Exact Fast Computation of Optimal Filter in Gaussian Switching Linear Systems

Stéphane Derrode and Wojciech Pieczynski

Abstract—**We consider triplet Markov Gaussian linear systems $(\mathbf{X}, \mathbf{R}, \mathbf{Y})$, where $\mathbf{X}$ is hidden continuous random sequence, $\mathbf{R}$ is hidden discrete Markov chain, $\mathbf{Y}$ is observed continuous random sequence, and $(\mathbf{X}, \mathbf{Y})$ is Gaussian conditionally on $\mathbf{R}$. In the classical "Conditionally Gaussian Linear State-Space Model" (CGLSSM), optimal filter is not workable with a reasonable complexity. The aim of the paper is to propose a new model, quite close to the CGLSSM, belonging to the general and recently proposed family of models, called "Conditionally Markov Switching Hidden Linear Models" (CMSHLMs), in which the computation of optimal filter with complexity linear in the number of observations is feasible. The new model and related filtering are immediately applicable in all situations where the classical CGLSSM is used via approximated filtering.**

Index Terms—**Conditionally Gaussian linear state-space model, switching systems, optimal statistical filter, Kalman filter.**

I. Introduction

Let us consider three random sequences $\mathbf{X}_1^N = (\mathbf{X}_1, \ldots, \mathbf{X}_N)$, $\mathbf{R}_1^N = (R_1, \ldots, R_N)$ and $\mathbf{Y}_1^N = (\mathbf{Y}_1, \ldots, \mathbf{Y}_N)$, 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 $\Omega = \{1, \ldots, K\}$. Both $\mathbf{X}_1^N$ and $\mathbf{R}_1^N$ are hidden, while $\mathbf{Y}_1^N$ is observed. The process $\mathbf{R}_1^N$ can be seen as modeling the random "switches" of the distributions linked with $(\mathbf{X}_1^N, \mathbf{Y}_1^N)$, which can be of utmost importance in non stationary situations. The "optimal filter" problem we deal with in this paper consists in the sequential search of $(\mathbf{R}_1^N, \mathbf{X}_1^N)$ from $\mathbf{Y}_1^N$. More precisely, with usual notations for conditional probabilities and conditional expectations and variances, we search $p\left(r_{n+1} \left| \mathbf{y}_1^{n+1}\right.\right)$, $\mathrm{E}\left[\mathbf{X}_{n+1} \left| r_{n+1}, \mathbf{y}_1^{n+1}\right.\right]$ and $\mathrm{E}\left[\mathbf{X}_{n+1}\mathbf{X}_{n+1}^T \left| r_{n+1}, \mathbf{y}_1^{n+1}\right.\right]$ from $p\left(r_n \left| \mathbf{y}_1^n\right.\right)$, $\mathrm{E}\left[\mathbf{X}_n \left| r_n, \mathbf{y}_1^n\right.\right]$, $\mathrm{E}\left[\mathbf{X}_n\mathbf{X}_n^T \left| r_n, \mathbf{y}_1^n\right.\right]$ and $\mathbf{y}_{n+1}$. The optimal filter is then given by $\mathrm{E}\left[\mathbf{X}_{n+1} \left| \mathbf{y}_1^{n+1}\right.\right] = \sum_{r_{n+1}} p\left(r_{n+1} \left| \mathbf{y}_1^{n+1}\right.\right) \mathrm{E}\left[\mathbf{X}_{n+1} \left| r_{n+1}, \mathbf{y}_1^{n+1}\right.\right]$ and its variance by $\mathrm{Var}\left[\mathbf{X}_{n+1} \left| \mathbf{y}_1^{n+1}\right.\right] = \mathrm{E}\left[\mathbf{X}_{n+1}\mathbf{X}_{n+1}^T \left| \mathbf{y}_1^{n+1}\right.\right] - \mathrm{E}\left[\mathbf{X}_{n+1} \left| \mathbf{y}_1^{n+1}\right.\right] \mathrm{E}\left[\mathbf{X}_{n+1}^T \left| \mathbf{y}_1^{n+1}\right.\right]$ with $\mathrm{E}\left[\mathbf{X}_{n+1}\mathbf{X}_{n+1}^T \left| \mathbf{y}_1^{n+1}\right.\right] = \sum_{r_{n+1}} p\left(r_{n+1} \left| \mathbf{y}_1^{n+1}\right.\right) \mathrm{E}\left[\mathbf{X}_{n+1}\mathbf{X}_{n+1}^T \left| r_{n+1}, \mathbf{y}_1^{n+1}\right.\right]$. Such a problem is of importance in numerous situations and hundreds of papers deal with different solutions for several decades. In this paper we deal with the simple "Conditionally Gaussian Linear State-Space Models" (CGLSSMs) [1], [2], though the presented results are likely to be extended to other more sophisticated models in references mentioned above [3]–[5]. There exists numerous applications, among which tracking problems are of importance [6].

In CGLSSMs the distribution of $\mathbf{T}_1^N = (\mathbf{X}_1^N, \mathbf{R}_1^N, \mathbf{Y}_1^N)$ is obtained by setting together two classical and widely used models that are "Hidden Markov Chains" (HMCs) and "Linear Gaussian State-Space Models" (LGSSMs). Roughly speaking, $(\mathbf{X}_1^N, \mathbf{R}_1^N)$ has the structure of an HMC and, conditionally on $\mathbf{R}_1^N$, $(\mathbf{X}_1^N, \mathbf{Y}_1^N)$ is a LGSSM. Then, when $\mathbf{R}_1^N$ is known, the problem is solved by the classical Kalman filter and, when $\mathbf{R}_1^N$ is not known, the problem has no known solution with a reasonable complexity and approximate methods are used. The aim of the paper is to introduce an alternative model, which is not more complicated than CGLSSM, which is close to it, and which does allow fast and exact optimal filtering.

More precisely, a CGLSSM $\mathbf{T}_1^N = (\mathbf{X}_1^N, \mathbf{R}_1^N, \mathbf{Y}_1^N)$ is given by the distribution $p\left(\boldsymbol{x}_1, r_1, \boldsymbol{y}_1\right)$ of $(\mathbf{X}_1, R_1, \mathbf{Y}_1)$ and the recursions $p\left(\boldsymbol{t}_{n+1} \left| \boldsymbol{t}_n\right.\right)$ verifying

$$\mathbf{R}_1^N \text{ Markov with } p\left(r_{n+1} \left| \mathbf{x}_1^n, \mathbf{r}_1^n, \mathbf{y}_1^n\right.\right) = p\left(r_{n+1} \left| r_n\right.\right); \quad (1)$$

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

$$\mathbf{Y}_{n+1} = \mathbf{B}_{n+1}(R_{n+1})\, \mathbf{X}_{n+1} + \mathbf{D}_{n+1}(R_{n+1})\, \mathbf{V}_{n+1}, \quad (3)$$

with $\mathbf{A}_{n+1}(R_{n+1})$, $\mathbf{B}_{n+1}(R_{n+1})$, $\mathbf{C}_{n+1}(R_{n+1})$, $\mathbf{D}_{n+1}(R_{n+1})$ appropriate matrices depending on switches, and $\mathbf{U}_{n+1}$, $\mathbf{V}_{n+1}$ white Gaussian noises independent each from the other and such that for each $n = 1, \ldots, N-1$, $(\mathbf{U}_{n+1}, \mathbf{V}_{n+1})$ is independent from $\mathbf{T}_1^N$. In such a model the marginal distributions $p\left(\mathbf{x}_n, r_n, \mathbf{y}_n\right)$ are, in the general case, mixtures of Gaussian distributions with a number of components exponentially increasing with n.

We propose two contributions:

1) we modify the CGLSSM above by replacing $\mathbf{C}_{n+1}(R_{n+1})$ with $\mathbf{C}_{n+1}(\mathbf{R}_n^{n+1})$ in such a way that in the modified model, the marginal distributions $p\left(\mathbf{x}_n, r_n, \mathbf{y}_n\right)$ give the conditional distributions $p\left(\mathbf{x}_n, \mathbf{y}_n \left| r_n\right.\right)$ Gaussian. Thus the general form of margins $p\left(\mathbf{x}_n, r_n, \mathbf{y}_n\right)$ does not depend on n –however the parameters can vary with n–, which seems to us to better suit real situations;
2) we associate with the modified model, called "Model 1", a new model, called "Model 2", which belongs to the "Conditionally Markov Switching Hidden Linear Models" (CMSHLMs) family introduced in [7] –

S. Derrode is with Centrale Marseille, Aix-Marseille Université, CNRS, Institut Fresnel, UMR 7249, 13013 Marseille, France. e-mail: stephane.derrode@centrale-marseille.fr

W. Pieczynski is with Telecom SudParis, CITI Dpt (CNRS UMR 5157) Evry, France. e-mail: wojciech.pieczynski@it-sudparis.eu

and thus in which fast exact filtering is possible– and which is "close" to the Model 1. In particular, for each $n = 1, \ldots, N-1$, $p(\mathbf{x}_{n+1} | \mathbf{x}_n)$ and $p(\mathbf{y}_n | \mathbf{x}_n)$ are identical in both models. Then we specify how the fast optimal filter runs.

Let us insist on the fact that we do not consider Model 2 as an approximation of a given Model 1, but rather as an alternative model, close to Model 1, but allowing fast optimal filtering.

II. Modified CGLSSM and associated CMSHLM

Let us denote $\mathbf{Z}_n = (\mathbf{X}_n, \mathbf{Y}_n)^T$ and $\mathbf{W}_n = (\mathbf{U}_n, \mathbf{V}_n)^T$, and let us consider the CGLSSM defined by (1)-(3) above. Conditionally on $\mathbf{R}_1^N = \mathbf{r}_1^N$, the covariance matrix $\mathbf{\Gamma}_{\mathbf{X}_n}(\mathbf{r}_1^n)$ of the Gaussian distribution of $\mathbf{X}_n$ depends on $\mathbf{r}_1^n = (r_1, \ldots, r_n)$. Indeed, we have classically the recursion $\mathbf{\Gamma}_{\mathbf{X}_{n+1}}(\mathbf{r}_1^{n+1}) = \mathbf{A}_{n+1}(r_{n+1})\mathbf{\Gamma}_{\mathbf{X}_n}(\mathbf{r}_1^n)\mathbf{A}_{n+1}^T(r_{n+1})$ $+ \mathbf{C}_{n+1}(r_{n+1})\mathbf{C}_{n+1}^T(r_{n+1})$ and thus the marginal distributions $p(\mathbf{x}_n, \mathbf{y}_n | r_n)$ are mixtures of K^{n-1} Gaussian distributions. Let us search for a model having desired Gaussian marginal distributions $p(\mathbf{x}_n, \mathbf{y}_n | r_n)$, and thus such that its covariance matrix $\mathbf{\Gamma}_{\mathbf{Z}_n}(r_n)$ only depends on r_n. Such models can be defined recursively: for a given desired sequence $\mathbf{\Gamma}_{\mathbf{Z}_1}(r_1), \ldots, \mathbf{\Gamma}_{\mathbf{Z}_N}(r_N)$, let us consider the following CGLSSM, called Model 1: $\mathbf{R}_1^N$ verifies (1) and

$$\mathbf{X}_{n+1} = \mathbf{A}_{n+1}(R_{n+1})\mathbf{X}_n + \mathbf{C}_{n+1}(\mathbf{R}_n^{n+1})\mathbf{U}_{n+1}; \tag{4}$$

$$\begin{aligned}\mathbf{C}_{n+1}(\mathbf{R}_n^{n+1})\mathbf{C}_{n+1}^T(\mathbf{R}_n^{n+1}) = \mathbf{\Gamma}_{\mathbf{X}_{n+1}}(R_{n+1}) - \\ \mathbf{A}_{n+1}(R_{n+1})\mathbf{\Gamma}_{\mathbf{X}_n}(R_n)\mathbf{A}_{n+1}^T(R_{n+1});\end{aligned} \tag{5}$$

$$\mathbf{Y}_{n+1} = \mathbf{B}_{n+1}(R_{n+1})\mathbf{X}_{n+1} + \mathbf{D}_{n+1}(R_{n+1})\mathbf{V}_{n+1}. \tag{6}$$

We can also say that Model 1 is a classic CGLSSM in which $\boldsymbol{C}_{n+1}$ depends on both R_n and R_{n+1} in such a way that $p(\mathbf{x}_n, \mathbf{y}_n | \mathbf{r}_1^n) = p(\mathbf{x}_n, \mathbf{y}_n | r_n)$.

Reporting $\mathbf{X}_{n+1}$ given by (4) into (6), (4)-(6) can be written

$$\mathbf{Z}_{n+1} = \mathbf{A}_{n+1}^1(R_{n+1})\mathbf{Z}_n + \mathbf{B}_{n+1}^1(\mathbf{R}_n^{n+1})\mathbf{W}_{n+1}, \tag{7}$$

with $\mathbf{A}_{n+1}^1(R_{n+1})$ and $\mathbf{B}_{n+1}^1(\mathbf{R}_n^{n+1})$ defined by

$$\mathbf{A}_{n+1}^1(R_{n+1}) = \begin{bmatrix} \mathbf{A}_{n+1}(R_{n+1}) & 0 \\ \mathbf{B}_{n+1}(R_{n+1})\mathbf{A}_{n+1}(R_{n+1}) & 0 \end{bmatrix} \tag{8}$$

$$\begin{aligned}&\mathbf{B}_{n+1}^1(\mathbf{R}_n^{n+1}) = \\ &\begin{bmatrix} \mathbf{C}_{n+1}(\mathbf{R}_n^{n+1}) & 0 \\ \mathbf{B}_{n+1}(R_{n+1})\mathbf{C}_{n+1}(\mathbf{R}_n^{n+1}) & \mathbf{D}_{n+1}(R_{n+1}) \end{bmatrix}\end{aligned} \tag{9}$$

We define the Model 2 associated with the Model 1 given by (1), (7)-(9) as the model verifying

$$\mathbf{Z}_{n+1} = \mathbf{A}_{n+1}^2(\mathbf{R}_n^{n+1})\mathbf{Z}_n + \mathbf{B}_{n+1}^2(\mathbf{R}_n^{n+1})\mathbf{W}_{n+1}, \tag{10}$$

with $\mathbf{A}_{n+1}^2(\mathbf{R}_n^{n+1})$ given by

$$\mathbf{A}_{n+1}^2(\mathbf{R}_n^{n+1}) = \begin{bmatrix} \mathbf{A}_{n+1}(R_{n+1}) & 0 \\ 0 & \mathbf{E}_{n+1}(\mathbf{R}_n^{n+1}) \end{bmatrix} \tag{11}$$

with

$$\begin{aligned}\mathbf{E}_{n+1}(\mathbf{R}_n^{n+1}) = \mathbf{B}_{n+1}(R_{n+1})\mathbf{A}_{n+1}(R_{n+1}) \\ \mathbf{\Gamma}_{\mathbf{X}_n\mathbf{Y}_n}(R_n)(\mathbf{\Gamma}_{\mathbf{Y}_n}(R_n))^{-1}\end{aligned} \tag{12}$$

and $\mathbf{B}_{n+1}^2(\mathbf{R}_n^{n+1})$ such that the covariance matrix $\mathbf{\Gamma}_{Z_{n+1}}(R_{n+1})$ is the same as in Model 1, which gives

$$\begin{aligned}\mathbf{\Gamma}_{Z_{n+1}}(R_{n+1}) =& \mathbf{A}_{n+1}^2(\mathbf{R}_n^{n+1})\mathbf{\Gamma}_{Z_n}(R_n)(\mathbf{A}_{n+1}^2(\mathbf{R}_n^{n+1}))^T \\ &+ \mathbf{B}_{n+1}^2(\mathbf{R}_n^{n+1})(\mathbf{B}_{n+1}^2(\mathbf{R}_n^{n+1}))^T.\end{aligned} \tag{13}$$

Thus $\mathbf{B}_{n+1}^2(\mathbf{R}_n^{n+1})(\mathbf{B}_{n+1}^2(\mathbf{R}_n^{n+1}))^T$ is recursively given from $\mathbf{A}_{n+1}^2(\mathbf{R}_n^{n+1})$ and $\mathbf{\Gamma}_{\mathbf{Z}_n}(R_n)$, $\mathbf{\Gamma}_{\mathbf{Z}_{n+1}}(R_{n+1})$, the covariance matrices common to Model 1 and Model 2.

Finally, for each $n = 1, \ldots, N-1$, the marginal Gaussian distributions $p(\mathbf{x}_n, \mathbf{y}_n | r_n)$ are the same in Model 1 and Model 2. To see the difference between them, let us compute

$$\mathbf{\Gamma}_{\mathbf{Z}_{n+1}\mathbf{Z}_n}(\mathbf{R}_n^{n+1}) = \begin{bmatrix} \mathbf{\Gamma}_{\mathbf{X}_{n+1}\mathbf{X}_n}(\mathbf{R}_n^{n+1}) & \mathbf{\Gamma}_{\mathbf{X}_{n+1}\mathbf{Y}_n}(\mathbf{R}_n^{n+1}) \\ \mathbf{\Gamma}_{\mathbf{Y}_{n+1}\mathbf{X}_n}(\mathbf{R}_n^{n+1}) & \mathbf{\Gamma}_{\mathbf{Y}_{n+1}\mathbf{Y}_n}(\mathbf{R}_n^{n+1}) \end{bmatrix}$$

We have for Model 1 and Model 2 respectively

$$\mathbf{\Gamma}_{\mathbf{Z}_{n+1}\mathbf{Z}_n}^1(\mathbf{R}_n^{n+1}) = \mathbf{A}_{n+1}^1(R_{n+1})\mathbf{\Gamma}_{\mathbf{Z}_n}(R_n) = \begin{bmatrix} \boldsymbol{\alpha}^1 & \boldsymbol{\beta}^1 \\ \boldsymbol{\chi}^1 & \boldsymbol{\delta}^1 \end{bmatrix}$$

$$\mathbf{\Gamma}_{\mathbf{Z}_{n+1}\mathbf{Z}_n}^2(\mathbf{R}_n^{n+1}) = \mathbf{A}_{n+1}^2(\mathbf{R}_n^{n+1})\mathbf{\Gamma}_{\mathbf{Z}_n}(R_n) = \begin{bmatrix} \boldsymbol{\alpha}^2 & \boldsymbol{\beta}^2 \\ \boldsymbol{\chi}^2 & \boldsymbol{\delta}^2 \end{bmatrix}$$

We find that $\boldsymbol{\alpha}^1 = \boldsymbol{\alpha}^2$, $\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2$, $\boldsymbol{\delta}^1 = \boldsymbol{\delta}^2$, but $\boldsymbol{\chi}^1 = \mathbf{B}_{n+1}(R_{n+1})\mathbf{A}_{n+1}(R_{n+1})\mathbf{\Gamma}_{\mathbf{X}_n}(R_n)$ wheras $\boldsymbol{\chi}^2 = \mathbf{B}_{n+1}(R_{n+1})\mathbf{A}_{n+1}(R_{n+1})\mathbf{\Gamma}_{\mathbf{X}_n\mathbf{Y}_n}(R_n)\mathbf{\Gamma}_{\mathbf{Y}_n}(R_n)^{-1}\mathbf{\Gamma}_{\mathbf{Y}_n\mathbf{X}_n}(R_n)$. Thus we can state the following result, which specifies the "closeness" of both models.

Proposition 1. *Let us consider the Model 1 given by eq.* (1), (7)-(9), *and the associated Model 2 given by eq.* (1), (10)-(13). *We can state:*

1) *For each $n = 1, \ldots, N-1$, the Gaussian conditional distributions $p(\mathbf{x}_n, \mathbf{y}_n | r_n)$, $p(\mathbf{x}_{n+1}, \mathbf{y}_{n+1} | r_{n+1})$, $p(\mathbf{x}_n^{n+1} | \mathbf{r}_n^{n+1})$, $p(\mathbf{y}_n^{n+1} | \mathbf{r}_n^{n+1})$, and $p(\mathbf{x}_{n+1}, \mathbf{y}_n | r_{n+1})$ are the same in models 1 and 2, and the only difference comes from $p(\mathbf{x}_n, \mathbf{y}_{n+1} | \mathbf{r}_n^{n+1})$. In particular, $p(\mathbf{x}_{n+1} | \mathbf{x}_n, r_n)$ and $p(\mathbf{y}_n | \mathbf{x}_n, r_n)$, which are used in building Model 1, are the same;*
2) *Distribution $p(\mathbf{x}_1^N | \mathbf{r}_1^N)$ is Markov in both models with identical distribution, while $p(\mathbf{y}_1^N | \mathbf{r}_1^N)$ is Markov in Model 2, but not in Model 1.*

Proof. Point 1) comes from the very construction of the Model 2 above. To show that $p(\mathbf{x}_1^N | \mathbf{r}_1^N)$ is Markov with identical distributions in Model 1 and Model 2, we remark that eq. (4) is valid for both of them. The fact that $p(\mathbf{y}_1^N | \mathbf{r}_1^N)$ is not Markov in Model 1 is a classic property. To show that it is Markov in Model 2, let us first remark that the second line in matrix $\mathbf{A}_{n+1}^2(\mathbf{R}_n^{n+1})$ is of the form $\begin{bmatrix} 0 & \mathbf{E}_{n+1}(\mathbf{R}_n^{n+1}) \end{bmatrix}$ and let us set $\begin{bmatrix} f_{n+1}(\mathbf{R}_n^{n+1}) & g_{n+1}(\mathbf{R}_n^{n+1}) \end{bmatrix}$ the second line of the matrix $\mathbf{B}_{n+1}^2(\mathbf{R}_n^{n+1})$. We have, according to (10), $\mathbf{Y}_{n+1} = \mathbf{E}_{n+1}(\mathbf{R}_n^{n+1})\mathbf{Y}_n + f_{n+1}(\mathbf{R}_n^{n+1})\mathbf{U}_{n+1} + g_{n+1}(\mathbf{R}_n^{n+1})\mathbf{V}_{n+1}$ where $(\mathbf{U}_{n+1}, \mathbf{V}_{n+1})$ is independent from $(\mathbf{X}_1^n, \mathbf{R}_1^n, \mathbf{Y}_1^n)$. This implies that $p(\mathbf{y}_1^N | \mathbf{r}_1^N)$ is Markov. □

Let us show that Model 2 is a "Conditionally Markov Switching Hidden Linear Model" (CMSHLM) introduced in [7]. The latter verifies:

$$\begin{aligned} &\mathbf{T}_1^N = (\mathbf{X}_1^N, \mathbf{R}_1^N, \mathbf{Y}_1^N) \text{ is Markov with} \\ &p\left(r_{n+1}, \mathbf{y}_{n+1} \left| \mathbf{x}_n, r_n, \mathbf{y}_n \right.\right) = p\left(r_{n+1}, \mathbf{y}_{n+1} \left| r_n, \mathbf{y}_n \right.\right) \quad (14) \\ &((\mathbf{R}_1^N, \mathbf{Y}_1^N) \text{ is then Markov}); \end{aligned}$$

$$\begin{aligned} \mathbf{X}_{n+1} = \mathbf{F}_{n+1}(\mathbf{R}_n^{n+1}, \mathbf{Y}_n^{n+1})\mathbf{X}_n + \\ \mathbf{G}_{n+1}(\mathbf{R}_n^{n+1}, \mathbf{Y}_n^{n+1})\mathbf{W}_{n+1} + \mathbf{H}_{n+1}(\mathbf{R}_n^{n+1}, \mathbf{Y}_n^{n+1}). \end{aligned} \quad (15)$$

with $\mathbf{F}_{n+1}(\mathbf{R}_n^{n+1}, \mathbf{Y}_n^{n+1})$, $\mathbf{G}_{n+1}(\mathbf{R}_n^{n+1}, \mathbf{Y}_n^{n+1})$ appropriate matrices, $\mathbf{W}_{n+1}$ appropriate white noise, and $\mathbf{H}_{n+1}(\mathbf{R}_n^{n+1}, \mathbf{Y}_n^{n+1})$ appropriate vectors. $p\left(r_{n+1} \left| \mathbf{y}_1^{n+1} \right.\right)$, $\mathrm{E}\left[\mathbf{X}_{n+1} \left| r_{n+1}, \mathbf{y}_1^{n+1} \right.\right]$ and $\mathrm{E}\left[\mathbf{X}_{n+1}\mathbf{X}_{n+1}^T \left| r_{n+1}, \mathbf{y}_1^{n+1} \right.\right]$ can then be computed from $p\left(r_{n+1}, \mathbf{y}_{n+1} \left| r_n, \mathbf{y}_n \right.\right)$, $\mathbf{F}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})$, $\mathbf{H}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})$, $p\left(r_n \left| \mathbf{y}_1^n \right.\right)$ and $\mathrm{E}\left[\mathbf{X}_n \left| r_n, \mathbf{y}_1^n \right.\right]$ with complexity independent from n as follows:

$$p\left(r_{n+1} \left| \mathbf{y}_1^{n+1} \right.\right) = \frac{\sum_{r_n} p\left(r_{n+1}, \mathbf{y}_{n+1} \left| r_n, \mathbf{y}_n \right.\right) p\left(r_n \left| \mathbf{y}_1^n \right.\right)}{\sum_{r_n, r_{n+1}^*} p\left(r_{n+1}^*, \mathbf{y}_{n+1} \left| r_n, \mathbf{y}_n \right.\right) p\left(r_n \left| \mathbf{y}_1^n \right.\right)}, \quad (16)$$

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

$$\begin{aligned} \mathrm{E}\left[\mathbf{X}_{n+1} \left| r_{n+1}, \mathbf{y}_1^{n+1} \right.\right] = \sum_{r_n} p\left(r_n \left| r_{n+1}, \mathbf{y}_1^{n+1} \right.\right) \\ \Big(\mathbf{F}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})\mathrm{E}\left[\mathbf{X}_n \left| r_n, \mathbf{y}_1^n \right.\right] + \mathbf{H}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})\Big), \end{aligned} \quad (18)$$

$$\begin{aligned} \mathrm{E}\left[\mathbf{X}_{n+1}\mathbf{X}_{n+1}^T \left| r_{n+1}, \mathbf{y}_1^{n+1} \right.\right] = \sum_{r_n} p\left(r_n \left| r_{n+1}, \mathbf{y}_1^{n+1} \right.\right) \\ \Big(\mathbf{F}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})\mathrm{E}\left[\mathbf{X}_n\mathbf{X}_n^T \left| r_n, \mathbf{y}_1^n \right.\right] \mathbf{F}_{n+1}^T(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}) + \\ \mathbf{F}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})\mathrm{E}\left[\mathbf{X}_n \left| r_n, \mathbf{y}_1^n \right.\right] \mathbf{H}_{n+1}^T(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}) + \\ \mathbf{H}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})\mathrm{E}^T\left[\mathbf{X}_n \left| r_n, \mathbf{y}_1^n \right.\right] \mathbf{F}_{n+1}^T(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}) + \\ \mathbf{G}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})\mathbf{G}_{n+1}^T(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}) + \\ \mathbf{H}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})\mathbf{H}_{n+1}^T(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})\Big), \end{aligned} \quad (19)$$

Proposition 2. *Model 2 defined with* (1) *and* (10)-(13) *is a CMSHLM* (14)-(15).

Proof. Let

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

Let us first verify (14). $\mathbf{T}_1^N$ is Markov and we can write $p\left(r_{n+1}, \mathbf{y}_{n+1} \left| \mathbf{x}_n, \mathbf{y}_n, r_n \right.\right) = p\left(r_{n+1} \left| \mathbf{x}_n, \mathbf{y}_n, r_n \right.\right) p\left(\mathbf{y}_{n+1} \left| \mathbf{x}_n, \mathbf{y}_n, \mathbf{r}_n^{n+1} \right.\right)$. According to (1) we have $p\left(r_{n+1} \left| \mathbf{x}_n, \mathbf{y}_n, r_n \right.\right) = p\left(r_{n+1} \left| r_n \right.\right)$ and, according to (10)-(11), we have $p\left(\mathbf{y}_{n+1} \left| \mathbf{x}_n, \mathbf{y}_n, \mathbf{r}_n^{n+1} \right.\right) = p\left(\mathbf{y}_{n+1} \left| \mathbf{y}_n, \mathbf{r}_n^{n+1} \right.\right)$. The two equalities give $p\left(r_{n+1}, \mathbf{y}_{n+1} \left| \mathbf{x}_n, r_n, \mathbf{y}_n \right.\right) = p\left(r_{n+1}, \mathbf{y}_{n+1} \left| r_n, \mathbf{y}_n \right.\right)$.

Let us then verify (15). According to (10) the distribution $p\left(\mathbf{x}_{n+1}, \mathbf{y}_{n+1} \left| \mathbf{x}_n, \mathbf{y}_n, \mathbf{r}_n^{n+1} \right.\right)$ is Gaussian with mean

$$\mathbf{A}_{n+1}^2(\boldsymbol{r}_n^{n+1}) \begin{bmatrix} \mathbf{x}_n \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{n+1}(r_{n+1})\mathbf{x}_n \\ \mathbf{E}_{n+1}(\mathbf{r}_n^{n+1})\mathbf{y}_n \end{bmatrix}$$

and variance $\mathbf{Q}_{n+1}(\mathbf{r}_n^{n+1})$ ($\mathbf{E}_{n+1}(\mathbf{r}_n^{n+1})$ is given in eq. (12)). Using the classical Gaussian conditioning rule we can say that the distribution $p\left(\mathbf{x}_{n+1} \left| \mathbf{x}_n, \mathbf{y}_n^{n+1}, \mathbf{r}_n^{n+1} \right.\right)$ is then Gaussian with mean and variance respectively given by $\mathbf{A}_{n+1}(r_{n+1})\mathbf{x}_n + \mathbf{Q}_{n+1}^2(\mathbf{r}_n^{n+1})(\mathbf{Q}_{n+1}^4(\mathbf{r}_n^{n+1}))^{-1}(\mathbf{y}_{n+1} - \mathbf{E}_{n+1}(\mathbf{r}_n^{n+1})\mathbf{y}_n)$ and $\mathbf{Q}_{n+1}^1(\mathbf{r}_n^{n+1}) - \mathbf{Q}_{n+1}^2(\mathbf{r}_n^{n+1})(\mathbf{Q}_{n+1}^4(\mathbf{r}_n^{n+1}))^{-1}\mathbf{Q}_{n+1}^3(\mathbf{r}_n^{n+1})$.

Then we can state, according to classic properties of Gaussian laws, that (15) is verified with

$$\begin{aligned} \mathbf{F}_{n+1}(\mathbf{R}_n^{n+1}, \mathbf{Y}_n^{n+1}) = \mathbf{A}_{n+1}(R_{n+1}), \\ \mathbf{H}_{n+1}(\mathbf{R}_n^{n+1}, \mathbf{Y}_n^{n+1}) = \mathbf{Q}_{n+1}^2(\mathbf{R}_n^{n+1}) \\ (\mathbf{Q}_{n+1}^4(\mathbf{R}_n^{n+1}))^{-1}(\mathbf{Y}_{n+1} - \mathbf{E}_{n+1}(\mathbf{R}_n^{n+1})\mathbf{Y}_n), \\ \mathbf{G}_{n+1}(\mathbf{R}_n^{n+1}, \mathbf{Y}_n^{n+1})(\mathbf{G}_{n+1}(\mathbf{R}_n^{n+1}, \mathbf{Y}_n^{n+1}))^T = \\ \mathbf{Q}_{n+1}^1(\mathbf{R}_n^{n+1}) - \mathbf{Q}_{n+1}^2(\mathbf{R}_n^{n+1}) \\ (\mathbf{Q}_{n+1}^4(\mathbf{R}_n^{n+1}))^{-1}\mathbf{Q}_{n+1}^3(\mathbf{R}_n^{n+1}). \end{aligned} \quad (20)$$

□

Finally, the optimal filter in the switching system (4)-(6) is: for given $\mathbf{\Gamma}_{\mathbf{Z}_n}(r_n)$, $p\left(r_n \left| \mathbf{y}_1^n \right.\right)$, $\mathrm{E}\left[\mathbf{X}_n \left| r_n, \mathbf{y}_1^n \right.\right]$, and $\mathbf{y}_{n+1}$

1) consider $\mathbf{\Gamma}_{\mathbf{X}_{n+1}}(r_{n+1})$ and $\mathbf{C}_{n+1}(\mathbf{R}_n^{n+1})$ verifying (5), which provides $\mathbf{\Gamma}_{\mathbf{Z}_n}(r_{n+1})$ with (6) and $\mathbf{A}_{n+1}^2(\mathbf{R}_n^{n+1})$ with (11) and (12);
2) compute $\mathbf{Q}_{n+1}(\mathbf{r}_n^{n+1}) = \mathbf{B}_{n+1}^2(\boldsymbol{r}_n^{n+1})(\mathbf{B}_{n+1}^2(\boldsymbol{r}_n^{n+1}))^T$ with (13);
3) compute $\mathbf{F}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})$, $\mathbf{G}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})(\mathbf{G}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}))^T$ and $\mathbf{H}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})$ with (20);
4) compute $p\left(r_{n+1}, \mathbf{y}_{n+1} \left| r_n, \mathbf{y}_n \right.\right) = p\left(r_{n+1} \left| r_n \right.\right) p\left(\mathbf{y}_{n+1} \left| \mathbf{r}_n^{n+1}, \mathbf{y}_n \right.\right)$, knowing that $p\left(\mathbf{y}_{n+1} \left| \mathbf{r}_n^{n+1}, \mathbf{y}_n \right.\right)$ is Gaussian with mean $\mathbf{E}_{n+1}(\mathbf{r}_n^{n+1})\mathbf{y}_n$ and covariance matrix $\mathbf{Q}_{n+1}^4(\mathbf{r}_n^{n+1})$;
5) compute $p\left(r_{n+1} \left| \mathbf{y}_1^{n+1} \right.\right)$, $\mathrm{E}\left[\mathbf{X}_{n+1} \left| r_{n+1}, \mathbf{y}_1^{n+1} \right.\right]$ and $\mathrm{E}\left[\mathbf{X}_{n+1}\mathbf{X}_{n+1}^T \left| r_{n+1}, \mathbf{y}_1^{n+1} \right.\right]$ with (16)-(20).

III. Experiments

Let us consider $m = q = 1$ (both $\mathbf{X}_1^N$ and $\mathbf{Y}_1^N$ are real valued processes), and the stationary case where the distributions of $(\mathbf{Z}_n, \mathbf{Z}_{n+1})$ are independent from $1, \ldots, N-1$ in both models. Thus $\boldsymbol{A}_{n+1} = \boldsymbol{A}$, $\boldsymbol{B}_{n+1} = \boldsymbol{B}$, $\boldsymbol{C}_{n+1} = \boldsymbol{C}$ and $\boldsymbol{D}_{n+1} = \boldsymbol{D}$. Each R_n takes its values in $\Omega = \{1, 2\}$ and we set, for both models, and for $i = 1, 2$: $a_i = \boldsymbol{A}(r_n = i)$, $b_i = \boldsymbol{B}(r_n = i)$ and $\sigma_i^2 = \mathbf{\Gamma}_{\mathbf{X}_n}(r_n = i) = \mathbf{\Gamma}_{\mathbf{Y}_n}(r_n = i)$. Then for $(R_n = i, R_{n+1} = j)$ we have for Model 1

$$\mathbf{A}^1(j) = \begin{bmatrix} a_j & 0 \\ a_j b_j & 0 \end{bmatrix},$$

and

$$\mathbf{B}^1(i,j)\ (\mathbf{B}^1(i,j))^T = \begin{bmatrix} \sigma_j^2 - a_j^2\sigma_i^2 & b_j(\sigma_j^2 - a_j^2\sigma_i^2) \\ b_j(\sigma_j^2 - a_j^2\sigma_i^2) & \sigma_j^2 - b_j^2 a_j^2 \sigma_i^2 \end{bmatrix}.$$

TABLE I
MSE ERROR OF F1, F2 AND F3 FILTERS WITH $b_1 = 0.3$.

$p_{11}+p_{22}$	σ_1^2	σ_2^2	F1	F2	F3
			First series of data		
0.98	0.5	2	0.8014	0.8014	0.8019 (11.9%)
	1.0	2	0.9270	0.9270	0.9282 (23.0%)
0.80	0.5	2	0.8147	0.8147	0.8162 (30.1%)
	1.0	2	0.9296	0.9296	0.9318 (39.2%)
			Second series of data		
0.98	0.5	2	0.8024	0.8023	0.8028 (11.9%)
	1.0	2	0.9236	0.9236	0.9249 (23.0%)
0.80	0.5	2	0.8147	0.8146	0.8162 (29.8%)
	1.0	2	0.9282	0.9281	0.9300 (39.2%)

TABLE II
MSE ERROR OF F1, F2 AND F3 FILTERS WITH $b_1 = 0.8$.

$p_{11}+p_{22}$	σ_1^2	σ_2^2	F1	F2	F3
			First series of data		
0.98	0.5	2	0.7024	0.7026	0.7159 (11.4%)
	1.0	2	0.7777	0.7778	0.8091 (21.7%)
0.80	0.5	2	0.7109	0.7113	0.7427 (29.6%)
	1.0	2	0.7813	0.7816	0.8320 (38.4%)
			Second series of data		
0.98	0.5	2	0.7010	0.7009	0.7140 (11.2%)
	1.0	2	0.7745	0.7744	0.8045 (21.2%)
0.80	0.5	2	0.7090	0.7086	0.7398 (29.5%)
	1.0	2	0.7777	0.7774	0.8288 (38.6%)

According to eq. (11) and (13) the associated Model 2 is given by

$$\mathbf{A}^2(i,j) = \begin{bmatrix} a_j & 0 \\ 0 & a_j b_j b_i \end{bmatrix},$$

and

$$\mathbf{B}^2(i,j)\,(\mathbf{B}^2(i,j))^T = \begin{bmatrix} \sigma_j^2 - a_j^2\sigma_i^2 & b_j(\sigma_j^2 - a_j^2\sigma_i^2 b_i^2) \\ b_j(\sigma_j^2 - a_j^2\sigma_i^2 b_i^2) & \sigma_j^2 - b_j^2 a_j^2 \sigma_i^2 b_i^2 \end{bmatrix},$$

which gives $\mathbf{F}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})$, $\mathbf{H}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})$ and $\mathbf{G}_{n+1}(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1})$ by (20). Finally, the two models are defined by the parameters σ_1^2, σ_2^2, a_1, a_2, b_1, b_2 and by the distribution $p_{ij} = p(R_1 = i, R_2 = j)$.

We present two series of experiments. In the first series, data are sampled according to Model 1, and in the second one, data are sampled according to Model 2. Both series were filtered according to three methods: (F1) is the Model 1 based optimal filter with known switches, (F2) is the Model 2 based optimal filter with known switches (conditionally on $\mathbf{R}_1^N$, Model 2 also is a classical system but more general situations, where it would be a Pairwise Markov Model and in which the classical Kalman filter is workable [8], [9], could be considered), and (F3) is the Model 2 based optimal filter with unknown switches. The sample size is $N = 1000$ and we consider $a_1 = 0.3$, $a_2 = 0.6$, $b_2 = 0.2$ and $\sigma_2^2 = 2$. We consider two cases: $b_1 = 0.3$ (Table I) and $b_1 = 0.8$ (Table II). Then we consider two possible values 0.98 and 0.80 for $p_{11} + p_{22} = p(R_1 = R_2)$ and two possible values 0.5 and 1.0 for σ_1^2. For filter F3, we also report the error rate while estimating the switches by maximizing $p(r_n | \mathbf{y}_1^n)$ (notice that these estimates of switches are not used in F3). The results, which are means of 300 independent experiments are expressed in term of the Mean Square Error (MSE).

The presented results, and different other results not reported here, show that it is difficult to obtain a significant difference between F1 and F2, which is our main conclusion.

IV. CONCLUSION

We proposed a new model, very close to the classic "Conditionally Gaussian Linear State-Space Model" (CGLSSM), but allowing, in spite of switches, a fast optimal statistical filter. This property is due to the fact that the model proposed belongs to the family of models recently introduced in [7]. The distribution of the switching sequence, that of the hidden continuous sequence, and that of each observation at time n, conditional on the switch and the state at n, are strictly the same in both models. So the new model can be immediately used in all situations the classical CGLSSM is. Simulation experiments numerically show that the new model is fairly identical to CGLSSM.

The main conclusion is that the two models are so close that it is difficult to see any difference at results level, at least in the case of real-valued sequences considered. In addition, the results obtained with the new model with known switches are very close to those obtained when the switches are unknown.

As perspective for further works let us mention that we consider the filtering problem in this paper, but the recent models can also be used to deal with prediction [10] and smoothing [11]. Another perspective would be the parameter estimation problem [12], [13].

REFERENCES

[1] C. Andrieu, M. Davy, and A. Doucet, "Efficient particle filtering for jump markov systems. application to time-varying autoregressions," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1762–1770, Jul 2003.

[2] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*. Springer-Verlag, 2005.

[3] O. L. V. Costa, M. D. Fragoso, and R. P. Marques, *Discrete Time Markov Jump Linear Systems*. Springer-Verlag, 2005.

[4] C.-J. Kim and C. R. Nelson, *State-Space Models With Regime Switching*. MIT Press, 1999.

[5] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman filter: particle filters for tracking applications*. Artech House, 2004.

[6] 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.

[7] W. Pieczynski, "Exact filtering in conditionally markov switching hidden linear models," *Comptes Rendus Mathematique*, vol. 349, no. 9-10, pp. 587–590, May 2011.

[8] W. Pieczynski and F. Desbouvries, "Kalman filtering using pairwise Gaussian models," in *Proc. IEEE Int. Conf on Acoustics, Speech and Signal Processing (ICASSP'03)*, Hong-Kong, Apr. 2003.

[9] V. Nemesin and S. Derrode, "Robust blind pairwise kalman algorithms using qr decompositions," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 5–9, Jan 2013.

[10] N. Bardel and F. Desbouvries, "Exact bayesian prediction in a class of markov-switching models," *Methodology and Computing in Applied Probability*, vol. 14, no. 1, pp. 125–134, Mar 2012.

[11] W. Pieczynski, "Exact smoothing in hidden conditionally markov switching linear models," *Communications in Statistics - Theory and Methods*, vol. 40, no. 16, pp. 2823–2829, Aug 2011.

[12] E. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, "Bayesian nonparametric methods for learning Markov switching processes," *IEEE Signal Process. Mag.*, vol. 27, no. 6, pp. 43–54, 2010.

[13] M. I. Jordan, *Learning in Graphical Models*. MIT Press, 1998.