An exact smoother in a fuzzy jump Markov switching model

Zied Bouyahia, Stéphane Derrode, Wojciech Pieczynski

An exact smoother in a fuzzy jump Markov switching model

Zied Bouyahia[1], Stéphane Derrode[2] and Wojciech Pieczynski[3]

[1] Department of Computer Science, CAAS, Dhofar University, Oman,
bouyahiazied@gmail.com,
[2] Ecole Centrale de Lyon, LIRIS, CNRS UMR 5205, Écully, France,
stephane.derrode@ec-lyon.fr
[3] Telecom Sudparis, SAMOVAR, CNRS UMR 5157, Évry, France,
wojciech.Pieczynski@telecom-sudparis.eu

Abstract. In this paper, we address the exact smoothing problem of Conditionally Gaussian Observed Markov Switching Model (CGOMSM). The proposed approach tackles the discontinuity feature in switching regime models by incorporating fuzzy switches instead of hard jumps. Fuzzy switched based approach is more adapted to real-world application in which regime continuity is an intrinsic property. We define, with respect to a measure, the local density of the switches and we show how fast smoothing equations should be adapted to cope with the fuzzy model. Finally, we show through several experiments the interest of the fuzzy switches model.

1 Introduction

Let $\mathbf{X}_1^N = \{X_1, \ldots, X_N\}$, $\mathbf{Y}_1^N = \{Y_1, \ldots, Y_N\}$ and $\mathbf{R}_1^N = \{R_1, \ldots, R_N\}$ be three random sequences taking values in $\mathbb{R}^m$, $\mathbb{R}^q$ and $\Omega = 1, \ldots, K$. Let $\mathbf{X}_1^N$ be a hidden process and $\mathbf{Y}_1^N$ be an observed process. We consider a switching regime model represented by the sequence of switches $\mathbf{R}_1^N$. We address the smoothing problem consisting in an recursive search of the unobserved process $\mathbf{X}_1^N$ and the switches sequence $\mathbf{R}_1^N$ knowing the observed sequence $\mathbf{Y}_1^N$. A fast Bayesian processing can be carried out by assuming that the distribution of $(\mathbf{X}_1^N, \mathbf{Y}_1^N)$ is within the framework of hidden Gaussian Markov Model. The non-linearity can be modeled by a switching regime system. Then, the idea is to approximate a non-linear non-Gaussian system by a regime switching Gaussian system. Some recent switching models have been proposed with efficient fast exact filtering schemes [9] and satisfactiry computation time. These switching Gaussian models include conditionally Markov switching hidden linear models (CMSHLM) [16] and the conditionally Gaussian observed Markov switching model (CGOMSM) which is a sub-model of CMSHLM defined as follows:

- $T_1^N = (X_1^N, R_1^N, Y_1^N)$ is Markov chain;

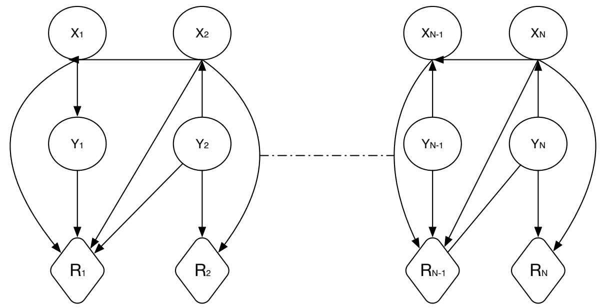

Fig. 1. Directed graph representing dependencies between random sequences $\mathbf{X}_1^N$ $\mathbf{Y}_1^N$ and $\mathbf{R}_1^N$. Circles represent continuous process and diamond represents discrete process.

– $p(r_{n+1}|x_n, r_n, y_n) = p(r_{n+1}|r_n)$ and

$$\begin{bmatrix} X_{n+1} \\ Y_{n+1} \end{bmatrix} = \begin{bmatrix} A^{xx}_{n+1}(R_n^{n+1}) & A^{xy}_{n+1}(R_n^{n+1}) \\ 0 & A^{yy}_{n+1}(R_n^{n+1}) \end{bmatrix} \begin{bmatrix} X_n \\ Y_n \end{bmatrix} + \begin{bmatrix} B^{xx}_{n+1}(R_n^{n+1}) & B^{xy}_{n+1}(R_n^{n+1}) \\ B^{yx}_{n+1}(R_n^{n+1}) & B^{yy}_{n+1}(R_n^{n+1}) \end{bmatrix} \begin{bmatrix} U_{n+1} \\ V_{n+1} \end{bmatrix} + \begin{bmatrix} M^X(R_{n+1}) - A^{xx}_{n+1}(R_n^{n+1})M^X(R_n) - A^{xy}_{n+1}(R_n^{n+1})M^Y(R_n) \\ M^Y(R_{n+1}) - A^{yy}_{n+1}(R_n^{n+1})M^Y(R_n) \end{bmatrix}$$

with U_1^N and V_1^N a Gaussian unit-variance white noise vector, $M^X(R_n)$ and $M^Y(R_n)$ are respective means of $\mathbf{X}_1^N$ and $\mathbf{Y}_1^N$ in each state (independently from n). The CGOMSM is then defined by matrices $A(\mathbf{R}_n^{n+1})$, $B(\mathbf{R}_n^{n+1})$, transition matrix denoted as t such that $t(i,j) = p(r_{n+1} = j|r_n = i)$ and vector $M = [M(R_n); M(R_{n+1})]$ such that $M(R_n) = [M^X(R_n); M^Y(R_n)]$. Figure 1 depicts the dependence graph of the CGOMSM model.
We assume that $\mathbf{R}_1^N$ takes its values in a discrete finite set of K switches $\Omega = \{1, ..., K\}$. This hard jumps model has been widely used in several contexts dealing with switching regime Markov systems. Its success comes from its ability to represent non-linear dynamic patterns which is an inherent property in several applications (analysis of economic and finance time series [11], sustainable energy [6], robotics [12] [7] [8], etc.).
However, this model does not take into account the intrinsic imprecision of the switches in real-world applications. In fact, hard jumps induce discontinuity in the dynamic behavior of the studied system. This transitory imprecision can be modeling fuzzy modeling which consists in allowing each switch to take its value

as a mixture of many components simultaneously. Fuzzy modelling has been already incorporated in several applications dealing with Markov models [14, 13, 15]. In this paper, we present a new method to approximate non-linear Markov systems using a new variant of CGOMSM using fuzzy switches (hereafter called CGOMFSM). The remaining of this paper is organized as follows. In the second section, we detail the formulation of fuzzy switching model. The third section describes the adaptation of CGOMSM algorithms for parameters estimation and for posterior marginal probabilities computation the fuzzy counterpart. The fourth section presents experimental results, and the last one draws conclusions and future work.

2 Fuzzy switching model with K hard classes

In the fuzzy switches system we assume that each jump r_n in the random process R_1^N is a mixture of the K classes. Let ε_n^i be the contribution of the i^{th} hard component ω_i such that $\varepsilon_n^i \in [0,1]$. Then each switch r_n is given by the vector ε_n :

$$r_n = \varepsilon_n = (\varepsilon_n^1, \varepsilon_n^2, ..., \varepsilon_n^K), \tag{1}$$

The normalization condition yields:

$$\sum_{k=1}^{K} \varepsilon^k = 1 \tag{2}$$

In the remainder, for the sake of simplicity, we refer to each hard component by its index in the discrete set i.e. class i corresponds to the hard component ω_i.

2.1 Measure of local density

The distribution of each switch is defined by a density h_r with respect to a Lebesgue measure υ. This measure includes K discrete components represented by K Dirac functions $\{\delta_{k,\ 1 \le k \le K}\}$ as well as the continuous components representing the combination of the hard switches.
Let S be a subset of $0 < l \le K$ hard components in Ω. If a switch r_n is a mixture of the hard components in $S \subset \Omega$ then r_n belongs to the hyperplane of dimension l corresponding to $\sum_{i=1}^{l} \varepsilon_n^i = 1$ such that $\varepsilon_n^i > 0,\ \forall 1 \le i \le l$. The corresponding measure υ is given by :

$$\upsilon = \sum_{k=1}^{K} \delta_k + \sum_{S \subseteq \Omega, S \neq \emptyset} \mu_S \tag{3}$$

where μ_S are Lebesgue measures on $[0,1]^l$.

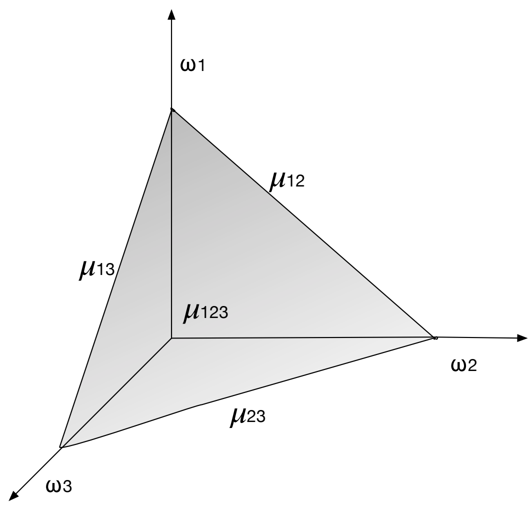

Fig. 2. Example of measure density for a fuzzy model with 3 discrete classes.

Example

In this example we consider the particular case $K = 3$ to illustrate the local density measure. The three dimensional space corresponding to the three hard components is depicted in figure 2.1. The hyperplane corresponding to the normalization condition (eq. (2)) coincides with the shaded triangle Δ. The edges of Δ represent the mixture of two hard classes while the vertices of Δ correspond to the three hard components. When $K = 3$, eq. (3) becomes:

$$\upsilon = \delta_{\omega_1} + \delta_{\omega_2} + \delta_{\omega_3} + \mu_{\omega_1\omega_2} + \mu_{\omega_1\omega_3} + \mu_{\omega_2\omega_3} + \mu_{\omega_1\omega_2\omega_3}$$

2.2 Joint densities

Let H be the hyperplane of dimension $K-1$ and $f(r_n, r_{n+1})$ be the joint density of the pair $(r_n, r_{n+1}) \in H \times H$. The joint density f is defined as follows:

$$f(r_n, r_{n+1}) = \begin{cases} \alpha_{ij} & \text{if both switches are hard} \\ \beta\phi(r_n, r_{n+1}) + \theta & \text{otherwise} \end{cases} . \tag{4}$$

We choose $\phi(r_n, r_{n+1}) = \left(1 - \frac{\|r_{n+1} - r_n\|}{\sqrt{2}}\right)^r$, $r \in \mathbb{R}$ with $\|r_{n+1} - r_n\|$ is the distance between two consecutive switches given by the quadratic norm. Then:

$$\phi(r_n, r_{n+1}) = \left(1 - \left(\frac{\sum_{i=1}^{K} (\varepsilon_n^i - \varepsilon_{n+1}^i)^2}{2}\right)^{\frac{1}{2}}\right)^r, \ r \in \mathbb{R} \tag{5}$$

The normalization condition yields:

$$\sum_{i=1}^{K} \alpha_{ii} + \oint_H \oint_H \theta + \beta \phi(r_n, r_{n+1})(\nu, \mu) d\nu d\mu = 1 \tag{6}$$

Example:
When the number of hard switches equals 3, the expression of normalization condition gives:

$$\sum_{i=1}^{3} \alpha_{ii} + (\beta_a + \beta_b + \beta_c) \int_0^1 \int_0^1 (1 - |u - v|)^r du dv$$
$$+ 2(\beta_a + \beta_b + \beta_c) \int_0^1 (1 - |\varepsilon|)^r + (1 - |1 - \varepsilon|)^r d\varepsilon = 1 \tag{7}$$

2.3 Parameters interpolation

The model matrices of the fuzzy switching model can be calculated by linear interpolation using the following formula:

$$A(\varepsilon_1, \varepsilon_2) = \sum_{1 \le i,j \le K} \varepsilon_1^i \varepsilon_2^j A(i,j),$$
$$B(\varepsilon_1, \varepsilon_2) = \sum_{1 \le i,j \le K} \varepsilon_1^i \varepsilon_2^j B(i,j),$$
$$M(\varepsilon_1, \varepsilon_2) = \sum_{1 \le i,j \le K} \varepsilon_1^i \varepsilon_2^j M(i,j),$$

where $A(i,j)$ and $B(i,j)$ are the model matrices corresponding to the hard components i and j. $M(i,j)$ is means model vector for hard switches i and j.

The implementation of the fuzzy switching model can be performed by adequate quantification of the interval $[0,1]$ into F discrete fuzzy levels. The larger F is, the more accurate the representation of data would be. However, choosing a large number of fuzzy levels will lead to high computation time. For example, when the number of crisp components equals three, setting $F = 3$ yields 15 switches and setting $F = 4$ gives 21 switches.

3 Fuzzy switching model with two hard components

In this remaining of the paper, we consider the case of two hard switches $\Omega = \{0,1\}$. To model fuzzy switches, we consider that each random variable R_n in $\mathbf{R}_1^N$ takes its values in the continuous interval $[0,1]$, instead of the set $\{0,1\}$. Let us denote the pair $(\varepsilon_n^0, \varepsilon_n^1) \in [0,1]$, in which ε_n^i represents the contribution of the hard component i to the switch r_n. Without loss of generality, let $\varepsilon_n = \varepsilon_n^1 = 1 - \varepsilon_n^0$. Then we have $R_n = \varepsilon_n$:

- $\varepsilon_n = 0$ if the switch is the hard component 0.
- $\varepsilon_n \in]0,1[$ if the switch is fuzzy.
- $\varepsilon_n = 1$ if the switch is the hard component 1.

So this model able to represent signals with both discrete (hard) and continuous (fuzzy) components. Let ν define the measure associated to random variable R_n (ν is defined by 2 diracs and a Lebesgue measure).

Let us now precisely define the joint a priori density $p\left(\mathbf{R}_n^{n+1}\right)$, where notation $\mathbf{R}_n^{n+1}$ represents the couple (R_n, R_{n+1}). $p\left(\mathbf{R}_n^{n+1}\right)$ is defined with respect to the measure product $\nu \otimes \nu$, under normalisation condition

$$\iint_{[0,1]^2} p\,(u,v)\, d(\nu \otimes \nu)(u,v) = 1. \tag{8}$$

This distribution can be represented by a density h[1] which involves three types of components:

- Hard components when both R_n and R_{n+1} are in $\{0,1\}$;
- Composite components when either R_n and R_{n+1} is in $\{0,1\}$ and the other is in $]0,1[$;
- Fuzzy components when both switches are in fuzzy area $]0,1[^2$.

Let assume that the density h is uniform in each of its components. It is so possible to define the density h by a set of 9 parameters explicitly expressing $h(r_n, r_{n+1}) = p\left(\mathbf{R}_n^{n+1} = \mathbf{r}_n^{n+1}\right) = \pi_{ij}$. But since we have

$$\pi_{0F} = h(0,u) = h(u,0), \tag{9}$$
$$\pi_{1F} = h(1,u) = h(u,1), \tag{10}$$

with $u \in]0,1[$, only 7 parameters are required to entirely define h : π_{00}, π_{01}, π_{10}, π_{11}, π_{0F}, π_{1F} and π_F (see Fig. 3). With these notations, the normalisation eq. (8) writes

$$\begin{aligned}\pi_{00} + \pi_{01} + \pi_{10} + \pi_{11} + 2\pi_{0F} + 2\pi_{1F} + \pi_F = \\ \pi_{00} + \pi_{01} + \pi_{10} + \pi_{11} + 2\int_{]0,1[} p(0,u)du + \\ 2\int_{]0,1[} p(1,u)du + \iint_{]0,1[^2} p(u,v)dudv = 1\end{aligned} \tag{11}$$

Let now assume that the shape of h is defined according to the simple following shape

$$h(\varepsilon_1, \varepsilon_2) = a\Phi(\varepsilon_1, \varepsilon_2) + b, \tag{12}$$

whatever $\varepsilon_1, \varepsilon_2 \in [0,1]$, with $(a,b) \in \mathbb{R}^2$ and

$$\Phi(\varepsilon_1, \varepsilon_2) = (1 - |\varepsilon_1 - \varepsilon_2|)^r, \tag{13}$$

[1] Density h does not depend on index n since the model we consider in this paper is assumed stationary.

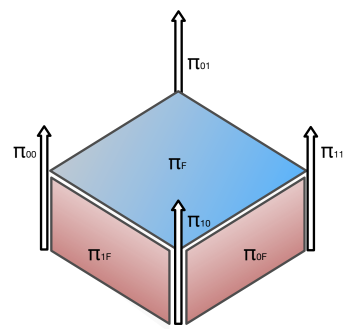

Fig. 3. Density of $p\left(\mathbf{R}_n^{n+1}\right)$ with respect to measure $\nu \otimes \nu$.

with $r \in \mathbb{R}$. Hence, this model for h relies 3 parameters: a, b, and r. Analytic computation of the joint prior densities can be worked out by quantifying the interval $[0, 1]$ into F equal-length sub-intervals $\left[\frac{i}{F}, \frac{i+1}{F}\right]$ as described in Figure 3. Using this scheme, the normalization condition in eq. (11) yields:

$$\begin{aligned}
&\pi_{00} + \pi_{01} + \pi_{10} + \pi_{11} + \\
&\frac{1}{2F}\sum_{i=0}^{F-1}\left(a(1-\varepsilon_i)^r + b\right) + \frac{1}{2F}\sum_{i=0}^{F-1}\left(a\varepsilon_i^r + b\right) + \\
&\frac{1}{2F^2}\sum_{i=0}^{F-1}\sum_{j=0}^{F-1}\left(a\left(1-|\varepsilon_i - \varepsilon_j|\right)^r + b\right) = 1.
\end{aligned} \tag{14}$$

Each sub-interval can be represented by its medium value $\frac{2i+1}{2F}$. So, in this discrete approximate scheme, the joint a priori density can be defined by a $(2+F) \times (2+F)$ matrix.

Under the assumptions of fuzzy switches, we can define the matrices of the incorporated model using a bi-linear function as follows:

$$\begin{aligned}
A(\varepsilon_1, \varepsilon_2) = {} & [(1-\varepsilon_1)A(0,0) + \varepsilon_1 A(1,0)](1-\varepsilon_2) \\
& + [(1-\varepsilon_1)A(0,1) + \varepsilon_1 A(1,1)]\varepsilon_2
\end{aligned} \tag{15}$$

$$\begin{aligned}
B(\varepsilon_1, \varepsilon_2) = {} & [(1-\varepsilon_1)B(0,0) + \varepsilon_1 B(1,0)](1-\varepsilon_2) \\
& + [(1-\varepsilon_1)B(0,1) + \varepsilon_1 B(1,1)]\varepsilon_2
\end{aligned} \tag{16}$$

The means vectors of the fuzzy model are calculated using the following equations:

$$M(\varepsilon_i) = [(1-\varepsilon_i)M(0) + \varepsilon_i M(1)] \tag{17}$$

(18)

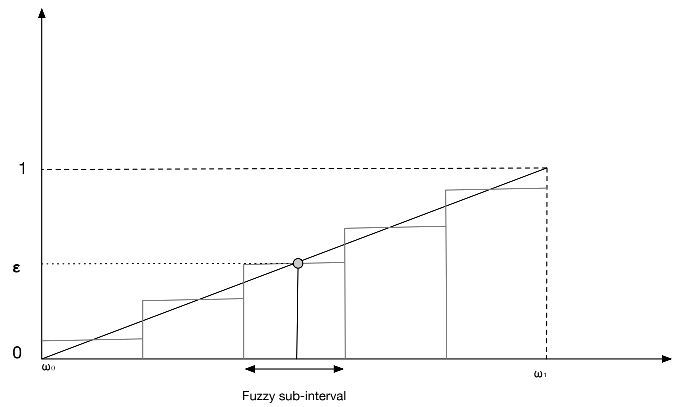

Fig. 4. Subdivision of the interval $[0, 1]$ into $F = 5$ equal-length fuzzy sub-intervals.

Hence, the CGOMFSM is entirely defined by

- the parameters of the corresponding deterministic hard switching model,
- the number of fuzzy levels F, and
- parameters a, b and r.

The parameter r specifies the homogeneity of the switching model. The larger r is, the larger the probability of having two similar consecutive switches is.
Figure 5 represents and example of simulation of (X,Y,R) using the set of parameters of a fuzzy switching model defined in table 1. Simulations were performed using the following transition matrix:

$$t = \begin{pmatrix} 0.99 & 0.01 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.99 & 0.01 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.99 & 0.01 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.99 & 0.01 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.99 & 0.01 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.99 & 0.01 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1.00 \end{pmatrix}$$

This simulation shows the imprecision between hard switches 0 and 1 as presented in the trajectory of $\mathbf{Y}_1^N$. The choice of the transition matrix allow a progressive regime switching from parameters set corresponding to hard switch 0 to hard switch 1.

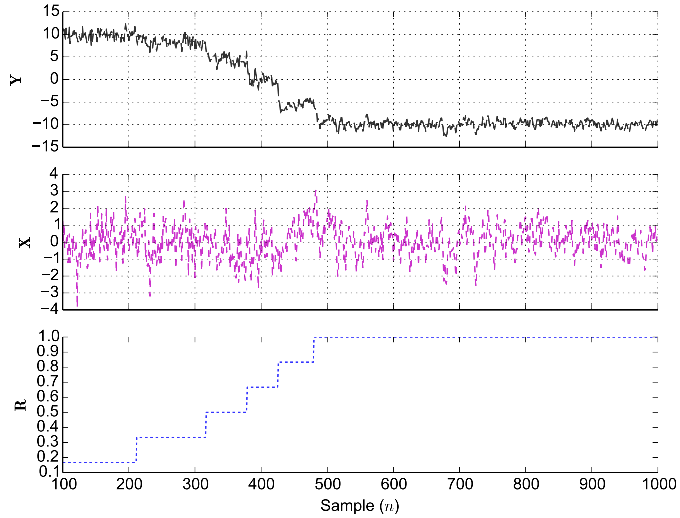

Fig. 5. Trajectories of simulated CGOMFSM $(\mathbf{X}_1^N, \mathbf{Y}_1^N, \mathbf{R}_1^N)$.

Table 1. Example of fuzzy switching model with 5 fuzzy switches.

(R_n, R_{n+1})	$(0,0)$	$(1,1)$
(MX, MY)	$(1,10)$	$(-1,-10)$
A	0.550000 0.350000 0.000000 0.830000	0.400000 0.200000 0.000000 0.460000
B	0.516720 0.000000 0.279261 0.482818	0.867179 0.000000 0.176203 0.870260

4 Fast smoothing in a CGOMFSM

Let us denote by $\mathbf{T}_1^N$ the triplet $(\mathbf{X}_1^N, \mathbf{R}_1^N, \mathbf{Y}_1^N)$. The smoothing problem consists in computing :

$$\mathbb{E}\left[\mathbf{X}_{n+1} \left| \mathbf{y}_1^N \right.\right] = \int_{[0,1]} p\left(r_{n+1} = \nu \left| \mathbf{y}_1^N \right.\right) \mathbb{E}\left[\mathbf{X}_{n+1} \left| r_{n+1} = \nu, \mathbf{y}_1^{n+1} \right.\right] d\nu, \quad (19)$$

from $p\left(r_{n+1} \left| \mathbf{y}_1^N \right.\right)$ and $\mathbb{E}\left[\mathbf{X}_{n+1} \left| r_{n+1}, \mathbf{y}_1^{n+1} \right.\right]$.

The optimal smoother computes recursively $p\left(r_{n+1} \left| \mathbf{y}_1^N \right.\right)$ and $\mathbb{E}\left[\mathbf{X}_{n+1} \left| r_{n+1}, \mathbf{y}_1^{n+1} \right.\right]$ from $p\left(r_n \left| \mathbf{y}_1^N \right.\right)$ and $\mathbb{E}\left[\mathbf{X}_n \left| r_n, \mathbf{y}_1^n \right.\right]$ and the model parameters using the procedure detailed in [9]. The main difference between CGOMSM and CGOMFSM is that in the case of fuzzy switches we involve continuous integration, requiring to be quantified with respect to the number of discrete fuzzy levels F.

Since $(\mathbf{R}_1^N, \mathbf{Y}_1^N)$ is a pairwise Markov chain in the model, we get

$$p\left(r_{n+1}\left|\mathbf{y}_1^{n+1}\right.\right) = \frac{\int_{[0,1]} p\left(r_{n+1}, \mathbf{y}_{n+1}\left|r_n = \nu, \mathbf{y}_n\right.\right) p\left(r_n = \nu\left|\mathbf{y}_1^n\right.\right) d\nu}{\int_{[0,1]}\int_{[0,1]} p(r_{n+1}^* = \upsilon, \mathbf{y}_{n+1}|r_n = \nu, \mathbf{y}_n) p\left(r_n = \nu\left|\mathbf{y}_1^n\right.\right) d\nu d\upsilon}, \quad (20)$$

and

$$p\left(r_n\left|r_{n+1}, \mathbf{y}_1^{n+1}\right.\right) = \frac{\int_{[0,1]} p\left(r_{n+1}, \mathbf{y}_{n+1}\left|r_n, \mathbf{y}_n\right.\right) p\left(r_n\left|\mathbf{y}_1^n\right.\right) d\nu}{\int_{[0,1]} p\left(r_{n+1}, \mathbf{y}_{n+1}\left|r_n^*, \mathbf{y}_n\right.\right) p\left(r_n^*\left|\mathbf{y}_1^n\right.\right) d\upsilon}. \quad (21)$$

Since

$$\mathbb{E}\left[\mathbf{X}_n\left|\mathbf{r}_n^{n+1}, \mathbf{y}_1^{n+1}\right.\right] = \mathbb{E}\left[\mathbf{X}_n\left|r_n, \mathbf{y}_1^n\right.\right], \quad (22)$$

and from (19), we can derive the following recursive equation:

$$\mathbb{E}\left[\mathbf{X}_{n+1}\left|r_{n+1}, \mathbf{y}_1^{n+1}\right.\right] = \int_{[0,1]} p\left(r_n\left|r_{n+1}, \mathbf{y}_1^{n+1}\right.\right) \times F_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})\mathbb{E}\left[\mathbf{X}_n\left|r_n, \mathbf{y}_1^n\right.\right] + H_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}) d\nu, \quad (23)$$

with $F_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})$ and $H_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})$ are adequate matrices. Probabilities $p(r_n|y_1^n)$ and $p(r_n, y_1^N)$ are recursively calculated in linear time using forward and backward probabilities in the Markov chain (Y_1^N, R_1^N) such that $\alpha_n(r_n) = p(r_n, y_1^n)$ and $\beta_n(r_n) = p(y_{n+1}^N|r_n, y_n)$.

$$\alpha_1(r_1) = p(r_1, y_1)$$

$$\alpha_{n+1}(r_{n+1}) = \int_{[0,1]} \alpha_n(\upsilon) p(r_{n+1}, \mathbf{y}_{n+1}|r_n, \mathbf{y}_n) d\upsilon \quad (24)$$

and

$$\beta_N(r_N) = 1$$

$$\beta_n(r_n) = \int_{[0,1]} \beta_{n+1}(\upsilon) p(r_{n+1}, \mathbf{y}_{n+1}|r_n, \mathbf{y}_n) d\upsilon \quad (25)$$

Using forward-backward probabilities, we can compute the smoothed and the filtered probabilities as follows::

$$p(r_n|\mathbf{y}_1^N) = \frac{\alpha_n(r_n)\beta_n(r_n)}{\int_{[0,1]} \alpha_n(\upsilon)\beta_n(\upsilon) d\upsilon} \quad (26)$$

$$p(r_n|\mathbf{y}_1^n) = \frac{\alpha_n(r_n)}{\int_{[0,1]} \alpha_n(\upsilon) d\upsilon} \quad (27)$$

Posterior marginal probabilities are calculated using the normalized Baum-Welch algorithm. The algorithm computes recursively the forward and backward probabilities. In the case of fuzzy switches, these probabilities are defined as follows:

$$\alpha_{n+1}(\delta) = \int_{[0,1]} \alpha_n(\theta) p\left(\mathbf{t}_{n+1}(\delta)\left|\mathbf{t}_n(\theta)\right.\right) d\theta \quad (28)$$

$$\beta_n(\delta) = \int_{[0,1]} \beta_{n+1}(\theta) p\left(\mathbf{t}_{n+1}(\delta)\left|\mathbf{t}_n(\theta)\right.\right) d\theta, \quad (29)$$

with $\mathbf{t}_n(\theta) = (\mathbf{x}_n, \mathbf{y}_n, r_n = \theta)$.
Then:

$$p(r_n, r_{n+1} | \mathbf{x}_1^N, \mathbf{y}_1^N) = \frac{\alpha_n(r_n) p(\mathbf{t}_{n+1} | \mathbf{t}_n) \beta_{n+1}(r_{n+1})}{\int_{[0,1]} \int_{[0,1]} \alpha_n(\delta) p(\mathbf{t}_{n+1} | \mathbf{t}_n) \beta_{n+1}(\theta) d\delta d\theta} \tag{30}$$

5 Experiments

In this section, we present two series of experiments to illustrate the smoother, in the case of scalar data ($m = q = 1$). In the first series we assess the performance of the fuzzy model with synthetic fuzzy signals; in the second series we apply our algorithm to smooth simulated Stochastic Volatility (SV) data. In both experiments, parameters estimation is carried out using EM algorithm using training samples denoted by $(\boldsymbol{x}_1^T, \boldsymbol{y}_1^T)$ of size T. Then we repeatedly generate, according to the considered model, synthetic sequences of size S denoted by $(\boldsymbol{x}_1^S, \boldsymbol{y}_1^S)$. Smoothing algorithm is then performed using estimated parameters to generate $\hat{\boldsymbol{x}}_1^S$ from the observed sequence $\boldsymbol{y}_1^S$. The criterion used to assess the efficiency of smoothing algorithms is the mean squared error (MSE) defined as follows:

$$MSE = \sum_{n=1}^{S} (x_n - \hat{x}_n)^2 \tag{31}$$

5.1 Smoothing synthetic fuzzy signals

We generate a fuzzy signal with 5 discrete fuzzy switches. Then we estimate the model considering different values of F ranging from 1 to 5. For each set of data, we consider 3 cases for the value of $r \in \{2, 5, 20\}$. For each value of r, we perform 10 independent experiments with generated signals of size $S = 1000$. For each experiment, we perform 100 EM iterations on independently generated samples $(\boldsymbol{X}_1^T, \boldsymbol{Y}_1^T)$ of size $T = 20000$. Figure 6 depicts an example of simulated data and the optimal (but approximated) smoothing output, and Table 2 reports the MSE results for different cases of fuzzy models and different values of r. The chief finding of this series of experiments is that when the number of fuzzy levels increases, the smoothed signal is closer to the "ground-truth" hidden signal: *e.g.* in case 3, for $r = 20$, the MSE decreases from 0.741 for $F = 0$ to 0.577 for $F = 5$.

5.2 Experiments on stochastic volatility models

Stochastic volatility (SV) models are widely used to highlight the variance of stochastic processes [10]. Several variants of SV models have been studied (Henston, CEV, GARCH, Chen, etc.). In this paper, we consider standard SV models defined as follows:

$$X_1 = \mu + U_1 \tag{32}$$

$$X_{n+1} = \mu + \phi(X_n - \mu) + \sigma U_{n+1} \tag{33}$$

$$Y_n = \beta \exp\left(\frac{X_n}{2}\right) V_n, \tag{34}$$

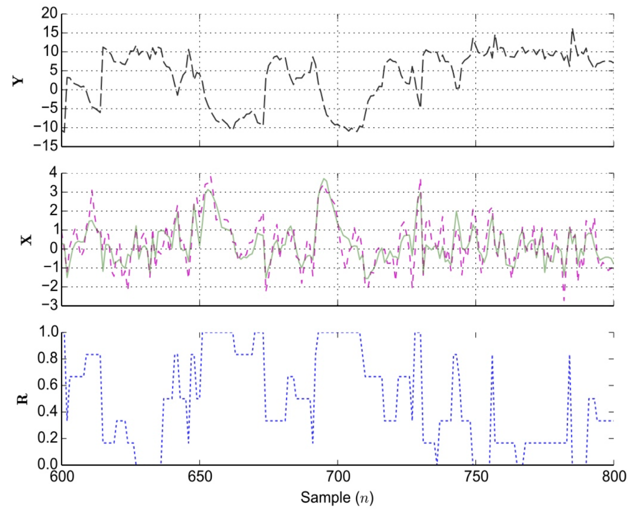

Fig. 6. A $(\mathbf{x}_1^N, \mathbf{R}_1^N, \mathbf{Y}_1^N)$ CGOMFSM trajectory, together with the automatically restored states in magenta (dashed).

Table 2. MSE results for different fuzzy signals with 5 discrete fuzzy levels and different values of r.

F	r	Case 1	Case 2	Case 3	Case 4	Case 5
0	2	1.395	1.155	0.854	1.009	1.291
	5	1.179	1.007	0.795	0.929	0.924
	20	0.688	0.72	**0.741**	0.954	0.565
1	2	1.264	1.029	0.797	0.93	1.146
	5	1.082	0.812	0.69	0.841	0.841
	20	0.574	0.697	0.654	0.812	0.518
2	2	1.197	0.977	0.767	0.888	1.098
	5	0.985	0.775	0.673	0.797	0.775
	20	0.552	0.613	0.623	0.722	0.484
3	2	1.138	0.931	0.745	0.845	1.093
	5	0.953	0.743	0.666	0.785	0.765
	20	0.542	0.58	0.579	0.72	0.468
5	2	1.121	0.913	0.738	0.838	1.064
	5	0.902	0.728	0.656	0.746	0.727
	20	0.519	0.489	**0.577**	0.698	0.404

where U_i, V_i are independent standard Gaussian vectors. The SV models is defined by the set of parameters σ, μ, and α. The main conclusion is that when the number of discrete fuzzy states increase, the model approaches the results of the optimal (but time consuming) particle smoother.

Table 3. MSE for SV models with $\mu = 0.5, \beta = 0.5$. To insure stationarity of models, $\phi^2 = 1 - \sigma^2$. PS column is the result from the particle smoother.

			F				
	ϕ	σ	0	1	3	5	PS
Case 1	0.99	0.141	0.395	0.233	0.144	0.124	0.12
Case 2	0.9	0.435	0.480	0.387	0.350	0.339	0.33
Case 3	0.8	0.6	0.557	0.496	0.474	0.467	0.46
Case 4	0.5	0.866	0.701	0.672	0.661	0.661	0.66

6 Conclusion

In this paper, we presented a novel approach to approximate non-linear Markov system using Conditionally Gaussian Observed Markov Fuzzy Switching Model (CGOMFSM). The chief novelty of this work is the introduction of fuzzy jumps instead of classical crisp states. This model still allows exact (up to required quantification) and fast smoothing equations. The fuzzy jumps allows transient modification of parameters, which is more appropriate for real-world applications. Future work includes the evaluation of the model for real data.

References

1. N. Abbassi, D. Benboudjema, S. Derrode, and W. Pieczynski, Optimal Filter Approximations in Conditionally Gaussian Pairwise Markov Switching Models, in IEEE Trans. on AC, Vol. 60(4), pp. 1104-1109, 2015.
2. F. Salzenstein, and C. Collet, Fuzzy Markov Random Fields versus Chains for Multispectral Image Segmentation, in IEEE Trans. on PAMI, Vol. 28(11), pp. 1753-1767, 2006.
3. H. Caillol, A. Hillion, and W. Pieczynski, Fuzzy Random Fields and Unsupervised Image Segmentation, in IEEE Trans. on GRS, Vol. 31(4), pp. 801-810, 1993.
4. I. Gorynin, S. Derrode, E. Monfrini, and W. Pieczynski, Exact Fast Smoothing in Switching Models with Application to Stochastic Volatility, EUSIPCO, Nice, France, pp. 924-928, 2015.
5. H. Caillol, W. Pieczynski, and A. Hillon, Estimation of Fuzzy Gaussian Mixture and Unsupervised Statistical Image Segmentation, IEEE Trans. on IP, Vol. 6(3), pp. 425-440, 1997.
6. Yang, Lei and He, Miao and Zhang, Junshan and Vittal, Vijay, Support Vector Machine Enhanced Markov Model for Short Term Wind Power Forecast, IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, VOL. 6, NO. 3, JULY 2015.
7. Artemiadis PK, Kyriakopoulos KJ. A switching regime model for the EMG-based control of a robot arm. IEEE Trans Syst Man Cybern B Cybern. 2011 Feb;41(1) 53-63. doi:10.1109/tsmcb.2010.2045120. PMID: 20403787.

8. S. Le Corff, G. Fort, and E. Moulines, Online Expectation Maximization algorithm to solve the SLAM problem, in Statistical Signal Processing Workshop (SSP),2011 IEEE, Nice, France, Jun., pp. 225 228.
9. I. Gorynin, S. Derrode, E. Monfrini and W. Pieczynski, Fast filtering in switching approximations of non-linear Markov switching systems with application to stochastic volatility, IEEE Trans on Automatic Control, accepted in May 2016.
10. E. Ghysels, A. Harvey, and E. Renault, Stochastic volatility, Handbook of Statistics, vol. 14, pp. 119 192, 1995.
11. Koko, M. , Application of Markov-Switching Model to Stock Returns Analysis, Dynamic Econometric Models, (2006) 7, 259268.
12. Baltzakis, Haris, Trahanias, Panos, A Hybrid Framework for Mobile Robot Localization: Formulation Using Switching State-Space Models, Autonomous Robots, 2003, Vol 15, N 2, 169-191.
13. F. Salzenstein and W. Pieczynski, Parameter estimation in hidden fuzzy Markov random fields and image segmentation, Graph. Mod. and Im. Proc., vol. 59, no. 4, pp. 205220, July 1997.
14. C. Carincotte, S. Derrode, G. Sicot, and J. M. Boucher, Unsupervised image segmentation based on a new fuzzy hidden Markov chain model, in IEEE Int. Conf. Acoust., Speech, Signal Processing, Montreal, Canada, May 17-21 2004.
15. C. Carincotte, S. Derrode et S. Bourennane, Unsupervised Change Detection on SAR Images using Fuzzy Hidden Markov Chains, in IEEE Trans. Geosci. Remote Sensing, vol. 44(2), p. 432-441, Feb. 2006
16. W. Pieczynski, Exact filtering in conditionally Markov switching hidden linear models, Comptes Rendus Mathématique, vol. 349, no. 9-10, pages 587 - 590, May 2011, 10.1016/j.crma.2011.02.007.
17. A. Gamal Eldin, F. Salzenstein and C. Collet, Hidden fuzzy Markov chain model with K discrete classes, Information Sciences Signal Processing and their Applications (ISSPA), mai 2010.