

# Filtrage optimal dans les systèmes linéaires gaussiens markoviens à sauts

NOUFEL ABBASSI<sup>1</sup>, DALILA BENBOUDJEMA<sup>2</sup>, WOJCIECH PIECZYNSKI<sup>1</sup>

<sup>1</sup>Institut Telecom, Telecom Sudparis, Département CITI, CNRS UMR 5157,  
9, rue Charles Fourier, 91000 Evry,

<sup>2</sup>Equipe ETIS, CNRS UMR 8051, ENSEA, Université Cergy-Pontoise,  
F- 95000 Cergy-Pontoise, France

<sup>1</sup>Noufel.Abbassi@it-sudparis.eu, <sup>1</sup>Wojciech.Pieczynski@it-sudparis.eu, <sup>2</sup>Dalila.Benboudjema@ensea.fr

Résumé – Nous proposons une extension du modèle linéaire conditionnellement Gaussien markovien et à sauts markoviens classique, et étendons le filtre particulaire au nouveau modèle. Par ailleurs, nous considérons une sous famille de modèles appartenant à cette extension dans laquelle un filtrage optimal est possible, malgré la présence des sauts, avec une complexité linéaire en temps. Des simulations montrent l'intérêt de cette sous famille de par sa rapidité.

Abstract – We propose an extension of the classical conditionally Gaussian Linear State-Space Model (CGLSSM) and adapt particle filtering to this new model. We consider a subset of this family in which exact optimal filtering is workable with complexity linear in time, in spite of the presence of switches. Computer experiments show the interest of this family of models, especially in terms of computational cost.

## 1 Modèle conditionnellement de Markov couple gaussien (MCMCG).

On considère trois processus stochastiques  $X_1^N = (X_1, \dots, X_N)$ ,  $R_1^N = (R_1, \dots, R_N)$ , et  $Y_1^N = (Y_1, \dots, Y_N)$ . Les processus  $X_1^N$  et  $Y_1^N$  sont à valeurs respectivement dans  $\mathbb{R}^m$  et  $\mathbb{R}^q$ , alors que le processus  $R_1^N$  est à valeurs dans un espace fini des « sauts »  $S = \{1, \dots, K\}$ . Les processus  $X_1^N$  et  $Y_1^N$  sont cachés et le processus  $Y_1^N$  est observé. On supposera tout au long de cet article que le couple  $(X_1^N, Y_1^N)$  est gaussien conditionnellement à  $R_1^N$ . Le problème est d'estimer séquentiellement  $(X_1^N, R_1^N)$  à partir de  $Y_1^N$ . Plus précisément, on souhaite calculer – ou estimer –  $p(r_{n+1}|y_1^{n+1})$  et  $E[X_{n+1}|r_{n+1}, y_1^{n+1}]$  à partir de  $p(r_n|y_1^n)$ ,  $E[X_n|r_n, y_1^n]$ , et  $y_{n+1}$ . D'une part, cela donne le filtrage optimal au sens de l'erreur quadratique moyenne et, d'autre part, cela permet d'estimer, par une démarche bayésienne, la réalisation de  $R_1^N$ , ce qui donne une estimation de l'emplacement des sauts.

Dans l'approche classique on considère que  $R_1^N$  est une chaîne de Markov et la loi du couple  $(X_1^N, Y_1^N)$  conditionnelle à  $R_1^N = r_1^N$  est celle d'un système linéaire classique :

$$X_{n+1} = \alpha(r_{n+1})X_n + \sigma_1(r_{n+1})U_{n+1}; \quad (1)$$

$$Y_{n+1} = \gamma'(r_{n+1})X_{n+1} + \sigma_2(r_{n+1})V_{n+1}, \quad (2)$$

avec  $\alpha(r_{n+1})$ ,  $\sigma_1(r_{n+1})$ ,  $\gamma'(r_{n+1})$ , et  $\sigma_2(r_{n+1})$  des matrices de taille adéquate et  $U_2, \dots, U_N, V_2, \dots, V_N$  une suite vérifiant les hypothèses précisées ci-après dans un cadre plus général.

En reportant  $X_{n+1}$  donnée par (1) dans (2) on obtient  $Y_{n+1} = \gamma'(r_{n+1})\alpha(r_{n+1})X_n + \gamma'(r_{n+1})\sigma_1(r_{n+1})U_{n+1} + \sigma_2(r_{n+1})V_{n+1}$ ; il en résulte que (1) implique l'écriture sous la forme matricielle suivante, avec  $\gamma(r_{n+1}) = \gamma'(r_{n+1})\alpha(r_{n+1})$  et  $\rho_2(r_{n+1}) = \gamma'(r_{n+1})\sigma_1(r_{n+1})$  :

$$\begin{bmatrix} X_{n+1} \\ Y_{n+1} \end{bmatrix} = \begin{bmatrix} \alpha(r_{n+1}) & 0 \\ \gamma(r_{n+1}) & 0 \end{bmatrix} \begin{bmatrix} X_n \\ Y_n \end{bmatrix} + \begin{bmatrix} \sigma_1(r_{n+1}) & 0 \\ \rho_2(r_{n+1}) & \sigma_2(r_{n+1}) \end{bmatrix} \begin{bmatrix} U_{n+1} \\ V_{n+1} \end{bmatrix} \quad (3)$$

Considérons la généralisation suivante de (3).  $R_1^N$  est une chaîne de Markov et la loi du couple  $(X_1^N, Y_1^N)$  conditionnelle à  $R_1^N$  est celle d'un couple vérifiant le système linéaire suivant :

$$\begin{bmatrix} X_{n+1} \\ Y_{n+1} \end{bmatrix} = \begin{bmatrix} \alpha(r_n^{n+1}) & \beta(r_n^{n+1}) \\ \gamma(r_n^{n+1}) & \delta(r_n^{n+1}) \end{bmatrix} \begin{bmatrix} X_n \\ Y_n \end{bmatrix} + \begin{bmatrix} \sigma_1(r_n^{n+1}) & \rho_1(r_n^{n+1}) \\ \rho_2(r_n^{n+1}) & \sigma_2(r_n^{n+1}) \end{bmatrix} \begin{bmatrix} U_{n+1} \\ V_{n+1} \end{bmatrix}, \quad (4)$$

avec  $r_n^{n+1} = (r_n, r_{n+1})$ ,  $(U_2, V_2)^T, \dots, (U_N, V_N)^T$  suite des vecteurs gaussiens indépendant entre eux, centrés, de

matrices de covariance unité et telle que pour tout  $n=1, \dots, N-1$ ,  $(U_{n+1}, V_{n+1})^T$  est indépendant de  $(X_1^n, Y_1^n)$ . Enfin, on suppose que pour tout  $n=1, \dots, N-1$  les variables  $R_{n+1}$  et  $(X_n, Y_n)$  sont indépendantes conditionnellement à  $R_n$ .

Un tel modèle sera dit « Modèle conditionnellement de Markov couple gaussien » (MCMCG).

La loi de  $(X_1^N, Y_1^N)$  conditionnellement à  $R_1^N$  est alors celle d'un modèle de Markov couple gaussien dans lequel le couple  $(X_1^N, Y_1^N)$  est markovien. Dans ce dernier cas la loi de  $X_1^N$  conditionnelle à  $Y_1^N$  est markovienne et c'est également le cas de la loi de  $Y_1^N$  conditionnelle à  $X_1^N$ ; cependant, ni la loi de  $X_1^N$  ni celle de  $Y_1^N$  ne sont nécessairement markoviennes. Cependant, comme montré dans [8], le filtrage de Kalman exact classique peut être étendu, en conservant les mêmes avantages en terme de complexité algorithmique, aux modèles de Markov couples gaussiens.

Finalement, considérons le cas particulier du modèle (3) suivant :

$$\begin{bmatrix} X_{n+1} \\ Y_{n+1} \end{bmatrix} = \begin{bmatrix} \alpha(r_n^{n+1}) & \beta(r_n^{n+1}) \\ 0 & \delta(r_n^{n+1}) \end{bmatrix} \begin{bmatrix} X_n \\ Y_n \end{bmatrix} + \begin{bmatrix} \sigma_1(r_n^{n+1}) & \rho_1(r_n^{n+1}) \\ \rho_2(r_n^{n+1}) & \sigma_2(r_n^{n+1}) \end{bmatrix} \begin{bmatrix} U_{n+1} \\ V_{n+1} \end{bmatrix}, \quad (5)$$

qui sera appelé « Modèle (5) », et le cas particulier symétrique suivant

$$\begin{bmatrix} X_{n+1} \\ Y_{n+1} \end{bmatrix} = \begin{bmatrix} \alpha(r_n^{n+1}) & 0 \\ \gamma(r_n^{n+1}) & \delta(r_n^{n+1}) \end{bmatrix} \begin{bmatrix} X_n \\ Y_n \end{bmatrix} + \begin{bmatrix} \sigma_1(r_n^{n+1}) & \rho_1(r_n^{n+1}) \\ \rho_2(r_n^{n+1}) & \sigma_2(r_n^{n+1}) \end{bmatrix} \begin{bmatrix} U_{n+1} \\ V_{n+1} \end{bmatrix}, \quad (6)$$

qui sera appelé « Modèle (6) ». Notons que le degré de généralité du Modèle (6) se situe entre celle du modèle classique (3) et le modèle MCMCG général vérifiant (4). A  $R_1^N = r_1^N$  donné le filtre de Kalman (FK) est faisable dans les modèles (4), (5), et (6); il est alors intéressant de se poser la question suivante. Supposons que les données sont simulées selon le modèle général (4). On peut alors les filtrer selon trois FKs : celui utilisant le modèle (4), qui donne une erreur quadratique moyenne (EQM) minimale, celui utilisant le modèle (5) obtenu à partir du modèle (4) en remplaçant  $\gamma(r_n^{n+1})$  par 0, qui donne une EQM5, et celui utilisant le modèle (6)

obtenu à partir du modèle (4) en remplaçant  $\beta(r_n^{n+1})$  par 0, qui donne une EQM6. Les premières simulations présentées dans [3] indiquent que, d'une part, EQM5 et EQM6 sont comparables et, d'autre part, leurs augmentations par rapport à EQM semblent acceptables. En revenant au cas où le processus des sauts  $R_1^N$  n'est pas observable, nous avons les propriétés suivantes:

(i) Le filtrage optimal exact avec une complexité raisonnable est impossible dans le modèle classique (3) et des approximations, comme celles utilisant le filtrage particulière, sont indispensables [2, 4, 5, 6, 7, 10];

(ii) Bien entendu, le modèle MCMCG (4) étant plus général que le Modèle (3), le filtrage optimal exact n'est pas non plus possible dans le MCMCG ;

(iii) Le filtrage optimal exact avec complexité linéaire en temps est possible dans le modèle (4) [1, 9]; nous en rappelons la formule principale dans la section suivante ;

(iv) Dans le modèle classique (3) et dans le Modèle (6) le couple  $(X_1^N, R_1^N)$  est markovien et le couple  $(R_1^N, Y_1^N)$  ne l'est pas ; dans le Modèle (5) la situation est inverse : le couple  $(R_1^N, Y_1^N)$  est markovien et le couple  $(X_1^N, R_1^N)$  ne l'est pas ;

(v) Dans le modèle général MCMCG (4) aucun des couples  $(X_1^N, R_1^N)$ ,  $(R_1^N, Y_1^N)$  n'est nécessairement markovien.

Dans ce contexte, nous précisons deux contributions. Nous étendons le filtrage particulière, couramment utilisé dans le modèle (2), au modèle général (4) et étudions son comportement dans ce nouveau contexte ; Nous considérons des données générées par le modèle général (4) et étudions deux méthodes de filtrage « approximatif ». La première est le filtrage particulière mentionné dans le point (1) ci-dessus. La deuxième consiste à considérer un modèle (5) obtenu à partir du modèle (4), qui a servi pour générer les données, en mettant  $\gamma(r_n^{n+1}) = 0$ .

## 2 Filtrage particulière dans le MCMCG

Considérons le modèle MCMCG vérifiant (4). Nous présentons brièvement dans cette section une extension du filtrage particulière utilisé dans les modèles classique (3) au modèle MCMCG. Le problème vient du fait que la loi  $p(r_{n+1} | y_1^{n+1})$  ne peut pas être calculée récursivement et doit être approchée. La méthode brièvement décrite ci-dessous est une extension relativement directe de celles proposées dans les modèles classiques dans les références [1, 5].

Supposons que  $R_1^N = r_1^N$  est donné. La loi de  $(X_1^N, Y_1^N)$  conditionnelle à  $R_1^N = r_1^N$  est alors celle d'une chaîne de Markov gaussienne couple; il en résulte que  $p(x_{n+1}|r_1^{n+1}, y_1^{n+1})$  est calculable à partir de  $p(x_n|r_1^{n+1}, y_1^n)$  et  $y_{n+1}$  en utilisant le FK généralisé à modèles couples, comme proposé dans [8]. Par ailleurs, nous avons la formule suivante, qui donne  $p(r_1^{n+1}|y_1^{n+1})$  à partir de  $p(r_1^n|y_1^n)$  et  $y_{n+1}$ :

$$p(r_1^{n+1}|y_1^{n+1}) = \frac{p(y_{n+1}|r_1^{n+1}, y_1^n)p(r_{n+1}|r_n)}{p(y_{n+1}|y_1^n)} p(r_1^n|y_1^n) = \frac{p(y_{n+1}|r_1^{n+1}, y_1^n)p(r_{n+1}|r_n)}{\sum_{r_1^{n+1}} p(y_{n+1}|r_1^{n+1}, y_1^n)p(r_{n+1}|r_n)p(r_1^n|y_1^n)} p(r_1^n|y_1^n) \quad (7)$$

L'égalité (7) est alors utilisée séquentiellement pour approcher  $p(r_{n+1}|y_1^{n+1})$  par un ensemble de masses de Dirac pondérées, selon la technique classique du filtrage particulière. Pour commencer,  $p(r_1^n|y_1^n)$  est approché avec

$$p(r_1^n|y_1^n) \approx \hat{p}_{N_p}(r_1^n|y_1^n) = \frac{1}{N_p} \sum_{i=1}^{N_p} \delta\{r_1^{n,i}\} \quad (8)$$

où  $N_p$  est le nombre de trajectoires simulées. Connaissant  $p(r_1^n|y_1^n)$ ,  $p(x_n|y_1^n)$  est donné par

$$p(x_n|y_1^n) = \sum_{r_1^n} p(r_1^n|y_1^n)p(x_n|r_1^n, y_1^n) \quad (9)$$

Reportant (8) dans (9) on obtient

$$p(x_n|y_1^n) \approx \hat{p}_{N_p}(x_n|y_1^n) = \frac{1}{N_p} \sum_{i=1}^{N_p} p(x_n|r_1^{n,i}, y_1^n). \quad (10)$$

Précisons que les paramètres des lois  $p(y_{n+1}|r_1^{n+1}, y_1^n)$  dans (7) ceux des lois  $p(x_n|r_1^{n,i}, y_1^n)$ , pour  $i \in \{1, \dots, N_p\}$ , dans (10) sont calculés séquentiellement par le FK classique [8].

### 3 Filtrage exact dans le modèle approché

Le problème est de calculer  $p(r_{n+1}|y_1^{n+1})$  et  $E[X_{n+1}|r_{n+1}, y_1^{n+1}]$  à partir de  $p(r_n|y_1^n)$ ,  $E[X_n|r_n, y_1^n]$ , et  $y_{n+1}$ . La loi  $p(r_{n+1}|y_1^{n+1})$  est obtenue à partir de  $p(r_n|y_1^n)$  et  $y_{n+1}$  classiquement, en utilisant le fait que  $(R_1^N, Y_1^N)$  est une chaîne de Markov cachée. Par ailleurs, l'égalité (5) implique la possibilité de l'écriture suivante :

$$X_{n+1} = A(R_n^{n+1})X_n + B(R_n^{n+1})Y_n + C(R_n^{n+1})Y_{n+1} + D(R_n^{n+1})W_{n+1}, \quad (11)$$

avec  $A(R_n^{n+1}), B(R_n^{n+1}), C(R_n^{n+1})$ , et  $D(R_n^{n+1})$  des matrices de taille adéquate dépendant de  $r_n^{n+1}$  (calculées classiquement, en utilisant le conditionnement dans le cas gaussien, à partir de (4)) et  $W_2, \dots, W_N$  suite des vecteurs gaussiens indépendant entre eux, centrés, de matrices de covariance unité et telle que pour tout  $n = 1, \dots, N-1$ ,  $W_{n+1}$  est indépendant de  $(X_1^n, R_1^{n+1}, Y_1^n)$ . En prenant l'espérance de (11) conditionnelle à  $(R_{n+1}, Y_1^{n+1}) = (r_{n+1}, y_1^{n+1})$  nous avons, en utilisant le fait que dans le Modèle (5) on a  $E[X_n|r_n, r_{n+1}, y_1^{n+1}] = E[X_n|r_n, y_1^n]$  (car  $(R_{n+1}, Y_{n+1})$  et  $X_n$  sont indépendantes conditionnellement à  $(R_n, Y_n)$ ):

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

ce qui donne le filtre (voir aussi [1, 9]).

Notons que le calcul séquentiel de la variance  $Var[X_n|r_n, y_1^n]$ , non précisé ici faute de place, est analogue à celui de  $E[X_{n+1}|r_{n+1}, y_1^{n+1}]$  ci-dessus.

### 4 Exemples de résultats

La méthode de référence (MR), optimale, utilise le  $r_1^N$  simulé. FP est la méthode utilisant le filtrage particulière (500 particules), et NM est la « nouvelle méthode » utilisant le Modèle (4). L'erreur, calculée sur les 50 derniers points, est donnée par

$$\frac{1}{N} \sqrt{\sum_{n=1}^N (x_n - \hat{x}_n)^2}.$$

On considère une chaîne triplet  $(X_1^N, R_1^N, Y_1^N)$  homogène avec deux sauts possibles. La chaîne de Markov  $R_1^N$ , supposée stationnaire et à deux états, vérifie  $p(r_1 = r_2) = 0.8$  et  $p(r_1 \neq r_2) = 0.2$ . La loi  $p(x_1^N, y_1^N|r_1^N)$  est alors donnée par les lois de  $(X_1, Y_1, X_2, Y_2)$  conditionnelle à  $(R_1, R_2)$ . On suppose

ces lois de moyennes nulles et de matrice de covariance

$$\begin{bmatrix} \Gamma & A^T(r_2) \\ A(r_2) & \Gamma \end{bmatrix}, \text{ avec } \Gamma = \begin{bmatrix} 1 & b \\ b & 1 \end{bmatrix}, A = \begin{bmatrix} a(r_2) & e(r_2) \\ d(r_2) & c(r_2) \end{bmatrix}.$$

On montre alors que  $\begin{bmatrix} \alpha(r_2) & \beta(r_2) \\ \gamma(r_2) & \delta(r_2) \end{bmatrix} = A\Gamma^{-1}A^T$ , ce qui donne, en particulier :

$$\gamma(r_2) = \frac{d(r_2) - bc(r_2)}{1 - b^2} \quad (12)$$

On considère alors cinq matrices  $A^1, \dots, A_5$  correspondant aux cinq cas du Tableau 1. Dans tous les cas on a  $a(1) = 0.1$ ,  $a(2) = 0.5$ ,  $c(1) = 0.4$ ,  $c(2) = 0.9$ ,  $e(1) = 0.75$ ,  $e(2) = 0.33$ . Les cinq cas correspondent aux cinq  $d_1(r_2), \dots, d_5(r_2)$  tels que les  $\gamma_1(r_2), \dots, \gamma_5(r_2)$  donnés par (12) varient de 0.1 à 0.4 avec un pas de 0.1.

**Tab 1 : comparaisons entre le filter de Kalman utilisant les vrais sauts (MR, le filtre particulaire (FP), et la nouvelle méthode (NM)**

	Cas 1	Cas 2	Cas 3	Cas 4	Cas 5
$\gamma$	0.0	0.1	0.2	0.3	0.4
	Erreur quadratique moyenne entre $x_1^N$ et $\hat{x}_1^N$				
MR	0.0580	0.0574	0.0539	0.0570	0.0619
FP	0.0595	0.0596	0.0542	0.0562	0.0642
NM	0.0588	0.0590	0.0537	0.0554	0.0648
	Taux d'erreur entre $r_1^N$ et $\hat{r}_1^N$				
MR	0 %	0 %	0 %	0 %	0 %
FP	22.0%	24.0%	23.0%	22.0%	25.5%
NM	21.5%	22.5%	22.5%	22.5%	24.5%
	Temps d'exécution informatique en secondes				
MR	0.15	0.14	0.15	0.16	0.12
FP	123.58	113.17	156.28	152.19	113.56
NM	0.49	0.48	0.75	0.66	0.47

On note la bonne robustesse, dans le cadre des expérimentations présentées, de la nouvelle méthode par rapport aux paramètres  $\gamma(r_n)$  : le cas 5 est « extrême » dans la mesure où au delà le modèle servant à générer les données n'est plus défini.

On note également le bon comportement des méthodes FP et MN par rapport à MR, avec un fort avantage de MN en rapidité d'exécution.

## 5 Conclusion et perspectives

On note que, dans le cadre de nos simulations, le très bon comportement du filtrage particulaire, avéré dans

les modèles classiques (3), est préservé dans le modèle général (4). Par ailleurs la qualité du filtrage, ainsi que celle de l'estimation des sauts, par la nouvelle méthode est comparable à celle obtenue avec le filtrage particulaire. L'intérêt de la nouvelle méthode réside dans la complexité calculatoire, environ deux cent fois plus rapide que celle utilisant le filtrage particulaire (pour 500 particules).

D'autres expérimentations sont nécessaires pour valider l'intérêt de la nouvelle méthode, en particulier dans les cas des données observées et cachées vectorielles.

## 6 Références

- [1] N. Abbassi, D. Benboudjema, and W. Pieczynski, Kalman filtering approximations in triplet Markov Gaussian switching models, Statistical Signal Processing Workshop (SSP 2011), Nice, France, June 28-30, 2011.
- [2] 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.
- [3] D. Benboudjema, M. Malainin, and W. Pieczynski, Exact Kalman filtering in pairwise Gaussian switching systems, Applied Stochastic Models and Data Analysis (ASMDA 2011), June 7-10, Roma, Italy, 2011.
- [4] O. Cappé, E. Moulines, and T. Ryden, Inference in hidden Markov models, Springer, 2005.
- [5] O. L. V. Costa, M. D. Fragoso, and R. P. Marques, *Discrete time Markov jump linear systems*, New York, Springer-Verlag, 2005.
- [6] A. Doucet, N. J. Gordon, and V. Krishnamurthy, Particle filters for state estimation of Jump Markov Linear Systems, *IEEE Trans. On signal Processing*, 49 pp.613-624, 2001.
- [7] P. Giordani, R. Kohn, and D. van Dijk, A unified approach to nonlinearity, structural change, and outliers, *Journal of Econometrics*, 137, 112-133, 2007.
- [8] W. Pieczynski and F. Desbouvries, Kalman Filtering using Pairwise Gaussian Models, International Conference on Acoustics, Speech and Signal Processing (ICASSP 2003), Hong-Kong, 2003.
- [9] W. Pieczynski, Exact filtering in Markov marginal switching hidden models, *Comptes Rendus Mathématique*, Vol. 349, No. 9-10, pp. 587-590, May 2011.
- [10] J. K. Tugnait, Adaptive estimation and identification for discrete systems with Markov jump parameters, *IEEE Trans. on Automatic Control*, AC-25, 1054-1065, 1982.