



**HAL**  
open science

# Exact filtering and smoothing in short or long memory stochastic switching systems

Wojciech Pieczynski, Noufel Abbassi

► **To cite this version:**

Wojciech Pieczynski, Noufel Abbassi. Exact filtering and smoothing in short or long memory stochastic switching systems. IEEE International Workshop on MACHINE LEARNING FOR SIGNAL PROCESSING, Sep 2009, Grenoble, France. hal-00408868

**HAL Id: hal-00408868**

**<https://hal.science/hal-00408868>**

Submitted on 3 Aug 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Exact filtering and smoothing in short or long memory stochastic switching systems

Wojciech Pieczynski and Noufel Abbassi

**Abstract**— Let  $X$  be a hidden real stochastic chain,  $R$  be a discrete finite Markov chain,  $Y$  be an observed stochastic chain. In this paper we address the problem of filtering and smoothing in the presence of stochastic switches where the problem is to recover both  $R$  and  $X$  from  $Y$ . In the classical conditionally Gaussian state space models, exact computing with polynomial complexity in the time index is not feasible and different approximations are used. Different alternative models, in which the exact calculations are feasible, have been recently proposed since 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 the classical ones. Another extension deals with the case in which the observed chain  $Y$  is not necessarily Markovian conditionally on  $(X, R)$  and, in particular, the long-memory distributions can be considered. The aim of this paper is to show that, in the context of these different recent models, it is possible to calculate any moments of the posterior marginal distribution, which makes it feasible to know these distributions with any desired precision.

## I. INTRODUCTION

Let  $X_1^N = (X_1, \dots, X_N)$ ,  $R_1^N = (R_1, \dots, R_N)$ , and  $Y_1^N = (Y_1, \dots, Y_N)$  be three sequences of random variables. Each  $X_n$  and  $Y_n$  takes its values from  $R$ , while each  $R_n$  takes its values from a finite set  $\Omega = \{\lambda_1, \dots, \lambda_k\}$ . Let us notice that the proposed study is likely to be extended to  $X_1^N$  and  $Y_1^N$  taking their values in  $R^q$  and  $R^m$ , respectively; however, we keep  $R$  for the sake of simplicity. The sequences  $X_1^N$  and  $R_1^N$  are hidden and the sequence  $Y_1^N$  is observed. We deal with two classical problems, which are the “filtering” problem and the “smoothing” one. The formulation of these problems considered in the present paper will be, respectively:

(F) : Computation of  $p(r_n | y_1^n)$  and  $E[X_n | r_n, y_1^n]$  with complexity polynomial in time;

(S) : Computation of  $p(r_n | y_1^N)$  and  $E[X_n | r_n, y_1^N]$  with complexity polynomial in time.

Let us consider a simple classical conditionally Gaussian state space model, which consists of considering that  $R_1^N$  is a Markov chain and, roughly speaking,  $(X_1^N, Y_1^N)$  is the

classical linear system conditionally on  $R_1^N$ . This is summarized in the following:

$$R_1^N \text{ is a Markov chain;} \quad (1)$$

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

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

where  $X_1, W_1, \dots, W_N, Z_1, \dots, Z_N$  are independent (conditionally on  $R_1^N$ ) Gaussian variables, and  $F_1(R_1), \dots, F_N(R_N), H_1(R_1), \dots, H_N(R_N), G_1(R_1), \dots, G_N(R_N), J_1(R_1), \dots, J_N(R_N)$  are real numbers depending on switches  $(R_n)$ , and (2), (3) hold for each  $n = 1, \dots, N$ . Such models are of interest in numerous situations [2], [3], [12], among others. However, it has been well known since [13] that 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 [1]-[3], [12].

To remedy this impossibility of exact computation different models have been recently proposed in [9]. These models, whose general idea is to consider the independence of  $X_1^N$  and  $Y_1^N$  conditionally on  $R_1^N$ , lead to exact filtering and smoothing. Two kinds of extensions of these models have then been proposed. The first one, called “Markov marginal switching hidden model” (MMSHM [10]), verifies:

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

$$X_{n+1} = F_{n+1}(R_{n+1}, Y_{n+1})X_n + G_{n+1}(R_{n+1})W_{n+1}, \quad (5)$$

where  $F_n(r_n, y_n), G_n(r_n)$  are real numbers depending on  $(r_n, y_n)$ , and  $W_1, \dots, W_N$  are independent centered real random variables such that  $W_n$  is independent from  $(R_1^N, Y_1^N)$  for each  $n = 1, \dots, N$ . In the second one, proposed in [11], the Markov chain  $(R_1^N, Y_1^N)$  is replaced by a “partially” Markov Gaussian chain recently introduced in [8], and (5) is kept. Both problems (F) and (S) can still be solved using these two models without any approximation.

The aim of this paper is to show that the calculations proposed in [10], [11] can be extended to calculations of  $E[(X_n)^k | r_n, y_1^n]$  and  $E[(X_n)^k | r_n, y_1^N]$ , where  $k$  is an arbitrary positive integer. In addition, we consider a third

model, in which the switching chain  $R_1^N$  has a semi-Markov distribution. This gives the moments of the distributions  $p(x_n|r_n, y_1^n)$  and  $p(x_n|r_n, y_1^N)$ , which means that these distributions can be known with any desired precision.

The oriented dependence graphs of the models (1)-(3), (4)-(5), and the ‘‘partially’’ Markov model (12)-(13) - which will be specified below - are given in Figure 1.

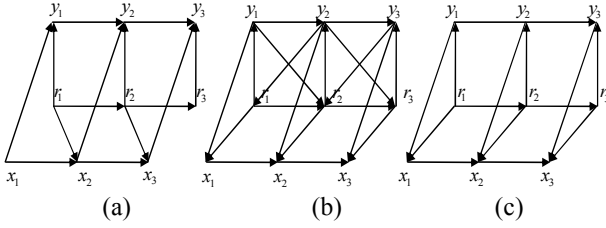


Fig. 1. (a): classical model (1)-(3), (b): short memory triplet Markov model (4)-(5) ; (c): long memory triplet partially Markov model.

The organization of the paper is the following. The next section is devoted to the MMSHM models, while the ‘‘hidden Markov switching conditionally linear model’’ (HMSCLM) models are dealt with in section three. The semi-Markov switching models are addressed in section four, and some experiments are provided in section five. The sixth section contains some conclusions.

## II. TRIPLET SHORT MEMORY MARKOV MODELS

### A. Filtering in a Short Memory Model

Let us consider the MMSHM model defined with (4)-(5). The computations of  $p(r_n|y_1^n)$  and  $E[(X_n)^m|r_n, y_1^n]$  are specified in the following Lemma 1 and Proposition 1, respectively.

#### Lemma 1

Let us consider a triplet Markov chain (TMM)  $(X_1^N, R_1^N, Y_1^N)$  verifying (4)-(5). Then  $p(r_{n+1}|y_1^{n+1})$  is given from  $p(r_n|y_1^n)$  by

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

Proof

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

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

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

$$\sum_{r_{n+1}} \sum_{r_n} p(r_n|y_1^n) p(r_{n+1}|r_n, y_n) p(y_{n+1}|r_n, r_{n+1}, y_n), \text{ which ends the proof.}$$

#### Proposition 1

Let us consider a TMM  $(X_1^N, R_1^N, Y_1^N)$  defined with (4)-(5).

Let  $C_j^k = (k!)/(j!)(k-j)!$ .

For each natural number  $m > 0$ , and each  $k = 1, \dots, m$ , the conditional expectation  $E[(X_{n+1})^k|r_{n+1}, y_1^{n+1}]$  is given from the conditional expectations  $E[X_n|r_n, y_1^n]$ ,  $E[(X_n)^2|r_n, y_1^n], \dots, E[(X_n)^k|r_n, y_1^n]$  by

$$\begin{aligned} E[(X_{n+1})^k|r_{n+1}, y_1^{n+1}] &= \\ &= \sum_{i=0}^k (C_i^k [F_{n+1}(r_{n+1}, y_{n+1})]^i E[(X_n)^i|r_{n+1}, y_1^{n+1}] + \\ &C_{k-i}^k [G_{n+1}(r_{n+1})]^{k-i} E[(W_{n+1})^{k-i}]) \end{aligned} \quad (7)$$

$$E[(X_n)^i|r_{n+1}, y_1^{n+1}] = \sum_{r_n} E[(X_n)^i|r_n, y_1^n] p(r_n|r_{n+1}, y_1^{n+1}), \quad (8)$$

and

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

As a consequence, for each fixed natural number  $m > 0$ ,  $E[(X_{n+1})^k|r_{n+1}, y_1^{n+1}]$  is computable with complexity polynomial in time.

Proof

Let  $m > 0$ , and let  $1 \leq k \leq m$ . According to (5) we have

$$(X_{n+1})^k = (F_{n+1}(R_{n+1}, Y_{n+1})X_n + G_{n+1}(R_{n+1})W_{n+1})^k = \sum_{i=0}^k C_i^k [F_{n+1}(R_{n+1}, Y_{n+1})]^i (X_n)^i + C_{k-i}^k [G_{n+1}(R_{n+1})]^{k-i} (W_{n+1})^{k-i}.$$

Taking the expectation of  $(X_{n+1})^k$  conditional on  $(R_{n+1}, Y_1^{n+1}) = (r_{n+1}, y_1^{n+1})$ , and taking into account the fact that the independence of  $W_{n+1}$  from  $(R_{n+1}, Y_1^{n+1})$  implies  $E[(W_{n+1})^i|r_{n+1}, y_1^{n+1}] = E[(W_{n+1})^i]$ , we may write:

$$\begin{aligned} E[(X_{n+1})^k|r_{n+1}, y_1^{n+1}] &= \sum_{i=0}^k (C_i^k [F_{n+1}(r_{n+1}, y_{n+1})]^i E[(X_n)^i|r_{n+1}, y_1^{n+1}] + \\ &+ C_{k-i}^k [G_{n+1}(r_{n+1})]^{k-i} E[(W_{n+1})^{k-i}]), \text{ which is (7).} \end{aligned}$$

Besides we have:

$$\begin{aligned}
 E[(X_n)^i | r_{n+1}, y_1^{n+1}] &= \sum_{r_n} E[(X_n)^i | r_n, r_{n+1}, y_1^{n+1}] p(r_n | r_{n+1}, y_1^{n+1}) = \\
 &= \sum_{r_n} E[(X_n)^i | r_n, y_1^n] p(r_n | r_{n+1}, y_1^{n+1}), \text{ the last equality being due} \\
 &\text{to the independence of } X_n \text{ and } (R_{n+1}, Y_{n+1}) \text{ conditionally on} \\
 &(R_n, Y_1^n) \text{ (see the dependence graph (b), Figure 1). Thus (8)} \\
 &\text{is verified. Finally, as } R_n \text{ is independent from } Y_{n+1} \\
 &\text{conditionally on } (R_{n+1}, Y_1^n), \text{ we have } p(r_n | r_{n+1}, y_1^{n+1}) = \\
 &= p(r_n | r_{n+1}, y_1^n) = \frac{p(r_n, r_{n+1} | y_1^n)}{\sum_{r_n} p(r_n, r_{n+1} | y_1^n)} = \frac{p(r_n | y_1^n) p(r_{n+1} | r_n)}{\sum_{r_n} (r_n | y_1^n) p(r_{n+1} | r_n)},
 \end{aligned}$$

which is (9) and ends the proof.

### B. Smoothing in a Short Memory Model

Let us consider a TMM  $(X, R, Y)$  defined with (4)-(5), and let  $m > 0$  be a natural fixed number. The problem in this section is to calculate  $p(r_n | y_1^N)$  and  $E[(X_n)^k | r_n, y_1^N]$  for each  $n = 1, \dots, N$ . We can state the following results.

#### Lemma 2

Let us consider a TMM  $(X_1^N, R_1^N, Y_1^N)$  verifying (4)-(5). The probabilities  $p(r_n | y_1^N)$ ,  $p(r_{n+1} | r_n, y_1^N)$ , and  $p(r_n | r_{n+1}, y_1^N)$  are computable with complexity linear in time.

Proof

As  $(R_1^N, Y_1^N)$  is a Markov chain,  $p(r_n | y_1^N)$  and  $p(r_{n+1} | r_n, y_1^N)$  are computable (see [8]); then  $p(r_n | r_{n+1}, y_1^N)$  is computed with (9).

#### Proposition 2

Let us consider a TMM  $(X, R, Y)$  defined with (4)-(5). For each natural number  $m > 0$ , and each  $k = 1, \dots, m$ , the conditional expectation  $E[(X_{n+1})^k | r_{n+1}, y_1^N]$  is given from

$$E[X_n | r_n, y_1^N], E[(X_n)^2 | r_n, y_1^N], \dots, E[(X_n)^k | r_n, y_1^N] \text{ by}$$

$$\begin{aligned}
 E[(X_{n+1})^k | r_{n+1}, y_1^N] &= \sum_{i=0}^k (C_i^k [F_{n+1}(r_{n+1}, y_{n+1})]^i E[(X_n)^i | r_{n+1}, y_1^N] + \\
 &+ C_{k-i}^k [G_{n+1}(r_{n+1})]^{k-i} E[(W_{n+1})^{k-i}]), \quad (10)
 \end{aligned}$$

with  $C_j^k = (k!)/(j!(k-j)!) \text{ and}$

$$E[(X_n)^i | r_{n+1}, y_1^N] = \sum_{r_n} E[(X_n)^i | r_n, y_1^N] p(r_n | r_{n+1}, y_1^N), \quad (11)$$

As a consequence, for each fixed natural number  $k > 0$ ,  $E[(X_n)^k | r_n, y_1^N]$  is computable with complexity polynomial in time.

Proof

To show (10), we have by assumption (5)

$$\begin{aligned}
 (X_{n+1})^k &= (F_{n+1}(R_{n+1}, Y_{n+1})X_n + G_{n+1}(R_{n+1})W_{n+1})^k = \\
 &= \sum_{i=0}^k (C_i^k [F_{n+1}(R_{n+1}, Y_{n+1})]^i (X_n)^i + C_{k-i}^k [G_{n+1}(R_{n+1})]^{k-i} (W_{n+1})^{k-i}).
 \end{aligned}$$

Taking the expectation of both sides conditional on  $(R_{n+1}, Y_1^N) = (r_{n+1}, y_1^N)$ , and using the fact that the noise  $W_{n+1}$  is independent from  $(R_{n+1}, Y_1^N)$ , we have (10).

$$\begin{aligned}
 \text{To show (11), we classically write } p(x_n | r_{n+1}, y_1^N) &= \\
 \sum_{r_n} p(x_n, r_n | r_{n+1}, y_1^N) &= \sum_{r_n} p(x_n | r_n, r_{n+1}, y_1^N) p(r_n | r_{n+1}, y_1^N) = \\
 \sum_{r_n} p(x_n | r_n, y_1^N) p(r_n | r_{n+1}, y_1^N), &\text{ the last equality coming from} \\
 p(x_n | r_n, r_{n+1}, y_1^N) = p(x_n | r_n, y_1^N). &
 \end{aligned}$$

## III. TRIPLET LONG MEMORY PARTIALLY MARKOV MODELS

### A. Filtering in a Long Memory Model

Let us consider the following model introduced in [11]. Let  $(X_1^N, R_1^N, Y_1^N)$  be the triplet of random sequences as above. The core point of the model is 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 [8]. Let us recall that the transitions of the distribution  $p(r_1^N, y_1^N)$  of a ‘‘partially’’ Markov chain  $(R_1^N, Y_1^N)$  verify  $p(r_{n+1}, y_{n+1} | r_n, y_1^n) = p(r_{n+1}, y_{n+1} | r_n, y_1^n)$ , which means that the chain  $(R_1^N, Y_1^N)$  is a Markov one with respect to the variables  $R_1^N$ , but is not necessarily Markovian with respect to the variables  $Y_1^N$  (which is at the origin of the name ‘‘partially’’ Markov chain). In addition, it will be assumed that  $p(r_{n+1}, y_{n+1} | r_n, y_1^n) = p(r_{n+1} | r_n) p(y_{n+1} | r_{n+1}, y_1^n)$  - which means that  $p(r_{n+1} | r_n, y_1^n) = p(r_{n+1} | r_n)$  - and it will be assumed that  $p(y_{n+1} | r_{n+1}, y_1^n)$  is the conditional distribution defined by a Gaussian distribution  $g_{r_{n+1}}(y_1^{n+1})$ . Let us focus on the last point. Recalling that each  $R_n$  takes its values from  $\Omega = \{\omega_1, \dots, \omega_K\}$ , let us consider, for each  $\omega_i$ , a Gaussian distribution  $g_{\omega_i}^N$  on  $\mathbb{R}^N$ . Then  $g_{r_{n+1}}(y_1^{n+1})$ , which is a Gaussian distribution on  $\mathbb{R}^{n+1}$ , is the marginal distribution of  $g_{\omega_i}^N$ . Finally, we see that the distribution of a PMGC  $(R_1^N, Y_1^N)$  is given by a Markov distribution of  $R_1^N$  and  $K$  Gaussian densities on  $\mathbb{R}^N$ .

Let us underline that such a model is quite different from the PMC model  $(R_1^N, Y_1^N)$  considered in the previous section. In fact, in PMC  $p(y_1^N | r_1^N)$  is a Markov distribution, but  $p(r_1^N)$  is not necessarily Markovian. In PMGC considered here,  $p(y_1^N | r_1^N)$  is not necessarily a Markov distribution, when  $p(r_1^N)$  is. Thus in a PMGC considered here we have a ‘‘Markov switching’’ model because of the Markov distribution  $p(r_1^N)$ .

Concerning the dependence graphs presented in Figure 1, let us 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 (5) includes different ‘‘long memory’’ distributions for  $p(y_1^N | r_1^N)$ , which are very useful in numerous situations [5].

Finally, the triplet  $(X_1^N, R_1^N, Y_1^N)$  is said to be a ‘‘hidden Markov switching conditionally linear model’’ (HMSCLM) if

$$(R_1^N, Y_1^N) \text{ is a PMGC}; \quad (12)$$

$$X_{n+1} = F_{n+1}(R_{n+1}, Y_{n+1})X_n + G_{n+1}(R_{n+1})W_{n+1}, \quad (13)$$

where  $F_n(r_n, y_n)$ ,  $G_n(r_n)$  are real numbers depending on  $(r_n, y_n)$ , and  $W_1, \dots, W_N$  are independent centered real random variables such that  $W_n$  is independent from  $(R_1^N, Y_1^N)$  for each  $n = 1, \dots, N$ .

Then in an HMSCLM  $(X_1^N, R_1^N, Y_1^N)$   $p(r_{n+1} | y_1^{n+1})$  is given from  $p(r_n | y_1^n)$  by [11]:

$$p(r_{n+1} | y_1^{n+1}) = \frac{\sum_{r_n} p(r_n | y_1^n) p(r_{n+1} | r_n) p(y_{n+1} | r_{n+1}, y_1^n)}{\sum_{r_{n+1}} \sum_{r_n} p(r_n | y_1^n) p(r_{n+1} | r_n) p(y_{n+1} | r_{n+1}, y_1^n)} \quad (14)$$

The following result shows the computability of  $E[(X_{n+1})^k | r_{n+1}, y_1^{n+1}]$  with complexity linear in time:

### Proposition 3

Let us consider a HMSCLM  $(X_1^N, R_1^N, Y_1^N)$ . Then for each natural number  $m > 0$ , and each  $k = 1, \dots, m$ , the conditional expectation  $E[(X_{n+1})^k | r_{n+1}, y_1^{n+1}]$  is given from the conditional expectations  $E[X_n | r_n, y_1^n]$ ,  $E[(X_n)^2 | r_n, y_1^n], \dots,$

$E[(X_n)^k | r_n, y_1^n]$  by (7), with  $C_j^k = (k!)/(j!)(k-j)!$ ,  $E[(X_n)^k | r_{n+1}, y_1^{n+1}]$  given by (8), and

$$p(r_n | r_{n+1}, y_1^{n+1}) = \frac{p(r_n | y_1^n) p(r_{n+1} | r_n)}{\sum_{r_n} (r_n | y_1^n) p(r_{n+1} | r_n)}. \quad (15)$$

The proof is analogous to the proof of the Proposition 1.

### B. Smoothing in a Long Memory Model

Let us consider a HMSCLM  $(X_1^N, R_1^N, Y_1^N)$  defined with (12)-(13), and let  $k > 0$  be a natural fixed number. As in the previous section, the problem is to compute  $p(r_n | y_1^N)$  and  $E[(X_n)^k | r_n, y_1^N]$ . According to the results presented in the probabilities  $p(r_n | y_1^N)$ ,  $p(r_{n+1} | r_n, y_1^N)$ , and  $p(r_n | r_{n+1}, y_1^N)$  are computable with complexity linear in time.

We can state:

### Proposition 4

Let us consider an HMSCLM  $(X_1^N, R_1^N, Y_1^N)$ .

For each natural number  $m > 0$ , and each  $k = 1, \dots, m$ , the conditional expectation  $E[(X_{n+1})^k | r_{n+1}, y_1^N]$  is given from  $E[X_n | r_n, y_1^N]$ ,  $E[(X_n)^2 | r_n, y_1^N], \dots,$   $E[(X_n)^k | r_n, y_1^N]$  by (10) and (11).

As a consequence, for each  $k > 0$ ,  $E[(X_n)^k | r_n, y_1^N]$  is computable with complexity polynomial in time.

The proof is similar to the proof of the Proposition 2.

## IV. SEMI-MARKOV SWITCHING MODELS

We have considered, in the two previous sections, two different distributions for the couple  $(R_1^N, Y_1^N)$ : pairwise Markov chains (PMC) and pairwise partially Markov chains (PPMC), the latter being considered with two additional hypotheses which are the Gaussianity of  $p(y_1^N | r_1^N)$  and the markovianity of  $p(r_1^N)$ . In the first case, the distribution of  $R_1^N$  is the marginal distribution of a Markov one, and thus, in the general case, it can be Markovian or not. However, in both cases the distribution  $p(r_1^N | y_1^N)$  is a Markov one. The aim of this section is to mention a more general case, in which  $p(r_1^N | y_1^N)$  is semi-Markovian. Let us first recall how one can introduce an auxiliary finite chain  $U_1^N$  in such a way that  $(R_1^N, U_1^N)$  is a Markov chain and  $R_1^N$  is a semi-Markov one. Following [9], we will assume that each  $U_i$  takes its values in a finite set  $A = \{0, 1, \dots, L\}$ , so that  $(R_1^N, U_1^N)$  is a finite Markov chain. For  $(R_n, U_n) = (r_n, u_n)$ ,

the number  $u_n$  denotes the minimal sojourn time of the next  $R_{n+1}, \dots, R_N$  in  $r_n$ . Therefore, if  $u_n = j > 0$ , we have  $(r_{n+1}, u_{n+1}) = (r_n, u_n - 1), \dots, (r_{n+j}, u_{n+j}) = (r_n, 0)$ . If  $u_n = 0$ , the distribution of  $R_{n+1}$  is a given transition  $p(r_{n+1}|r_n, u_n = 0)$ . Finally, the transition  $p(r_{n+1}, u_{n+1}|r_n, u_n) = p(r_{n+1}|r_n, u_n)p(u_{n+1}|r_{n+1}, r_n, u_n)$  of the Markov chain  $(R_1^N, U_1^N)$  is defined by

$$p(r_{n+1}|r_n, u_n) = \begin{cases} \delta_{r_n}(r_{n+1}) & \text{if } u_n > 0 \\ p(r_{n+1}|r_n) & \text{if } u_n = 0 \end{cases} \quad (16)$$

$$p(u_{n+1}|r_{n+1}, r_n, u_n) = \begin{cases} \delta_{u_n-1}(u_{n+1}) & \text{if } u_n > 0 \\ p(u_{n+1}|r_{n+1}) & \text{if } u_n = 0 \end{cases} \quad (17)$$

with  $\delta_a(b) = 1$  for  $a = b$ , and  $\delta_a(b) = 0$  otherwise.

Once the distribution of  $(R_1^N, U_1^N)$  is defined as a Markov chain distribution, we may consider the chain  $(X_1^N, R_1^N, U_1^N, Y_1^N)$ , and we may make  $(R_1^N, U_1^N)$  play a similar role to that played by  $R_1^N$  in the previous sections. For example, we can consider the following model:

$$(R_1^N, U_1^N) \text{ verifies (16)-(17);} \quad (18)$$

$$p(y_1^N|r_1^N, u_1^N) = \prod_{n=1}^N p(y_n|r_n); \quad (19)$$

$$X_{n+1} = F_{n+1}(R_{n+1}, Y_{n+1})X_n + G_{n+1}(R_{n+1})W_{n+1}, \quad (20)$$

where each  $F_n(r_n, y_n)$ ,  $G_n(r_n)$  are a real numbers depending on  $(r_n, y_n)$ , and  $W_1, \dots, W_N$  are independent centered real random variables such that  $W_n$  is independent from  $(R_1^N, Y_1^N)$  for each  $n = 1, \dots, N$ . Then we have a system with semi-Markov switches, in which the probabilities  $p(r_n|y_1^N)$ ,  $p(r_n|y_1^N)$  and the conditional expectations  $E[(X_n)^k|r_n, y_1^N]$ ,  $E[(X_n)^k|r_n, y_1^N]$  are computable in similar ways that in the previous sections. The oriented dependence graph of the model (18)-(20) is presented in Figure 2.

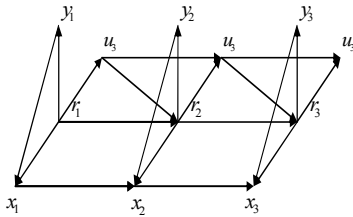


Fig. 2. Oriented graph dependence of a semi-Markov switching model (18)-(20).

## V. PARAMETER ESTIMATION AND EXPERIMENTS

The common point in the different models presented in the previous sections is that the distribution of the triplet  $(X_1^N, R_1^N, Y_1^N)$  is defined in two stages: first, one defines the distribution  $p(r_1^N, y_1^N)$  of  $(R_1^N, Y_1^N)$ , and then the distribution  $p(x_1^N|r_1^N, y_1^N)$  is defined by the linear equation (20). When  $(R_1^N, Y_1^N)$  is a Markov chain with Gaussian  $p(y_1^N|r_1^N)$ , all the parameters can be estimated by ‘‘Iterative Conditional Estimation’’ (ICE) as described in [4]. When  $(R_1^N, Y_1^N)$  is a partially Markov chain with Gaussian  $p(y_1^N|r_1^N)$ , all the parameters can still be estimated by ICE, as described in [8]. Thus we can consider ‘‘partially’’ unsupervised processing in both (4)-(5) (MMSHM) and (12)-(13) (HMSCLM) models: filtering and smoothing can be performed once the parameters of  $p(x_1^N|r_1^N, y_1^N)$  defined with (5) – or with (13), which is identical – are known.

The interest of these different models in real situations has to be tested by experiments, some of which are in progress. However, as shown by the following simple simulation study, they seem to be very powerful and promising.

Let us consider the following model

$$R_1^N \text{ is a Markov chain;} \quad (21)$$

$$X_n = F_n(R_n)X_{n-1} + G_n(R_n)W_n; \quad (22)$$

$$p(y_1^N|r_1^N, x_1^N) = \prod_{n=1}^N p(y_n|r_n), \quad (23)$$

which is a very simple particular case of the model (4)-(5).

We simulate data according to the classical model (1)-(3) and then they are filtered by two methods. The first one is the classical particle filtering based method explained in [1]. The second one is the exact method based on the MMSHM (21)-(23). The parameters of the hidden Markov chain  $(R_1^N, Y_1^N)$  needed in the MMSHM used are estimated with an ‘‘adaptive’’ ICE and the parameters in (22) are the same as the parameters in (2) used for simulation. Notice that the distribution of  $R_1^N$  is known; however, to make the filtering more unsupervised in the second method we use the distribution of  $R_1^N$  which is estimated with ICE.

The model (1)-(3) considered is given by a two-state stationary Markov chain taking its values from  $\Omega = \{\lambda_1, \lambda_2\}$ . The parameters used to simulate realizations of the triplet  $T_1^N = (X_1^N, R_1^N, Y_1^N)$  are the following. The initial distribution of  $p(r_1^N)$  is  $p_1 = (0.5, 0.5)$ , and the transitions are  $p(r_2|r_1 = r_2) = 1 - \rho$ ,  $p(r_2|r_1 \neq r_2) = \rho$ . The remaining parameters are  $F_n(\lambda_1) = -0.5$ ,  $F_n(\lambda_2) = 0.5$ ,  $H_n(\lambda_1) = -2$ ,  $H_n(\lambda_2) = 2$ ,  $G_n(\lambda_1) = 0.5$ ,  $G_n(\lambda_2) = 1$ ,  $J_n(\lambda_1) = 1$ , and

$J_n(\lambda_2) = 2$ . We consider  $N = 1000$  for the sample size and the results for four different values of  $\rho$  are presented in Table 1, where the difference between  $x_n$  and the filtered  $\hat{x}_n$  is measured by  $[\sum_{n=1}^N (x_n - \hat{x}_n)^2] / N$ . One example of simulated trajectory and filtered results (last fifty points), corresponding to  $\rho = 0.35$ , is presented in Figure 3.

The presented results, and others experiments results we performed, show that the semi-supervised (21)-(23) based method always takes upper hand over the particle filtering based one.

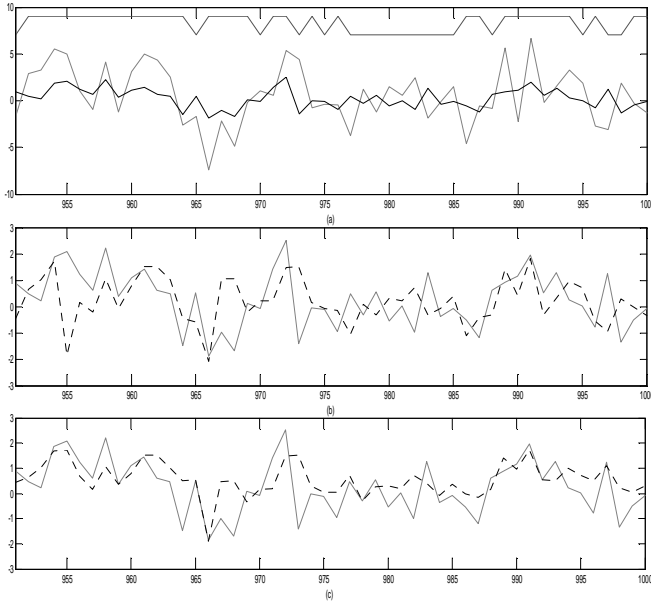


Fig. 3. (a): simulated  $r_1^N$  (top),  $x_1^N$  (dotted line), and  $y_1^N$  (double dotted line); (b):  $x_1^N$  (dotted) and PF based  $\hat{x}_1^N$  (double dotted) (c):  $x_1^N$  (dotted) and MMSHM based  $\hat{x}_1^N$  (double dotted).

$\rho$	0.10	0.35	0.65	0.90
PF	0.96	0.99	0.99	1.05
MMSHM	0.78	0.89	0.89	0.95

Table 1. Errors of filters based on PF and MMSHM measured by  $[\sum_{n=1}^N (x_n - \hat{x}_n)^2] / N$ .

## VI. CONCLUSION

We have considered the problem of filtering and smoothing in the presence of stochastic switches. Thus there is a triplet  $T_1^N = (X_1^N, R_1^N, Y_1^N)$  of random chains, where  $X_1^N$  is hidden continuous,  $R_1^N$  is hidden discrete, and  $Y_1^N$  is observed continuous. Usually, the distribution  $p(t_1^N)$  of  $T_1^N$  is given by  $p(x_1^N, r_1^N)$  and  $p(y_1^N | x_1^N, r_1^N)$ , which makes the computations of  $E[X_n | y_1^N]$  and  $E[X_n | y_1^N]$  with complexity polynomial in time unfeasible and different approximations

methods are needed. Here we exploited recent ideas according to which  $p(t_1^N)$  can also be defined by  $p(r_1^N, y_1^N)$  and  $p(x_1^N | y_1^N, r_1^N)$ . This key point makes filtering and smoothing computations feasible in different general models [9]-[11]. The contribution of this paper was to extend the computation of the conditional expectations  $E[X_n | y_1^N]$  and  $E[X_n | y_1^N]$  to the computation of any conditional moments  $E[(X_n)^m | y_1^N]$  and  $E[(X_n)^m | y_1^N]$ , which makes the computation of the marginal posterior distributions  $p(x_n | y_1^N)$ ,  $p(x_n | y_1^N)$  feasible with any desired precision. In addition, some experiments based on the very simple model proposed in [9] show the interest of the new models with respect to the classical particle filtering based methods.

As perspectives, we can mention the study extensions of the different models to general Bayesian networks [7].

## REFERENCES

- [1] C. Andrieu, C. M. Davy, and A. Doucet, "Efficient particle filtering for jump Markov systems. Application to time-varying autoregressions", *IEEE Trans. on Signal Processing*, vol. 51, no. 7, pp. 1762-1770, 2003.
- [2] O. Cappé, E. Moulines, and T. Ryden, "Inference in hidden Markov models", Springer, 2005.
- [3] O. L. V Costa, M. D. Fragoso, and R. P. Marques, "Discrete time Markov jump linear systems", New York, Springer-Verlag, 2005.
- [4] S. Derrode and W. Pieczynski, "Signal and image segmentation using pairwise Markov chains", *IEEE Trans. on Signal Processing*, vol. 52, no. 9, pp. 2477-2489, 2004.
- [5] P. Doukhan, G. Oppenheim, and M. S. Taqqu, "Long-Range Dependence", Birkhauser, 2003.
- [6] A. Doucet, N. J. Gordon, and V. Krishnamurthy, "Particle filters for state estimation of Jump Markov Linear Systems", *IEEE Trans. on Signal Processing*, vol. 49, no. 3, pp. 613-624, 2001.
- [7] F. V. Jensen, *Introduction to Bayesian network*, Springer-Verlag, 1996.
- [8] P. Lanchantin, J. Lapuyade-Lahorgue, and W. Pieczynski, "Unsupervised segmentation of triplet Markov chains hidden with long-memory noise", *Signal Processing*, vol. 88, no. 5, pp. 1134-1151, 2008.
- [9] W. Pieczynski, "Exact calculation of optimal filter in semi-Markov switching model", In *Proc. Fourth World Conference of the International Association for Statistical Computing (IASC 2008)*, December 5-8, Yokohama, Japan, 2008.
- [10] W. Pieczynski and F. Desbouvries, "Exact Bayesian smoothing in triplet switching Markov chains", In *Proc. Complex data modelling and computationally intensive statistical methods for estimation and prediction (S. Co 2009)*, September 14-16, Milan, Italy, 2009.
- [11] W. Pieczynski, N. Abbassi, and M. Ben Mabrouk, "Exact filtering and smoothing of Markov switching linear system hidden with Gaussian long-memory Noise", In *Proc. XIII International Conference Applied Stochastic Models and Data Analysis (ASMDA 2009)*, June 30- July 3, Vilnius, Lithuania, 2009.
- [12] B. Ristic, S. Arulampalam, and N. Gordon, "Beyond the Kalman Filter - Particle filters for tracking applications", Artech House, Boston, 2004.
- [13] J. K. Tugnait, "Adaptive estimation and identification for discrete systems with Markov jump parameters", *IEEE Trans. on Automatic Control*, vol. AC-25, pp. 1054-1065, 1982.
- [14] O. Zoeter and T. Heskes, "Deterministic approximate inference techniques for conditionally Gaussian state space models", *Statistical Computation*, vol. 16, pp. 279-292, 2006.