XIII International Conference Applied Stochastic Models and Data Analysis, (ASMDA 2009), June 30- July 3, Vilnius Lithuania, 2009

Exact Filtering and Smoothing in Markov Switching Systems Hidden with Gaussian Long Memory Noise

Wojciech Pieczynski[1], Noufel Abbassi[1], and Mohamed Ben Mabrouk[1]

[1] *Institut Telecom, Telecom SudParis,*
Dept. CITI, CNRS UMR 5157
8, rue Charles Fourier, 91 000 Evry, France
E-mail: Wojciech.Pieczynski@it-sudparis.eu

Abstract. Let X be a hidden real stochastic process, R be a discrete finite Markov chain, Y be an observed chain. The problem of filtering and smoothing is the problem of recovering both R and X from Y. In the classical models the exact computing with linear - or even polynomial - complexity in time index is not feasible and different approximations are used. Different alternative models, in which the exact calculations are feasible, have been recently proposed (2008). The core difference between these models and the classical ones is that the couple (R, Y) is a Markov one in the recent models, while it is not in all the classical ones. Here we propose a further extension of these models. The core point of the extension is the fact that the observed chain Y is not necessarily Markovian conditionally on (X, R) and, in particular, the long-memory distributions can be considered. We show that both filtering and smoothing are computable with complexity polynomial in the number of observations in the new model.
Keywords: exact filtering, exact smoothing, Markov switches, hidden linear system, long-memory noise

1 Introduction

Let $X_1^N = (X_1, ..., X_N)$ and $Y_1^N = (Y_1, ..., Y_N)$ be two sequences of random variables, and let $R_1^N = (R_1, ..., R_N)$ be a finite-value Markov chain. Each X_n takes its values from R^q, while each Y_n takes its values from R^m. The sequences X_1^N and R_1^N are hidden and the sequence $Y_1^N = (Y_1, ..., Y_N)$ is observed. We deal with two problems: the filtering problem and the smoothing one, whose formulation considered in this paper are

(i) calculation of $p(r_n | y_1^n)$ and $E[X_n | r_n, y_1^n)$;

(ii) calculation $p(r_n | y_1^N)$ and $E[X_n | r_n, y_1^N)$,

respectively. Let us consider a simple classical Gaussian state-space system, which consists of considering that R is a Markov chain and, roughly speaking, that (X, Y) is the classical linear system conditionally on R . This is summarized in the following:

$$R \text{ is a Markov chain;} \quad (1)$$

$$X_{n+1} = F_n(R_n)X_n + W_n \; ; \quad (2)$$

$$Y_n = H_n(R_n)X_n + Z_n \, , \quad (3)$$

where X_1, W_1, …, W_N , Z_1, …, Z_N are independent (conditionally on R_1^N) Gaussian vectors, $F_1(R_1)$, …, $F_N(R_N)$ are matrices of size $q \times q$ depending on switches, and $H_1(R_1)$, …, $H_N(R_N)$ are matrices of size $q \times m$ also depending on switches. The exact filtering and smoothing are not feasible with linear - or even polynomial - complexity in time in such models, and different approximations must be used. Many papers deal with this approximation problem and a rich bibliography can be seen in recent books (Cappe *et al.* 2005, Costa *et al.* 2005). Roughly speaking, there are two families of approximating methods: the stochastic ones, based on the Monte Carlo Markov Chains (MCMC) principle (Andrieu *et al.* 2003, Doucet *et al.* 2001, Cappe *et al.* 2005, Giordani *et al.* 2007, among others), and deterministic ones (Costa *et al.* 2005, Zoeter *et al.* 2006, among others).

To remedy this impossibility of exact computation two different models have been recently proposed in (Abbassi and Pieczynski 2008, Pieczynski 2008). Based on the general triplet Markov chains (Pieczynski and Desbouvries 2005), they make the exact computation of optimal Kalman-like filters possible, and the exact calculation of smoothing is also possible, as shown in (Bardel *et al.* 2009). The general idea leading to these models is to consider the independence of the X_1^N and Y_1^N conditionally on R_1^N.

Then these early models have been extended to more general ones, in which the independence of X_1^N and Y_1^N conditionally on R_1^N is no longer required (Pieczynski 2009, Pieczynski and Desbouvries 2009). Called "Markov marginal switching hidden model" (MMSHM), they verify:

$$(R_1^N, Y_1^N) \text{ is a Markov chain;} \quad (4)$$

$$X_{n+1} = F_n(R_n, Y_n) X_n + W_n, \quad (5)$$

with X_1 given and W_1, …, W_N independent random centered vectors in R^q.

The oriented dependence graphs of the models (1)-(3), (4)-(5), and the new one are given in Figure 1.

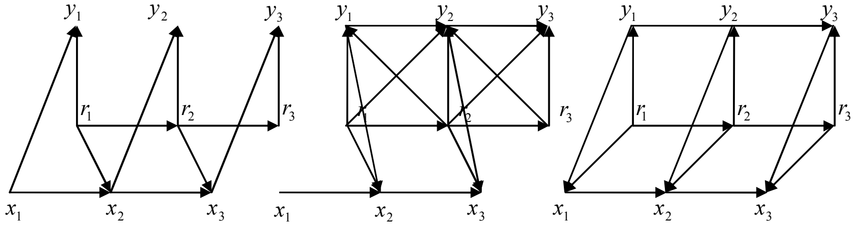

Fig. 1. Classical model (a), recent model making exact filtering and smoothing feasible (b); proposed long correlation model (c)

2 Markov switching state model with Gaussian correlated noise

Let us consider the triplet (X_1^N, R_1^N, Y_1^N) as above. The core point of the model we propose is to consider that the distribution of the couple (R_1^N, Y_1^N) is the distribution of the "partially Markov Gaussian chain" (PMGC) recently introduced in (Lanchantin *et al.* 2007). A PMGC verifies

$$p(y_{n+1}, r_{n+1} | y_1^n, r_1^n) = p(r_{n+1} | r_n) p(y_{n+1} | r_{n+1}, y_1^n) \quad (6)$$

where $p(y_{n+1} | r_{n+1}, y_1^n)$ are assumed Gaussian. Important in these models is that the conditional distributions $p(y_{n+1} | r_{n+1}, y_1^n)$ are computable: see (Lanchantin *et al.* 2007) for details.

We see that (R_1^N, Y_1^N) is Markovian with respect to R_1^N, but is not necessarily Markovian with respect to Y_1^N, which is at the origin of its appellation "partially" Markov.

Finally, the model we propose is the following

Definition

A triplet (X_1^N, R_1^N, Y_1^N) will be said to be a "hidden Markov switching conditionally linear model" (HMSCLM) if:

(R_1^N, Y_1^N) is a PMGC; (7)

$$X_{n+1} = F_{n+1}(R_{n+1}, Y_{n+1}) X_n + W_{n+1} , \quad (8)$$

where each $F_n(r_n, y_n)$ is a matrix of size $q \times q$ depending on (r_n, y_n), and W_1, …, W_N independent centered vectors in R^q such that W_n is independent from (R_1^N, Y_1^N) for each $n = 1$, …, N.

The oriented dependence graph of the new model (7)-(8) is given in Figure 1, (c). Let us underline the fact that there are no arrows going from y_1 to y_2, which means that conditionally on (X_1^N, R_1^N) the chain Y_1^N is not necessarily Markovian. Let us also highlight that the main difference between the classical models of kind (a) and the models of kind (b) or (c) consists of the fact that in (a) the arrows go from x_1, x_2, and x_3 to y_1, y_2, and y_3, while in the models (b) and (c) they go from y_1, y_2, and y_3 to x_1, x_2, and x_3.

Let us also notice that, as described in (Lanchantin *et al.* 2007), (7) includes different "long memory" distributions for $p(y_1^N | r_1^N)$.

According to the results presented in (Lanchantin *et al.* 2007), we have the following

Lemma

Let (R_1^N, Y_1^N) be a PMGC. Then the posterior margins $p(r_n | y_1^N)$ and transitions $p(r_{n+1} | r_n, y_1^N)$ are computable with complexity linear in time.

As a consequence, $p(r_n | r_{n+1}, y_1^N)$ are also computable.

3 Exact filtering

We can state the following result:

Proposition 1

Let us consider an HMSCLM (X_1^N, R_1^N, Y_1^N). Then $E[X_{n+1} | r_{n+1}, y_1^{n+1}]$ and $p(r_{n+1} | y_1^{n+1})$ are given from $E[X_n | r_n, y_1^n]$ and $p(r_n | y_1^n)$ by

$$p(r_{n+1} | y_1^{n+1}) = \frac{\sum_{r_n} p(r_n | y_1^n) p(r_{n+1} | r_n) p(y_{n+1} | r_{n+1}, y_1^n)}{\sum_{r_{n+1}} \sum_{r_n} p(r_n | y_1^n) p(r_{n+1} | r_n) p(y_{n+1} | r_{n+1}, y_1^n)} \tag{9}$$

$$E[X_{n+1} | r_{n+1}, y_1^{n+1}] = F_{n+1}(r_{n+1}, y_{n+1}) \sum_{r_n} E[X_n | r_n, y_1^n][\frac{p(r_n | y_1^n) p(r_{n+1} | r_n)}{\sum_{r_n} (r_n | y_1^n) p(r_{n+1} | r_n)}] \tag{10}$$

Proof

To show (9), we write

$$p(r_{n+1} | y_1^{n+1}) = \frac{p(r_{n+1}, y_{n+1} | y_1^n)}{p(y_{n+1} | y_1^n)} = \frac{\sum_{r_n} p(r_n, r_{n+1}, y_{n+1} | y_1^n)}{p(y_{n+1} | y_1^n)} = \frac{\sum_{r_n} p(r_n, r_{n+1} | y_1^n) p(y_{n+1} | r_{n+1}, y_1^n)}{p(y_{n+1} | y_1^n)} =$$

$$\frac{\sum_{r_n} p(r_n | y_1^n) p(r_{n+1} | r_n) p(y_{n+1} | r_{n+1}, y_1^n)}{p(y_{n+1} | y_1^n)}$$, and $p(y_{n+1} | y_1^n) = \sum_{r_{n+1}} p(y_{n+1}, r_{n+1} | y_1^n) =$

$$\sum_{r_{n+1}} \sum_{r_n} p(r_n | y_1^n) p(r_{n+1} | r_n) p(y_{n+1} | r_{n+1}, y_1^n) .$$

To show (10), let us take the conditional expectation of (8). We get

$$E[X_{n+1} | r_{n+1}, y_1^{n+1}] = F_{n+1}(r_{n+1}, y_{n+1}) E[X_n | r_{n+1}, y_1^{n+1}] = F_{n+1}(r_{n+1}, y_{n+1}) \sum_{r_n} E[X_n, r_n | r_{n+1}, y_1^{n+1}] =$$

$$F_{n+1}(r_{n+1}, y_{n+1}) \sum_{r_n} E[X_n | r_n, r_{n+1}, y_1^{n+1}] p(r_n | r_{n+1}, y_1^{n+1}) =$$

$$F_{n+1}(r_{n+1}, y_{n+1}) \sum_{r_n} E[X_n | r_n, y_1^n] p(r_n | r_{n+1}, y_1^{n+1}) ,$$

the last equality being due to the independence of X_n and (R_{n+1}, Y_{n+1}) conditionally on (R_n, Y_1^n) (see the dependence graph (c), Figure 1).

Otherwise, $p(r_n | r_{n+1}, y_1^{n+1}) = p(r_n | r_{n+1}, y_1^n) = \dfrac{p(r_n, r_{n+1} | y_1^n)}{\sum_{r_n} p(r_n, r_{n+1} | y_1^n)} = \dfrac{p(r_n | y_1^n) p(r_{n+1} | r_n)}{\sum_{r_n} (r_n | y_1^n) p(r_{n+1} | r_n)}$.

Reporting this quantity to the expression of $E[X_{n+1} | r_{n+1}, y_1^{n+1}]$ above gives (10).

4 Exact smoothing

We can state the following result:

Proposition 2

Let (X_1^N, R_1^N, Y_1^N) be an HMSCLM with given transitions $p(r_{n+1}, y_{n+1} | r_n, y_n)$. Then $E[X_{n+1}, r_{n+1} | y_1^N]$ can be computed from $E[X_n, r_n | y_1^N]$ by:

$$E[X_{n+1} | r_{n+1}, y_1^N] = F_{n+1}(r_{n+1}, y_{n+1}) \sum_{r_n} E[X_n | r_n, y_1^N] p(r_n |, r_{n+1}, y_1^N) \cdot \qquad (11)$$

If, in addition, the covariance matrices Σ_1, …, Σ_N of W_1, …, W_N exist, then $E[X_{n+1} X_{n+1}^T, r_{n+1} | y_1^N]$ satisfies

$$E[X_{n+1} X_{n+1}^T | r_{n+1}, y_1^N] = F_{n+1}(r_{n+1}, y_{n+1}) [\sum_{r_n} E[X_n X_n^T | r_n, y_1^N] p(r_n | r_{n+1}, y_1^N)] F_{n+1}^T(r_{n+1}, y_{n+1}) + \Sigma_{n+1}, \qquad (12)$$

and thus $Cov[X_n | y_1^N]$ can also be computed with complexity linear in time.

Proof.
By assumption, $X_{n+1} = F_{n+1}(R_{n+1}, Y_{n+1}) X_n + W_{n+1}$. Since W_{n+1} and (R_n, Y_1^N) are independent, and W_{n+1} is zero-mean, we have by taking the expectation of the both sides conditional on $(R_n, Y_1^N) = (r_n, y_1^N)$

$$\begin{aligned} E[X_{n+1} | r_{n+1}, y_1^N] &= F_{n+1}(r_{n+1}, y_{n+1}) E[X_n | r_{n+1}, y_1^N] \\ &= F_{n+1}(r_{n+1}, y_{n+1}) \sum_{r_n} E[X_n | r_n, r_{n+1}, y_1^N] p(r_n |, r_{n+1}, y_1^N) \cdot \end{aligned}$$

On the other hand, from model $E[X_n | r_n, r_{n+1}, y_1^N] = E[X_n | r_n, y_1^N]$, which gives (11).
Equation (2.7) is shown similarly: the independence of W_1, …, W_N implies that X_n and W_n are independent conditionally on (R_1^N, Y_1^N), so (1.5) gives

$$\begin{aligned} E[X_{n+1} X_{n+1}^T | r_{n+1}, y_1^N] &= F_{n+1}(r_{n+1}, y_{n+1}) E[X_{n+1} X_{n+1}^T | r_{n+1}, y_1^N] F_{n+1}^T(r_{n+1}, y_{n+1}) + E[W_{n+1} W_{n+1}^T | r_{n+1}, y_1^N] \\ &= F_{n+1}(r_{n+1}, y_{n+1}) E[X_{n+1} X_{n+1}^T | r_{n+1}, y_1^N] F_{n+1}^T(r_{n+1}, y_{n+1}) + \Sigma_{n+1} \cdot \qquad (13) \end{aligned}$$

On the other hand $E[X_{n+1}X_{n+1}^T | r_{n+1}, y_1^N] = \sum_{r_n} E[X_{n+1}X_{n+1}^T, r_n | r_{n+1}, y_1^N] = \sum_{r_n} E[X_{n+1}X_{n+1}^T | r_n, y_1^N] p(r_n | r_{n+1}, y_1^N)$. Combining it with (13) gives (12) and ends the proof.

References

Abbassi, N. and Pieczynski, W. 2008. Exact Filtering In Semi-Markov Jumping System, Sixth International Conference of Computational Methods in Sciences and Engineering, September 25-30, Hersonissos, Crete, Greece.

Andrieu, C., Davy, C. M., and Doucet, A. 2003. Efficient particle filtering for jump Markov systems. Application to time-varying autoregressions, *IEEE Trans. on Signal Processing*, 51(7): 1762-1770.

Bardel, N., Desbouvries, F., Pieczynski, W., and Barbaresco, F. 2009. An exact non-Gaussian Markov switching Bayesian smoothing algorithm with application to Track-Before-Detect, *RADAR 2009*, submitted.

Cappé, O., Moulines, E., and Ryden, T. 2005. *Inference in hidden Markov models*, Springer.

Costa, O. L. V., Fragoso, M. D., and Marques, R. P. 2005. *Discrete time Markov jump linear systems*, New York, Springer-Verlag.

Doucet, A., Gordon, N. J., and Krishnamurthy, V. 2001. Particle filters for state estimation of Jump Markov Linear Systems, *IEEE Trans. on Signal Processing*, 49: 613-624,

Giordani, P., Kohn, R., and van Dijk, D. 2007. A unified approach to nonlinearity, structural change, and outliers, *Journal of Econometrics*, 137: 112-133.

Lanchantin, P., Lapuyade-Lahorgue, J., and Pieczynski, W. 2008. Unsupervised segmentation of triplet Markov chains hidden with long-memory noise, *Signal Processing*, 88(5): 1134-1151.

Pieczynski W., and Desbouvries, F. 2005. On triplet Markov chains, International Symposium on Applied Stochastic Models and Data Analysis, (ASMDA 2005), Brest, France.

Pieczynski, W. 2008. Exact calculation of optimal filter in semi-Markov switching model, Fourth World Conference of the International Association for Statistical Computing (IASC 2008), December 5-8, Yokohama, Japan.

Pieczynski, W. 2009. Exact filtering in Markov marginal switching hidden models. Submitted to *Comptes Rendus Mathématique*.

Pieczynski, W. and Desbouvries, F. 2009. Exact Bayesian smoothing in triplet switching Markov chains, Complex data modeling and computationally intensive statistical methods for estimation and prediction (S. Co 2009), September 14-16, Milan, Italy, 2009.

Zoeter O., and Heskes, T. 2006. Deterministic approximate inference techniques for conditionally Gaussian state space models, *Statistical Computation*, 16: 279-292.