



# Optimal Filter Approximations in Conditionally Gaussian Pairwise Markov Switching Models

N Abbassi, D Benboudjema, Stéphane Derrode, W Pieczynski

## ► To cite this version:

N Abbassi, D Benboudjema, Stéphane Derrode, W Pieczynski. Optimal Filter Approximations in Conditionally Gaussian Pairwise Markov Switching Models. IEEE Transactions on Automatic Control, Institute of Electrical and Electronics Engineers, 2015, 60 (4), pp.1104-1109. 10.1109/TAC.2014.2340591 . hal-01157814

HAL Id: hal-01157814

<https://hal.archives-ouvertes.fr/hal-01157814>

Submitted on 28 May 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Optimal Filter Approximations in Conditionally Gaussian Pairwise Markov Switching Models

N. Abbassi, D. Benboudjema, S. Derrode and W. Pieczynski

**Abstract**—We consider a general triplet Markov Gaussian linear system  $(\mathbf{X}, \mathbf{R}, \mathbf{Y})$ , where  $\mathbf{X}$  is an hidden continuous random sequence,  $\mathbf{R}$  is an hidden discrete Markov chain,  $\mathbf{Y}$  is an observed continuous random sequence. When the triplet  $(\mathbf{X}, \mathbf{R}, \mathbf{Y})$  is a classical “Conditionally Gaussian Linear State-Space Model” (CGLSSM), the mean square error optimal filter is not workable with a reasonable complexity and different approximate methods, e.g. based on particle filters, are used. We propose two contributions. The first one is to extend the CGLSSM to a new, more general model, called the “Conditionally Gaussian Pairwise Markov Switching Model” (CGPMSM), in which  $\mathbf{X}$  is not necessarily Markov given  $\mathbf{R}$ . The second contribution is to consider a particular case of CGPMSM in which  $(\mathbf{R}, \mathbf{Y})$  is Markov and in which an exact filter, optimal in the sense of mean square error, can be performed with linear-time complexity. Some experiments show that the proposed method and the suited particle filter have comparable efficiency, while the second one is much faster.

**Index Terms**—Gaussian switching system, Exact optimal filtering, Hidden Markov models, Conditionally Gaussian Linear State-Space Model, Conditionally Gaussian Pairwise Markov Switching Model.

## I. INTRODUCTION

LET us consider three random sequences  $\mathbf{X}_1^N = (\mathbf{X}_1, \dots, \mathbf{X}_N)$ ,  $\mathbf{R}_1^N = (R_1, \dots, R_N)$  and  $\mathbf{Y}_1^N = (\mathbf{Y}_1, \dots, \mathbf{Y}_N)$ , where the sequences  $\mathbf{X}_1^N$  and  $\mathbf{Y}_1^N$  take their values in  $\mathbb{R}^m$  and  $\mathbb{R}^q$  respectively, while  $\mathbf{R}_1^N$  is discrete finite, each  $R_n$  taking its values in  $\Omega = \{0, \dots, K-1\}$ . Both  $\mathbf{X}_1^N$  and  $\mathbf{R}_1^N$  are hidden, while  $\mathbf{Y}_1^N$  is observed. The problem we deal with in this paper is the sequential search of  $(\mathbf{R}_1^N, \mathbf{X}_1^N)$  from  $\mathbf{Y}_1^N$ . The optimal filter is given by

$$\mathbb{E}[\mathbf{X}_{n+1} | \mathbf{y}_1^{n+1}] = \sum_{r_{n+1}} p(r_{n+1} | \mathbf{y}_1^{n+1}) \mathbb{E}[\mathbf{X}_{n+1} | r_{n+1}, \mathbf{y}_1^{n+1}],$$

and its variance is

$$\text{Var}[\mathbf{X}_{n+1} | \mathbf{y}_1^{n+1}] = \sum_{r_{n+1}} p(r_{n+1} | \mathbf{y}_1^{n+1}) \text{Var}[\mathbf{X}_{n+1} | r_{n+1}, \mathbf{y}_1^{n+1}].$$

So, we search to express  $p(r_{n+1} | \mathbf{y}_1^{n+1})$ ,  $\mathbb{E}[\mathbf{X}_{n+1} | r_{n+1}, \mathbf{y}_1^{n+1}]$  and  $\mathbb{E}[\mathbf{X}_{n+1} \mathbf{X}_{n+1}^t | r_{n+1}, \mathbf{y}_1^{n+1}]$  from  $p(r_n | \mathbf{y}_1^n)$ ,  $\mathbb{E}[\mathbf{X}_n | r_n, \mathbf{y}_1^n]$ ,  $\mathbb{E}[\mathbf{X}_n \mathbf{X}_n^t | r_n, \mathbf{y}_1^n]$  and  $\mathbf{y}_{n+1}$ . We assume that the random switch sequence  $\mathbf{R}_1^N$  is a Markov chain. Also, to simplify, we consider that random variables  $\mathbf{Y}_n$  and  $\mathbf{X}_n$  conditionally on switches have zero mean.

Our work is related to the so-called “Conditionally Gaussian Linear State-Space Models” (CGLSSMs), which have been widely applied in different situations. Due to their popularity in many different fields and, in particular, in different “tracking” problems [1], [2], speech processing problems [3], or biomedical engineering ones [4], these models are known under different names as “switching linear

dynamic systems” [4], “jump Markov linear systems” [5], “switching linear state-space models” [6], “conditional linear Gaussian models” [7], or still “conditionally Gaussian linear state-space models” [8].

These models combine two well-known and efficient models: a hidden Markov chain (with correlated noise) for the couple  $(\mathbf{R}_1^N, \mathbf{X}_1^N)$  and a Gaussian linear system for the distribution of  $(\mathbf{X}_1^N, \mathbf{Y}_1^N)$  conditional on  $\mathbf{R}_1^N$ . However, in spite of its intuitive and simple formulation, exact optimal filtering or smoothing is not workable with a reasonable computational time. Thus, a great deal of effort has been devoted to propose different approximate solutions, whether deterministic [9], [10] or stochastic [11], [12]. Among the latter, methods based on sequential Monte Carlo algorithms became very popular and can present quite satisfactory efficiency in numerous situations [5]. They are shown to be asymptotically optimal; however, a good choice of the importance measure, which is used to sample particles, can be difficult in complex situations. In addition, the requested number of particles quickly increases with the dimension  $m$  of the state space  $\mathbb{R}^m$ . Here we propose an alternative method, which consists in taking a different model for which the exact optimal filter can be computed with a complexity linear with respect to the observation time  $n$  by using the recent models in [13]. The difference between these recent models and the classical ones is the following. In the classical models the triplet  $(\mathbf{X}_1^N, \mathbf{R}_1^N, \mathbf{Y}_1^N)$  and the couple  $(\mathbf{X}_1^N, \mathbf{R}_1^N)$  are Markov, while the couple  $(\mathbf{R}_1^N, \mathbf{Y}_1^N)$  is not necessarily Markov. In the recent models the triplet  $(\mathbf{X}_1^N, \mathbf{R}_1^N, \mathbf{Y}_1^N)$  and the couple  $(\mathbf{R}_1^N, \mathbf{Y}_1^N)$  are Markov, while the couple  $(\mathbf{X}_1^N, \mathbf{R}_1^N)$  is not necessarily Markov.

More precisely, extending result in [14], we propose the two following contributions:

- (i) The CGLSSM is a linear Gaussian system conditionally on  $\mathbf{R}_1^N$  which verifies the following conditions:

$$\mathbf{R}_1^N \text{ is a Markov chain;} \quad (1)$$

$$\mathbf{X}_{n+1} = \mathbf{C}_{n+1}^1(R_{n+1}) \mathbf{X}_n + \mathbf{C}_{n+1}^2(R_{n+1}) \mathbf{U}_{n+1}; \quad (2)$$

$$\mathbf{Y}_{n+1} = \mathbf{C}_{n+1}^3(R_{n+1}) \mathbf{X}_{n+1} + \mathbf{C}_{n+1}^4(R_{n+1}) \mathbf{V}_{n+1}; \quad (3)$$

where  $\mathbf{C}_j^i(R_j)$ ,  $i = 1, \dots, 4$ ,  $j = 0, \dots, N-1$ , are appropriate matrices depending on the switches  $\mathbf{R}_1^N$ ,  $\mathbf{X}_0$  is given, and  $\mathbf{U}_1, \dots, \mathbf{U}_N$ ,  $\mathbf{V}_1, \dots, \mathbf{V}_N$  are sequences of independent centred Gaussian random vectors with unit variance-covariance matrices and such that  $\mathbf{U}_{n+1}$  and  $\mathbf{V}_{n+1}$  are independent from  $\mathbf{R}_1^n$  for each  $n = 1, \dots, N-1$ . Thus, given  $\mathbf{R}_1^N = \mathbf{r}_1^N$ , the couple  $(\mathbf{X}_1^N, \mathbf{Y}_1^N)$  is a classical Gaussian linear system in which the classical optimal Kalman filter can be applied. In such a system  $\mathbf{X}_1^N$  is linear Gaussian and Markovian, and the distribution of  $\mathbf{Y}_1^N$  given  $\mathbf{X}_1^N$  is very simple. We propose to extend this model to a more general one, in which, given  $\mathbf{R}_1^N = \mathbf{r}_1^N$ , the couple  $(\mathbf{X}_1^N, \mathbf{Y}_1^N)$  is a “Gaussian Pairwise Markov Model” (GPMM). Such a GPMM model, which can be of interest with respect to the classical model and in which neither  $\mathbf{X}_1^N$  nor  $\mathbf{Y}_1^N$  is necessarily Markovian, is more general

N. Abbassi is with Telecom SudParis (TSP), CITI Dpt (CNRS UMR 5157) Evry, France. noufel\_abbassi@yahoo.fr

D. Benboudjema is with TSP, CITI Dpt (CNRS UMR 5157) Evry, France. dbenboudjema@gmail.com

S. Derrode is with École Centrale Lyon, LIRIS, CNRS UMR 5205, 69134 Ecully, France. stephane.derrode@ec-lyon.fr

W. Pieczynski is with TSP, CITI Dpt (CNRS UMR 5157) Evry, France. wojciech.pieczynski@it-sudparis.eu

but still allows the use of exact Kalman filter [15], [16]. The general model  $(\mathbf{X}_1^N, \mathbf{R}_1^N, \mathbf{Y}_1^N)$  so obtained will be called ‘‘Conditionally Gaussian Pairwise Markov Switching Model’’ (CGPMSM);

- (ii) The classical CGLSSM is thus a particular case of the CGPMSM such that, given  $\mathbf{R}_1^N = \mathbf{r}_1^N$ ,  $\mathbf{X}_1^N$  is Markov and  $\mathbf{Y}_1^N$  is not necessarily Markov. We consider another recent particular case of the CGPMSM in which, given  $\mathbf{R}_1^N = \mathbf{r}_1^N$ ,  $\mathbf{Y}_1^N$  is Markov and  $\mathbf{X}_1^N$  is not necessarily Markov, which belongs to the family of models in [13]. The interest of this model, called ‘‘Conditionally Gaussian Observed Markov Switching Model’’ (CGOMSM), is that the exact optimal filtering is workable with linear complexity in time, even if  $\mathbf{R}_1^N = \mathbf{r}_1^N$  is not observed.

The organization of the paper is as follows. The next section is devoted to the new model we propose, extending the classical CGLSSM. Section III contains the description of a particular case called CGOMSM, which allows exact filtering in the presence of unknown switches. Experiments are presented in section IV and the last section contains conclusions and perspectives.

## II. THE CGPMSM AND THE GPMM

Let us note that the classical CGLSSM given with (1)-(3) can be reformulated as:

$$\begin{aligned} \mathbf{R}_1^N \text{ is a Markov chain;} \quad (4) \\ \begin{bmatrix} \mathbf{X}_{n+1} \\ \mathbf{Y}_{n+1} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{n+1}^1(R_{n+1}) & \mathbf{0} \\ \mathbf{A}_{n+1}^3(R_{n+1}) & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{bmatrix} + \\ \begin{bmatrix} \mathbf{B}_{n+1}^1(R_{n+1}) & \mathbf{0} \\ \mathbf{B}_{n+1}^3(R_{n+1}) & \mathbf{B}_{n+1}^4(R_{n+1}) \end{bmatrix} \begin{bmatrix} \mathbf{U}_{n+1} \\ \mathbf{V}_{n+1} \end{bmatrix}, \quad (5) \end{aligned}$$

with  $\mathbf{A}_{n+1}^1(R_{n+1}) = \mathbf{C}_{n+1}^1(R_{n+1})$ ,  $\mathbf{B}_{n+1}^1(R_{n+1}) = \mathbf{C}_{n+1}^2(R_{n+1})$ ,  $\mathbf{A}_{n+1}^3(R_{n+1}) = \mathbf{C}_{n+1}^3(R_{n+1})\mathbf{C}_{n+1}^1(R_{n+1})$ ,  $\mathbf{B}_{n+1}^3(R_{n+1}) = \mathbf{C}_{n+1}^3(R_{n+1})\mathbf{C}_{n+1}^2(R_{n+1})$  and  $\mathbf{B}_{n+1}^4(R_{n+1}) = \mathbf{C}_{n+1}^4(R_{n+1})$ .

Setting  $\mathbf{R}_n^{n+1} = (R_n, R_{n+1})$ , the classical CGLSSM given by (4)-(5) can be extended to the new CGPMSM according to:

$$\begin{aligned} (\mathbf{X}_1^N, \mathbf{R}_1^N, \mathbf{Y}_1^N) \text{ is Markovian and} \\ p(r_{n+1} | r_n, \mathbf{x}_n, \mathbf{y}_n) = p(r_{n+1} | r_n); \quad (6) \\ \begin{bmatrix} \mathbf{X}_{n+1} \\ \mathbf{Y}_{n+1} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{n+1}^1(\mathbf{R}_n^{n+1}) & \mathbf{A}_{n+1}^2(\mathbf{R}_n^{n+1}) \\ \mathbf{A}_{n+1}^3(\mathbf{R}_n^{n+1}) & \mathbf{A}_{n+1}^4(\mathbf{R}_n^{n+1}) \end{bmatrix} \begin{bmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{bmatrix} + \\ \begin{bmatrix} \mathbf{B}_{n+1}^1(\mathbf{R}_n^{n+1}) & \mathbf{B}_{n+1}^2(\mathbf{R}_n^{n+1}) \\ \mathbf{B}_{n+1}^3(\mathbf{R}_n^{n+1}) & \mathbf{B}_{n+1}^4(\mathbf{R}_n^{n+1}) \end{bmatrix} \begin{bmatrix} \mathbf{U}_{n+1} \\ \mathbf{V}_{n+1} \end{bmatrix}, \quad (7) \end{aligned}$$

with given  $\mathbf{X}_1, \mathbf{Y}_1$  and Gaussian distributions  $p(\mathbf{x}_1, \mathbf{y}_1 | r_1)$ , with  $\mathbf{A}_j^i(\mathbf{R}_j^{j+1}), \mathbf{B}_j^i(\mathbf{R}_j^{j+1})$  appropriate matrices depending on the switches and with  $\mathbf{W}_1^N = (\mathbf{W}_1, \dots, \mathbf{W}_N)$ , with  $\mathbf{W}_n = [\mathbf{U}_n \ \mathbf{V}_n]^t$  a sequence of independent centred Gaussian random vectors with unit variances and such that  $\mathbf{W}_{n+1}$  is independent from  $(\mathbf{X}_1^n, \mathbf{R}_1^n, \mathbf{Y}_1^n)$  for each  $n = 1, \dots, N-1$ .

*Remark:* There exists situations in which the observed chain and the hidden chain play symmetrical roles, and no one is a ‘‘natural’’ noisy version of the other. For example, let  $\mathbf{X}_1^N$  be unemployment and let  $\mathbf{Y}_1^N$  be inflation. It could be of interest to estimate  $\mathbf{X}_1^N$  from  $\mathbf{Y}_1^N$ , and it also could be of interest to estimate  $\mathbf{Y}_1^N$  from  $\mathbf{X}_1^N$ . Thus the general CGPMSM, which is symmetrical with respect to  $\mathbf{X}_1^N$  and  $\mathbf{Y}_1^N$ , would be better suited than the classical CGLSSM in such situation.

Given  $\mathbf{R}_1^N = \mathbf{r}_1^N$  the couple  $(\mathbf{X}_1^N, \mathbf{Y}_1^N)$  is then a ‘‘Gaussian Pairwise Markov Model’’ (GPMM) in which optimal Kalman filter can still be applied, as studied in a general context in [15], [16]. Let us briefly recall how it runs in a GPMM considered here. Forgetting the dependence of  $(\mathbf{X}_1^N, \mathbf{Y}_1^N)$  on the switches, the filtering problem

is to compute  $p(\mathbf{x}_{n+1} | \mathbf{y}_1^{n+1})$  from  $p(\mathbf{x}_n | \mathbf{y}_1^n)$  and  $\mathbf{y}_{n+1}$ . Let  $p(\mathbf{x}_n | \mathbf{y}_1^n) = \mathcal{N}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$ ,  $p(\mathbf{x}_{n+1} | \mathbf{y}_1^{n+1}) = \mathcal{N}(\boldsymbol{\mu}_{n+1}, \boldsymbol{\Sigma}_{n+1})$ , and  $\mathbf{W}_{n+1} = \begin{bmatrix} \mathbf{U}_{n+1} \\ \mathbf{V}_{n+1} \end{bmatrix}$ . For fixed  $\mathbf{y}_1^n$ , (7) means that the variables  $\mathbf{X}_n, \mathbf{X}_{n+1}$ , and  $\mathbf{Y}_{n+1}$  verify

$$\begin{bmatrix} \mathbf{X}_{n+1} \\ \mathbf{Y}_{n+1} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{n+1}^1 \\ \mathbf{A}_{n+1}^3 \end{bmatrix} \mathbf{X}_n + \begin{bmatrix} \mathbf{A}_{n+1}^2 \\ \mathbf{A}_{n+1}^4 \end{bmatrix} \mathbf{y}_n + \mathbf{B}_{n+1} \mathbf{W}_{n+1},$$

and thus  $p(\mathbf{x}_{n+1}, \mathbf{y}_{n+1} | \mathbf{y}_1^n)$  is Gaussian with mean  $\mathbf{m}_{n+1}$  and variance  $\mathbf{S}_{n+1}$  given by

$$\begin{bmatrix} \mathbf{m}_{n+1}^x \\ \mathbf{m}_{n+1}^y \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{n+1}^1 \\ \mathbf{A}_{n+1}^3 \end{bmatrix} \boldsymbol{\mu}_n + \begin{bmatrix} \mathbf{A}_{n+1}^2 \\ \mathbf{A}_{n+1}^4 \end{bmatrix} \mathbf{y}_n, \quad (8)$$

$$\begin{bmatrix} \mathbf{S}_{n+1}^{xx} & \mathbf{S}_{n+1}^{xy} \\ \mathbf{S}_{n+1}^{yx} & \mathbf{S}_{n+1}^{yy} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{n+1}^1 \\ \mathbf{A}_{n+1}^3 \end{bmatrix} \boldsymbol{\Sigma}_n \begin{bmatrix} \mathbf{A}_{n+1}^1 \\ \mathbf{A}_{n+1}^3 \end{bmatrix}^t + \mathbf{B}_{n+1} \mathbf{B}_{n+1}^t \quad (9)$$

Then  $p(\mathbf{x}_{n+1} | \mathbf{y}_1^{n+1}) = p(\mathbf{x}_{n+1} | \mathbf{y}_{n+1}, \mathbf{y}_1^n)$  is Gaussian with mean and variance given by

$$\begin{aligned} \boldsymbol{\mu}_{n+1} &= \mathbf{S}_{n+1}^{xy} (\mathbf{S}_{n+1}^{yy})^{-1} (\mathbf{y}_{n+1} - \mathbf{m}_{n+1}^y) + \mathbf{m}_{n+1}^x; \quad (10) \\ \boldsymbol{\Sigma}_{n+1} &= \mathbf{S}_{n+1}^{xx} - \mathbf{S}_{n+1}^{xy} (\mathbf{S}_{n+1}^{yy})^{-1} \mathbf{S}_{n+1}^{yx}. \quad (11) \end{aligned}$$

Thus (8)-(11) define the Kalman filter in GPMM, and it can be applied in CGPMSM (6)-(7) if the switches  $\mathbf{R}_1^N = \mathbf{r}_1^N$  are known. Once they are not known,  $p(r_{n+1} | \mathbf{y}_1^{n+1})$  cannot be computed sequentially and has to be approximated using, *e.g.*, a particle filter.

## III. THE CGOMSM AND RELATED EXACT FILTER

In this section we introduce a new model, which is a particular case of the CGPMSM given by (6)-(7) and in which the exact optimal filter can be computed with linear complexity in time. Called ‘‘Conditionally Gaussian Observed Markov Switching Model’’ (CGOMSM), this new model is a CGPMSM such that  $\mathbf{A}_{n+1}^3(\mathbf{R}_n^{n+1}) = \mathbf{0}$ . More precisely, we define a CGOMSM as a triplet Markov chain  $(\mathbf{X}_1^N, \mathbf{R}_1^N, \mathbf{Y}_1^N)$  verifying :

$$\begin{aligned} (\mathbf{X}_1^N, \mathbf{R}_1^N, \mathbf{Y}_1^N) \text{ is Markovian and} \\ p(r_{n+1} | r_n, \mathbf{x}_n, \mathbf{y}_n) = p(r_{n+1} | r_n); \quad (12) \\ \begin{bmatrix} \mathbf{X}_{n+1} \\ \mathbf{Y}_{n+1} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{n+1}^1(\mathbf{R}_n^{n+1}) & \mathbf{A}_{n+1}^2(\mathbf{R}_n^{n+1}) \\ \mathbf{0} & \mathbf{A}_{n+1}^4(\mathbf{R}_n^{n+1}) \end{bmatrix} \begin{bmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{bmatrix} + \\ \underbrace{\begin{bmatrix} \mathbf{B}_{n+1}^1(\mathbf{R}_n^{n+1}) & \mathbf{B}_{n+1}^2(\mathbf{R}_n^{n+1}) \\ \mathbf{B}_{n+1}^3(\mathbf{R}_n^{n+1}) & \mathbf{B}_{n+1}^4(\mathbf{R}_n^{n+1}) \end{bmatrix}}_{\mathbf{B}_{n+1}(\mathbf{R}_n^{n+1})} \begin{bmatrix} \mathbf{U}_{n+1} \\ \mathbf{V}_{n+1} \end{bmatrix}. \quad (13) \end{aligned}$$

Let

$$\begin{aligned} \mathbf{Q}_{n+1}(\mathbf{r}_n^{n+1}) &= \begin{bmatrix} \mathbf{Q}_{n+1}^1(\mathbf{r}_n^{n+1}) & \mathbf{Q}_{n+1}^2(\mathbf{r}_n^{n+1}) \\ \mathbf{Q}_{n+1}^3(\mathbf{r}_n^{n+1}) & \mathbf{Q}_{n+1}^4(\mathbf{r}_n^{n+1}) \end{bmatrix} \\ &= \mathbf{B}_{n+1}(\mathbf{r}_n^{n+1}) \mathbf{B}_{n+1}^t(\mathbf{r}_n^{n+1}). \quad (14) \end{aligned}$$

Let us derive how the exact fast (with complexity linear in time) optimal filter can be computed in any CGOMSM. The computation of initial  $p(r_1 | \mathbf{y}_1)$ ,  $E[\mathbf{X}_1 | r_1, \mathbf{y}_1]$ ,  $E[\mathbf{X}_1 \mathbf{X}_1^t | r_1, \mathbf{y}_1]$ , and  $\text{Var}[\mathbf{X}_1 | \mathbf{y}_1]$  from  $p(r_1)$  and  $p(\mathbf{x}_1, \mathbf{y}_1 | r_1)$  is trivial. For  $n = 1, \dots, N-1$ , we can express  $p(r_{n+1} | \mathbf{y}_1^{n+1})$ ,  $E[\mathbf{X}_{n+1} | r_{n+1}, \mathbf{y}_1^{n+1}]$  and  $E[\mathbf{X}_{n+1} \mathbf{X}_{n+1}^t | r_{n+1}, \mathbf{y}_1^{n+1}]$  from  $p(r_n | \mathbf{y}_1^n)$ ,  $E[\mathbf{X}_n | r_n, \mathbf{y}_1^n]$  and  $E[\mathbf{X}_n \mathbf{X}_n^t | r_n, \mathbf{y}_1^n]$ ,  $p(r_{n+1}, \mathbf{y}_{n+1} | r_n, \mathbf{y}_n)$  and  $\mathbf{y}_{n+1}$ , using simple relations given in introduction.

An important point, which makes the difference between CGOMSM and classical models, is that, according to (12)-(13),  $(\mathbf{R}_1^N, \mathbf{Y}_1^N)$  is a Markov chain. In fact, as we are going to see, this allows the exact computation of  $p(r_{n+1} | \mathbf{y}_1^{n+1})$ , which is not possible in classical models and which needs to be approximated with a particle filter for example.

According to (12)-(13),  $p(r_{n+1} | r_n, \mathbf{y}_n) = p(r_{n+1} | r_n)$ , and thus we have

$$p(r_{n+1}, \mathbf{y}_{n+1} | r_n, \mathbf{y}_n) = p(r_{n+1} | r_n) p(\mathbf{y}_{n+1} | \mathbf{r}_n^{n+1}, \mathbf{y}_n), \quad (15)$$

where

$$p(\mathbf{y}_{n+1} | \mathbf{r}_n^{n+1}, \mathbf{y}_n) = \mathcal{N}(\mathbf{A}_{n+1}^4(\mathbf{r}_n^{n+1})\mathbf{y}_n, \mathbf{Q}_{n+1}^4(\mathbf{r}_n^{n+1})). \quad (16)$$

Besides, since  $(\mathbf{R}_1^N, \mathbf{Y}_1^N)$  is Markov, we have

$$p(r_n, r_{n+1} | \mathbf{y}_1^{n+1}) = \frac{p(r_{n+1}, \mathbf{y}_{n+1} | r_n, \mathbf{y}_n) p(r_n | \mathbf{y}_1^n)}{\sum_{r_n, r_{n+1}} p(r_{n+1}, \mathbf{y}_{n+1} | r_n, \mathbf{y}_n) p(r_n | \mathbf{y}_1^n)},$$

and thus

$$p(r_n | r_{n+1}, \mathbf{y}_1^{n+1}) = \frac{p(r_{n+1}, \mathbf{y}_{n+1} | r_n, \mathbf{y}_n) p(r_n | \mathbf{y}_1^n)}{\sum_{r_n} p(r_{n+1}, \mathbf{y}_{n+1} | r_n, \mathbf{y}_n) p(r_n | \mathbf{y}_1^n)}. \quad (17)$$

According to (13), conditionally on  $\mathbf{R}_1^N = \mathbf{r}_1^N$ , the vector  $(\mathbf{X}_{n+1}, \mathbf{Y}_{n+1})$  depends on  $(\mathbf{X}_n, \mathbf{Y}_n)$  linearly and is Gaussian (please note that  $(\mathbf{X}_n, \mathbf{Y}_n)$  is not Gaussian as its distribution is a Gaussian mixture). This implies that the Gaussian  $\mathbf{X}_{n+1}$  also depends on  $(\mathbf{X}_n, \mathbf{Y}_n, \mathbf{Y}_{n+1})$  linearly and thus can be written as  $\mathbf{X}_{n+1} = \mathbf{F}_{n+1}(\mathbf{R}_n^{n+1})\mathbf{X}_n + \mathbf{I}_{n+1}(\mathbf{R}_n^{n+1})\mathbf{Y}_n + \mathbf{J}_{n+1}(\mathbf{R}_n^{n+1})\mathbf{Y}_{n+1} + \mathbf{G}_{n+1}(\mathbf{R}_n^{n+1})\mathbf{W}_{n+1}$ , with  $\mathbb{E}[\mathbf{W}_{n+1} | r_{n+1}, \mathbf{y}_1^{n+1}] = \mathbb{E}[\mathbf{W}_{n+1}] = 0$  and  $\mathbf{F}_{n+1}, \mathbf{I}_{n+1}, \mathbf{J}_{n+1}, \mathbf{G}_{n+1}$  expressed from parameters in (13) by eq. (22)-(25) specified below. To simplify notations let us set

$$\mathbf{H}_{n+1}(\mathbf{R}_n^{n+1}, \mathbf{Y}_n^{n+1}) = \mathbf{I}_{n+1}(\mathbf{R}_n^{n+1})\mathbf{Y}_n + \mathbf{J}_{n+1}(\mathbf{R}_n^{n+1})\mathbf{Y}_{n+1}, \quad (18)$$

so that we finally have

$$\mathbf{X}_{n+1} = \mathbf{F}_{n+1}(\mathbf{R}_n^{n+1})\mathbf{X}_n + \mathbf{G}_{n+1}(\mathbf{R}_n^{n+1})\mathbf{W}_{n+1} + \mathbf{H}_{n+1}(\mathbf{R}_n^{n+1}, \mathbf{Y}_n^{n+1}). \quad (19)$$

So the filter is given by

$$\mathbb{E}[\mathbf{X}_{n+1} | r_{n+1}, \mathbf{y}_1^{n+1}] = \sum_{r_n} \left( p(r_n | r_{n+1}, \mathbf{y}_1^{n+1}) [\mathbf{F}_{n+1}(\mathbf{r}_n^{n+1})\mathbb{E}[\mathbf{X}_n | \mathbf{r}_n^{n+1}, \mathbf{y}_1^{n+1}] + \mathbf{H}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})] \right).$$

Using (13), we can show that  $(R_{n+1}, \mathbf{Y}_{n+1})$  and  $\mathbf{X}_n$  are independent conditionally on  $(R_n, \mathbf{Y}_n)$ . Thus we have  $\mathbb{E}[\mathbf{X}_n | \mathbf{r}_n^{n+1}, \mathbf{y}_1^{n+1}] = \mathbb{E}[\mathbf{X}_n | r_n, \mathbf{y}_1^n]$ , and finally get

$$\mathbb{E}[\mathbf{X}_{n+1} | r_{n+1}, \mathbf{y}_1^{n+1}] = \sum_{r_n} \left( p(r_n | r_{n+1}, \mathbf{y}_1^{n+1}) [\mathbf{F}_{n+1}(\mathbf{r}_n^{n+1})\mathbb{E}[\mathbf{X}_n | r_n, \mathbf{y}_1^n] + \mathbf{H}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})] \right) \quad (20)$$

$\mathbb{E}[\mathbf{X}_{n+1}\mathbf{X}_{n+1}^t | r_{n+1}, \mathbf{y}_1^{n+1}]$  is computed in a similar way:

$$\mathbb{E}[\mathbf{X}_{n+1}\mathbf{X}_{n+1}^t | r_{n+1}, \mathbf{y}_1^{n+1}] = \sum_{r_n} \left( p(r_n | r_{n+1}, \mathbf{y}_1^{n+1}) [\mathbf{F}_{n+1}(\mathbf{r}_n^{n+1})\mathbb{E}[\mathbf{X}_n\mathbf{X}_n^t | r_n, \mathbf{y}_1^n] \mathbf{F}_{n+1}^t(\mathbf{r}_n^{n+1}) + \mathbf{F}_{n+1}(\mathbf{r}_n^{n+1})\mathbb{E}[\mathbf{X}_n | r_n, \mathbf{y}_1^n] \mathbf{H}_{n+1}^t(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}) + \mathbf{H}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})\mathbb{E}[\mathbf{X}_n^t | r_n, \mathbf{y}_1^n] \mathbf{F}_{n+1}^t(\mathbf{r}_n^{n+1}) + \mathbf{G}_{n+1}(\mathbf{r}_n^{n+1})\mathbf{G}_{n+1}^t(\mathbf{r}_n^{n+1}) + \mathbf{H}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})\mathbf{H}_{n+1}^t(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})] \right). \quad (21)$$

Finally, let us specify  $\mathbf{F}_{n+1}, \mathbf{I}_{n+1}, \mathbf{J}_{n+1}$ , and  $\mathbf{G}_{n+1}$ . According to (13),  $p(\mathbf{x}_{n+1}, \mathbf{y}_{n+1} | \mathbf{x}_n, \mathbf{r}_n^{n+1}, \mathbf{y}_n)$  is Gaussian with mean

$\left[ \mathbf{A}_{n+1}^1(\mathbf{r}_n^{n+1})\mathbf{x}_n + \mathbf{A}_{n+1}^2(\mathbf{r}_n^{n+1})\mathbf{y}_n \right]$  and variance-covariance matrix  $\mathbf{Q}_{n+1}(\mathbf{r}_n^{n+1})$  in eq. (14).

Then  $p(\mathbf{x}_{n+1} | \mathbf{r}_n^{n+1}, \mathbf{x}_n, \mathbf{y}_n, \mathbf{y}_{n+1})$  is Gaussian with mean  $\mathbf{A}_{n+1}^1(\mathbf{r}_n^{n+1})\mathbf{x}_n + \mathbf{A}_{n+1}^2(\mathbf{r}_n^{n+1})\mathbf{y}_n + \mathbf{Q}_{n+1}^2(\mathbf{r}_n^{n+1})(\mathbf{Q}_{n+1}^4(\mathbf{r}_n^{n+1}))^{-1}(\mathbf{y}_{n+1} - \mathbf{A}_{n+1}^4(\mathbf{r}_n^{n+1})\mathbf{y}_n)$  and variance-covariance matrix  $\mathbf{Q}_{n+1}^1(\mathbf{r}_n^{n+1}) - \mathbf{Q}_{n+1}^2(\mathbf{r}_n^{n+1})(\mathbf{Q}_{n+1}^4(\mathbf{r}_n^{n+1}))^{-1}\mathbf{Q}_{n+1}^3(\mathbf{r}_n^{n+1})$ . We see that

$$\mathbf{F}_{n+1}(\mathbf{r}_n^{n+1}) = \mathbf{A}_{n+1}^1(\mathbf{r}_n^{n+1}); \quad (22)$$

$$\mathbf{J}_{n+1}(\mathbf{r}_n^{n+1}) = \mathbf{Q}_{n+1}^2(\mathbf{r}_n^{n+1})(\mathbf{Q}_{n+1}^4(\mathbf{r}_n^{n+1}))^{-1}; \quad (23)$$

$$\mathbf{I}_{n+1}(\mathbf{r}_n^{n+1}) = \mathbf{A}_{n+1}^2(\mathbf{r}_n^{n+1}) - \mathbf{J}_{n+1}(\mathbf{r}_n^{n+1})\mathbf{A}_{n+1}^4(\mathbf{r}_n^{n+1}); \quad (24)$$

$$\mathbf{G}_{n+1}(\mathbf{r}_n^{n+1})\mathbf{G}_{n+1}^t(\mathbf{r}_n^{n+1}) = \mathbf{Q}_{n+1}^1(\mathbf{r}_n^{n+1}) - \mathbf{J}_{n+1}(\mathbf{r}_n^{n+1})\mathbf{Q}_{n+1}^3(\mathbf{r}_n^{n+1}). \quad (25)$$

So eq. (18) writes

$$\mathbf{H}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}) = \mathbf{A}_{n+1}^2(\mathbf{r}_n^{n+1})\mathbf{y}_n + \mathbf{J}_{n+1}(\mathbf{r}_n^{n+1})(\mathbf{y}_{n+1} - \mathbf{A}_{n+1}^4(\mathbf{r}_n^{n+1})\mathbf{y}_n). \quad (26)$$

Finally, let us consider the CGOMSM model (12)-(13). The proposed filter runs as follows. For  $n = 1, \dots, N - 1$

- 1) Compute  $\mathbf{Q}_{n+1}(\mathbf{r}_n^{n+1})$  matrix from (14);
- 2) Compute  $p(r_{n+1}, \mathbf{y}_{n+1} | r_n, \mathbf{y}_n) = p(r_{n+1} | r_n) p(\mathbf{y}_{n+1} | \mathbf{r}_n^{n+1}, \mathbf{y}_n)$  from  $p(r_{n+1} | r_n)$  and the Gaussian distribution  $p(\mathbf{y}_{n+1} | \mathbf{r}_n^{n+1}, \mathbf{y}_n)$  given by (16);
- 3) Compute  $p(r_n | r_{n+1}, \mathbf{y}_1^{n+1})$  from (17);
- 4) Compute  $\mathbf{F}_{n+1}(\mathbf{r}_n^{n+1})$  from (22),  $\mathbf{J}_{n+1}(\mathbf{r}_n^{n+1})$  from (23),  $\mathbf{G}_{n+1}(\mathbf{r}_n^{n+1})\mathbf{G}_{n+1}^t(\mathbf{r}_n^{n+1})$  from (25) and  $\mathbf{H}_{n+1}(\mathbf{R}_n^{n+1}, \mathbf{Y}_n^{n+1})$  from (26);
- 5) Compute  $\mathbb{E}[\mathbf{X}_{n+1} | r_{n+1}, \mathbf{y}_1^{n+1}]$  from (20) and  $\mathbb{E}[\mathbf{X}_{n+1}\mathbf{X}_{n+1}^t | r_{n+1}, \mathbf{y}_1^{n+1}]$  from (21).

As no computation above depends on  $n$ , we see that the filter running time is linear in the number of observations.

#### IV. EXPERIMENTS

We present two series of experiments, where, in both of them,  $\mathbf{X}_1^N$  and  $\mathbf{Y}_1^N$  are real valued. In the first series, we try to understand how far -when the optimal filtering is concerned- the corresponding CGOMSM, obtained from CGPMSM by modifying some parameters, is from the true CGPMSM. To study this point, data are sampled with a CGPMSM and then filtered using the same CGPMSM on the one hand, and a CGOMSM on the other hand. In the second series of experiments, we examine the behaviour of the CGOMSM based filter with respect to the classical CGLSSM based one.

Therefore, consider  $\mathbf{X}_1^N$  and  $\mathbf{Y}_1^N$  as scalar-valued processes ( $m = q = 1$ ). In all experiments, we consider homogeneous systems, which means that matrices  $\mathbf{A}_{n+1}^i, \mathbf{B}_{n+1}^i$ , for  $i = 1, \dots, 4$ , in (7) do not depend on  $n$ . Let us briefly examine the differences between a general CGPMSM and the ‘‘corresponding’’ CGOMSM, obtained from the considered CGPMSM in some way. The distributions of both of them are thus defined by the distribution of  $(T_1, T_2) = (\mathbf{X}_1, R_1, \mathbf{Y}_1, \mathbf{X}_2, R_2, \mathbf{Y}_2)$ , and this distribution can be written as  $p(t_1, t_2) = p(r_1, r_2) p(\mathbf{x}_1, \mathbf{y}_1, \mathbf{x}_2, \mathbf{y}_2 | r_1, r_2)$ . As the distribution of the Markov chain  $\mathbf{R}_1^N$  will be the same in the CGPMSM and in the corresponding CGOMSM, we have to compare  $p(\mathbf{x}_1, \mathbf{y}_1, \mathbf{x}_2, \mathbf{y}_2 | r_1, r_2)$ . Finally, let us temporarily forget the conditioning by  $(R_1, R_2)$  and compare the Gaussian distributions  $p(\mathbf{x}_1, \mathbf{y}_1, \mathbf{x}_2, \mathbf{y}_2)$ .

Let  $\mathbf{Z}_1^2 = (\mathbf{Z}_1, \mathbf{Z}_2) = (\mathbf{X}_1, \mathbf{Y}_1, \mathbf{X}_2, \mathbf{Y}_2)$ . To simplify, let assume that all means are zero and all variances are 1. The Gaussian distribution of  $\mathbf{Z}_1^2$  is then given by the covariance matrix

$$\mathbf{\Gamma}^{\mathbf{Z}_1^2} = \begin{bmatrix} 1 & b & a & d \\ b & 1 & e & c \\ a & e & 1 & b \\ d & c & b & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{\Gamma} & \mathbf{\Sigma}^t \\ \mathbf{\Sigma} & \mathbf{\Gamma} \end{bmatrix}. \quad (27)$$

Thus the distribution of  $\mathbf{Z}_1^2$  is defined by five co-variances  $a, b, c, d$ , and  $e$ , with the condition that  $\mathbf{\Gamma}^{\mathbf{Z}_1^2}$  is definite positive. The dependence graph is presented in Fig. 1.

The corresponding equation (5) is here of the simple form

$$\begin{bmatrix} \mathbf{X}_{n+1} \\ \mathbf{Y}_{n+1} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^1 & \mathbf{A}^2 \\ \mathbf{A}^3 & \mathbf{A}^4 \end{bmatrix} \begin{bmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{bmatrix} + \begin{bmatrix} \mathbf{B}^1 & \mathbf{B}^2 \\ \mathbf{B}^3 & \mathbf{B}^4 \end{bmatrix} \begin{bmatrix} \mathbf{U}_{n+1} \\ \mathbf{V}_{n+1} \end{bmatrix}, \quad (28)$$

with

$$\begin{bmatrix} \mathbf{A}^1 & \mathbf{A}^2 \\ \mathbf{A}^3 & \mathbf{A}^4 \end{bmatrix} = \mathbf{\Sigma} \mathbf{\Gamma}^{-1} = \frac{1}{1-b^2} \begin{bmatrix} a-eb & e-ab \\ d-cb & c-db \end{bmatrix}, \quad (29)$$

$$\begin{bmatrix} \mathbf{B}^1 & \mathbf{B}^2 \\ \mathbf{B}^3 & \mathbf{B}^4 \end{bmatrix} \begin{bmatrix} \mathbf{B}^1 & \mathbf{B}^2 \\ \mathbf{B}^3 & \mathbf{B}^4 \end{bmatrix}^t = \mathbf{\Gamma} - \mathbf{\Sigma} \mathbf{\Gamma}^{-1} \mathbf{\Sigma}^t. \quad (30)$$

Thus we can say that the CGPMSM model is a CGOMSM model when  $\mathbf{A}^3 = 0$ , which is equivalent to  $d-cb = 0$ . To obtain the latter, we can modify either  $b, c$ , or  $d$ .

Consider now the complete homogeneous CGPMSM whose distribution is defined by  $p(r_1, r_2)$  and  $\mathbf{\Gamma}^{\mathbf{Z}_1^2}(r_1^2)$ , with  $a, b, c, d$ , and  $e$  replaced by  $a(r_1^2), b(r_1^2), c(r_1^2), d(r_1^2)$ , and  $e(r_1^2)$ . Given  $(R_1, R_2) = (r_1, r_2)$ , the co-variances  $a(r_1^2), b(r_1^2), c(r_1^2)$  and  $e(r_1^2)$  can be the same in both CGPMSM and CGOMSM distributions, and the only difference can be at the  $d(r_1^2)$  value level: it is of any value in CGPMSM while it is equal to  $c(r_1^2)b(r_1^2)$  in CGOMSM. In other words, the distributions of  $(R_1, R_2), (\mathbf{X}_1, \mathbf{Y}_1, \mathbf{X}_2)$ , and  $(\mathbf{Y}_1, \mathbf{X}_2, \mathbf{Y}_2)$  are strictly the same in the CGPMSM considered and in its ‘‘approximate’’ CGOMSM version.

In all experiments below, we will consider two switches  $\Omega = \{0, 1\}$ , and the following simplified parameter set

- (i)  $p(r_1, r_2) = p(r_1)p(r_2|r_1)$  with  $p(r_1=0) = p(r_1=1) = 0.5$ , and  $p(r_2=0|r_1=0) = p(r_2=1|r_1=1) = q$ ;
- (ii)  $b(r_1^2) = b(r_2), a(r_1^2) = a(r_2), c(r_1^2) = c(r_2), d(r_1^2) = d(r_2)$  and  $e(r_1^2) = e(r_2)$ .

Finally, each CGPMM used in experiments will be defined by eleven parameters:  $q, b(0), a(0), c(0), d(0), e(0), b(1), a(1), c(1), d(1)$ , and  $e(1)$ . We will consider the following slightly simplified model

$$\begin{bmatrix} \mathbf{A}^1(\mathbf{r}_n^{n+1}) & \mathbf{A}^2(\mathbf{r}_n^{n+1}) \\ \mathbf{A}^3(\mathbf{r}_n^{n+1}) & \mathbf{A}^4(\mathbf{r}_n^{n+1}) \end{bmatrix} = \begin{bmatrix} \alpha(r_{n+1}) & \beta(r_{n+1}) \\ \gamma(r_{n+1}) & \delta(r_{n+1}) \end{bmatrix}. \quad (31)$$

In the two series of experiments below, the number of samples was set to  $N = 2000$  and results are means of 200 independent experiments.

#### A. First series of experiments

The aim of this series is to simulate realizations of  $\mathbf{T}_1^N = (\mathbf{X}_1^N, \mathbf{R}_1^N, \mathbf{Y}_1^N)$  according to the general CGPMSM, and to estimate  $(\mathbf{R}_1^N, \mathbf{X}_1^N)$  from  $\mathbf{Y}_1^N$  with different particular methods, seen as approximate methods of the optimal one. We considered five methods:

- (i) the ‘‘Reference Method’’, denoted by RM-KS (KS stands for ‘‘Known Switches’’), where  $\mathbf{R}_1^N = \mathbf{r}_1^N$  is considered as known and where the estimated  $\mathbf{X}_1^N = \hat{\mathbf{x}}_1^N$  is obtained by the optimal Kalman filter given by (8)-(11);
- (ii) the particle filter suited to CGPMSM, by extension of [5], denoted by RM-US (US stands for ‘‘Unknown Switches’’);

- (iii) the new alternative method we propose, denoted by M1-US, by considering that  $\gamma(0) = \gamma(1) = 0$ , and by applying the exact filter described in section III;
- (iv) the Kalman filter based on CGOMSM and known switches, denoted by M1-KS;
- (v) the Kalman filter based on the classic CGLSSM (see below) and known switches, denoted by M3-KS.

To further simplify, we consider here that  $b(0) = b(1) = b$ . To obtain the classical CGLSSM-based filter in (v), we consider, in (29) and (30), that  $e(r_n) - ba(r_n) = 0, c(r_n) - bd(r_n) = 0$ , and  $c(r_n) = a(r_n)b^2$ , which are its classical properties (see Fig. 1).

In all models studied here, we set  $q = 0.9, b = 0.3, a(0) = 0.1, a(1) = 0.5, c(0) = 0.4, c(1) = 0.9, e(0) = 0.75$ , and  $e(1) = 0.33$ . Then we chose 9 different couples  $(d^i(0), d^i(1))$ , defining 9 different CGPMSM, in such a way that the corresponding  $\gamma = \gamma(0) = \gamma(1)$  (see (31)) vary from 0 to 0.4 with a step of 0.05. Therefore we increase  $\gamma$  on purpose: the larger  $\gamma$  the greater the difference between the model CGPMSM used to simulate data and the model used to restore them with M1-US (*i.e.*, the model CGOMSM obtained by setting  $\gamma = 0$ ). We also filter data with M3-KS, where the CGLSSM used is obtained in the following way. First we set  $q = 0.9, b = 0.3, a(0) = 0.1$ , and  $a(1) = 0.5$ . Then, as known in the classical CGLSSM (see Fig. 1), we also set  $e(0) = d(0) = a(0)b, c(0) = a(0)b^2, e(1) = d(1) = a(1)b$ , and  $c(1) = a(1)b^2$ .

Thus, in each case,  $\mathbf{t}_1^N = (\mathbf{x}_1^N, \mathbf{r}_1^N, \mathbf{y}_1^N)$  have been simulated according to the real parameters and  $\mathbf{x}_1^N$  is estimated using RM-KS, RM-US, M1-KS, M1-US, and M3-KS. The difference between  $\mathbf{x}_1^N$  and the estimated  $\hat{\mathbf{x}}_1^N$  is calculated with the Mean Squared Error (MSE)  $\frac{1}{N} \sum_{n=1}^N (\mathbf{x}_n - \hat{\mathbf{x}}_n)^2$ . The number of particle for the particle filter method depends on the model and on the parameters of the model. Experiments not reported here lead to a number  $N_P = 200$  of particles. We make use of Sequential Importance Resampling (SIR) when  $\hat{N}_{eff} = \frac{1}{\sum_{j=1}^{N_P} (w_j)^2} < \frac{1}{3}$ , where  $w_j$  is the weight of particle  $j$ .

Results are presented in Fig. 2. Computation time for the particle filter method (RM-US) is approximately 1.47 seconds, whereas it is 0.014 second for the exact CGOMSM method (M1-US). Generally speaking, the computation time is about half time the number of particles longer for RM-US than for M1-US. From these results, we can put forth the following conclusions:

- 1) The main conclusion is that when  $\gamma$  increases, *i.e.* when the CGOMSM used by method M1-KS and the reference CGPMSM used to sample data break away, the results obtained using M1-KS remain very close to those obtained using RM-KS based on known switches (curves are nearly superposed). This means that M1-KS algorithm is very robust with respect to  $\gamma$ ;
- 2) The particle filter applied to the true CGPMSM model gives quite satisfactory results. They are nearly equivalent to those obtained with M1-US method but with a computation time about 100 times shorter;
- 3) Comparing M1-KS with M1-US shows that the new method seems quite robust with respect to the knowledge / ignorance of the switches (even if the switches are rather poorly estimated);
- 4) The results obtained with M3-KS based on the classical CGLSSM and known switches are quite poor, which means that, at least in the context considered, CGLSSM is a rather poor approximation of the general CGPMSM;
- 5) Both RM-US and M1-US (in which  $(\mathbf{R}_1^N, \mathbf{Y}_1^N)$  is Markovian) allow to estimate  $\mathbf{r}_1^N$ : we find about 20% of error ratios in both estimators, and the results vary little with  $\gamma$ .
- 6) Variation of  $\gamma$  has very little influence on the results.

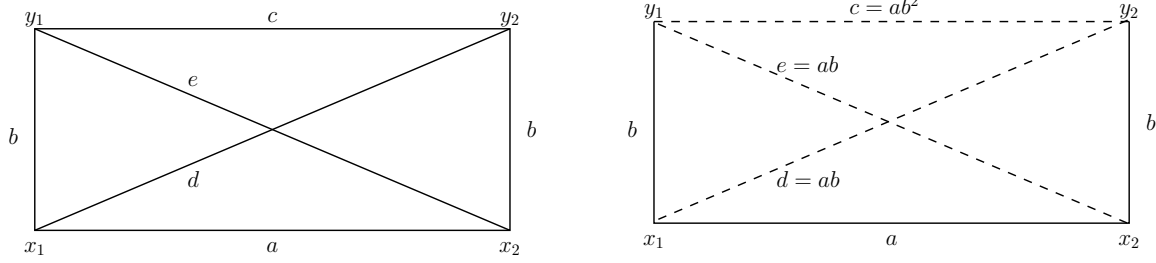


Fig. 1. Dependence graph of  $p(\mathbf{x}_1, \mathbf{y}_1, \mathbf{x}_2, \mathbf{y}_2)$  in the general CGPMSM (left - 5 parameters) and in the classical CGLSSM (right - 2 parameters).

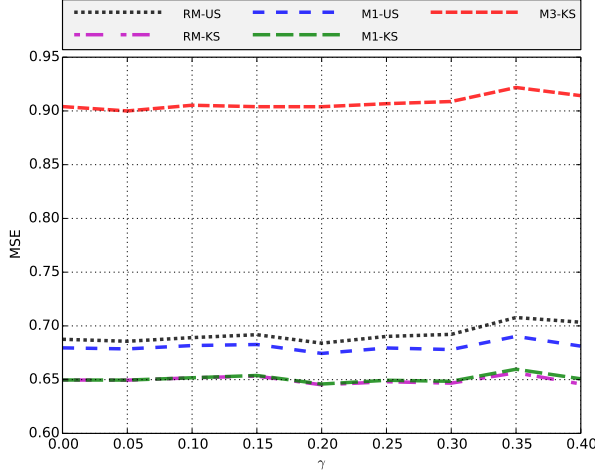


Fig. 2. MSE for the five methods studied in the first set of experiments.

In experiments not reported, we evaluated if the parameters of the switching chain  $p(r_2 = 0 | r_1 = 0) = p(r_2 = 1 | r_1 = 1) = q$  modify the conclusions obtained above. We showed that, as  $q$  increases, the MSE grows for all methods except for CGLSSM. When  $q$  is small (*i.e.* switches are frequent), the RM algorithms are more performing than the M1 ones.

### B. Second series

The aim of this series of experiments is to study which one among CGOMSM and CGLSSM models is “closer” to the general CGPMSM model, when optimal filtering is concerned. In these experiments, we set  $b(0) \neq b(1)$ . To ensure that the results are not possibly due to a dissymmetry of the CGPMSM, we used a CGPMSM in which  $\mathbf{X}_1^N$  and  $\mathbf{Y}_1^N$  play symmetric roles. Therefore we considered  $a(r_n) = c(r_n)$  and  $d(r_n) = e(r_n)$ . There is no classical CGLSSM of interest verifying this condition; indeed, in CGLSSM we would have  $c(r_n) = a(r_n)^2 b(r_n)$ , and thus either  $b(r_n) = 1$  or  $b(r_n) = -1$ , which are cases with no interest. Thus we used the “Extended CGLSSM” (ECGLSSM), obtained from (7) by setting  $\mathbf{A}_{n+1}^2(\mathbf{R}_n^{n+1}) = 0$  (thus  $\beta(r_n) = 0$  in (31), but  $\delta(r_n)$  are not necessarily zero). Thus the CGPMSM (7) extends the ECGLSSM, which itself extends the classical CGLSSM (2)-(3). ECGLSSM and CGOMSM can be seen as two symmetrical approximations of the CGPMSM with the same number of free parameters.

Finally, data are sampled with respect to a symmetrical CGPMSM and restored with methods based on the true CGPMSM (reference method RM-KS), with methods based on the CGOMSM obtained by setting  $\gamma(r_n) = 0$  (methods M1-KS and M1-US), and with methods

	Case 1	Case 2	Case 3	Case 4	Case 5
$b(0)$	0.9000	0.7000	0.5000	0.3000	0.8000
$\alpha(0)$	0.2000	0.2000	0.2000	0.2000	0.5500
$\beta(0)$	0.1254	0.0817	0.0536	0.0307	0.3500
$a(0)$	0.3128	0.2571	0.2268	0.2092	0.8300
$d(0)$	0.3054	0.2217	0.1536	0.0907	0.7900
$b(1)$	0.4000	0.4000	0.1000	0.8000	0.3000
$\alpha(1)$	0.6000	0.6000	0.6000	0.6000	0.4000
$\beta(1)$	0.3761	0.2450	0.1608	0.0921	0.2000
$a(1)$	0.7504	0.6980	0.6160	0.6736	0.4600
$d(1)$	0.6161	0.4850	0.2208	0.5721	0.3200

TABLE I  
PARAMETERS OF THE FIVE SYMMETRICAL CGPMM CASES STUDIED IN THE SECOND SERIES OF EXPERIMENTS.

	Case 1	Case 2	Case 3	Case 4	Case 5
Square error between $\mathbf{x}_1^N$ and $\hat{\mathbf{x}}_1^N$					
RM-KS	0.27182	0.60697	0.82975	0.64165	0.57802
M1-KS	0.31054	0.60792	0.82992	0.64190	0.58090
M2-KS	0.50170	0.66560	0.86202	0.64394	0.63725
RM-US	0.51635	0.67744	0.90203	0.73091	0.61710
M1-US	0.48351	0.66294	0.88792	0.68142	0.60033
M2-US	0.66988	0.78122	0.94102	0.77297	0.87519
Error ratio between $r_1^N$ and $\hat{r}_1^N$ (%)					
RM-US	30.911	32.782	35.697	33.159	30.308
M1-US	29.392	31.275	34.671	32.019	29.046
M2-US	38.094	35.026	35.756	33.265	36.984

TABLE II  
SQUARED ERROR AND ERROR RATIOS OBTAINED WITH THREE MODELS AND SIX METHODS (RM-KS, RM-US, M1-KS, M1-US, M2-KS AND M2-US). PARAMETERS USED ARE REPORTED IN THE TEXT.

based on the ECGLSSM obtained by setting  $\beta(r_n) = 0$  (method M2-KS). The symmetrical CGLSSM chosen to sample data is given by (31) with  $\alpha(r_n) = \delta(r_n)$ ,  $\beta(r_n) = \gamma(r_n)$ ,  $a(r_n) = c(r_n)$ , and  $d(r_n) = e(r_n)$ . Applying (29)-(30) for two switches  $r_n = 0, 1$  leads to

$$\begin{bmatrix} a(r_n) \\ d(r_n) \end{bmatrix} = \begin{bmatrix} 1 & b(r_n) \\ b(r_n) & 1 \end{bmatrix} \begin{bmatrix} \alpha(r_n) \\ \beta(r_n) \end{bmatrix}. \quad (32)$$

We set  $q = 0.4$  and thus each CGOMSM studied is defined by ten parameters  $b(0)$ ,  $a(0)$ ,  $d(0)$ ,  $\alpha(0)$ ,  $\beta(0)$ ,  $b(1)$ ,  $a(1)$ ,  $d(1)$ ,  $\alpha(1)$ , and  $\beta(1)$  verifying (32), which leads to five free parameters. We consider five cases whose parameters are given in Table I, whose results are specified in Table II.

According to the results, we propose the following conclusions:

- 1) The main conclusion is that, in both “known switches” and “unknown switches” cases, algorithm M1 always gives better results than M2 one, and the difference can be significant (Case 1). This means that, when the optimal filtering is considered, CGOMSM better approaches the general CGPMSM than ECGLSSM does;
- 2) Comparing the difference of the results obtained with M1-US with respect to those obtained with M1-KS, with the analogous

difference when replacing M1 by RM, we see that the new method is a bit more robust than the particle filter based ones;

- 3) Regarding the estimation of the switches, which can be an objective in itself, the method M1-US appears more efficient than M2-US. As in the first series, the computation-time is about 100 times shorter for M1-US or for M2-US than for RM-US. This is a crucial advantage of the new method with respect to the particle filter based ones.
- 4) The error ratio in switches estimation could seem being large, but this is normal because of the low signal-to-noise ratio. In fact, as the means are all set to zero, the influence of switches on the distribution of  $\mathbf{Y}_1^N$  is reduced to variances and covariance.

## V. CONCLUSION

Extending some preliminary results from [14], we proposed an exact filter called “Conditionally Gaussian Observed Markov Switching Model” (CGOMSM). The CGOMSM family of models is very flexible, so that it was possible to build a model as a close approximation to the CGPMSM (also see theoretical justifications in [17] of its closeness with the classical CGLSSM). We showed that, at least in the context of this study, our filter showed comparable efficiency with respect to a suited particle filter, while being much faster. Indeed, its computation times is nearly comparable to that of Kalman filter applied in absence of switches. In addition, problems like weights degeneracy can appear in the particle filter based methods, which is not the case in the proposed CGOMSM-based method.

Let us mention some perspectives for further works. First, it will be interesting to pursue comparison with an extension of the Interacting Multiple Model [12] to the pairwise context considered here. Also, parameter estimation and related unsupervised filtering provide a natural perspective for further works.

## REFERENCES

- [1] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking,” *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.
- [2] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman filter: particle filters for tracking applications*. Artech House, 2004.
- [3] N. S. Kim, T. G. Kang, S. J. Kang, C. W. Han, and D. H. Hong, “Speech feature mapping based on switching linear dynamic system,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 2, pp. 620–631, 2012.
- [4] J. F. Smith, A. S. Pillai, K. Chen, and B. Horwitz, “Identification and validation of effective connectivity networks in functional magnetic resonance imaging using switching linear dynamic systems,” *NeuroImage*, vol. 52, no. 3, pp. 1027–1040, 2010.
- [5] C. Andrieu, M. Davy, and A. Doucet, “Efficient particle filtering for jump Markov systems. Application to time-varying autoregressions,” *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1762–1770, Jul. 2003.
- [6] B. Ait-El-Fquih and F. Desbouvries, “Fixed-interval Kalman smoothing algorithms in singular state-space systems,” *The Journal of Signal Processing Systems*, vol. 65, no. 3, pp. 469–478, 2011.
- [7] M. R. Morelande and B. Moran, “An unscented transformation for conditionally linear models,” in *Proc. IEEE Int. Conf on Acoustics, Speech and Signal Processing (ICASSP’07)*, Honolulu, Hawaii, USA, Apr. 2007.
- [8] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*. Springer-Verlag, 2005.
- [9] O. Zoeter and T. Heskes, “Deterministic approximate inference techniques for conditionally Gaussian state space models,” *Statistics and Computing*, vol. 16, no. 3, pp. 279–292, 2006.
- [10] G. Kotsalis and A. Rantzer, “Balanced truncation for discrete time Markov jump linear systems,” *IEEE Trans. Autom. Control*, vol. 55, no. 11, pp. 2606–2611, 2010.
- [11] A. I. Petrov and A. G. Zubov, “On applicability of the imm approach to state estimation for systems with sojourn-time-dependent Markov model switching,” *IEEE Trans. Autom. Control*, vol. 41, pp. 136–140, 1996.
- [12] H. A. P. Blom and Y. Bar-Shalom, “The interacting multiple model algorithm for systems with Markovian switching coefficients,” *Automatic Control, IEEE Transactions on*, vol. 33, no. 8, pp. 780–783, Aug 1988.
- [13] W. Pieczynski, “Exact filtering in conditionally Markov switching hidden linear models,” *Comptes Rendus Mathématique*, vol. 349, no. 9-10, pp. 587–590, May 2011.
- [14] N. Abbassi, D. Benboudjema, and W. Pieczynski, “Kalman filtering approximations in triplet Markov Gaussian switching models,” in *Proc. IEEE Int. Workshop on Statistical Signal Processing (SSP’11)*, Nice, France, Jun. 2011, pp. 290–294.
- [15] R. S. Lipster and A. N. Shiryaev, *Statistics of Random Processes: I. General theory*, 2nd ed. Springer, 2001.
- [16] F. Carravetta, A. Germani, R. Liptser, and C. Manes, “Filtering of nonlinear stochastic feedback systems,” *SIAM Journal on Control & Optimization*, vol. 40, no. 5, pp. 1576–1584, Feb. 2002.
- [17] S. Derrode and W. Pieczynski, “Exact fast computation of optimal filter in Gaussian switching linear systems,” *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 701–704, Jul. 2013.