Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Signal Processing 86 (2006) 1049–1054

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

Kalman filtering in pairwise Markov trees

François Desbouvries*, Jean Lecomte, Wojciech Pieczynski

GET/INT/dépt. CITI and CNRS UMR 5157, 9 rue Charles Fourier, 91011 Evry, France

Received 23 December 2004; accepted 23 July 2005
Available online 30 August 2005

Abstract

An important problem in multiresolution analysis of signals or images consists in estimating hidden random variables $\mathbf{x} = \{\mathbf{x}_s\}_{s\in S}$ from observed ones $\mathbf{y} = \{\mathbf{y}_s\}_{s\in S}$. This is done classically in the context of hidden Markov trees (HMT). HMT have been extended recently to the more general context of pairwise Markov trees (PMT). In this note, we propose an adaptive filtering algorithm which is an extension to PMT of the Kalman filter (KF).

Keywords: Hidden Markov trees; Pairwise Markov trees; Multiresolution signal and image analysis; Multiscale algorithms; Kalman filtering

1. Introduction

An important problem in signal and image processing consists in recursively estimating a set of hidden variables $\mathbf{x} = \{\mathbf{x}_s\}_{s\in S}$ from a set of observed variables $\mathbf{y} = \{\mathbf{y}_s\}_{s\in S}$. To that end, Bayesian estimation algorithms have been developed, mainly in the framework of hidden Markov models (HMM).

Now, it is well-known that if $(\mathbf{x}, \mathbf{y})$ is a classical HMM, then the pair $(\mathbf{x}, \mathbf{y})$ itself is Markovian. Conversely, starting from the sole assumption that $(\mathbf{x}, \mathbf{y})$ is a Markov chain (MC), i.e. that $(\mathbf{x}, \mathbf{y})$ is a so-called pairwise Markov model (PMM), is a more general point of view which nevertheless enables the development of similar restoration algorithms. More precisely, some of the classical Bayesian restoration algorithms used in hidden Markov fields (HMF), hidden Markov chains (HMC) or hidden Markov trees (HMT), have been generalized recently to the frameworks of pairwise Markov fields (PMF) [1], pairwise Markov chains (PMC) with discrete [2] or continuous [3,4] hidden process, and of pairwise Markov trees (PMT) with discrete [5,6] or continuous [6,7] hidden variables.

Let us turn back with more details to some estimation algorithms used in HMM. In the

*Corresponding author. Tel.: +33 1 60 76 45 27; fax: +33 1 60 76 44 33.
E-mail addresses: Francois.Desbouvries@int-evry.fr (F. Desbouvries), jean.lecomte@m4x.org (J. Lecomte), Wojciech.Pieczynski@int-evry.fr (W. Pieczynski).

0165-1684/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2005.07.026

context of linear Gaussian HMC, a classical adaptive filtering solution is provided by the celebrated KF. This method was first introduced in the control engineering community [8]. Since then, the KF has been extended in various directions and a rich family of estimation algorithms have been developed (see e.g. [9,10]). The KF has now become a major tool in signal processing and automatic control [10] as well as in the statistical and time series communities (see e.g. [11–14]).

Efficient restoration algorithms have also been developed in the context of HMT (see e.g. [15–18], as well as the tutorial [19]). These smoothing-like algorithms enable to compute the conditional law $p(\mathbf{x}_s|\{\mathbf{y}_\sigma\}_{\sigma\in S})$ of a hidden variable $\mathbf{x}_s$ at an arbitrary node $s \in S$, given all observations $\{\mathbf{y}_\sigma\}_{\sigma\in S}$. In particular, the algorithm which was developed in [17] in the context of Gaussian HMT is a two-step procedure: firstly, a filtering sweep in the backward (fine-to-coarse) direction computes the conditional law of the root node $\mathbf{x}_r$ given $\{\mathbf{y}_\sigma\}_{\sigma\in S}$; and then a smoothing sweep in the forward (coarse-to-fine) direction eventually computes $p(\mathbf{x}_s|\{\mathbf{y}_\sigma\}_{\sigma\in S})$, via a computational procedure which iterates along the path relating the root node r to node s (see Fig. 1).

In this note, we will see that in the cases where the laws of interest are the conditional laws of the last generation leaves, these two sweeps can indeed be replaced by just one (block) filtering sweep in the forward (coarse-to-fine) direction. Such a procedure is feasible because in an MT the successive subsets of variables belonging to a given generation S_n (see Fig. 1) form an MC, and one can thus adapt to tree structures the KF which originally was derived for HMCs. In general, the algorithm we get is no longer linear (it however remains polynomial) in the number of nodes. However, such a filtering solution presents the advantage of being adaptive, and is thus well-suited to situations where the observations become available progressively, generation after generation.

In this note, we thus extend the KF to PMT. More precisely, the algorithm we obtain is an extension to PMT of the algorithm derived in [3] in the context of PMC, which itself was an extension of a particular Kalman-like algorithm. The rest of this note is organized as follows. In Section 2 we briefly recall some embedded Markovian models that are used in tree-based structures. An extension to the PMT model of the KF is given in Section 3.

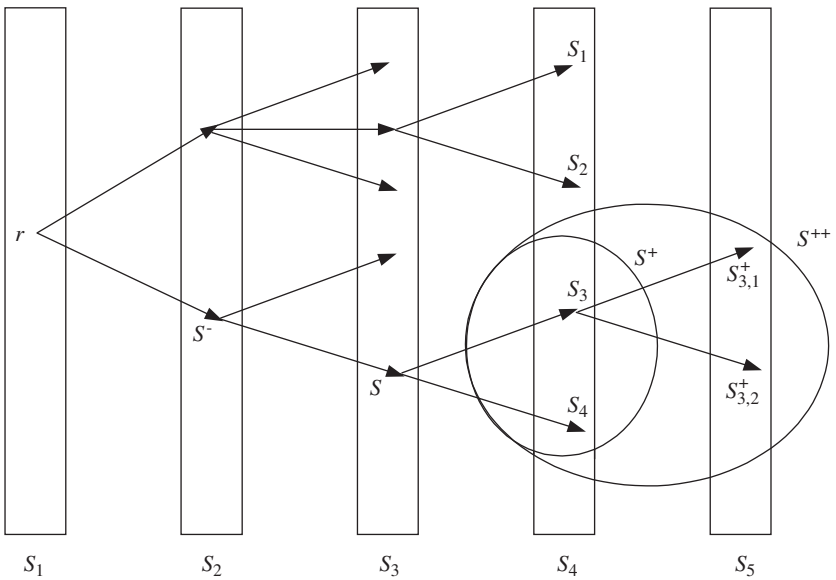

Fig. 1. The tree structure.

2. Hidden vs. pairwise Markov trees

Let S be a finite set of indices, and let us consider a tree structure with nodes indexed on S. Let us partition S in terms of its successive generations $S_1, \ldots, S_N$. So, S_1 is made of the root node r, S_2 gathers the children of node r, and so on. Each node s (apart from the root node r) has one father denoted by s^-. The children of a node s are denoted by s^+, and the set of all descendants of a node s is denoted by s^{++} (see Fig. 1).

Let now $\mathbf{x} = \{\mathbf{x}_s\}_{s\in S}$ and $\mathbf{y} = \{\mathbf{y}_s\}_{s\in S}$ be two sets of random variables indexed on S. Each $\mathbf{x}_s$ (resp. $\mathbf{y}_s$) belongs to $\mathbb{R}^p$ (resp. to $\mathbb{R}^q$). Let $p(\mathbf{x}_s)$ (resp. $p(\mathbf{y}_s)$) denote the probability density function (pdf) of $\mathbf{x}_s$ (resp. of $\mathbf{y}_s$) w.r.t. Lebesgue measure, and let $p(\mathbf{x}_s|\{\mathbf{y}_\sigma\}_{\sigma\in\Sigma})$ denote the conditional pdf of $\mathbf{x}_s$ given $\{\mathbf{y}_\sigma\}_{\sigma\in\Sigma}$. Other pdf's or conditional pdf's of interest are defined similarly.

The classical HMT model is widely used for modeling $p(\mathbf{x}, \mathbf{y})$. In this model, $\mathbf{x}$ is a Markov tree (MT), and conditionally on $\mathbf{x}$, the variables $\mathbf{y}_s$ are independent and satisfy $p(\mathbf{y}_s|\mathbf{x}) = p(\mathbf{y}_s|\mathbf{x}_s)$. For reasons to become clear below, in the sequel we will no longer call such a model an HMT, but rather a hidden Markov tree with independent noise (HMT-IN). In an HMT-IN the pdf of the pair $(\mathbf{x}, \mathbf{y})$ can thus be written as

$$p(\mathbf{x},\mathbf{y}) = \underbrace{p(\mathbf{x}_r)\prod_{i=2}^{N}\prod_{s\in S_i} p(\mathbf{x}_s|\mathbf{x}_{s^-})}_{p(\mathbf{x})} \times \underbrace{\prod_{s\in S} p(\mathbf{y}_s|\mathbf{x}_s)}_{p(\mathbf{y}|\mathbf{x})}. \tag{1}$$

Now, let us introduce the pair $\mathbf{z}_s = (\mathbf{x}_s, \mathbf{y}_s)$, and let $\mathbf{z} = \{\mathbf{z}_s\}_{s\in S}$. A PMT is a model in which we only assume that $\mathbf{z}$ is an MT:

$$p(\mathbf{z}) = p(\mathbf{z}_r)\prod_{i=2}^{N}\prod_{s\in S_i} p(\mathbf{z}_s|\mathbf{z}_{s^-}). \tag{2}$$

One can check easily that (1) implies (2), so any HMT-IN is a PMT. However, the converse is not true. More precisely, in a PMT the transition pdf $p(\mathbf{z}_s|\mathbf{z}_{s^-})$ reads

$$\begin{aligned} p(\mathbf{z}_s|\mathbf{z}_{s^-}) &= p(\mathbf{x}_s,\mathbf{y}_s|\mathbf{x}_{s^-},\mathbf{y}_{s^-}) \\ &= p(\mathbf{x}_s|\mathbf{x}_{s^-},\mathbf{y}_{s^-})p(\mathbf{y}_s|\mathbf{x}_s,\mathbf{x}_{s^-},\mathbf{y}_{s^-}); \end{aligned}$$

so an HMT-IN is a particular PMT in which $p(\mathbf{x}_s|\mathbf{x}_{s^-},\mathbf{y}_{s^-})$ reduces to $p(\mathbf{x}_s|\mathbf{x}_{s^-})$ and $p(\mathbf{y}_s|\mathbf{x}_s,\mathbf{x}_{s^-},\mathbf{y}_{s^-})$ to $p(\mathbf{y}_s|\mathbf{x}_s)$. These simplifications are rather rough, and we can see that a lot of information is lost when making use of an HMT-IN rather than of a PMT.

Let us also mention an intermediate model, which we call a hidden Markov tree (HMT), in which both $\mathbf{x}$ and $(\mathbf{x}, \mathbf{y})$ are MT but the observations $\mathbf{y}_s$ are not necessarily independent conditionally on $\mathbf{x}$. Of course, any HMT-IN is an HMT, and any HMT is a PMT. However, PMT are more general than HMT, because if (2) holds, $\mathbf{x}$ is not necessarily an MT, as we see from the following result, proved in [7] (an analogous result for the case where $\mathbf{x}$ is discrete can be found in [5]):

Proposition 1. *Let* $\mathbf{z}$ *be a PMT satisfying* (2). *Assume that the tree is dyadic*, (*i.e.*, *that each node* $s \in S\backslash S_n$ *has exactly two children* s_1 *and* s_2), *and that*

$$\text{For all } s \in S\backslash S_1, \quad p(\mathbf{x}_s|\mathbf{x}_{s^-},\mathbf{y}_{s^-}) = p(\mathbf{x}_s|\mathbf{x}_{s^-}). \tag{3}$$

Then $\mathbf{x}$ *is an MT. Conversely*, *assume that* $\mathbf{z}$ *and* $\mathbf{x}$ *are MT*, *and that for all* $s \in S\backslash S_n$, $p(\mathbf{z}_{s_1}|\mathbf{z}_s) = p(\mathbf{z}_{s_2}|\mathbf{z}_s)$, *i.e. that conditionally on the father*, *the laws of the children are equal. Then* (3) *holds.*

3. Linear Gaussian PMT and Kalman filtering

In this section, we develop a Kalman-like adaptive filtering algorithm for PMT. To that end, we gather all variables $\mathbf{x}_s$ belonging to a same level S_n into a vector $\mathbf{X}_n$. Let also $\mathbf{X}_{1:n} \overset{\text{def}}{=} (\mathbf{X}_1, \ldots, \mathbf{X}_n)$, and let $\mathbf{Y}_n$, $\mathbf{Y}_{1:n}$, $\mathbf{Z}_n$ and $\mathbf{Z}_{1:n}$ be defined similarly. Since $\mathbf{z} = \{\mathbf{z}_s\}_{s\in S}$ is an MT, the time-varying sequence $\{\mathbf{Z}_n\}_{1\leqslant n\leqslant N}$ is an MC. This observation enables us to adapt to PMT the Kalman filtering methodology which is valid in the context of HMC. More precisely, our aim consists in recursively estimating (as new data become available) the pdf of the last leaves $\mathbf{X}_n$ given all observed variables up to level n, i.e. we want to compute $p(\mathbf{X}_{n+1}|\mathbf{Y}_{1:n+1})$ in terms of $p(\mathbf{X}_n|\mathbf{Y}_{1:n})$ and of $\mathbf{Y}_{n+1}$.

Our hypotheses are as follows. We assume that the model is linear and Gaussian[1]:

$$\underbrace{\begin{bmatrix}\mathbf{x}_s\\ \mathbf{y}_s\end{bmatrix}}_{\mathbf{z}_s}=\underbrace{\begin{bmatrix}\mathbf{F}_s^1 & \mathbf{F}_s^2\\ \mathbf{H}_s^1 & \mathbf{H}_s^2\end{bmatrix}}_{\mathbf{F}_s}\begin{bmatrix}\mathbf{x}_{s^-}\\ \mathbf{y}_{s^-}\end{bmatrix}+\underbrace{\begin{bmatrix}\mathbf{u}_s\\ \mathbf{v}_s\end{bmatrix}}_{\mathbf{w}_s}, \tag{4}$$

in which $\{\mathbf{w}_s\}_{s\in S\setminus S_1}$ are random vectors which are zero-mean, independent and independent of $\mathbf{z}_r$. We also assume that the process $\mathbf{w}=\{\mathbf{w}_s\}_{s\in S\setminus S_1}$ is Gaussian and that $p(\mathbf{z}_r)\sim N(\overline{\mathbf{z}}_r,\mathbf{P}_r)$. Then $\mathbf{Z}$ is Gaussian and consequently the pdf $p(\mathbf{X}_n|\mathbf{Y}_{1:n})$ and $p(\mathbf{X}_{n+1}|\mathbf{Y}_{1:n})$ are also Gaussian. Let us set

$$p(\mathbf{X}_n|\mathbf{Y}_{1:n})\sim N(\widehat{\mathbf{X}}_{n|n},\mathbf{P}_{n|n}), \tag{5}$$

$$p(\mathbf{X}_{n+1}|\mathbf{Y}_{1:n})\sim N(\widehat{\mathbf{X}}_{n+1|n},\mathbf{P}_{n+1|n}), \tag{6}$$

and let

$$E(\mathbf{w}_s\mathbf{w}_s^{\mathrm{T}})=\mathbf{Q}_s=\begin{bmatrix}\mathbf{Q}_s^{11} & \mathbf{Q}_s^{12}\\ \mathbf{Q}_s^{21} & \mathbf{Q}_s^{22}\end{bmatrix}. \tag{7}$$

We shall also need the following notation (see Fig. 1): for n fixed, let $S_n=(s_1,\ldots,s_{k(n)})$, and let $s_i^+=\{s_{i,p}^+\}_{p=1}^{j(i)}$ (i.e., $s_{i,p}^+$ is the pth son of node s_i). For $l,m\in\{1,2\}$, let

$$\mathbf{F}_{s_i^+}^l=\begin{bmatrix}\mathbf{F}_{s_{i,1}^+}^l\\ \vdots\\ \mathbf{F}_{s_{i,j(i)}^+}^l\end{bmatrix},\quad \mathbf{H}_{s_i^+}^l=\begin{bmatrix}\mathbf{H}_{s_{i,1}^+}^l\\ \vdots\\ \mathbf{H}_{s_{i,j(i)}^+}^l\end{bmatrix}\quad\text{and}$$

$$\mathbf{Q}_{s_i^+}^{l,m}=\begin{bmatrix}\mathbf{Q}_{s_{i,1}^+}^{l,m} & & \mathbf{0}\\ & \ddots & \\ \mathbf{0} & & \mathbf{Q}_{s_{i,j(i)}^+}^{l,m}\end{bmatrix}, \tag{8}$$

and let F_{n+1}^l, H_{n+1}^l and $Q_{n+1}^{l,m}$ be the following block-diagonal matrices:

$$F_{n+1}^l=\mathrm{diag}(\mathbf{F}_{s_1^+}^l,\ldots,\mathbf{F}_{s_{k(n)}^+}^l), \tag{9}$$

$$H_{n+1}^l=\mathrm{diag}(\mathbf{H}_{s_1^+}^l,\ldots,\mathbf{H}_{s_{k(n)}^+}^l), \tag{10}$$

$$Q_{n+1}^{l,m}=\mathrm{diag}(\mathbf{Q}_{s_1^+}^{l,m},\ldots,\mathbf{Q}_{s_{k(n)}^+}^{l,m}). \tag{11}$$

The following algorithm is an extension of the classical KF [8,9]:

Proposition 2 (*KF for PMT*). *Let us assume that* $\mathbf{z}$ *is a PMT and that model* (4) *holds. Suppose that* $p(\mathbf{z}_r)\sim N(\overline{\mathbf{z}}_r,\mathbf{P}_r)$ *and that* $p(\mathbf{w}_s)\sim N(\mathbf{0},\mathbf{Q}_s)$ *for* $s\in S\setminus S_1$. *Assume that*

$$\mathbf{P}_r \text{ and } \{\mathbf{Q}_s\}_{s\in S} \text{ are positive definite.} \tag{12}$$

Then $(\widehat{\mathbf{X}}_{n+1|n+1},\mathbf{P}_{n+1|n+1})$ *can be computed from* $(\widehat{\mathbf{X}}_{n|n},\mathbf{P}_{n|n})$ *and* $\mathbf{Y}_{n+1}$ *via*:

Time-update equations:

$$\widehat{\mathbf{X}}_{n+1|n}=F_{n+1}^1\widehat{\mathbf{X}}_{n|n}+F_{n+1}^2\mathbf{Y}_n, \tag{13}$$

$$\mathbf{P}_{n+1|n}=Q_{n+1}^{11}+F_{n+1}^1\mathbf{P}_{n|n}(F_{n+1}^1)^{\mathrm{T}}. \tag{14}$$

Measurement-update equations:

$$\widetilde{\mathbf{Y}}_{n+1}=\mathbf{Y}_{n+1}-H_{n+1}^1\widehat{\mathbf{X}}_{n|n}-H_{n+1}^2\mathbf{Y}_n, \tag{15}$$

$$\mathbf{L}_{n+1}=Q_{n+1}^{22}+H_{n+1}^1\mathbf{P}_{n|n}(H_{n+1}^1)^{\mathrm{T}}, \tag{16}$$

$$\mathbf{K}_{n+1|n+1}=(Q_{n+1}^{12}+F_{n+1}^1\mathbf{P}_{n|n}(H_{n+1}^1)^{\mathrm{T}})\mathbf{L}_{n+1}^{-1}, \tag{17}$$

$$\widehat{\mathbf{X}}_{n+1|n+1}=\widehat{\mathbf{X}}_{n+1|n}+\mathbf{K}_{n+1|n+1}\widetilde{\mathbf{Y}}_{n+1}, \tag{18}$$

$$\mathbf{P}_{n+1|n+1}=\mathbf{P}_{n+1|n}-\mathbf{K}_{n+1|n+1}\mathbf{L}_{n+1}\mathbf{K}_{n+1|n+1}^{\mathrm{T}}, \tag{19}$$

Proof 1. Since $\mathbf{z}=\{\mathbf{z}_s\}_{s\in S}$ is a PMT, the time-varying sequence $\{\mathbf{Z}_n\}_{1\leqslant n\leqslant N}$ is an MC. So

$$\begin{aligned}p(\mathbf{X}_{n+1},\mathbf{Y}_{n+1}|\mathbf{Y}_{1:n})&=\int p(\mathbf{X}_{n+1},\mathbf{Y}_{n+1}|\mathbf{X}_n,\mathbf{Y}_{1:n})p(\mathbf{X}_n|\mathbf{Y}_{1:n})\,\mathrm{d}\mathbf{X}_n\\&=\int p(\mathbf{X}_{n+1},\mathbf{Y}_{n+1}|\mathbf{X}_n,\mathbf{Y}_n)p(\mathbf{X}_n|\mathbf{Y}_{1:n})\,\mathrm{d}\mathbf{X}_n.\end{aligned} \tag{20}$$

On the other hand, from (4) and (7) we get

$$p(\mathbf{z}_s|\mathbf{z}_{s^-})\sim N(\mathbf{F}_s\mathbf{z}_{s^-},\mathbf{Q}_s). \tag{21}$$

[1]Our algorithm could of course also be obtained as a recursive linear minimum mean square error restoration procedure; we chose to adopt the Gaussian point of view because the proofs are obtained in a simpler and more direct way.

Since $\mathbf{z}$ is a PMT,

$$p(\mathbf{Z}_{n+1}|\mathbf{Z}_n) = p(\mathbf{z}_{s_1^+},\ldots,\mathbf{z}_{s_{k(n)}^+}|\mathbf{z}_{s_1},\ldots,\mathbf{z}_{s_{k(n)}}) = \prod_{i=1}^{k(n)}\prod_{p=1}^{j(i)} p(\mathbf{z}_{s_{i,p}^+}|\mathbf{z}_{s_1},\ldots,\mathbf{z}_{s_{k(n)}}) = \prod_{i=1}^{k(n)}\prod_{p=1}^{j(i)} p(\mathbf{z}_{s_{i,p}^+}|\mathbf{z}_{s_i}). \quad (22)$$

Injecting (22), (21) and (5) into (20), and using Proposition 4 (see the Appendix), we get

$$p(\mathbf{X}_{n+1},\mathbf{Y}_{n+1}|\mathbf{Y}_{1:n}) \sim N\left(\begin{bmatrix} F_{n+1}^1\widehat{\mathbf{X}}_{n|n} + F_{n+1}^2\mathbf{Y}_n \\ H_{n+1}^1\widehat{\mathbf{X}}_{n|n} + H_{n+1}^2\mathbf{Y}_n \end{bmatrix}, \begin{bmatrix} Q_{n+1}^{1,1} & Q_{n+1}^{1,2} \\ Q_{n+1}^{2,1} & Q_{n+1}^{2,2} \end{bmatrix} + \begin{bmatrix} \widetilde{F}_{n+1}^1 \\ \widetilde{H}_{n+1}^1 \end{bmatrix} \mathbf{P}_{n|n}[(\widetilde{F}_{n+1}^1)^{\mathrm{T}} \ (\widetilde{H}_{n+1}^1)^{\mathrm{T}}]\right), \quad (23)$$

whence (13) and (14). Lastly, by conditioning with respect to $\mathbf{Y}_{n+1}$ (see Proposition 3) we get (15)–(19) (note that condition (12) is a simple sufficient condition ensuring that all equations are valid). □

Remarks.

- It was implicitly assumed in the proof that each node has at least one child. In case some node(s) has(ve) no child, one can check easily that Eqs. (13)–(19) still hold, provided however that when defining the internal variables, only the reduced set of nodes which have at least one child are taken into account. More precisely, let $\{s_{\alpha_i}\}$ be the subset of $\{s_i\}_{i=1}^{k(n)}$ made of the indices of the nodes which have at least one child. Then, F_{n+1}^l (and similarly H_{n+1}^l) should be replaced by $F_{n+1}^{l,\mathrm{red.}} = \mathrm{diag}(\mathbf{F}_{s_{\alpha_i}^+}^l)$, $\widehat{\mathbf{X}}_n = [E(\mathbf{x}_{s_i}|\mathbf{Y}_{1:n})]_{i=1}^{k(n)}$ by $\widehat{\mathbf{X}}_n^{\mathrm{red.}} = [E(\mathbf{x}_{s_{\alpha_i}}|\mathbf{Y}_{1:n})]$ (and consequently $\mathbf{P}_{n|n}$ by $\mathbf{P}_{n|n}^{\mathrm{red.}}$), and $\mathbf{Y}_n = [\mathbf{y}_{s_i}]_{i=1}^{k(n)}$ by $\mathbf{Y}_n^{\mathrm{red.}} = [\mathbf{y}_{s_{\alpha_i}}]$.
- If each node (beginning with root node r) has exactly one child, then the PMT reduces to a particular case of the PMC Model introduced in [20] (see also [3, Corollary 1, p. 72]), and the algorithm of Proposition 1 reduces to the algorithm of Lipster and Shiryaev (see [3, Eqs. (13.56) and (13.57)]) which has been developed for that model.
- The algorithm is valid in the general case where each node s has an arbitrary number of children, and is well-suited to asymmetric tree structures since the number of children may vary from node to node.
- The algorithm requires the inversion of the square matrix $\mathbf{L}_{n+1}$ defined in (16), the dimension of which is proportional to the number of variables in generation $n+1$ of the tree. This can become a severe computational bottleneck in trees where each node has at least two children, in which case the number of variables in S_n grows exponentially with n. However, this full-size matrix inversion can easily be avoided by conditioning w.r.t. each variable in $\mathbf{Y}_{n+1}$ one after the other, which is another aspect (but now within the last generation) of the adaptive character of the algorithm; in terms of matrix computations, this amounts to sequentially using the quotient property of Schur complements, see e.g. [21]. More precisely, we can compute $p(\mathbf{X}_n|\mathbf{Y}_{1:n})$ from $p(\mathbf{X}_n|\mathbf{Y}_{1:n-1})$ by conditioning w.r.t. each variable $\mathbf{y}_{s_i}$ in $\mathbf{Y}_n = \{\mathbf{y}_{s_i}\}_{i=1}^{k(n)}$, one after the other: we first compute $p(\mathbf{X}_n|\mathbf{Y}_{1:n-1},\mathbf{y}_{s_1})$ from $p(\mathbf{X}_n|\mathbf{Y}_{1:n-1})$, then incorporate the measure $\mathbf{y}_{s_2}$ by computing $p(\mathbf{X}_n|\mathbf{Y}_{1:n-1},\mathbf{y}_{s_1},\mathbf{y}_{s_2})$ from $p(\mathbf{X}_n|\mathbf{Y}_{1:n-1},\mathbf{y}_{s_1})$, and so on until we get $p(\mathbf{X}_n|\mathbf{Y}_{1:n-1},\{\mathbf{y}_{s_i}\}_{i=1}^{k(n)}) = p(\mathbf{X}_n|\mathbf{Y}_{1:n})$. If this procedure is used, the inversion of a $k(n)\times n_{\mathbf{y}}$ square matrix is replaced by the inversion of $k(n)$ $n_{\mathbf{y}}\times n_{\mathbf{y}}$ square matrices (assuming a constant size $n_{\mathbf{y}}$ of the random vectors $\mathbf{y}_{s_i}$), which is of interest, in particular in the case of large $k(n)$ and small $n_{\mathbf{y}}$.

Appendix

The following two properties of Gaussian random variables are recalled for convenience of the reader.

Proposition 3. *Let* $p(\mathbf{u}_1,\mathbf{u}_2)\sim N(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\Sigma_{1,1} & \Sigma_{1,2}\\ \Sigma_{2,1} & \Sigma_{2,2}\end{bmatrix})$. *Then* $p(\mathbf{u}_1|\mathbf{u}_2)\sim N(\mu_{1|2},\Sigma_{1|2})$, *with* $\mu_{1|2} = \mu_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(\mathbf{u}_2-\mu_2)$, $\Sigma_{1|2} = \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}$.

Proposition 4. *Let* $p(\mathbf{u}_1) \sim N(\mu_1, \boldsymbol{\Sigma}_1)$ *and* $p(\mathbf{u}_2|\mathbf{u}_1) \sim N(\mathbf{A}\mathbf{u}_1 + \mathbf{b}, \boldsymbol{\Sigma}_{2|1})$. *Then*

$$p(\mathbf{u}_1, \mathbf{u}_2) \sim N\left(\begin{bmatrix} \mu_1 \\ \mathbf{A}\mu_1 + \mathbf{b} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_1\mathbf{A}^{\mathrm{T}} \\ \mathbf{A}\boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{2|1} + \mathbf{A}\boldsymbol{\Sigma}_1\mathbf{A}^{\mathrm{T}} \end{bmatrix}\right).$$

References

[1] W. Pieczynski, A.N. Tebbache, Pairwise Markov random fields and segmentation of textured images, Machine Graphics and Vision 9 (3) (2000) 705–718.

[2] W. Pieczynski, Pairwise Markov chains, IEEE Trans. Pattern Anal. Machine Intell. 25 (5) (May 2003) 634–639.

[3] R.S. Lipster, A.N. Shiryaev, Statistics of random processes, vol. 2: applications, Conditionally Gaussian Sequences: Filtering and Related Problems, Springer, Berlin, 2001 (Chapter 13).

[4] W. Pieczynski, F. Desbouvries, Kalman filtering using pairwise Gaussian models, in: Proceedings of the Icassp, Hong-Kong, vol. 6, 2003, pp. VI-57–VI-60.

[5] W. Pieczynski, Arbres de Markov couple—pairwise Markov trees, C.R.A.S.-Mathématiques 335 (2002) 79–82.

[6] E. Monfrini, J. Lecomte, F. Desbouvries, W. Pieczynski, Image and signal restoration using pairwise Markov trees, in: Proceedings of the 2003 IEEE Workshop on Statistical Signal Processing, St. Louis, MI, September 2003.

[7] F. Desbouvries, J. Lecomte, Multiscale Bayesian restoration in pairwise Markov trees, IEEE Trans. Automat. Control 50 (8) (August 2005) 1185–1190.

[8] R.E. Kalman, A new approach to linear filtering and prediction problems, J. Basic Eng., Trans. ASME, Ser. D 82 (1) (1960) 35–45.

[9] B.D.O. Anderson, J.B. Moore, Optimal Filtering, Prentice-Hall, Englewood Cliffs, NJ, 1980.

[10] T. Kailath, A.H. Sayed, B. Hassibi, Linear Estimation, Prentice-Hall Information and System Sciences Series, Prentice-Hall, Upper Saddle River, NJ, 2000.

[11] R.J. Meinhold, N.D. Singpurwalla, Understanding the Kalman filter, Amer. Statist. 37 (2) (1983) 123–127.

[12] P.E. Caines, Linear Stochastic Systems, Wiley Series in Probability and Mathematical Statistics, Wiley, New York, 1988.

[13] A.C. Harvey, Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press, Cambridge, 1989.

[14] M. West, J. Harrison, Bayesian forecasting and dynamic models, Springer Series in Statistics, second ed., Springer, New York, 1997.

[15] M. Basseville, A. Benveniste, K.C. Chou, S.A. Golden, R. Nikoukhah, A.S. Willsky, Modeling and estimation of multiresolution stochastic processes, IEEE Trans. Inform. Theory 38 (2) (March 1992) 766–784.

[16] C.A. Bouman, M. Shapiro, A multiscale random field model for Bayesian image segmentation, IEEE Trans. Image Process. 3 (3) (March 1994) 162–177.

[17] K.C. Chou, A.S. Willsky, A. Benveniste, Multiscale recursive estimation, data fusion, and regularization, IEEE Trans. Automat. Control 39 (3) (March 1994) 464–478.

[18] J.M. Laferté, P. Perez, F. Heitz, Discrete Markov image modeling and inference on the quadtree, IEEE Trans. Image Process. 9 (3) (March 2000) 390–404.

[19] A.S. Willsky, Multiresolution Markov models for signal and image processing, Proc. IEEE 90 (8) (August 2002) 1396–1458.

[20] R.S. Lipster, A.N. Shiryaev, Statistics of conditionally Gaussian random sequences, in: Proceedings of the Sixth Berkeley Symposium on Mathematics, Statistics and Probability, vol. 2, University of California Press, Berteley, CA, 1972, pp. 389–422.

[21] D.V. Ouelette, Schur complements and statistics, Linear Algebra Appl. 36 (1981) 187–295.