Proceedings of the International Symposium on Applied Stochastic Models and Data Analysis (ASMDA 2005), Brest, France, May 2005.

On Triplet Markov Chains

Wojciech Pieczynski and François Desbouvries

GET/INT/Département CITI and CNRS UMR 5157
9 rue Charles Fourier
91011 Evry, France
(e-mail: `wojciech.pieczynski@int-evry.fr`,
`francois.desbouvries@int-evry.fr`)

Abstract. The restoration of a hidden process X from an observed process Y is often performed in the framework of hidden Markov chains (HMC). HMC have been recently generalized to triplet Markov chains (TMC). In the TMC model one introduces a third random chain U and assumes that the triplet $T = (X, U, Y)$ is a Markov chain (MC). TMC generalize HMC but still enable the development of efficient Bayesian algorithms for restoring X from Y. This paper lists some recent results concerning TMC; in particular, we recall how TMC can be used to model hidden semi-Markov Chains or deal with non-stationary HMC.

Keywords: hidden Markov chains, hidden semi-Markov chains, pairwise Markov chains, triplet Markov chains, Bayesian segmentation, Kalman filtering and smoothing, iterative conditional estimation.

1 Introduction

An important problem in statistical data restoration consists in estimating a hidden random chain $X = \{X_i\}_{i=1}^n$ from an observed random chain $Y = \{Y_i\}_{i=1}^n$. Let X_i be discrete and Y_i continuous. Many Bayesian methods are available once the distribution of $Z = (X, Y)$ is simple enough. In particular, HMC with independent noise (HMC-IN), in which[1] $p(z) = p(x_1)p(x_2|x_1)\cdots p(x_n|x_{n-1})\ p(y_1|x_1)\cdots p(y_n|x_n)$ have been widely used and studied (see e.g. [Ephraim and Merhav, 2002] for a recent tutorial).

The pairwise Markov chains (PMC) model has been proposed recently [Pieczynski, 2003] [Derrode and Pieczynski, 2004]. In a PMC one assumes that $Z = (X, Y)$ is an MC, i.e. that $p(z) = p(z_1)p(z_2|z_1)\cdots p(z_n|z_{n-1})$. Any

[1] in this formula $p(z)$ denotes the probability density function (pdf) of Z w.r.t. $\kappa^n \otimes \mu^n$, $p(x_i)$ the pdf of X_i w.r.t. κ, and $p(y_i|x_i)$ the conditional pdf (w.r.t. μ) of Y_i given X_i, where κ denotes the counting measure and μ denotes the Lebesgue measure. Later on, other pdf or conditional pdf w.r.t. Lebesgue measure, counting measure, or product measures involving the Lebesgue and/or the counting measure(s) will also be considered; the true meaning of $p(.)$ or of $p(.|.)$ is easily deduced from the context.

HMC-IN is a PMC, but the converse is not true, because in a PMC X is no longer necessarily an MC; however, conditionally on Y, X remains an MC, and in turn this key computational property enables the development of analogous Bayesian restoration algorithms [Lipster and Shiryaev, 2001, corollary 1 p. 72] [Pieczynski, 2003] [Pieczynski and Desbouvries, 2003] [Desbouvries and Pieczynski, 2003b]. PMC have been further extended to TMC. In the TMC model one introduces a third chain $U = \{U_i\}_{i=1}^n$ (which can be physically meaningful or not) and assumes that the triplet $T = (X, U, Y)$ is an MC [Pieczynski *et al.*, 2002] [Pieczynski, 2002]. TMC generalize some classical models in the sense that none of the chains X, U, Y, $V = (X, U)$, $Z = (X, Y)$ or (U, Y) needs to be an MC.

The wider generality of PMC w.r.t. HMC and of TMC w.r.t. PMC can also be seen through the expression of $p(y|x)$. In an HMC-IN $p(y|x) = p(y_1|x_1) \cdots p(y_n|x_n)$, which is very simple, and undoubtedly too simple in some applications, including speech recognition [Wellekens, 1987] [Ostendorf *et al.*, 1996]; in a PMC $p(y|x)$ is an MC, which is much richer; and in a TMC $p(y|x)$ is the marginal distribution of the MC $p(u, y|x)$, which is still much richer than an MC. In such applications as image processing, these increasingly complex models are likely to meet the growing need for a better modeling of the noise [Pérez, 2003].

Apart from this general discussion, the contribution of the TMC model (w.r.t. other possible extensions of the HMC-IN model) appears when describing how they encompass and extend some well known stochastic models. This is better appreciated at the local level, as we now see from a simple example. By definition, a TMC distribution is defined by $p(t_1)$ and by $p(t_{i+1}|t_i)$, which itself can be written by different expressions. In particular, the following factorizations will prove useful in the sequel :

$$p(t_{i+1}|t_i) = p(x_{i+1}|t_i)p(u_{i+1}|x_{i+1}, t_i)p(y_{i+1}|x_{i+1}, u_{i+1}, t_i) \tag{1}$$
$$= p(u_{i+1}|t_i)p(x_{i+1}|u_{i+1}, t_i)p(y_{i+1}|x_{i+1}, u_{i+1}, t_i). \tag{2}$$

The HMC-IN model is obtained from (1) if $p(x_{i+1}|t_i)$ reduces to $p(x_{i+1}|x_i)$, $p(u_{i+1}|x_{i+1}, t_i)$ to $\delta_{x_{i+1}}(u_{i+1})$ (with $\delta_{x_{i+1}}$ the Dirac mass, which simply means that $u_{i+1} = x_{i+1}$), and $p(y_{i+1}|x_{i+1}, u_{i+1}, t_i)$ to $p(y_{i+1}|x_{i+1})$. Other (non-trivial) examples will be given below.

The aim of this paper is to summarize some recent results (some of which are still under review) concerning the large family of TMC. In particular, we will see that the TMC model gathers some well known dynamical stochastic models (and thus provides a unifying framework for these models), as well as some new extensions of these models, and yet still enables the development of efficient hidden chain restoration and parameter estimation algorithms.

The rest of this paper is organized as follows. We will say that X (resp. U, Y) is discrete (resp. continuous) if each X_i (resp. U_i, Y_i) takes discrete (resp. continuous) values, and in this paper X and U can be either discrete or continuous (Y will be assumed to be continuous). So we have four possible situations, which are discussed in sections 2 to 5; as we will see, depending on the situation U admits a physical interpretation (see e.g. §2, item (iii), or §3, item (ii)) or not (see e.g. §2, item (i), or §3, item (i)). Finally section 6 is devoted to parameter estimation.

2 Discrete hidden chain with discrete auxiliary chain

Let X and U be discrete, with $X_i \in \Omega$ and $U_i \in \Lambda$. In this section we shall briefly recall why some classical Bayesian methods like Maximum Posterior Mode (MPM) can be used in TMC. Let $T = (X, U, Y)$ be an MC. The conditional law of $V = (X, U)$ given Y is then an MC, with initial pdf and transitions given by

$$p(v_1|y) = \frac{p(t_1)\beta_1(v_1)}{\sum_{v_1 \in \Omega \times \Lambda} p(t_1)\beta_1(v_1)}, \quad p(v_{i+1}|v_i, y) = \frac{p(t_{i+1}|t_i)\beta_{i+1}(v_{i+1})}{\beta_i(v_i)}, \quad (3)$$

in which β_i can be computed via the classical backward recursions : $\beta_n(v_n) = 1$ and $\beta_i(v_i) = \sum_{v_{i+1} \in \Omega \times \Lambda} p(t_{i+1}|t_i)\beta_{i+1}(v_{i+1})$ for $1 \leq i \leq n-1$. Once $p(v_1|y)$ has been computed, the a posteriori marginals are computed recursively via $p(v_{i+1}|y) = \sum_{v_i \in \Omega \times \Lambda} p(v_i|y)p(v_{i+1}|v_i, y)$. Finally $p(x_i|y) = \sum_{u_i \in \Lambda} p(v_i|y)$, and thus the MPM estimate, which is defined by

$$[\hat{x}_{MPM}(y) = \{\hat{x}_i\}_{i=1}^n] \Longleftrightarrow [\text{for all } i, 1 \leq i \leq n, \hat{x}_i = \arg\max_{x_i} p(x_i|y)],$$

can be computed.

Let us now describe five particular applications of TMC in which this MPM restoration algorithm can be used.

(i) *Mixture approximation.* Assume that a given PMC (X, Y) is stationary, i.e. that $p(x_i, x_{i+1}, y_i, y_{i+1})$ does not depend on i. Then the distribution of (X, Y) is given by $p(x_1, x_2, y_1, y_2) = p(x_1, x_2)\ p(y_1, y_2|x_1, x_2)$. If $p(y_1, y_2|x_1, x_2)$ is not known exactly, one can approximate it by a mixture distribution (for instance a Gaussian one)

$$p(y_1, y_2|x_1, x_2) = \sum_{u_1, u_2 \in \Lambda \times \Lambda} p(u_1, u_2)p(y_1, y_2|x_1, x_2, u_1, u_2),$$

and in this case the model we implicitely deal with is actually a stationary TMC model, the distribution of which is defined by $p(t_1, t_2) = p(u_1, u_2)$ $p(x_1, x_2)\ p(y_1, y_2|x_1, x_2, u_1, u_2)$.

(ii) *"Switching" or "jumping" models".* One way to model non stationary hidden chains is to assume that for each i, $1 \leq i \leq n-1$, there are m possible transitions $p(x_{i+1}|x_i, u_i)$ with $u_i \in \Lambda = \{\lambda_j\}_{j=1}^m$. One usually considers that u_i is a realization of U_i, and $(U_1, \cdots, U_n)$ is an MC. If we directly assume that (X, U) is an MC then we obtain a more general model since U does not need to be an MC any longer. This model has been successfully applied in non stationary image segmentation [Lanchantin and Pieczynski, 2004a]. A further generalization consists in assuming that (X, U, Y) is a general TMC.

(iii) *Hidden semi-Markov chains (HSMC).* When X is an MC, the distribution of the sojourn duration in a given state is exponential, which is restrictive in some situations. In HSMC this distribution can be of any form; these models thus extend HMC, and yet still enable analogous processing, see e.g. [Yu and Kobayashi, 2003] [Moore and Savic, 2004] [Guédon, 2005]. Let q be a pdf on $\mathbb{N}^*$ modeling the probability distribution of the state duration, let $U_n \in \mathbb{N}^*$ be the time during which X_n remains in the same state, and let $\delta_{x_i}(.)$ be the Dirac mass on x_i. Then the semi-MC model can be written as

$$p(u_{i+1}|u_i) = \begin{cases} \delta_{u_i-1}(u_{i+1}) & \text{if } u_i > 1 \\ q(u_{i+1}) & \text{if } u_i = 1 \end{cases}; \tag{4}$$

$$p(x_{i+1}|x_i, u_i) = \begin{cases} \delta_{x_i}(x_{i+1}) & \text{if } u_i > 1 \\ p(x_{i+1}|x_i) & \text{if } u_i = 1 \end{cases}, \tag{5}$$

with $p(x_{i+1}|x_i) = 0$ for $x_{i+1} = x_i$. Consequently HSMC happen to be particular TMC (with auxiliary chain U), in which the three transition pdf in the r.h.s. of factorization (2) reduce respectively to $p(u_{i+1}|t_i) = p(u_{i+1}|u_i)$ given by (4), $p(x_{i+1}|u_{i+1}, t_i) = p(x_{i+1}|x_i, u_i)$ given by (5), and $p(y_{i+1}|x_{i+1}, u_{i+1}, t_i) = p(y_{i+1}|x_{i+1})$. Notice that the fact that HSMC are particular TMC enables to consider a lot of TMC models generalizing HSMC [Pieczynski, 2004].

(iv) *Non-stationary hidden chain X.* Let us consider the problem of unsupervised restoration using the classical HMC-IN $Z = (X, Y)$. The assumption that X is stationary cannot always be done, and yet this assumption is required when estimating the model parameters. However, the possible non stationarity of X can also be modeled by "mass functions", which can be seen as an extension of the probability distribution on discrete finite sets, and then the computation of the posterior distribution of X becomes a particular "Dempster-Shafer" fusion. Now, one can show that introducing mass functions is mathematically equivalent to considering some TMC, which in turn enables one to use different Bayesian algorithms. In particular, using TMC in unsupervised image segmentation enables to improve the results obtained with classical HMC [Lanchantin and Pieczynski, 2004b].

(v) *Vector auxiliary chain.* In a TMC $T = (X, U, Y)$ the chain U can be a vector one. For instance, it is possible to deal with non-stationary HSMC by introducing the pair $U = (W, S)$, in which W models the fact that an HSMC is a TMC, and S models the fact that the TMC (X, W, Y), which is seen as a PMC (V', Y) with $V' = (X, W)$, is not stationary.

3 Discrete hidden chain with continuous auxiliary chain

Let us now give two examples of TMC models with a discrete hidden chain and a continuous auxiliary chain; the first one, in which (U, Y) is Gaussian conditionally on X, enables to model complex noise distributions; while the second one, in which (U, Y) is not Gaussian conditionally on X, appears in radar signal or images modeling.

(i) Consider the following model : let $T = (X, U, Y)$ be an MC, X be an MC, and (U, Y) be Gaussian conditionnally on X. Since T is an MC, the conditional law of (U, Y) given X is an MC as well. However the conditional distribution of Y given X remains Gaussian but is no longer necessarily an MC (the proof of this result is an adaptation of the proof in [Pieczynski and Desbouvries, 2003] [Desbouvries and Pieczynski, 2003a]), so these simple assumptions can lead to "noise" models (i.e., $p(y|x)$) which are significantly more complex than those one usually deals with. Unfortunately, computing $p(x_i|y)$ exactly is not feasible and approximate methods are needed, as we now briefly explain. Let $x_i \in \Omega$. Since T is an MC, the distribution of (X, U) conditionally on Y is also an MC, the transitions of which can be computed by the "backward" recursion (with the difference that now U_i is continuous). As in section 2, let us classically set $\beta_n(v_n) = 1$ and

$$\beta_i(v_i) = \sum_{x_{i+1} \in \Omega} \int_{\mathbb{R}} p(t_{i+1}|t_i) \beta_{i+1}(v_{i+1}) du_{i+1} \text{ for } 1 \leq i \leq n-1. \quad (6)$$

Then $p(v_{i+1}|v_i, y) = \frac{p(t_{i+1}|t_i)\beta_{i+1}(v_{i+1})}{\beta_i(v_i)}$, so $p(v_{i+1}|v_i, y)$ can be computed if $\beta_i(v_i)$ can be computed. But we see from (6) that $\beta_i(v_i)$ is a rather rich mixture, containing, for k classes, k^{n-i} components.

(ii) *Speckle distribution in SAR images.* TMC with a discrete hidden chain and a continuous auxiliary chain are encountered for instance in radar signal or images, as we see from the following example. Let us consider a TMC $T = (X, U, Y)$ such that X is an MC, and $p(u, y|x) = \prod_{i=1}^{n} p(u_i, y_i|x_i)$. Let also $p(u_i, y_i|x_i) = p(u_i|x_i)p(y_i|u_i, x_i)$, in which $p(u_i|x_i)$ are Gamma distributions, and $p(y_i|u_i, x_i)$ are Gaussian distributions with mean $\mu(x_i)$ and variance $\sigma^2(u_i, x_i) = u_i \sigma^2(x_i)$. Then the distributions $p(y_i|x_i)$ are the so-called "K-distributions", and the chain U is the "speckle" process [Barnard and Weiner, 1996] [Delignon and Pieczynski, 2002] [Brunel and Pieczynski, 2005].

4 Continuous hidden chain with discrete auxiliary chain

In this section we assume that T is a TMC in which both X and Y are continuous, and U is discrete with $u_i \in \Lambda$. As in section 2, switching or jump-Markov models, i.e. models in which U is assumed to be an MC, and (X, Y) is an HMC-IN conditionally on U, are well known simple examples of such TMC; for such models the 3 factors in the r.h.s. of (2) reduce respectively to $p(u_{i+1}|t_i) = p(u_{i+1}|u_i)$, $p(x_{i+1}|u_{i+1}, t_i) = p(x_{i+1}|u_{i+1}, x_i)$, and $p(y_{i+1}|x_{i+1}, u_{i+1}, t_i) = p(y_{i+1}|x_{i+1}, u_{i+1})$.

Let us now consider the restoration problem. Although the physical meanings of the TMC models we deal with in this section are very different of those of section 3, the mathematical modeling and computational difficulties are indeed quite similar. Let us for instance consider the filtering problem, which consists in computing $p(x_i|y_{0:i})$. A recursive solution is given by

$$p(x_i|y_{0:i}) = \frac{\sum_{u_{i-1} \in \Lambda} \int p(x_i, u_i, y_i|x_{i-1}, u_{i-1}, y_{i-1}) p(x_{i-1}, u_{i-1}|y_{0:i-1}) dx_{i-1}}{p(y_i|y_{0:i-1})}$$

which, in general, cannot be computed in closed form.

This computational problem is already encountered in the context of jump-Markov models. In particular, the linear Gaussian case has been studied for a long time, and as is well known the exact computation of the posterior filtered or smoothed estimates leads to a computational cost which grows exponentially with time (see e.g. [Tugnait, 1982] and the references therein). So approximate solutions have been proposed, see e.g. [Tugnait, 1982] [Kim, 1994] [Bar-Shalom and Li, 1995] [Doucet *et al.*, 2001]. Reformulating the jump-Markov model as a particular TMC does not help in solving the filtering problem; however, it can lead to interesting generalizations, to which the classical approximate methods designed for jump-Markov systems could be extended. For instance, in the TMC above U is a discrete MC and thus T can be viewed as a "hidden" MC. Such an HMC could then be extended to an HSMC, as specified in section 2, item (iii).

5 Continuous hidden chain with continuous auxiliary chain

TMC with continuous processes X, U and Y are used in some applications, including the extensions of the classical linear state-space system (7) to colored process and/or measurement noise. Let

$$\begin{cases} X_{n+1} = F_n X_n + G_n \eta_n \\ Y_n \quad\;\; = H_n X_n + J_n \xi_n \end{cases}, \tag{7}$$

in which η_n is the process noise and ξ is the measurement noise. F_n, G_n, H_n and J_n are known deterministic matrices, and processes $\eta = \{\eta_n\}_{n \in \mathbb{N}}$ and $\xi = \{\xi_n\}_{n \in \mathbb{N}}$ are assumed to be independent, jointly independent and independent of X_0. As a consequence, (X, Y) is an HMC-IN. The filtering problem consists in computing the posterior pdf $p(x_n|y_{0:n})$. From (7),

$p(x_i|y_{0:i})$ can be computed recursively as

$$p(x_{i+1}|y_{0:i+1}) = \frac{p(y_{i+1}|x_{i+1}) \int p(x_{i+1}|x_i)p(x_i|y_{0:i})dx_i}{\int p(y_{i+1}|x_{i+1})[\int p(x_{i+1}|x_i)p(x_i|y_{0:i})dx_i]dx_{i+1}}. \quad (8)$$

If furthermore X_0 and (η_n, ξ_n) are Gaussian, then $p(x_n|y_{0:n})$ is also Gaussian and is thus described by its mean and covariance matrix. Propagating $p(x_n|y_{0:n})$ amounts to propagating these parameters, and (8) reduces to the celebrated Kalman filter [Kalman, 1960] see also [Ho and Lee, 1964] [Anderson and Moore, 1979] [Kailath *et al.*, 2000].

It happens that some classical extensions of model (7) are particular TMC. Consider for instance model (7), but in which we now assume that

$$\begin{bmatrix} \eta_{n+1} \\ \xi_{n+1} \end{bmatrix} = \underbrace{\begin{bmatrix} A_n^{\eta,\eta} & 0 \\ 0 & A_n^{\xi,\xi} \end{bmatrix}}_{A_n} \underbrace{\begin{bmatrix} \eta_n \\ \xi_n \end{bmatrix}}_{u_n} + \underbrace{\begin{bmatrix} \epsilon_n^{\eta} \\ \epsilon_n^{\xi} \end{bmatrix}}_{\epsilon_n}, \quad (9)$$

where $\epsilon^{\eta} = \{\epsilon_n^{\eta}\}_{n\in\mathbb{N}}$ (resp. $\epsilon^{\xi} = \{\epsilon_n^{\xi}\}_{n\in\mathbb{N}}$ is zero-mean, independent and independent of η_0 (resp. of ξ_0), and ϵ^{η} and ϵ^{ξ} are independent. Each one of the two processes $\eta = \{\eta_n\}_{n\in\mathbb{N}}$ and $\xi = \{\xi_n\}_{n\in\mathbb{N}}$ is thus an MC, and η is independent of ξ. Such a model has been introduced by Sorenson [Sorenson, 1966] (see also [Chui and Chen, 1999, ch. 5]). It is no longer an HMC (X is not an MC), but the whole model $T_n = (X_n, U_n, Y_{n-1})$ can be rewritten as

$$\underbrace{\begin{bmatrix} X_{n+1} \\ U_{n+1} \\ Y_n \end{bmatrix}}_{T_{n+1}} = \underbrace{\begin{bmatrix} F_n & \overline{G}_n & 0 \\ 0 & A_n & 0 \\ H_n & \overline{J}_n & 0 \end{bmatrix}}_{\mathcal{F}_n} \begin{bmatrix} X_n \\ U_n \\ Y_{n-1} \end{bmatrix} + \underbrace{\begin{bmatrix} 0 \\ \epsilon_n \\ 0 \end{bmatrix}}_{W_n} \quad (10)$$

(with $\overline{G}_n = [G_n, 0]$ and $\overline{J}_n = [0, J_n]$), and so $T = \{T_n\}$ is a TMC.

Model (10) is indeed a particular case of a linear TMC, defined by $T_{n+1} = \mathcal{F}_n T_n + W_n$, with $T_n = (X_n, U_n, Y_{n-1})$, and W_n independent and independent of T_0. $p(x_n|y_{0:n})$ is obtained by marginalizing $p(v_n|y_{0:n})$ which, in the Gaussian case, can be computed efficiently by a Kalman-like filtering algorithm [Desbouvries and Pieczynski, 2003a] [Ait-el-Fquih and Desbouvries, 2005b]. Kalman-like smoothing algorithms, extending to linear Gaussian TMC the two-filter and RTS smoothers, have also been derived [Ait-el-Fquih and Desbouvries, 2005a].

6 Parameter estimation

Let us finally mention that the model parameters can be estimated from the observed data Y, either by using the well-known "Expectation-Maximization" (EM) method [McLachlan and Krishnan, 1997] or the "Iterative conditional estimation" (ICE) method (some relationships between ICE and EM can be found in [Delmas, 1997]).

As an illustrative example, let us see how the model parameters can be estimated by ICE, which we first briefly recall. Parameter estimation according to the ICE principle can be performed once

(i) an estimator $\hat{\theta}(X,Y)$ of the parameters θ from the complete data (X,Y) is available; and
(ii) one can sample X according to $p(x|y)$.

Then ICE is described by the recursion $\theta^{q+1} = E(\hat{\theta}(X,Y)|Y=y,\theta^q)$, starting with some initial value θ^0. If for some components θ_j of θ this expectation cannot be computed, one samples $x^1, \cdots, x^l$ according to $p(x|y,\theta^q)$ and sets $\theta_j^{q+1} = \frac{1}{l}\sum_{i=1}^{l}\hat{\theta}_j(x^i,y)$.

Let us turn to parameter estimation in PMC and TMC. Let us first remark that the problem is identical in both cases, since a TMC $T=(X,U,Y)$ can be seen as a PMC (V,Y) with $V=(X,U)$. Let us as an illustrative example consider the case of a stationary PMC $Z=(X,Y)=(X_1,Y_1,\cdots X_n,Y_n)$ in which $p(z_i,z_{i+1})$ does not depend on i. So the distribution of Z is given by $p(z_1,z_2) = p(x_1,x_2)p(y_1,y_2|x_1,x_2)$. Assume that $X_i \in \Omega = \{\omega_1,\omega_2\}$ and that $p(y_1,y_2|x_1,x_2)$ are Gaussian. Then the model parameter consists of $\theta = (\alpha,\beta)$, where α gathers the four parameters $\alpha = \{\alpha_{i,j} = p(x_1=\omega_i,x_2=\omega_j)\}_{i,j=1}^2$, and $\beta = \{\beta_i\}_{i=1}^{20}$ the twenty parameters of the four Gaussian densities $\{p(y_1,y_2|x_1,x_2)\}_{x_1,x_2\in\Omega\times\Omega}$ on $\mathbb{R}^2$.

Let us now apply ICE to this model. Let $\hat{\theta}(X,Y) = (\hat{\alpha}(X),\hat{\beta}(X,Y))$. $\hat{\alpha}(X)$ can be chosen as the classical frequency estimator, and $\hat{\beta}(X,Y)$ as the classical empirical means and variance-covariance matrices. Then $\alpha_i^{q+1} = E(\hat{\alpha}_i(X)|Y=y,\theta^q)$ can be computed, but $\beta_i^{q+1} = E(\hat{\beta}_i(X,Y)|Y=y,\theta^q)$ cannot. In practice, the interest of PMC over HMC-IN in unsupervised segmentation using the ICE principle has been proven by different experiments [Derrode and Pieczynski, 2004]. On the other hand, using copulas enables to extend ICE to the case where the exact nature of the noise distribution is not known (it can take different possible forms) [Brunel and Pieczynski, 2003].

References

[Ait-el-Fquih and Desbouvries, 2005a]B. Ait-el-Fquih and F. Desbouvries. Bayesian smoothing algorithms in pairwise and triplet Markov chains. In *submitted to the 2005 IEEE Workshop on Statistical Signal Processing*, Bordeaux, France, July 2005.

[Ait-el-Fquih and Desbouvries, 2005b]B. Ait-el-Fquih and F. Desbouvries. Kalman filtering for triplet Markov chains : Applications and extensions. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP 05)*, Philadelphia, USA, March 19-23 2005.

[Anderson and Moore, 1979]B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Prentice Hall, Englewood Cliffs, New Jersey, 1979.

[Bar-Shalom and Li, 1995]Y Bar-Shalom and X. R. Li. *Multitarget-multisensor tracking : principles and techniques*. YBS, 1995.

[Barnard and Weiner, 1996]T. J. Barnard and D. D. Weiner. Non-Gaussian clutter modeling with generalized spherically invariant random vectors. *IEEE Transactions on Signal Processing*, 44(10):2384–2390, 1996.

[Brunel and Pieczynski, 2003]N. Brunel and W. Pieczynski. Unsupervised signal restoration using copulas and pairwise Markov chains. In *Proceedings of the 2003 IEEE Workshop on Statistical Signal Processing*, St. Louis, MI, September 2003.

[Brunel and Pieczynski, 2005]N. Brunel and W. Pieczynski. Modeling temporal dependence of spherically invariant random vectors with triplet Markov chains. In *submitted to the 2005 IEEE Workshop on Statistical Signal Processing*, Bordeaux, France, July 2005.

[Chui and Chen, 1999]C.K. Chui and G. Chen. *Kalman Filtering with Real-Time Applications*. Berlin, DE: Springer, 1999.

[Delignon and Pieczynski, 2002]Y. Delignon and W. Pieczynski. Modeling non-Rayleigh speckle distribution in SAR images. *IEEE Transactions on Geoscience and Remote sensing*, 40(6):1430–1435, 2002.

[Delmas, 1997]J.-P. Delmas. An equivalence of the EM and ICE algorithm for exponential family. *IEEE Transactions on Signal Processing*, 45(10):2613–15, 1997.

[Derrode and Pieczynski, 2004]S. Derrode and W. Pieczynski. Signal and image segmentation using pairwise Markov chains. *IEEE Transactions on Signal Processing*, 52(9):2477–89, 2004.

[Desbouvries and Pieczynski, 2003a]F. Desbouvries and W. Pieczynski. Modèles de Markov triplet et filtrage de Kalman. *Comptes Rendus de l'Académie des Sciences - Mathématiques*, 336(8), 2003. in French.

[Desbouvries and Pieczynski, 2003b]F. Desbouvries and W. Pieczynski. Particle filtering in pairwise and triplet Markov chains. In *Proceedings of the IEEE - EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP 2003)*, Grado-Gorizia, Italy, June 8-11 2003.

[Doucet *et al.*, 2001]A. Doucet, N. J. Gordon, and V. Krishnamurthy. Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, 49(3):613–24, March 2001.

[Ephraim and Merhav, 2002]Y. Ephraim and N. Merhav. Hidden Markov processes. *IEEE Transactions on Information Theory*, 48(6):1518–69, June 2002.

[Guédon, 2005]Y. Guédon. Hidden hybrid Markov/semi-Markov chains. *Computational Statistics and Data Analysis*, 2005. to appear.

[Ho and Lee, 1964]Y. C. Ho and R. C. K. Lee. A Bayesian approach to problems in stochastic estimation and control. *IEEE Transactions on Automatic Control*, 9:333–339, October 1964.

[Kailath *et al.*, 2000]T. Kailath, A. H. Sayed, and B. Hassibi. *Linear estimation*. Prentice Hall Information and System Sciences Series. Prentice Hall, Upper Saddle River, New Jersey, 2000.

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

[Kim, 1994]C-J. Kim. Dynamic linear models with Markov switching. *J. of Econometrics*, 60:1–22, 1994.

[Lanchantin and Pieczynski, 2004a]P. Lanchantin and W. Pieczynski. Unsupervised non stationary image segmentation using triplet Markov chains. In *Advanced Concepts for Intelligent Vision Systems (ACVIS 04)*, Brussels, Belgium, August 31 - September 3 2004.

[Lanchantin and Pieczynski, 2004b]P. Lanchantin and W. Pieczynski. Unsupervised restoration of hidden non stationary Markov chain using evidential priors. *accepted for publication, IEEE Transactions on Signal Processing*, 2004.

[Lipster and Shiryaev, 2001]R. S. Lipster and A. N. Shiryaev. *Statistics of Random Processes, Vol. 2 : Applications*, chapter 13 : "Conditionally Gaussian Sequences : Filtering and Related Problems". Springer Verlag, Berlin, 2001.

[McLachlan and Krishnan, 1997]G. J. McLachlan and T. Krishnan. *EM Algorithm and Extensions*. Wiley, 1997.

[Moore and Savic, 2004]M. D. Moore and M. I. Savic. Speech reconstruction using a generalized HSMM (GHSMM). *Digital Signal Processing*, 14(1):37–53, 2004.

[Ostendorf *et al.*, 1996]M. Ostendorf, V.V. Digalakis, and O. A. Kimball. From HMMs to segment models: a unified view of stochastic modeling for speech recognition. *IEEE Transactions on Speech and Audio Processing*, 4(5):360–378, September 1996.

[Pérez, 2003]P. Pérez. *Modèles et algorithmes pour l'analyse probabiliste des images*. Université de rennes I. Habilitation à diriger les recherches, 2003. (in French).

[Pieczynski and Desbouvries, 2003]W. Pieczynski and F. Desbouvries. Kalman filtering using pairwise Gaussian models. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP 03)*, Hong-Kong, 2003.

[Pieczynski *et al.*, 2002]W. Pieczynski, C. Hulard, and T. Veit. Triplet Markov chains in hidden signal restoration. In *SPIE International Symposium on Remote Sensing*, Crete, Grece, September 22-27 2002.

[Pieczynski, 2002]W. Pieczynski. Chaînes de Markov triplet. *Comptes Rendus de l'Académie des Sciences - Mathématiques*, 335:275–278, 2002. in French.

[Pieczynski, 2003]W. Pieczynski. Pairwise Markov chains. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):634–39, May 2003.

[Pieczynski, 2004]W. Pieczynski. Chaînes semi-Markoviennes cachées et chaînes de Markov triplet. *submitted to : Comptes Rendus de l'Académie des Sciences - Mathématiques*, décembre 2004.

[Sorenson, 1966]H. W. Sorenson. Kalman filtering techniques. In C. T. Leondes, editor, *Advances in Control Systems Theory and Appl.*, volume 3, pages 219–92. Acad. Press, 1966.

[Tugnait, 1982]J. K. Tugnait. Adaptive estimation and identification for discrete systems with Markov jump parameters. *IEEE Transactions on Automatic Control*, 27(5):1054–65, October 1982.

[Wellekens, 1987]C. J. Wellekens. Explicit time correlation in hidden Markov models for speech recognition. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP 87)*, volume 12, pages 384–86, 1987.

[Yu and Kobayashi, 2003]S.-Z. Yu and H. Kobayashi. A hidden semi-Markov model with missing data and multiple observation sequences for mobility tracking. *Signal Processing*, 83(2):235–250, 2003.