

and for all $p = (p^j) \in W$ and $E \in B(\mathbb{R}^d)$

$$P(p, E) = \sum_{j \in S} p^j \int_E b^j(y) \lambda(dy). \tag{2.5}$$

As in [6] (page 5), for any $n \geq 1$, we define recursively the maps $u^{(n)} : W \times (\mathbb{R}^d)^n \to W$ and the transition probability functions P_n, and we get

$$u^{(n)}(p, (y_1, \ldots, y_n)) = py^{(n)} = f[y_n, \ldots, y_1, p], y_i \in \mathbb{R}^d, p \in W, \tag{2.6}$$

and

$$P_n(p, E) = \sum_{i_1, \ldots, i_n \in S} p^{i_1} q^{i_1, i_2} \cdots q^{i_{n-1}, i_n} \int_E b^{i_1}(y_1) \cdots b^{i_n}(y_n) \lambda^{(n)}(dy^{(nd)}), \tag{2.7}$$

for any $E \in B(\mathbb{R}^{nd})$ and $p \in W$.

Comparing (2.5) with (2.3) and (2.7) with (2.4), we notice that

$$Ug(p) = \int_{\mathbb{R}^d} g(f[y, p]) P(p, dy),$$
$$U^n g(p) = \int_{\mathbb{R}^{nd}} g(f[y_n, \ldots, y_1, p]) P_n(p, dy^{(n)}).$$

Furthermore, using (2.4)-(2.5) it is easy to show that, for any probability distribution $\mathbf{p}_\bullet \in W$, the random sequences $\{p_{n+1}, n \geq 0\}$ with $p_1 = \mathbf{p}_\bullet$ and $\{Y_n, n \geq 1\}$ are associated with the previously defined RSCC, in the sense of theorem 1.1.2, page 6 of [6], on the probability space $(\Omega, F, \mathbf{P}_\bullet(\cdot | p_1 = \mathbf{p}_\bullet)$.

As in [4], let denote by $\min^+$ the minimum over positive elements, and put

$$\delta(y) = \frac{\max_{i \in S} b^i(y)}{\min_{i \in S} b^i(y)}, \quad \Delta_{-1} = \min_{i \in S} \int_{\mathbb{R}^d} \delta^{-1}(y) b^i(y) \lambda(dy),$$
$$\epsilon = \min_{i,j \in S}{}^{+} q^{i,j}, \quad \Delta = \max_{i \in S} \int_{\mathbb{R}^d} \delta(y) b^i(y) \lambda(dy), \quad R = \epsilon^r \Delta_{-1}^{r-1}.$$

Throughout this section, we suppose that the matrix Q is primitive with index of primitivity r. Hence, by definition, the matrix Q^r is positive, and r is the smallest integer with this property. As a consequence, the Markov chain $\{S_n, n \geq 1\}$ is geometrically ergodic with a unique invariant probability distribution $\pi_\bullet = (\pi^i)$ on S. Thus, (theorem

4.3 in [5], page 126) there exist two positive constant a_1 and $c_1 < 1$ such that, $Q^n = (q_n^{i,j})$, the n-th power of the matrix Q, satisfies

$$|q_n^{i,j} - \pi^j| \leq a_1 c_1^n, \tag{2.8}$$

for any $n \geq 1$ and any $i, j \in S$. We now establish two important properties of the previously defined RSCC.

Proposition 2.1. *The RSCC* $\{(W, W), (\mathbb{R}^d, B(\mathbb{R}^d)), u, P\}$ *is uniformly ergodic ([6], page 42, definition 2.1.4).*

Proof. For any integers $n, k \geq 1$, every $p \in W$ and $E \in B(\mathbb{R}^{kd})$, from (2.7) we get

$$P_k^n(p, E) = P_{n+k-1}(p, \mathbb{R}^{(n-1)d} \times E) = \sum_{i_1,\dots,i_{k+1} \in S} p^{i_1} q_{n-1}^{i_1,i_2} q^{i_2,i_3} \cdots q^{i_k,i_{k+1}}$$
$$\times \int_E b^{i_2}(y_1) \cdots b^{i_{k+1}}(y_k) \lambda^{(k)}(dy^{(k)}).$$

Let

$$P_k^\infty(E) = \sum_{i_1,\dots,i_k \in S} \pi^{i_1} q^{i_1,i_2} \cdots q^{i_{k-1},i_k} \int_E b^{i_1}(y_1) \cdots b^{i_k}(y_k) \lambda^{(k)}(dy^{(k)}).$$

Using (2.8), it is easy to prove that

$$\lim_{n\to\infty} P_k^n(p, E) = P_k^\infty(E),$$

uniformly with respect to p, E and k. Thus the RSCC is uniformly ergodic. □

Proposition 2.2. *If* $\Delta < \infty$ *then* $\{(W, W), (\mathbb{R}^d, B(\mathbb{R}^d)), u, P\}$ *is an RSCC with contraction ([6], page 79, definition 3.1.15).*

Proof. Replacing in (2.5) and using the definition of the $\|\cdot\|_1$, we obtain

$$R_1 = \sup_{E \in B(\mathbb{R}^d)} \sup_{p \neq p', p, p' \in W} \frac{|P(p, E) - P(p', E)|}{\|p - p'\|_1} \leq 1 < \infty. \tag{2.9}$$

From lemma 2.2 in [3], we get

$$\|f[y, p] - f[y, p']\|_1 \leq \delta(y) \|p - p'\|_1, \quad p, p' \in W, y \in \mathbb{R}^d.$$

Hence,

$$r_1 = \sup_{p \neq p', p, p' \in W} \int_{\mathbb{R}^d} \frac{\|f[y, p] - f[y, p']\|_1}{\|p - p'\|_1} P(p, dy) \leq \Delta < \infty. \tag{2.10}$$

For $n > 2r$, from the first inequality in theorem 2.1 and the inequality 5 in [4], we get

$$r_n = \sup_{p \neq p', p, p' \in W} \int_{\mathbb{R}^{dn}} \frac{\|f[y_n, \dots, y_1, p] - f[y_n, \dots, y_1, p']\|_1}{\|p - p'\|_1} P_n(p, dy^{(n)})$$
$$\leq \epsilon^{-r} \Delta^r (1 - R)^{n/r-2}.$$

Thus, for n sufficiently large, $r_n < 1$, and together with (2.9) and (2.10) this implies that we have an RSCC with contraction. □

Remark 2.1. If the observation conditional densities b^i are Gaussian for any $i \in S$ (as in example 2.1), then $\Delta < \infty$ (Example 4.2 in [4]).

Now, we return to the Markov chain $\{p_n\}$ associated with the RSCC. Using the previous proposition, theorem 3.1.16, page 80 in [6] and the fact that $(W, \|\cdot\|_1)$ is a compact space, we obtain that $\{p_n\}$ is a compact Markov chain (according to the definition 3.2.1, page 93 in [6]). Furthermore, in what follow we show that it is geometrically ergodic.

Let $L(W)$ denote the set of real-valued, bounded and Lipschitz continuous functions defined on W. Then $L(W)$ is a Banach space for the norm $\|\cdot\|_{BL}$ defined by

$$\|g\|_{BL} = \|g\| + s(g),$$

where

$$\|g\| = \sup_{p\in W} |g(p)|, \quad s(g) = \sup_{p\neq p', p,p'\in W} \frac{|g(p) - g(p')|}{\|p - p'\|_1}.$$

For any bounded linear operator V from $L(W)$ to $L(W)$, let denote

$$\|V\|_{BL} = \sup_{\|g\|_{BL}=1} \|Vg\|_{BL}.$$

We then have the following results.

Lemma 2.1. *There exists a positive constant K such that for any positive integer $n > r+1$, all $f \in L(W)$ and any p_1, $p_2 \in W$*

$$|U^n g(p_1) - U^n g(p_2)| \le nKc^n \|g\|_{BL},$$

where $c = \max\{c_1, c_2\}$, with $c_2 = (1-R)^{1/r}$ and c_1 as defined in (2.8).

Proof. We follow the same ideas as in the proof of theorem 3.5 in [4]. Using the second inequality in theorem 2.1 and inequality (5) in [4], we obtain a result similar to the proposition 3.7 in [4]:

$$\begin{aligned} \max_{i_l,\ldots,i_n\in S} \int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d} &|g(f[y_n,\ldots,y_l,p_1]) - g(f[y_n,\ldots,y_l,p_2])| \\ &\times b^{i_l}(y_l)\ldots b^{i_n}(y_n)\lambda(dy_l)\ldots\lambda(dy_n) \le 2s(g)c_2^{n-l+1-r}, \end{aligned} \tag{2.11}$$

for any positive integers n, l such that $n \ge l + r - 1$, and any function $g \in L(W)$. Notice that we can express

$$\begin{aligned} U^n g(p_1) - U^n g(p_2) &= \sum_{i_1,\ldots,i_n\in S} p_1^{i_1} q^{i_1,i_2}\cdots q^{i_{n-1},i_n} \int_{\mathbb{R}^{nd}} b^{i_1}(y_1)\cdots b^{i_n}(y_n) \\ &\times (g(f[y_n,\ldots,y_1,p_1]) - g(f[y_n,\ldots,y_1,p_2])\,\lambda^{(n)}\left(dy^{(n)}\right) + \sum_{i_1,\ldots,i_n\in S}(p_1^{i_1} - p_2^{i_1}) \\ &\times q^{i_1,i_2}\cdots q^{i_{n-1},i_n} \int_{\mathbb{R}^{nd}} b^{i_1}(y_1)\ldots b^{i_n}(y_n) g(f[y_n,\ldots,y_1,p_2])\lambda^{(n)}\left(dy^{(n)}\right). \end{aligned}$$

Let T_1 denotes the first term. Using (2.11), we get:

$$|T_1| \leq 2s(g)c_2^{n-r}. \tag{2.12}$$

For the second term T_2, we decompose $g(f[y_n, \dots, y_1, p_2])$ as in the proof of theorem 3.5 in [4]:

$$\begin{aligned}
&|T_2| \leq \sum_{k=2}^{l} \sum_{i_k,\dots,i_n \in S} \left| \sum_{i_1,\dots,i_{k-1} \in S} (p_1^{i_1} - p_2^{i_1}) q^{i_1,i_2} \cdots q^{i_{k-1},i_k} \right| q^{i_k,i_{k+1}} \cdots q^{i_{n-1},i_n} \\
&\int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} |g(f[y_n, \dots, y_k, z_{k-1}, \dots, z_1, p_2]) - g(f[y_n, \dots, y_{k+1}, z_k, \dots, z_1, p_2])| \\
&b^{i_k}(y_k) \cdots b^{i_n}(y_n) \lambda(dy_k) \dots \lambda(dy_n) + \sum_{i_1,\dots,i_n \in S} |p_1^{i_1} - p_2^{i_1}| q^{i_1,i_2} \cdots q^{i_{n-1},i_n} \\
&\int_{\mathbb{R}^{nd}} b^{i_1}(y_1) \cdots b^{i_n}(y_n) \, |g(f[y_n, \dots, y_1, p_2]) - g(f[y_n, \dots, y_2, z_1, p_2])| \, \lambda^{(n)} \left(dy^{(n)} \right) \\
&+ \sum_{i_{l+1},\dots,i_n \in S} \left| \sum_{i_1,\dots,i_l \in S} (p_1^{i_1} - p_2^{i_1}) q^{i_1,i_2} \cdots q^{i_l,i_{l+1}} \right| q^{i_{l+1},i_{l+2}} \cdots q^{i_{n-1},i_n} \\
&\int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} |g(f[y_n, \dots, y_{l+1}, z_l, \dots, z_1, p_2])| \, b^{i_{l+1}}(y_{l+1}) \cdots b^{i_n}(y_n) \\
&\lambda(dy_{l+1}) \dots \lambda(dy_n),
\end{aligned}$$

for any positive integer $l \leq n-1$ and any sequence $z_1, \dots, z_l \in \mathbb{R}^d$. The inequality (2.11) and p_1, $p_2 \in W$ yields

$$\begin{aligned}
|T_2| \leq\ & 2s(g) \sum_{k=2}^{l} c_2^{n-k-r} \sum_{i_k \in S} \left| \sum_{i_1 \in S} (p_1^{i_1} - p_2^{i_1})(q_{k-1}^{i_1,i_k} - \pi^{i_k}) \right| \\
&+ 2s(g)c_2^{n-1-r} \|p_1 - p_2\|_1 + \|g\| \sum_{i_{l+1} \in S} \left| \sum_{i_1 \in S} (p_1^{i_1} - p_2^{i_1})(q_l^{i_1,i_{l+1}} - \pi^{i_{l+1}}) \right|
\end{aligned}$$

Using (2.8) and $l = n - r$, we get

$$\begin{aligned}
|T_2| \leq \|p_1 - p_2\|_1 \Big(& 2s(g) M a_1 (n-r-1) c^{n-r-1} + 2s(g) c_2^{n-1-r} \\
& + \|g\| M a_1 c_1^{n-r} \Big).
\end{aligned} \tag{2.13}$$

Since $\|p_1 - p_2\|_1 \leq 2$, adding (2.12) and (2.13), we get the conclusion. □

Let us define

$$U_n = \frac{1}{n} \sum_{k=1}^{n} U^n, n \geq 1.$$

We can now formulate the main result concerning the Markov chain $\{p_n, n \geq 1\}$.

Theorem 2.1. *The Markov chain* $\{p_n, n \geq 1\}$ *is geometrically ergodic. Moreover, if* $\Delta < \infty$ *and* Q^∞ *is the unique invariant probability distribution for the chain* $\{p_n, n \geq 1\}$, *then we have*

1. $\lim_{n\to\infty} \|U_n - U^\infty\|_{BL} = 0$.
2. *There exists two positive constants* C *and* $\theta < 1$, *such that for any function* $g \in L(W)$

$$\|U^n g - U^\infty g\|_{BL} \leq C\theta^n \|g\|_{BL}.$$

Here, U^∞ *is the linear operator defined by*

$$U^\infty g = \int_W g(p) Q^\infty(dp),$$

for any bounded and measurable real-valued function g.

Proof. Using lemma 1 and proceedings as in the proof of corollary 3.6 in [4], it can be shown that there exists a unique invariant probability distribution Q^∞ for the Markov chain $\{p_n, n \geq 1\}$, and we have

$$|U^n g(z) - U^\infty g| \leq K\|g\|_{BL} \frac{nc^n}{(1-c)^2},$$

for any positive integer $n \geq 1$, all $g \in L(W)$, and any $z \in W$. If $\Delta < \infty$, the properties of the associated RSCC allow us to state stronger results concerning the convergence of the iterates of the transition operator U. As we have already mentioned, the Markov chain $\{p_n, n \geq 1\}$ is compact and by theorem 3.2.2, page 93 in [6], the Ionescu Tulcea-Marinescu ergodic theorem applies ([6], theorem A2.4, page 263).

Moreover, lemma 1 implies that any eigenfunction $g \in L(W)$ of U, corresponding to a eigenvalue γ with $|\gamma| = 1$, is a constant function. Hence, $\gamma = 1$ is the only eigenvalue of modulus 1 of U and the subspace

$$E(1) = \{g \in L(W) : Ug = g\}$$

is one dimensional. Thus, we can get the stated conclusions from the Ionescu Tulcea-Marinescu ergodic theorem and theorem 3.2.4, page 94 in [6]. □

3 Concluding remarks.

In practice, the matrix Q or the initial probability distribution $\mathbf{p}_\bullet$ may be unknown, and only some estimations of them may be used in the formulas for p_n or w_n. An important question is how these misspecifications influence the results, for a large n. For the prediction filter $\{p_n^*, n \geq 1\}$, with a wrong transition matrix and initial distribution, Le Gland and Mevel have proved in [4] that we have an $\mathbf{P}_\bullet$-a.s. exponential rate of forgetting of the initial distribution. They suppose that both the real and the wrong transition matrices are primitive. Under these assumptions, they have also proved that the Markov chains $\{S_n, Y_n, p_n^*, n \geq 1\}$ is geometrically ergodic.

A simple extension is to consider a HMM for which the mutually independence of the observations $\{Y_n, n \geq 1\}$, given the sequence $\{S_n, n \geq 1\}$, is replaced by a Markovian dependence. For example, in [8] we generalize the example 2.1 and we consider a hybrid model with observations $\{Y_n, n \geq 1\}$

$$Y_n = C(S_n)X_n + v_n, \tag{3.14}$$

and two sequences of hidden states: $\{S_n, n \geq 1\}$, as in the previously presented HMM, and $\{X_n, n \geq 1\}$ with $X_0 \sim N(\mu, \Sigma)$ and

$$X_n = AX_{n-1} + w_n, \quad w_n \sim N(0, H), n \geq 1. \tag{3.15}$$

Here A is the $m \times m$ transition matrix, $C(i)$, $i \in S$, are the $d \times m$ output matrices for the state-space model, and H is the symmetric and positive definite covariance matrix of the normal distribution corresponding to the independent random vectors w_n, $n \geq 1$. The noise sequences $\{w_n,\ n \geq 1\}$ and $\{v_n,\ n \geq 1\}$ are independent. Moreover, the sequences $\{S_n, n \geq 1\}$ and $\{X_0, w_n, n \geq 1\}$ are independent, X_0, w_n, $n \geq 1$ are independent, and conditional on $\{S_n, n \geq 1\}$, v_n, $n \geq 1$, are Gaussian and mutually independent such that, if $S_n = i$, $i \in S$, then $v_n \sim N(0, R(i))$, where $R(i)$ is the $d \times d$ symmetric and positive definite covariance matrix. For any $n \geq 1$, $i \in S$, $A_x \in B(\mathbb{R}^m)$ and $A_y \in B(\mathbb{R}^d)$, we have

$$\mathbf{P}_\bullet(Y_n \in A_y | X_n = x, S_n = i) = \int_{A_y} b^i(x, y)\lambda_d(dy), \tag{3.16}$$

$$\mathbf{P}_\bullet(X_n \in A_x | X_{n-1} = x) = \int_{A_x} a(y, x)\lambda_m(dy), \tag{3.17}$$

where

$$b^i(x, y) = (2\pi)^{-d/2}[\det R(i)]^{-1/2} \exp[-(y - C(i)x)^t R(i)^{-1}(y - C(i)x)/2],$$
$$a(y, x) = (2\pi)^{-m/2}[\det H]^{-1/2} \exp[-(y - Ax)^t H^{-1}(y - Ax)/2],$$

and λ_d and λ_m are the Lebesgue measures on $\mathbb{R}^d$ and $\mathbb{R}^m$, respectively.

Since the model (3.14) - (3.15) is a mixture model with an exponentially increasing number of components, deterministic inference algorithms become intractable rapidly. Usually, to overcome this difficulty, the Gibbs sampler ([2]) is applied. This means to implement an iterative algorithm and to draw samples $\{X_n(l), n \geq 0\}$, $l = 1, \ldots, L$ and then, for each sample, to calculate the prediction filter $w_n(l) = (w_n^i(l))$, $n \geq 1$ in order to draw a sample for the hidden sequence $\{S_n, n \geq 1\}$, too. Here, $w_1 = \mathbf{p}_\bullet$, and for any $n \geq 2$

$$w_n^i(l) = \mathbf{P}_\bullet(S_n = i | Y_{n-1}, X_{n-1}(l), \ldots, Y_1, X_1(l)).$$

In [8] we prove the $\mathbf{P}_\bullet$-a.s. exponential rate of forgetting of the initial distribution. Dependence with complete connections can be used to study the ergodic properties of the prediction filter $w_n(l) = (w_n^i(l))$, $n \geq 1$ (the details will be reported elsewhere).

A version of the generalized model (3.14) - (3.15) is also obtained in [2] when applying the Gibbs sampler for a hybrid model, formed by a state-space model and a HMM. These hybrids, which combine the discrete switching structure of the HMMs with the linear Gaussian dynamics of the state-space models, are used in various fields, from control engineering to econometrics.

Acknowledgment

The author would like to thank Y.S. Wong and I. Mizera for many useful discussions. This work has been supported by a Natural Sciences and Engineering Research Council of Canada grant

References

[1] Arapostathis, A. and Marcus, S.I., *Analysis of an identification algorithm arising in the adaptive estimation of Markov chains*, Math. Control Signals Systems. **3** (1990), 1-29.

[2] Carter, C.K. and Kohn, R., *On Gibbs sampling for state space models*, Biometrika **81** (1994), 541-553.

[3] Le Gland, F and Mevel, L., *Basic properties of the projective product, with application to products of column-allowable non-negative matrices*, Math. Control Signals Systems. **13** (2000), 41-62.

[4] Le Gland, F. and Mevel, L. , *Exponential forgetting and geometric ergodicity in hidden Markov models*, Math. Control Signals Systems **13** (2000), 63-93.

[5] Iosifescu, M., *Finite Markov Processes and Their Applications*, John Wiley, New York., 1980.

[6] Iosifescu, M. and Grigorescu, S., *Dependence with Complete Connections and Its Applications*, Cambridge University Press, Cambridge, 1990.

[7] Kaijser, T., *A limit theorem for partially observed Markov chains*, Ann. Probab. **3** (1975), 677-696.

[8] Popescu, C. and Wong, Y.S., *Nested Monte Carlo EM algorithm for switching state-space models*, IEEE Transactions on Knowledge and Data Engineering. **17** (2005), 1653-1664.

[9] Kunita, H., *Asymptotic behavior of the nonlinear filtering errors of Markov processes*, J. Multivariate Anal. **1** (1971), 365-393.

[10] Stettner, L., *On invariant measures of filtering processes*, Stochastic Differential Systems, Bad Honnef, 1988 (Eds. N. Christopeit, K. Helmes, and M. Kohlmann) Lecture Notes in Control and Inform. Sci., **126** Berlin, Springer, 279-292, 1989.