Exact Bayesian Prediction in a class of Markov-switching models

Noémie Bardel, François Desbouvries
Telecom SudParis, Département CITI and CNRS UMR 5157,
91011, Evry, France
{ Noemie.Bardel,Francois.Desbouvries }@it-sudparis.eu

Abstract

Jump-Markov state-space systems (JMSS) are widely used in statistical signal processing. However as is well known Bayesian restoration in JMSS is an NP-hard problem, so in practice all inference algorithms need to resort to some approximations. In this paper we focus on the computation of the conditional expectation of the hidden variable of interest given the available observations, which is optimal from the Bayesian quadratic risk viewpoint. We show that in some stochastic systems, namely the Partially Pairwise Markov-switching Chains (PPMSC) and Trees (PPMST), no approximation scheme is actually needed since the conditional expectation of interest (be it either in a filtering or prediction problem) can be computed exactly and in a number of operations linear in the number of observations.

Index Terms

Bayesian restoration, Jump-Markov state-space systems, Partially Pairwise Markov switching models, NP-hard problems.

I. Introduction

Let $X_{0:n} = (X_0, ..., X_n)$ be a hidden random sequence with values in $\mathbb{R}^q$, $Y_{1:n} = (Y_1, ..., Y_n)$ an observed random sequence with values in $\mathbb{R}^m$, and $R_{1:n} = (R_1, ..., R_n)$ a discrete random

sequence with values in a finite set $\Omega = \{\omega_1, \cdots, \omega_K\}$ which usually models the random changes of regime, or switches of the distribution of $(X_{1:n}, Y_{1:n})$. The three chains are linked via some probability distribution $p(x_{0:n}, y_{1:n}, r_{1:n})$. We assume a perfect knowledge of $p(x_{0:n}, y_{1:n}, r_{1:n})$ and we address the restoration of the hidden random sequence X. From a Bayesian point of view, the best we can do is to compute a posterior probability density function (pdf) of interest, say $p(x_k|y_{1:n})$, for some value of k and n. On the other hand, in many applications only a point estimate of x_k is of interest, and a commonly used estimator is $\mathrm{E}(X_k|y_{1:n})$, i.e. the solution of the Bayesian restoration problem with quadratic loss.

However, in many stochastic models neither $p(x_k|y_{1:n})$ nor $\mathrm{E}(X_k|y_{1:n})$ can be computed exactly, because closed-form formulas are either unavailable, or involve a computational load exponential in the number of observations, and thus cannot be implemented exactly. In such cases one needs to resort to some kind of approximations.

Let us consider for instance the classical conditionally linear Gaussian model, also called JMSS, described by the following equations :

$$R_n \text{ is a Markov Chain (MC)}, \tag{1}$$

$$X_{n+1} = F_{n+1}(R_{n+1})X_n + W_{n+1}, \tag{2}$$

$$Y_n = H_n(R_n)X_n + Z_n, \tag{3}$$

where X_0,W_1,...W_n are Gaussian vectors in $\mathbb{R}^q$, and Z_1,...Z_n are Gaussian vectors in $\mathbb{R}^m$, which are independent and independent of R_1,...,R_n. So model (1)-(3) is nothing but a classical linear and Gaussian state-space system, except that its dynamics (given by matrices $F_n(R_n)$ and $H_n(R_n)$) depend on the realization of an MC R.

For fixed $R_1 = r_1$,...,$R_n = r_n$, $p(x_k|y_{1:n})$ is Gaussian with parameters which can be computed (whatever k and n) by standard Kalman-like techniques. However, it has been well known since [1] that exact computation of $p(x_k|y_{1:n})$ (or even of the conditional expectation $\mathrm{E}(X_k|y_{1:n}) = \sum_{r_{1:n}} \mathrm{E}(X_k|r_{1:n}, y_{1:n})p(r_{1:n}|y_{1:n})$) is no longer possible with random Markov R, because such a computation would involve a number of operations proportional to K^n, where K is the number of states of the Markov chain R. The Bayesian restoration problem in (1)-

(3) is thus NP-hard and different approximation schemes have been proposed, such as limiting in some way the number of components in the Gaussian mixture [1], [2], [3], [4], or using particle filtering methods [5], [6], [7]. Finally the point estimator of x_k is an approximation of $E(X_k|y_{1:n})$.

It is thus of interest to propose stochastic models $p(x_{0:N}, y_{1:N}, r_{1:N})$ for which the exact computation of $\mathrm{E}(X_k|y_{1:n})$ is computable in practice, i.e. with a number of operations linear (or at least polynomial) in the number n of observations.

Let us now turn to the contents of this paper. The Bayesian prediction problem which we address consists in computing efficiently the conditional expectation $\mathrm{E}[X_{n+p}|y_{1:n}]$ and associated conditional covariance matrix $\mathrm{Cov}(X_{n+p}|y_{1:n})$ in a particular class of stochastic dynamical models with Markov regime. More precisely, in section II we show that $\mathrm{E}[X_{n+p}|y_{1:n}]$ and $\mathrm{Cov}(X_{n+p}|y_{1:n})$ can be computed exactly, with complexity linear in time, in a recent switching-Markov system proposed in [8]. The extension of this algorithm to switching-Markov trees is considered in section III.

II. Exact filtering and prediction in PPMSC

In this paper we thus consider the following PPMSC model :

$$R_n \text{ is an MC}, \tag{4}$$

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

$$(R_n, Y_n) \text{ is a Partially Pairwise Markov Chain}, \tag{6}$$

where X_0,W_1,...,W_n are independent zero-mean random vectors, which are independent and independent of $(R_{1:n}, Y_{1:n})$. We suppose that (R_n, Y_n) is a Partially Pairwise MC (PPMC) [9], i.e. that

$$p(r_{n+1}, y_{n+1}|r_{1:n}, y_{1:n}) = p(r_{n+1}, y_{n+1}|r_n, y_{1:n}). \tag{7}$$

Note that in (5) (as compared to (2)) vectors W_n are not necessarily Gaussian. Also, by contrast with eq. (3), there is no longer any direct stochastic relation between Y_n and X_n; but of course $X_{1:n}$ and $Y_{1:n}$ remain dependent, and are linked via the MC $R_{1:n}$.

In this section we extend the restoration algorithm presented in [10] to the case where (R_n, Y_n) is a PPMC. Let us first introduce some notations. For any integers k and n, let us set

$$M_k(r_k, y_{1:n}) = \int_{\mathbb{R}^q} x_k p(x_k, r_k|y_{1:n}) dx_k. \tag{8}$$

We will assume either that $k = n$ (for filtering) or $k > n$ (if prediction is assumed). If the covariance matrix Σ_n of W_n exists for all n, let us set

$$V_k(r_k, y_{1:n}) = \int_{\mathbb{R}^q} x_k x_k^T p(x_k, r_k|y_{1:n}) dx_k. \tag{9}$$

Of course, $\mathrm{E}[X_k|y_{1:n}]$ and $\mathrm{Cov}(X_k|y_{1:n})$ can be computed from $M_k(r_k, y_{1:n})$ and $V_k(r_k, y_{1:n})$ as:

$$\mathrm{E}[X_k|y_{1:n}] = \sum_{r_k} M_k(r_k, y_{1:n}), \tag{10}$$

$$\mathrm{Cov}[X_k|y_{1:n}] = \sum_{r_k} V_k(r_k, y_{1:n}) - (\sum_{r_k} M_k(r_k, y_{1:n}))(\sum_{r_k} M_k(r_k, y_{1:n}))^T. \tag{11}$$

In the following we thus focus on the (recursive) computation of $M_k(r_k, y_{1:n})$ and $V_k(r_k, y_{1:n})$.

A. The filtering problem in PPMSC

For convenience of the reader we recall here a (slightly modified) version of the filtering algorithm presented in [8] because it will be needed in §II-B. Let $(X_{1:n}, R_{1:n}, Y_{1:n})$ satisfy (4)-(6), with given transitions $p(r_{n+1}|r_n)$ and $p(r_{n+1}, y_{n+1}|r_{1:n}, y_{1:n}) = p(r_{n+1}, y_{n+1}|r_n, y_{1:n})$. Then we have:

$$M_{n+1}(r_{n+1}, y_{1:n+1}) = \frac{1}{p(y_{n+1}|y_{1:n})} F_{n+1}(r_{n+1}) \sum_{r_n} M_n(r_n, y_{1:n}) p(r_{n+1}, y_{n+1}|r_n, y_{1:n}) \tag{12}$$

with

$$p(y_{n+1}|y_{1:n}) = \frac{p(y_{1:n+1})}{p(y_{1:n})} = \frac{\sum_{r_{n+1}} p(r_{n+1}, y_{1:n+1})}{\sum_{r_n} p(r_n, y_{1:n})}, \tag{13}$$

in which

$$p(r_{n+1}, y_{1:n+1}) = \sum_{r_n} p(r_n, y_{1:n}) p(r_{n+1}, y_{n+1}|r_n, y_{1:n}). \tag{14}$$

(Note that (14) is nothing but the extension to PPMC of the classical forward formula of the Forward-Backward or BCJR algorithm, originally designed for hidden Markov chains, see e.g. [11], [12].) Furthermore, if the covariance matrix Σ_n of W_n exists for all n we have :

$$\begin{aligned} V_{n+1}(r_{n+1}, y_{1:n+1}) &= \frac{1}{p(y_{n+1}|y_{1:n})} \sum_{r_n} p(r_{n+1}, y_{n+1}|r_n, y_{1:n}) \\ &\times [F_{n+1}(r_{n+1}) V_n(r_n, y_{1:n}) F_{n+1}(r_{n+1})^T + \Sigma_{n+1}]. \end{aligned} \tag{15}$$

B. The prediction problem in PPMSC

Proposition.

Let $(X_{1:N}, R_{1:N}, Y_{1:N})$ satisfy (4)-(6), with given transitions $p(r_{n+1}|r_n)$ and $p(r_{n+1}, y_{n+1}|r_n, y_{1:n})$. Then $M_{n+p}(r_{n+p}, y_{1:n})$ can be recursively computed with linear complexity in time $n+p$ by the following scheme :

- compute $M_n(r_n, y_{1:n})$ with the filtering algorithm recalled above;
- for each integer $p \geq 0$, compute

$$M_{n+p+1}(r_{n+p+1}, y_{1:n}) = \sum_{r_{n+p}} F_{n+p+1}(r_{n+p+1}) M_{n+p}(r_{n+p}, y_{1:n}) p(r_{n+p+1}|r_{n+p}). \tag{16}$$

Furthermore, if the covariance matrix Σ_n of W_n exists for all n, then $V_{n+p}(r_{n+p}, y_{1:n})$ can be computed as follows:

- compute $V_n(r_n, y_{1:n})$ with the filtering algorithm recalled above;
- for each integer $p \geq 0$, compute

$$\begin{aligned} V_{n+p+1}(r_{n+p+1}, y_{1:n}) &= \sum_{r_{n+p}} p(r_{n+p+1}|r_{n+p}) \\ &\times [F_{n+p+1}(r_{n+p+1}) V_{n+p}(r_{n+p}, y_{1:n}) F_{n+p+1}(r_{n+p+1})^T + \Sigma_{n+p+1}]. \end{aligned} \tag{17}$$

Proof.

We have

$$\begin{aligned} p(x_{n+p+1}, r_{n+p+1}|y_{1:n}) &= \int_{\mathbb{R}^q} \sum_{r_{n+p}} p(x_{n+p+1}, r_{n+p+1}, x_{n+p}, r_{n+p}|y_{1:n}) dx_{n+p} \\ &= \int_{\mathbb{R}^q} \sum_{r_{n+p}} p(x_{n+p}, r_{n+p}|y_{1:n}) p(x_{n+p+1}, r_{n+p+1}|x_{n+p}, r_{n+p}, y_{1:n}) dx_{n+p}. \end{aligned} \quad (18)$$

On the other hand, using (4) and (5),

$$\begin{aligned} p(x_{n+p+1}, r_{n+p+1}|x_{n+p}, r_{n+p}, y_{1:n}) &= p(x_{n+p+1}|r_{n+p+1}, x_{n+p}, r_{n+p}, y_{1:n}) p(r_{n+p+1}|x_{n+p}, r_{n+p}, y_{1:n}) \\ &= p(x_{n+p+1}|r_{n+p+1}, x_{n+p}) p(r_{n+p+1}|r_{n+p}). \end{aligned}$$

We then multiply (18) by x_{n+p+1} and integrate with respect to x_{n+p+1} to get :

$$\begin{aligned} M_{n+p+1}(r_{n+p+1}, y_{1:n}) &= \int_{\mathbb{R}^q} x_{n+p+1} p(x_{n+p+1}, r_{n+p+1}|y_{1:n}) dx_{n+p+1} \\ &= \int_{\mathbb{R}^q} \sum_{r_{n+p}} p(x_{n+p}, r_{n+p}|y_{1:n}) \mathrm{E}[X_{n+p+1}|r_{n+p+1}, x_{n+p}] p(r_{n+p+1}|r_{n+p}) dx_{n+p}. \end{aligned}$$

Since the $\{W_n\}$ are independent, zero-mean and independent from $(R_{1:n}, Y_{1:n})$, we have

$$\mathrm{E}[X_{n+p+1}|r_{n+p+1}, x_{n+p}] = F_{n+p+1}(r_{n+p+1}) x_{n+p},$$

and finally

$$\begin{aligned} M_{n+p+1}(r_{n+p+1}, y_{1:n}) &= \int_{\mathbb{R}^q} \sum_{r_{n+p}} p(x_{n+p}, r_{n+p}|y_{1:n}) F_{n+p+1}(r_{n+p+1}) x_{n+p} p(r_{n+p+1}|r_{n+p}) dx_{n+p} \\ &= \sum_{r_{n+p}} F_{n+p+1}(r_{n+p+1}) M_{n+p}(r_{n+p}, y_{1:n}) p(r_{n+p+1}|r_{n+p}), \end{aligned}$$

which completes the proof of (14). (15) is obtained similarly.

Remark.

In the proof only (4) and (5) have been used, which implies that our proposition also holds for JMSS; the prediction problem in JMSS however remains NP-hard since, of course, in that case $M_n(r_n, y_{1:n})$ cannot be computed exactly.

III. EXTENSION TO PPMST

A. Introduction

Multiresolution signal and image analysis and multiscale algorithms are of interest in many fields [13], [14]. In particular, efficient restoration algorithms in statistical models defined on Hidden Markov trees (HMT) have been developed (see e.g. [15], [16]).

Let us first briefly recall the definition of a Markov Tree (MT). Let $\mathcal{S}$ be a finite set of indices and let us consider a tree with nodes indexed by $\mathcal{S}$. Let us consider a partition $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, ..., \mathcal{S}_N\}$, where $\mathcal{S}_n$ are the generations of the tree : $\mathcal{S}_1$ is the root node r, $\mathcal{S}_2$ is the set of its children, and so on. Each node s except the root node r has exactly one parent s^-, the set of the children of s is denoted by s^+, the set of all descendants of s by s^{++} and the set of all ancestors of s by s^{--}. We also denote by $a(s)$ the set of all ancestors of s and s itself (i.e. $a(s) = \{s^{--}, s\}$).Without loss of generality we consider here the case of dyadic trees: each node $s \notin S_N$ has exactly two children s_1 and s_2 (i.e. $s^+ = \{s_1, s_2\}$) (see fig. 1). Each node s is associated with a random variable $X(s)$. Also we introduce the notation $X_{\mathcal{S}} = \{X(s), s \in \mathcal{S}\}$. The tree is a Markov one if

$$p(x_{\mathcal{S}}) = p(x_r) \prod_{s \in \mathcal{S} \setminus \mathcal{S}_1} p(x_s | x_{s^-}).$$

Let now $X_{\mathcal{S}} = \{X(s), s \in \mathcal{S}\}$ and $Y_{\mathcal{S}} = \{Y(s), s \in \mathcal{S}\}$ be two sets of variables defined on the same set $\mathcal{S}$. Variables $X(s)$ (resp. $Y(s)$) are hidden (resp. observed). $(X_{\mathcal{S}}, Y_{\mathcal{S}})$ is an HMT if their joint distribution satisfies:

$$p(x_{\mathcal{S}}, y_{\mathcal{S}}) = p(x_r) \prod_{s \in \mathcal{S} \setminus \mathcal{S}_1} p(x_s | x_{s^-}) \prod_{s \in \mathcal{S}} p(y_s | x_{\mathcal{S}}),$$

i.e. x is an MT and $p(y_{\mathcal{S}}|x_{\mathcal{S}}) = \prod_{s \in \mathcal{S}} p(y_s|x_{\mathcal{S}})$. As we can see HMT are a natural extension of well-known Hidden Markov Chains to trees. HMT have been extended to Pairwise Markov Trees (PMT) [17] [18] defined by :

$$p(z_{\mathcal{S}}) = p(z_r) \prod_{s \in \mathcal{S} \setminus \mathcal{S}_1} p(z_s | z_{s^-}),$$

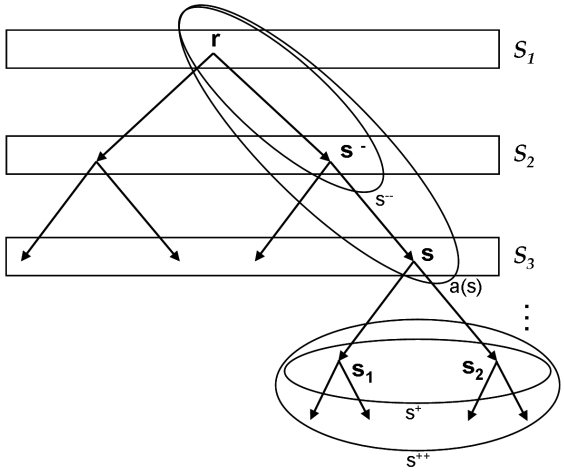

Fig. 1. Example of dyadic tree

in which $z_s = (x_s, y_s)$ and $z_{\mathcal{S}} = (x_{\mathcal{S}}, y_{\mathcal{S}})$. Any HMT is a PMT, but the converse is not true, since in a PMT, $x_{\mathcal{S}}$ is not necessarily an MT. Finally PMT can be extended to Partially Pairwise Markov Trees (PPMT), which we define by :

$$p(z_{\mathcal{S}}) = p(z_r) \prod_{s \in \mathcal{S} \setminus \mathcal{S}_1} p(z_s | z_{s^-}, y_{s^{--}}) = p(x_r, y_r) \prod_{s \in \mathcal{S} \setminus \mathcal{S}_1} p(x_s, y_s | x_{s^-}, y_{a(s^-)}). \tag{19}$$

Let us now introduce a third latent process $R_{\mathcal{S}}$ taking its values in a finite set $\Omega = \{\omega_1, ..., \omega_K\}$. We will say that $(X_{\mathcal{S}}, R_{\mathcal{S}}, Y_{\mathcal{S}})$ is a Triplet Markov Tree (TMT) if it is an MT. Since $R_{\mathcal{S}}$ monitors the changes of characteristics of the model, we will call it the "switching process" and the TMT involving such a process will be called a Markov-Switching Tree (MST). The aim of this last section is to extend the previous Bayesian filtering and prediction algorithms to some particular MST.

B. Exact filtering on PPMST

Let $X = \{X_s\}_{s \in \mathcal{S}}$, $Y = \{Y_s\}_{s \in \mathcal{S}}$ and $R = \{R_s\}_{s \in \mathcal{S}}$ be sets of random variables indexed by $\mathcal{S}$. Each X_s (resp. Y_s) takes its values in $\mathbb{R}^q$ (resp. $\mathbb{R}^m$) and R_s takes its values in $\Omega =$

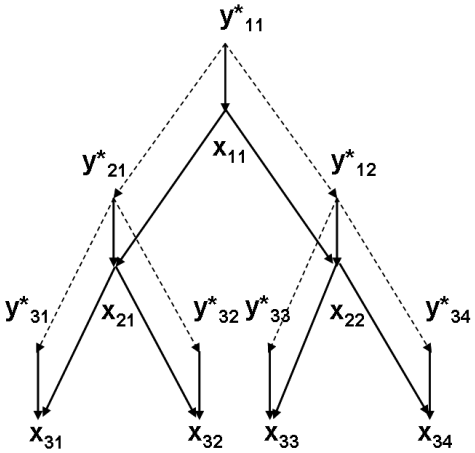

Fig. 2. Markov-Switching Tree, with $y^* = (r, y)$

$\{\omega_1, ..., \omega_K\}$. We consider the following particular PPMST (see Fig. 2) :

$$(R_{\mathcal{S}}, Y_{\mathcal{S}}) \text{ is a PPMT}, \tag{20}$$

$$X_s = F_s(R_s, Y_s)X_{s^-} + W_s, \tag{21}$$

where $\{W_s\}_{s\in\mathcal{S}}$ are independent zero-mean random vectors, such that for each $s \in \mathcal{S}$, W_s is independent from $(R_{\mathcal{S}}, Y_{\mathcal{S}})$ and from X_r. Again, note that in (21) vectors W_s are not necessarily Gaussian.

In this section we aim at computing $\mathrm{E}[X_s|Y_{a(s)} = y_{a(s)}]$ and $\mathrm{Cov}(X_s|Y_{a(s)} = y_{a(s)})$ for any $s \in \mathcal{S}$. As above we focus on the computation of $M_s(r_s, y_{a(s)})$ and $V_s(r_s, y_{a(s)})$ as defined in (8) and (9).

Proposition.

Let $(X_{\mathcal{S}}, R_{\mathcal{S}}, Y_{\mathcal{S}})$ satisfy (20)-(21), with given transition $p(r_s, y_s|r_{s^-}, y_{a(s^-)})$. Then $M_s(r_s, y_{a(s)})$ can be recursively computed with linear complexity in number of nodes by the following way:

$$M_s(r_s, y_{a(s)}) = \frac{1}{p(y_s|y_{a(s^-)})} \sum_{r_{s^-}} F_s(r_s, y_s) M_{s^-}(r_{s^-}, y_{a(s^-)}) p(r_s, y_s|r_{s^-}, y_{a(s^-)}), \tag{22}$$

with

$$p(y_s|y_{a(s^-)}) = \frac{p(y_{a(s)})}{p(y_{a(s^-)})} = \frac{\sum_{r_s} p(r_s, y_{a(s)})}{\sum_{r_{s^-}} p(r_{s^-}, y_{a(s^-)})}$$

and

$$p(r_s, y_{a(s)}) = \sum_{r_{s^-}} p(r_{s^-}, y_{a(s^-)}) p(r_s, y_s|r_{s^-}, y_{a(s^-)}).$$

Furthermore if the covariance matrix Σ_s of W_s exists for all $s \in \mathcal{S}$ then $V_s(r_s, y_{a(s)})$ can be computed as :

$$V_s(r_s, y_{a(s)}) = \frac{1}{p(y_s|y_{a(s^-)})} \sum_{r_{s^-}} [F_s(r_s, y_s) V_{s^-}(r_{s^-}, y_{a(s^-)}) F_s(r_s, y_s)^T + \Sigma_s] p(r_s, y_s|r_{s^-}, y_{a(s^-)}). \tag{23}$$

C. Exact prediction in PPMST

As in the case of Markov chains we have to introduce some new constraints on model (20)-(21) in order to solve the prediction problem. Let us thus consider the following particular PPMST :

$$R_{\mathcal{S}} \text{ is an MT;} \tag{24}$$

$$(R_{\mathcal{S}}, Y_{\mathcal{S}}) \text{ is an MT;} \tag{25}$$

$$X_s = F_s(R_s) X_{s^-} + W_s, \tag{26}$$

where $\{W_s\}_{s \in \mathcal{S}}$ are independent zero-mean random vectors, such that for each $s \in \mathcal{S}$, U_s is independent from $(R_{\mathcal{S}}, Y_{\mathcal{S}})$ and from X_r. In (26) (as in (21)) vectors W_s are not necessarily Gaussian.

In this section we aim at computing $\mathrm{E}[X_p|y_{a(s)}]$ and $\mathrm{Cov}(X_p|y_{a(s)})$ for any $s \in \mathcal{S}$ and $p \in s^{++}$.

Proposition.

Let $(X_{\mathcal{S}}, R_{\mathcal{S}}, Y_{\mathcal{S}})$ satisfy (24)-(26), with given transition $p(r_s, y_s|r_{s^-}, y_{a(s^-)})$ and $p(r_s|r_{s^-})$. Then $M_p(r_p, y_{a(s)})$ can be recursively computed with linear complexity in time by the following scheme:

- Compute $M_s(r_s, y_{a(s)})$ with the algorithm presented in the last section;
- For each $p \in s^{++}$ compute

$$M_p(r_p, y_{a(s)}) = \sum_{r_{p^-}} p(r_p|r_{p^-}) F_p(r_p) M_{p^-}(r_{p^-}, y_{a(s)}). \tag{27}$$

Furthermore if the covariance matrix Σ_s of W_s exists for all $s \in \mathcal{S}$, then $V_p(r_p, y_{a(s)})$ can be computed as follows:

- Compute $V_s(r_s, y_{a(s)})$ with the algorithm presented in the last section;
- For each $p \in s^{++}$ compute

$$V_p(r_p, y_{a(s)}) = \sum_{r_{p^-}} p(r_p|r_{p^-}) [F_p(r_p) V_{p^-}(r_{p^-}, y_{a(s)}) F_p(r_p)^T + \Sigma_p]. \tag{28}$$

Remark.

Let us finally remark that if one defines a partially Markovian tree by generations (as was done in [19]), then the tree can be considered as a generations-wise chain, in which the filtering and prediction algorithms of section II can be applied. However in that case the complexity, even if it remains linear in the number of nodes, becomes (at least in the case of a dyadic tree) exponential in the number of generations.

IV. CONCLUSION

In this paper we have proposed a class a dynamic stochastic models, namely PPMSC and PPMST, in which the optimal (from the quadratic loss viewpoint) filter and predictor can be computed *exactly* (in particular, without resorting to any Monte-Carlo procedure) and with a computational cost *linear* in the number of observations.

REFERENCES

[1] G. A. Ackerson and K. S. Fu. On state estimation in switching environments. *IEEE Transactions on Automatic Control*, AC-15(1):10–17, February 1970.

[2] J. K. Tugnait. Detection and estimation for abruptly changing systems. *Automatika*, 18(5):607–615.

[3] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, editors. *Estimation with Applications to Tracking and Navigation*. New York: John Wiley & Sons, 2001.

[4] H.A.P. Blom and Y. Bar-Shalom. The interacting multiple model algorithm for systems with markovian switching coefficients. *IEEE Transactions on Automatic Control*, 33:780–783, 1988.

[5] C. Andrieu, M. Davy, and A. Doucet. Efficient particle filtering for jump Markov systems. application to time-varying autoregressions. *IEEE Trans. on Signal Processing*, 51(7):1762–70, 2003.

[6] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman Filter - Particle filters for tracking applications*. Artech House, Boston, MA., 2004.

[7] O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models*. Springer-Verlag, 2005.

[8] W. Pieczynski, N. Abbassi, and M. Ben Mabrouk. Exact filtering and smoothing in Markov switching systems hidden with Gaussian long memory noise. In *XIII Conference Applied Stochastic Models and Data Analysis (ASMDA)*, Vilnius, Lithuania, June 30 - July 3 2009.

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

[10] N. Bardel and F. Desbouvries. Exact Bayesian prediction in non-Gaussian Markov-switching model. In *XIII Conference Applied Stochastic Models and Data Analysis (ASMDA)*, Vilnius, Lithuania, June 30 - July 3 2009.

[11] L. E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Stat.*, 37:1554–63, 1966.

[12] L. E. Baum and J. A. Eagon. An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a model for ecology. *Bull. Amer. Meteorol. Soc.*, 73:360–63, 1967.

[13] A. S. Willsky. Multiresolution Markov models for signal and image processing. *Proceedings of the IEEE*, 90(8):1396–1458, August 2002.

[14] F.M.J.-L. Starck and A. Bijaoui. *Multiscale Image Processing and Data Analysis*. Cambridge Univ. Press, 1998.

[15] K. C. Chou, A. S. Willsky, and A. Benveniste. Multiscale recursive estimation, data fusion, and regularization. *IEEE Transactions on Automatic Control*, 39(3):464–78, March 1994.

[16] J. M. Laferté, P. Perez, and F. Heitz. Discrete Markov image modeling and inference on the quadtree. *IEEE Transactions on Image Processing*, 9(3):390–404, March 2000.

[17] W. Pieczynski. Arbres de Markov couple - pairwise Markov trees. *Comptes Rendus de l'Académie des Sciences - Mathématiques*, 335:79–82, 2002. in French.

[18] F. Desbouvries J. Lecomte and W. Pieczynski. Kalman filtering in pairwise Markov trees. *Signal Processing*, 86(5):1049–1054, 2006.

[19] W. Pieczynski. Triplet partially Markov chains and trees. In *International Symposium on Image/Video Communications (ISIVC04)*, Brest, France, July 7-9 2004.