EXACT BAYESIAN RESTORATION IN NON-GAUSSIAN MARKOV-SWITCHING TREES

Noémie Bardel and François Desbouvries

Telecom SudParis / CITI department & CNRS UMR 5157
9, rue Charles Fourier, 91011 Evry, France
(e-mail: {noemie.bardel,francois.desbouvries}@it-sudparis.eu)

ABSTRACT. Multiresolution signal and image analysis and multiscale algorithms are of interest in many fields. In particular, efficient Bayesian restoration algorithms have been proposed for some tree-structured Markovian models. In this paper we show that Bayesian filtering and prediction can be performed exactly, with complexity linear in time index, in a particular class of Triplet Markov Trees.

1 INTRODUCTION

Multiresolution signal and image analysis and multiscale algorithms are of interest in many fields (Daubechies *et al.* (1992), Krim *et al.* (1999) and Starck and Bijaoui (1998)). In particular, efficient restoration algorithms in statistical models defined on Hidden Markov trees (HMT) have been developed (see e.g. Chou *et al.* (1994), Laferté *et al.* (2000) or Willsky (2002)).

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

$$p(x_{\mathcal{S}}) = p(x_r) \prod_{s \in \mathcal{S} \setminus \mathcal{S}_1} p(x_s | x_{s^-}). \tag{1}$$

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:

$$p(x_{\mathcal{S}}, y_{\mathcal{S}}) = p(x_r) \prod_{s \in \mathcal{S} \setminus \mathcal{S}_1} p(x_s | x_{s^-}) \prod_{s \in \mathcal{S}} p(y_s | x_{\mathcal{S}}), \tag{2}$$

i.e. x is an MT and $p(y_{\mathcal{S}} | x_{\mathcal{S}}) = \prod_{s \in \mathcal{S}} p(y_s | x_{\mathcal{S}})$. HMT have been extended to Pairwise Markov Trees (PMT) (see Pieczynski (2002) and Desbouvries *et al.* (2006)) defined by:

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

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.

We now introduce a third latent process $r_{\mathcal{S}}$ taking its values in a finite set $\Omega = \{\omega_1, ..., \omega_t\}$ which can monitor for example the change of characteristics of the model. We will say that $(x_{\mathcal{S}}, r_{\mathcal{S}}, y_{\mathcal{S}})$ is a Triplet Markov Tree (TMT) if it is an MT. Bayesian restoration of a hidden variable x_s from (some of the) observed variables $\{y_s\}$ is in general a difficult problem. For instance, as is well known (see Tugnait (1982)) Bayesian inference in Jump-Markov State-Space (JMSS) systems is an NP-hard problem. JMSS systems are conditionally linear and Gaussian dynamic systems, defined as:

$$\begin{aligned} x_{n+1} &= F_n(r_n)x_n + G_n(r_n)u_n \\ y_n &= H_n(r_n)x_n + v_n \end{aligned}$$

in which r_n is a Markov Chain, and $\{u_n\}_{n\in\{1,...,N\}}$ and $\{v_n\}_{n\in\{1,...,N\}}$ are independent and mutually independent zero-mean random vectors, and independent from $\{r_n\}_{n\in\{1,...,N\}}$ and x_0. Such a model is a particular Triplet Markov Chain (x_n, r_n, y_n), and thus a particular Triplet Markov Tree (TMT) (x_s, r_s, y_s) (a tree reduces to a chain if each node has exactly one child).

On the other hand, in most situations we are indeed more interested by some moment $E[g(x_k)|y_{1:n}]$ than by pdf $p(x_k|y_{1:n})$ itself. In particular, the conditional expectation $E[x_k|y_{1:n}]$ is of particular interest since it is the solution to the Bayesian estimation problem with quadratic loss.

The aim of this paper is to show that for some particular TMT Bayesian filtering (see section 2) and prediction (see section 3) can be performed with complexity linear in time index.

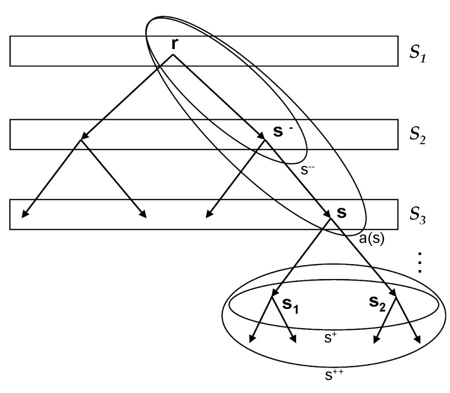

Figure 1. Example of dyadic tree

2 EXACT FILTERING ON SWITCHING-MARKOV TREES

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 $\Omega = \{\omega_1, ..., \omega_t\}$.

We consider the following particular TMT model(see Fig.2):

$$(R_{\mathcal{S}}, Y_{\mathcal{S}}) \text{ is a Markov Tree;} \tag{3}$$

$$X_s = F_{s^-}(R_{s^-}, Y_{s^-})X_{s^-} + U_{s^-}; \tag{4}$$

where $\{U_s\}_{s\in\mathcal{S}}$ 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. Note that in (4) vectors U_s are not necessarily Gaussian.

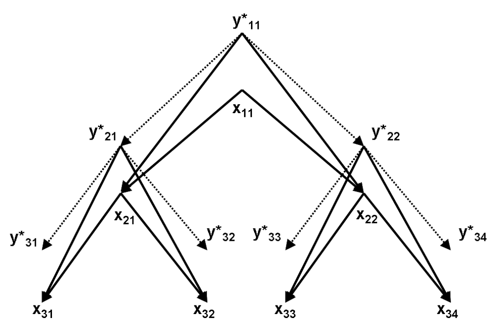

Figure 2. Switch-Markov Tree, with $y^* = (r, y)$

In this section we aim at computing $E[X_s|Y_{a(s)} = y_{a(s)}]$ and $Cov(X_s|Y_{a(s)} = y_{a(s)})$ for any $s \in \mathcal{S}$.

Notation

For each $p \in \mathcal{S}$ and $s \in \mathcal{S}$ let us first set:

$$M_p(r_p, y_{a(s)}) = \int_{\mathbb{R}^q} x_p p(x_p, r_p|y_{a(s)}) dx_p \tag{5}$$

If the covariance matrix Σ_p of U_p exists for all p, let us set:

$$V_p(r_p, y_{a(s)}) = \int_{\mathbb{R}^q} x_p {x_p}^T p(x_p, r_p|y_{a(s)}) dx_p \tag{6}$$

Of course, $E[X_s|Y_{a(s)} = y_{a(s)}]$ and $Cov(X_s|Y_{a(s)} = y_{a(s)})$ can be computed from $M_s(r_s, y_{a(s)})$ and $V_s(r_s, y_{a(s)})$ as:

$$E[X_s|Y_{a(s)} = y_{a(s)}] = \sum_{r_s} M_s(r_s, y_{a(s)}) \text{ and}$$

$$Cov(X_s|Y_{a(s)}=y_{a(s)})=\sum_{r_s}V_s(r_s,y_{a(s)})-(\sum_{r_s}M_s(r_s,y_{a(s)}))(\sum_{r_s}M_s(r_s,y_{a(s)}))^T.$$

In the following we thus focus on the computation of $M_s(r_s,y_{a(s)})$ and $V_s(r_s,y_{a(s)})$.

Proposition

Let $(X_{\mathcal{S}},R_{\mathcal{S}},Y_{\mathcal{S}})$ satisfy (3)-(4), with given transition $p(r_s,y_s|r_{s^-},y_{s^-})$. Then $M_s(r_s,y_{a(s)})$ can be recursively computed with linear complexity in time by the following way:

$$M_s(r_s,y_{a(s)})=\frac{1}{p(y_s|y_{a(s^-)})}\sum_{r_{s^-}}p(r_s,y_s|r_{s^-},y_{s^-})F_{s^-}(r_{s^-},y_{s^-})M_{s^-}(r_{s^-},y_{a(s^-)}) \tag{7}$$

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

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

Furthermore if the covariance matrix Σ_s of U_s exists for all $s\in\mathcal{S}$ it is possible to compute $V_s(r_s,y_{a(s)})$ as:

$$V_s(r_s,y_{a(s)})=\frac{1}{p(y_s|y_{a(s^-)})}\sum_{r_{s^-}}p(r_s,y_s|r_{s^-},y_{s^-})[F_{s^-}(r_{s^-},y_{s^-})V_{s^-}(r_{s^-},y_{a(s^-)})F_{s^-}(r_{s^-},y_{s^-})^T+\Sigma_{s^-}] \tag{8}$$

Proof

By using the Bayes formula, the fact that $(X_{\mathcal{S}},Y_{\mathcal{S}},R_{\mathcal{S}})$ is a Markov Tree and the model (3)-(4) we have:

$$\begin{aligned}
p(x_s,r_s|y_{a(s)}) &= \sum_{r_{s^-}}\int p(x_s,r_s,x_{s^-},r_{s^-}|y_{a(s^-)},y_s)dx_{s^-}\\
&= \frac{1}{p(y_s|y_{a(s^-)})}\sum_{r_{s^-}}\int p(x_s,r_s,x_{s^-},r_{s^-},y_s|y_{a(s^-)})dx_{s^-}\\
&= \frac{1}{p(y_s|y_{a(s^-)})}\sum_{r_{s^-}}\int p(x_{s^-},r_{s^-}|y_{a(s^-)})p(r_s,y_s|r_{s^-},y_{s^-})p(x_s|x_{s^-},r_{s^-},y_{s^-})dx_{s^-}
\end{aligned} \tag{9}$$

We next multiply (9) by x_s and integrate with respect to x_s. Since $\{U_s\}$ are independent, zero-mean and independent from $(R_{\mathcal{S}},Y_{\mathcal{S}})$:

$$M_s(r_s,y_{a(s)})=\frac{1}{p(y_s|y_{a(s^-)})}\sum_{r_{s^-}}\int p(x_{s^-},r_{s^-}|y_{a(s^-)})p(r_s,y_s|r_{s^-},y_{s^-})F_{s^-}(r_{s^-},y_{s^-})x_{s^-}dx_{s^-}$$

Finally:

$$M_s(r_s, y_{a(s)}) = \frac{1}{p(y_s|y_{a(s^-)})} \sum_{r_{s^-}} p(r_s, y_s|r_{s^-}, y_{s^-}) F_{s^-}(r_{s^-}, y_{s^-}) \int x_{s^-} p(x_{s^-}, r_{s^-}|y_{a(s^-)}) dx_{s^-}$$
$$= \frac{1}{p(y_s|y_{a(s^-)})} \sum_{r_{s^-}} p(r_s, y_s|r_{s^-}, y_{s^-}) F_{s^-}(r_{s^-}, y_{s^-}) M_{s^-}(r_{s^-}, y_{a(s^-)})$$

which brings us to the end of the proof. (8) is obtained similarly.

3 EXACT PREDICTION IN SWITCHING-MARKOV TREES

We consider the following particular TMT:

$$R_{\mathcal{S}} \text{ is a Markov Tree;} \tag{10}$$
$$(R_{\mathcal{S}}, Y_{\mathcal{S}}) \text{ is a Markov Tree;} \tag{11}$$
$$X_s = F_{s^-}(R_{s^-}) X_{s^-} + W_{s^-}; \tag{12}$$

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. Note that in (12) (as in (4)) vectors W_s are not necessarily Gaussian.

In this section we aim at computing $E[x_p|y_{a(s)}]$ and $Cov(x_p|y_{a(s)})$ for any $s \in \mathcal{S}$ and $p \in s^{++}$. As above we focus on the computation of $M_p(r_p, y_{a(s)})$ and $V_p(r_p, y_{a(s)})$, defined in (5) and (6).

Proposition

Let $(X_{\mathcal{S}}, R_{\mathcal{S}}, Y_{\mathcal{S}})$ satisfy (10)-(12), with given transition $p(r_s, y_s|r_{s^-}, y_{s^-})$. 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

$$M_p(r_p, y_{a(s)}) = \sum_{r_{p^-}} p(r_p|r_{p^-}) F_{p^-}(r_{p^-}) M_{p^-}(r_{p^-}, y_{a(s)}). \tag{13}$$

Furthermore if the covariance matrix Σ_s of W_s exists for all $s \in \mathcal{S}$ it is possible to compute $V_p(r_p, y_{a(s)})$ as follows:

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

$$V_p(r_p, y_{a(s)}) = \sum_{r_{p^-}} p(r_p|r_{p^-}) [F_{p^-}(r_{p^-}) V_{p^-}(r_{p^-}, y_{a(s)}) F_{p^-}(r_{p^-})^T + \Sigma_{p^-}]. \tag{14}$$

Proof

We have:

$$p(x_p, r_p | y_{a(s)}) = \sum_{r_{p^-}} \int p(x_p, r_p, x_{p^-}, r_{p^-} | y_{a(s)}) dx_{p^-}$$

$$= \sum_{r_{p^-}} \int p(x_{p^-}, r_{p^-} | y_{a(s)}) p(x_p, r_p | x_{p^-}, r_{p^-}, y_{a(s)}) dx_{p^-}$$

$$= \sum_{r_{p^-}} \int p(x_{p^-}, r_{p^-} | y_{a(s)}) p(x_p | r_p, x_{p^-}, r_{p^-}, y_{a(s)}) p(r_p | x_{p^-}, r_{p^-}, y_{a(s)}) dx_{p^-} \qquad (15)$$

Then from (10) and (12), $p(x_p | r_p, x_{p^-}, r_{p^-}, y_{a(s)})$ reduces to $p(x_p | x_{p^-}, r_{p^-})$ and $p(r_p | x_{p^-}, r_{p^-}, y_{a(s)})$ reduces to $p(r_p | r_{p^-})$. We next multiply (15) by x_p and integrate with respect to x_p to get (13). (14) is obtained similarly.

REFERENCES

BARDEL, N., DESBOUVRIES, F. (2009): Exact bayesian prediction in non-Gaussian Markov-switching model, *Proceedings of the XIIIth International Conference on Applied Stochastic Models and Data Analysis (ASMDA)*, June 30 - July 3, Vilnius, Lithuania.

CHOU, K.C., WILLSKY, A.S., BENVENISTE, A. (1994): Multiscale recursive estimation, data fusion, and regularization, *IEEE Trans. Autom. Control*, vol. 39, pp. 464-478.

DAUBECHIES, I., MALLAT, S.M., WILLSKY, A.S. (1992): Special issue on wavelet transforms and multiresolution signal analysis (Eds.), *IEEE Trans. Inform. Theory*, vol. 38, pp. 529-860.

DESBOUVRIES, F., LECOMTE, J., PIECZYNSKI, W. (2006): Kalman Filtering in pairwise Markov trees, *Signal Processing*, vol. 86, Number 5, pp. 1049-1054.

KRIM, H., WILLINGER, W., JUDITSKI, A., TSE, D. (1999): Special issue on multiscale statistical signal analysis and its applications (Eds.), *IEEE Trans. Inform. Theory*, vol. 45, pp. 825-1062.

LAFERTE, J.M., PEREZ, P., HEITZ, F. (2000): Discrete Markov image modeling and inference on the quadtree, *IEEE Trans. Image Process.*, vol. 9, pp. 390-404.

PIECZYNSKI, W. (2002): Arbres de Markov Couple - Pairwise Markov Trees, *C.R.A.S - Mathématiques*, Vol. 335, pp. 79-82.

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.

STARCK, F.M.J.-L., BIJAOUI, A. (1998): Multiscale Image Processing and Data Analysis. Cambridge, U.K.: Cambridge Univ. Press.

TUGNAIT, J.K. (1982): Adaptive estimation and identification for discrete systems with Markov jump parameters, *IEEE Transactions on Automatic Control*, vol.25, pp. 1054-1065.

WILLSKY, A.S. (2002): Multiresolution Markov models for signal and image processing, *Proc. IEEE*, vol. 90, no.8, pp. 1396-1458.