*Applied Stochastic Models and Data Analysis, ASMDA 2011, June 7-10, Rome, Italy, 2011*

# Exact Kalman filtering in pairwise Gaussian switching systems

Dalila Benboudjema[1], Mohamed El Bouchraya Malainin[2], and Wojciech Pieczynski[2]

[1] ENSEA, Université Cergy-Pontoise, Equipe ETIS, CNRS UMR 8051, F- 95000 Cergy-Pontoise, France (E-mail : Dalila.Benboudjema@ensea.fr)
[2] Institut Telecom, Telecom Sudparis, Département CITI, CNRS UMR 5157, 9, rue Charles Fourier, 91000 Evry, France (E-mail : Wojciech.Pieczynski@it-sudparis.eu)

**Abstract.** We consider a general pairwise Markov Gaussian linear system (X,Y), where X is hidden and Y is observed, in which an exact Kalman filter (KF) is workable. There are two kinds of particular cases: either X is Markov and Y is not, or vice versa. We show that when the processed data suit the general model, the KF based on both particular cases produce similar approximate results. This is of importance when introducing stochastic Markovian switches. In fact, it is well known that the KF is no longer workable in the first case, while it is, as detailed in the paper, in the second one.
**Keywords :** Gaussian switching system, exact Kalman filtering.

## 1 Introduction

Let us consider a couple of random variables $X_1^N = (X_1, ..., X_N)$, $Y_1^N = (Y_1, ..., Y_N)$, and let us set $Z_1^N = (Z_1, ..., Z_N)$, with $Z_n = (X_n, Y_n)$ for each $n = 1, ..., N$. To simplify, we will consider all variables real valued. The chain $Z_1^N = (Z_1, ..., Z_N)$ will be assumed to be a Gaussian « Pairwise Markov Chain" (GPMC), which means that it is Gaussian and Markovian. Let us underline the fact that in GPMC the hidden chain $X_1^N$ is not necessarily a Markov one, which makes GPMC different from other known models in which the hidden chain is Markovian. However, as studied in Pieczynski and Desbouvries[5], Kalman filtering (KF) remains possible in GPMC. Now, there are two possible particular cases of GPMC. In the first one the hidden chain $X_1^N$ is a Markov one and thus this particular case is the very classical Gaussian "Hidden Markov Chain" (GHMC) model. The second one is a "symmetrical" case : the observed chain $Y_1^N$ is Markovian. These two models, which can thus be seen as being two "approximate" submodels of GPMC, are "symmetrical" and each of them is obtained from the other by inverting $X_1^N$ and $Y_1^N$. However, the interest of these two approximations is very different in the presence of switches. In fact, while using the classical GHMC and while modeling the random switches by a third Markov chain $R_1^N = (R_1, ..., R_N)$, it has been well known since Tugnait[7] that KF can no longer be performed with complexity linear in time and different approximations must be used. These can be stochastic (Andrieu *et al.*[1], Cappé *et al.*[2]), or deterministic (Costa *et al.*[3], Giordani *et al.*[4]). In particular, different particle filters are being widely studied and applied at present. On the contrary, according to different recent results in Pieczynski and Abbassi[6], KF can still be performed with complexity linear in time in the symmetrical approximation.

Finally, we have three switching "triplet" Markov models $T_1^N = (T_1, ..., T_N)$, with $T_n = (X_n, R_n, Y_n)$ for each $n = 1, ..., N$:

- Model 1: the switching GPMC, in which $R_1^N$ is Markovian and $(X_1^N, Y_1^N)$ is a PMC conditionally on $R_1^N$ ;

- Model 2: the classical switching GHMC, in which $R_1^N$ is Markovian, $X_1^N$ Markovian conditionally on $R_1^N$, and $Y_1^N$ is Markovian conditionally on $(R_1^N, X_1^N)$ ;

- Model 3: the recent switching model, in which $R_1^N$ is Markovian, $Y_1^N$ Markovian conditionally on $R_1^N$, and $X_1^N$ is Markovian conditionally on $(R_1^N, Y_1^N)$.

As said above, the advantage of Model 3 over Model 2 is the possibility of exact KF computation with complexity linear in time.

The aim of this paper is to study whether one among Models 2 and 3 is better suited to approximate Model 1.

As the switching chain $R_1^N$ plays the same role in Models 2 and 3, we begin with comparing the two approximations without switches.

## 2 Kalman filtering in a Gaussian Pairwise Markov Chain

Let us consider $X_1^N = (X_1, ..., X_N)$, $Y_1^N = (Y_1, ..., Y_N)$, and $Z_1^N = (Z_1, ..., Z_N)$, with $Z_n = (X_n, Y_n)$ for each $n = 1, ..., N$. We will assume that the chain $Z_1^N$ is Gaussian, Markovian, and stationary. Moreover, all means are null and all variances are equal to 1: $E[X_n] = E[Y_n] = 0$ and $Var[X_n] = Var[Y_n] = 1$ for each $n = 1, ..., N$.

The Gaussian distribution of $Z_1^N$ is then defined by the distribution of $Z_1^2 = (Z_1, Z_2)$, which is given by the covariance matrix

$$\Gamma^{Z_1^2} = \begin{matrix} & X_1 & Y_1 & X_2 & Y_2 \\ X_1 & 1 & b & a & d \\ Y_1 & b & 1 & e & c \\ X_2 & a & e & 1 & b \\ Y_2 & d & c & b & 1 \end{matrix} = \begin{bmatrix} \Gamma & \mathrm{A}^T \\ \mathrm{A} & \Gamma \end{bmatrix}, \text{ with } \Gamma = \begin{bmatrix} 1 & b \\ b & 1 \end{bmatrix} \text{ and } \mathrm{A} = \begin{bmatrix} a & e \\ d & c \end{bmatrix} \quad (2.1)$$

Thus the distribution of $Z_1^N$ is defined by the five parameters $a,b,c,d,e$.
For $p(w)$ the probability distribution of a Gaussian vector $W$ of mean vector $m_w$ and covariance matrix $\Gamma_w$, we will set $p(w) \sim N(m_w, \Gamma_w)$. Let us recall two classical results of Gaussian distributions which will be useful to derive KF formulae, which in this case are somewhat different from KF proposed in Pieczynski and Desbouvries[5]. For two random Gaussian vectors $U,V$ we have :

Property 1. If $p(u) \sim N(m_u, \Gamma_u)$ and $p(v|u) \sim N(Au+b, \Gamma_{v|u})$,

then $p(u,v) \sim N(\begin{pmatrix} m_u \\ Am_u + b \end{pmatrix}, \begin{pmatrix} \Gamma_u & (\Gamma_u)^T A^T \\ A\Gamma_u & \Gamma_{v|u} + A(\Gamma_u)^T A^T \end{pmatrix}$, and thus

$p(v) \sim N(Am_u + b, \Gamma_{v|u} + A(\Gamma_u)^T A^T$;

Property 2. If $p(u,v) \sim N(\begin{pmatrix} m_u \\ m_v \end{pmatrix}, \begin{pmatrix} \Gamma_u & (\Gamma_{uv})^T \\ \Gamma_{uv} & \Gamma_v \end{pmatrix})$, then

$p(v|u) \sim N(\Gamma_{uv}(\Gamma_u)^{-1}(u - m_u) + m_v, \Gamma_v - \Gamma_{uv}(\Gamma_u)^{-1})(\Gamma_{uv})^T)$.

Using (2.1) and applying Property 2 to $U = (X_n, Y_n)$ and $V = (X_{n+1}, Y_{n+1})$ gives:

$$p(x_{n+1}, y_{n+1} | x_n, y_n) \sim N(\mathrm{B}\begin{bmatrix} x_n \\ y_n \end{bmatrix}, \Sigma), \quad (2.2)$$

with

$$\mathrm{B} = \mathrm{A}\Gamma^{-1} = \begin{bmatrix} a & e \\ d & c \end{bmatrix} \begin{bmatrix} 1 & b \\ b & 1 \end{bmatrix}^{-1} = \frac{1}{1-b^2} \begin{bmatrix} a - eb & -ab + e \\ d - cb & -db + c \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}, \quad (2.3)$$

and

$$\Sigma = \Gamma - \mathrm{A}\Gamma^{-1}\mathrm{A}^T = \quad (2.4)$$
$$\frac{1}{1-b^2} \begin{bmatrix} 1 - b^2 - a^2 - e^2 + 2abe & b(1 - b^2 + ed + ac) - ad - ec \\ b(1 - b^2 + ed + ac) - ad - ec & 1 - b^2 - d^2 - c^2 + 2dbc \end{bmatrix}$$

KF consists of computing $p(x_{n+1}|y_1^{n+1})$ from $p(x_n|y_1^n)$ and $y_{n+1}$. Let $p(x_{n+1}|y_1^{n+1}) \sim N(m_{n+1}, \sigma_{n+1}^2)$ and $p(x_n|y_1^n) \sim N(m_n, \sigma_n^2)$. First, we compute $p(x_{n+1}, y_{n+1}|y_1^n)$ from $p(x_n|y_1^n)$. We apply Property 1 to $u = x_n$ and $v = (x_{n+1}, y_{n+1})$, with the distribution $p(.)$ replaced by the conditional distribution $p(.|y_1^n)$. Knowing that $p(x_{n+1}, y_{n+1}|x_n, y_1^n) = p(x_{n+1}, y_{n+1}|x_n, y_n)$ according to the Markovianity of $Z_1^N$, and setting $m_u = m_n$, $\Gamma_u = \sigma_n^2$, $Au + b = \begin{bmatrix}\alpha \\ \gamma\end{bmatrix}[x_n] + \begin{bmatrix}\beta \\ \delta\end{bmatrix}[y_n]$, $\Gamma_{v|u} = \Sigma$, we have: $p(x_{n+1}, y_{n+1}|y_1^n) \sim N(\begin{bmatrix}\alpha \\ \gamma\end{bmatrix}[m_n] + \begin{bmatrix}\beta \\ \delta\end{bmatrix}[y_n], \Sigma + \begin{bmatrix}\alpha \\ \gamma\end{bmatrix}\sigma_n^2[\alpha \quad \gamma])$, or still

$$p(x_{n+1}, y_{n+1}|y_1^n) \sim N(\begin{bmatrix}\alpha m_n + \beta y_n \\ \gamma m_n + \delta y_n\end{bmatrix}, \begin{bmatrix}\alpha^* & \beta^* \\ \beta^* & \gamma^*\end{bmatrix}), \quad (2.5)$$

with

$$\begin{bmatrix}\alpha^* & \beta^* \\ \beta^* & \gamma^*\end{bmatrix} = \Sigma + \sigma_n^2 \begin{bmatrix}\alpha^2 & \alpha\gamma \\ \alpha\gamma & \gamma^2\end{bmatrix} \quad (2.6)$$

Then we obtain $p(x_{n+1}|y_1^{n+1}) = p(x_{n+1}|y_{n+1}, y_1^n)$ from $p(x_{n+1}, y_{n+1}|y_1^n)$ by applying Property 2 to $x_{n+1} = v$, $y_{n+1} = u$, $m_u = \gamma m_n + \delta y_n$, $m_v = \alpha m_n + \beta y_n$, $\alpha^* = \Gamma_v$, $\beta^* = \Gamma_{uv}$, $\gamma^* = \Gamma_u$, and $p(.)$ replaced by $p(.|y_1^n)$. We have: $p(x_{n+1}|y_1^{n+1}) \sim N(\frac{\beta^*}{\gamma^*}(y_{n+1} - \gamma m_n - \delta y_n) + \alpha m_n + \beta y_n, \alpha^* - \frac{(\beta^*)^2}{\gamma^*})$, which finally gives KF :

$$m_{n+1} = (\alpha - \frac{\beta^*}{\gamma^*}\gamma) m_n + (\beta - \frac{\beta^*}{\gamma^*}\delta) y_n + \frac{\beta^*}{\gamma^*} y_{n+1} \;; \quad (2.8)$$

$$\sigma_{n+1}^2 = \alpha^* - \frac{(\beta^*)^2}{\gamma^*}) \,. \quad (2.9)$$

## 3 Two particular cases of a Gaussian Pairwise Markov Chain

Let $Z_1^N = (X_1^N, Y_1^N)$ be a GPMC introduced in the previous section. It is possible to show that in the general case neither $X_1^N$ nor $Y_1^N$ is necessarily Markovian. As specified in the introduction, we will consider two particular "symmetrical" cases : in the first one $X_1^N$ is Markovian and $Y_1^N$ is not, while in the second one the situation is the opposite: $Y_1^N$ is Markovian and $X_1^N$ is not.
Let us consider the dependence graph of the four variables $(X_1, Y_1, X_2, Y_2)$ presented in Figure 1.

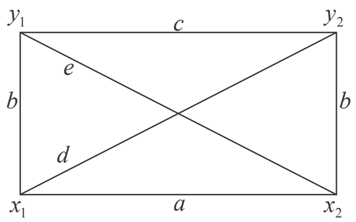


Fig. 1 Dependence graph of $(X_1, Y_1, X_2, Y_2)$

The first case, which will be called "Model 2", is obtained by considering that $e = ab$. In fact, this implies that $p(x_{n+1} | x_n, y_n) = p(x_{n+1} | x_n)$ and this equality leads to the fact that the distribution $p(x_1^N)$, which is obtained from $p(x_1^N, y_1^N)$ by integration with respect to $y_1^N$, is of the form $p(x_1^N) = p(x_1)p(x_2|x_1)...p(x_N|x_{N-1})$.
Similarly, the second case, called "Model 3", is obtained by considering $d = bc$.
The main aim of this paper is to study, via simulations, whether there is any noticeable difference in the degradation of KF quality, when data suit a GPMC model and when replacing the use of the true GPMC model with Model 2 on the one hand, and Model 3, on the other hand. Thus different data will be sampled according to GPMC and the optimal KF will be performed with the true model. Then two "approximate" KF, based on Model 2 and Model 3, will be performed and the results obtained will be compared.
Before presenting the experiments results let us make some remarks concerning the two approximations.
(i) The distributions $p(x_n, x_{n+1})$, $p(y_n, y_{n+1})$, and $p(x_n, y_n)$, are identical, for each $n = 1,..., N$, in the three models. This is of importance because the defenders of the classical Model 2 sometimes put forward the fact that $p(x_n, x_{n+1})$ often has a physical meaning and thus has to be defined first, and then the "noise" distribution $p(y_n | x_n)$ is given, also with a physical meaning. As these

distributions are equal in Models 2 and 3, using the latter does not contradict this viewpoint;
(ii) According to (2.2) and (2.3) GPMC can also be defined with the system

$$\begin{bmatrix} X_{n+1} \\ Y_{n+1} \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} X_n \\ Y_n \end{bmatrix} + \Sigma * \begin{bmatrix} U_{n+1} \\ V_{n+1} \end{bmatrix}, \tag{3.1}$$

with $\Sigma * (\Sigma *)^T = \Sigma$ and $\begin{bmatrix} U_{n+1} \\ V_{n+1} \end{bmatrix} \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix})$ independent from $(X_1^n, Y_1^n)$ for each $n = 1, \ldots, N$.
Model 2 consists of taking $\beta = 0$, and Model 3 consists of taking $\gamma = 0$. We can see that the very classical model, in which $\beta = 0$ and $\delta = 0$, is a particular case of the Model 2.
We performed numerous experiments and the results of some of them, expressed in the mean squared errors $MSE = \frac{1}{N}\sum_{n=1}^{N}(x_n - \hat{x}_n)^2$, are presented in Table 1.

| $d$ | $e$ | M1 | M2 | M3 | $d$ | $e$ | M1 | M2 | M3 |
|---|---|---|---|---|---|---|---|---|---|
| .05 | .05 | .37 | .67 | .47 | .35 | .05 | .50 | .60 | .51 |
| .05 | .15 | .49 | .55 | .49 | .35 | .15 | .51 | .55 | .51 |
| .05 | .25 | .46 | .48 | .47 | .35 | .25 | .48 | .49 | .49 |
| .05 | .35 | .41 | .41 | .41 | .35 | .35 | .44 | .44 | .45 |
| .05 | .45 | .36 | .37 | .36 | .35 | .45 | .38 | .39 | .40 |
| .05 | .55 | .28 | .33 | .28 | .35 | .55 | .30 | .35 | .32 |
| .15 | .05 | .49 | .62 | .49 | .45 | .05 | .50 | .59 | .51 |
| .15 | .15 | .50 | .54 | .50 | .45 | .15 | .50 | .54 | .50 |
| .15 | .25 | .47 | .48 | .47 | .45 | .25 | .48 | .49 | .50 |
| .15 | .35 | .43 | .43 | .43 | .45 | .35 | .44 | .44 | .46 |
| .15 | .45 | .37 | .38 | .37 | .45 | .45 | .38 | .39 | .41 |
| .15 | .55 | .30 | .34 | .30 | .45 | .55 | .29 | .35 | .33 |
| .25 | .05 | .50 | .60 | .50 | .55 | .05 | .50 | .59 | .52 |
| .25 | .15 | .50 | .54 | .50 | .55 | .15 | .50 | .55 | .51 |
| .25 | .25 | .48 | .49 | .48 | .55 | .25 | .48 | .50 | .50 |
| .25 | .35 | .43 | .43 | .44 | .55 | .35 | .44 | .44 | .47 |
| .25 | .45 | .38 | .39 | .39 | .55 | .45 | .38 | .39 | .42 |
| .25 | .55 | .30 | .35 | .31 | .55 | .55 | .27 | .33 | .34 |

Tab. 1 In all experiments $a = 0.5$, $b = 0.7$, and $c = 0.15$. True $d$ and $e$ corresponding to the true Model 1 (M1) are presented in columns 1, 2, 6, and 7, and the optimal MSE based on them is in columns 3 and 8. The MSE obtained with Model 2 ($e = ab$ and $X_1^N$ Markov) is in columns 4 and 9, while the MSE obtained with Model 3 ($d = bc$ and $Y_1^N$ Markov) is in columns 5 and 10. The sample size is $N = 1000$.

Thus considering Model 2 consists of taking $e = ab = 0.35$, and considering Model 3 consists of taking $d = cb = 0.105$. According to the results in table 1, and different other results we obtained, we can put forward the following conclusions:
(i) neither of the approximate models has the upper hand over the other in all cases considered ;
(ii) when the true $d$ varies between .05 and .45, the approximation with the Model 3, in which $d = cb = 0.105$, produces acceptable results, which are close to the optimal results ;
(iii) when the true $e$ is small (.05 or .15) or large (.45 or .55), the Model 2 approximation, in which $e = ab = 0.35$, produces less acceptable results than above. This is especially the case for $e = 0.5$, where the difference between the optimal MSE and the MSE obtained with the Model 2 is systematically about 0.1.

## 4 Switching models

Let us consider the random chain of switches $R_1^N = (R_1, ..., R_N)$, each $R_n$ taking its values in a finite set $\{1, ..., K\}$. Let $T_1^N = (T_1, ..., T_N)$ be a stationary TMC, with $T_n = (X_n, R_n, Y_n)$, whose distribution is given by $p(t_1, t_2) = p(r_1, r_2) p(x_1, y_1, x_2, y_2 | r_1, r_2)$. The distributions $p(x_1, y_1, x_2, y_2 | r_1, r_2)$ will be assumed to be defined by the matrix (2.1) depending on $(r_1, r_2)$. Applying Property 2 to $u = (x_n, y_n, y_{n+1})$ and $v = x_{n+1}$ we have $E[X_{n+1} | x_n, r_n, y_n, r_{n+1} y_{n+1}] = A(r_n^{n+1}) x_n + B(r_n^{n+1}) y_n + C(r_n^{n+1}) y_{n+1}$, which means that

$$X_{n+1} = A(r_n^{n+1}) X_n + B(r_n^{n+1}) Y_n + C(r_n^{n+1}) Y_{n+1} + D(r_n^{n+1}) W_{n+1} \qquad (4.1)$$

Assuming that $d(r_1, r_2) - b(r_1) c(r_1, r_2) = 0$ for each $r_1, r_2$ (we have Model 3 conditionally on $r_1^N$), let us specify how to compute $p(r_{n+1} | y_1^{n+1})$ and $E[X_{n+1} | r_{n+1}, y_1^{n+1}]$ from $p(r_n | y_1^n)$, $E[X_n | r_n, y_1^n]$, and $y^{n+1}$.

First, $(R_1^N, Y_1^N)$ being a hidden Markov chain we have $p(r_{n+1} | y_1^{n+1}) = [\sum_{r_n} p(r_n | y_1^n) p(r_{n+1}, y_{n+1} | r_n, y_n)] / [\sum_{r_n, r_2} p(r_n | y_1^n) p(r_{n+1}, y_{n+1} | r_n, y_n)]$. Second, in Model 3 we have $p(x_n | r_n, r_{n+1}, y_1^{n+1}) = p(x_n | r_n, y_1^n)$, which implies $E[X_n | r_n, r_{n+1}, y_1^{n+1}] = E[X_n | r_n, y_1^n]$. Thus taking the conditional expectation of (4.1) we obtain

$E[X_{n+1} | r_n, r_{n+1}, y_1^{n+1}] = A(r_n^{n+1}) E[X_n | r_n, r_{n+1}, y_1^{n+1}] + B(r_n^{n+1}) y_n + C(r_n^{n+1}) y_{n+1} =$

$A(r_n^{n+1})E[X_n | r_n, y_1^n] + B(r_n^{n+1})y_n + C(r_n^{n+1})y_{n+1}$, and thus

$$E[X_{n+1} | r_{n+1}, y_1^{n+1}] = \sum_{r_n} E[X_{n+1} | r_n, r_{n+1}, y_1^{n+1}] p(r_n | y_1^{n+1}) =$$

$$A(r_n^{n+1})[\sum_{r_n} E[X_n | r_n, y_1^n] p(r_n | y_1^{n+1}) + B(r_n^{n+1})y_n + C(r_n^{n+1})y_{n+1} =$$

$$A(r_n^{n+1}) \frac{\sum_{r_n} E[X_n | r_n, y_1^n] p(r_n | y_1^n) p(y_{n+1} | r_n, y_n)}{\sum_{r_n} p(r_n | y_1^n) p(y_{n+1} | r_n, y_n)} + B(r_n^{n+1})y_n + C(r_n^{n+1})y_{n+1}.$$

## 5 Conclusions

We considered a general pairwise Markov Gaussian linear system $(X_1^N, Y_1^N)$, in which an exact Kalman filter (KF) is workable, and we showed that there are two kinds of particular cases, which can be seen as two kinds of approximations of the general case. In the first classical case $X_1^N$ is Markovian while $Y_1^N$ is not, and in the second $Y_1^N$ is Markovian while $X_1^N$ is not. The main contribution of this paper was to show that when the processed data suit the general model, both particular cases produce similar approximate results. This is of importance when introducing stochastic Markovian switches. In fact, it is well known that KF is no longer workable in the first case, while it is in the second one, as recalled in section 4 above.

## References

1. C. Andrieu, C. M. Davy, and A. Doucet, Efficient particle filtering for jump Markov systems. Application to time-varying autoregressions, IEEE Trans. on Signal Processing, 51, 7, 1762-1770, 2003.
2. O. Cappé, E. Moulines, and T. Ryden, *Inference in hidden Markov models*, Springer, 2005.
3. O. L. V. Costa, M. D. Fragoso, and R. P. Marques, *Discrete time Markov jump linear systems*, New York, Springer-Verlag, 2005.
4. P. Giordani, R. Kohn, and D. van Dijk, A unified approach to nonlinearity, structural change, and outliers, *Journal of Econometrics*, 137, 112-133, 2007.
5. W. Pieczynski and F. Desbouvries, Kalman Filering using Pairwise Gaussian Models, International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hong-Kong, April 2003.
6. W. Pieczynski and N. Abbassi, Exact filtering and smoothing in short or long memory stochastic switching systems, *2009 IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2009)*, September 2-4, Grenoble, France, 2009.
7. J. K. Tugnait, Adaptive estimation and identification for discrete systems with Markov jump parameters, *IEEE Trans. on Automatic Control*, AC-25, 1054-1065, 1982.