

Filtrage exact partiellement non supervisé dans les modèles cachés à sauts markoviens

NOUFEL ABBASSI, WOJCIECH PIECZYNSKI

Laboratoire Communications, Images et Traitement de l'information
9 rue Charles Fourier, 91000 Evry, France

Noufel.Abbassi@it-sudparis.eu, Wojciech.Pieczynski@it-sudparis.eu

Résumé – On considère le problème de filtrage optimal dans les systèmes linéaires gaussiens à sauts markoviens. Ce problème étant NP-complet, différentes approximations ont été proposées, en particulier des algorithmes de filtrage de Kalman mixte qui utilisent partiellement le filtrage particulaire. Nous proposons deux méthodes alternatives d'approximation du filtre optimal, qui présentent des formules analytiques et une complexité linéaire en N , N étant le nombre d'observations. Les résultats de simulations montrent l'existence de situations dans lesquelles ces méthodes sont compétitives par rapport à certaines solutions classiques fondées partiellement sur le filtrage particulaire.

Abstract – We consider the optimal filtering problem in Gaussian Markov switching linear systems. This problem is NP-hard, so a number of approximate methods have been proposed, in particular mixture Kalman filters which are based partially on particle filtering. We propose two methods for approximating the optimal filter, which are based on closed-form formulas and have a computational cost linear in N , where N is the number of observations. Some simulation results provided in the paper show the existence of situations in which the proposed methods prove more efficient than some of the classical solutions partially based on particle filtering.

1 Introduction

On considère $X_1^N = (X_1, \dots, X_N)$, $Y_1^N = (Y_1, \dots, Y_N)$ deux processus stochastiques, chaque X_n et chaque Y_n prenant leurs valeurs dans l'ensemble des nombres réels \mathfrak{R} , et $R_1^N = (R_1, \dots, R_N)$ un processus tel que chaque R_n prend ses valeurs dans un ensemble fini $\Omega = \{\lambda_1, \dots, \lambda_m\}$. On suppose que les réalisations de Y_1^N sont observables alors que ceux de R_1^N et X_1^N ne le sont pas. Nous considérons le problème de filtrage optimal, au sens de l'erreur quadratique moyenne, avec complexité linéaire en N . Dans cet article, nous considérerons que ce problème consiste en calcul $p(r_{n+1}|y_1^{n+1})$ et $E[X_{n+1}|r_{n+1}, y_1^{n+1}]$ à partir de $p(r_n|y_1^n)$ et $E[X_n|r_n, y_1^n]$. En l'absence du processus des sauts R_1^N on est confronté à un problème de filtrage classique qui est résolu dans le cas des systèmes linéaires et gaussiens par le filtre de Kalman. L'introduction de R_1^N , qui peut s'interpréter comme modélisant les "changements de régime" du modèle classique, présente au moins deux types d'intérêts :

- (i) modélisation des situations dans lesquelles il existe un nombre fini de stationnarité différentes dans le modèle linéaire gaussien classique ;
- (ii) l'approximation des modèles non gaussiens et non linéaires par un mélange de modèles linéaires gaussiens.

Le modèle probabiliste classique simple définissant la loi du triplet (X_1^N, R_1^N, Y_1^N) est le suivant:

$$R_1^N \text{ est une chaîne de Markov ;} \quad (1)$$

$$X_n = F(R_n)X_{n-1} + H(R_n)W_n ; \quad (2)$$

$$Y_n = G(R_n)X_n + K(R_n)V_n, \quad (3)$$

où (W_n) et (V_n) sont des variables indépendantes gaussiennes. Conditionnellement à R_1^N , la loi du couple (X_1^N, Y_1^N) est celle du modèle classique auquel le filtrage de Kalman est applicable [2, 6, 7]. Il s'agit donc là d'une extension « naturelle » au modèle à sauts. Cependant, lorsque R_1^N dans le modèle (1)-(3) n'est pas observé, il n'est plus possible de calculer les quantités $p(r_n|y_1^n)$ et $E[X_n|r_n, y_1^n]$ avec une complexité linéaire (ou même polynomiale) en N [1, 2, 3, 5, 10]. Il existe alors un certain nombre de techniques d'approximation de la solution exacte, qui peuvent être déterministes [3] ou stochastiques [1, 2, 5].

L'objet de l'article est de proposer deux nouvelles méthodes d'approximation du filtrage exact dans le modèle (1)-(3) et d'évaluer leur intérêt en les comparant à la méthode fondée partiellement sur le filtrage particulaire [1, 5].

La première consiste en remplacement de la loi de (R_1^N, Y_1^N) , dont la forme dans le modèle (1)-(3), non explicite, est à l'origine des difficultés, par la loi d'une chaîne de Markov cachée classique. Le filtrage exact dans un tel modèle approché devient alors possible et, les paramètres de la loi $p(r_1^N, y_1^N)$ du couple (R_1^N, Y_1^N) étant estimés par la méthode ICE [4], nous aboutissons à une méthode partiellement non-supervisée.

La deuxième méthode consiste en remplacement de $p(r_1^N, y_1^N)$ dans le modèle (1)-(3) par la loi d'une chaîne couple « partiellement » de Markov (CCPM [8]). Comme dans la première méthode, le filtrage exact dans un tel modèle approché devient alors possible.

Nous menons une série de simulations visant à comparer les trois méthodes en présence. Il s'avère que la première méthode améliore légèrement les résultats du filtrage partiellement particulière classique, alors que la deuxième peut les améliorer sensiblement.

2 Estimation de la trajectoire cachée

Considérons le système (1)-(3). Nous cherchons à calculer $p(r_{n+1}|y_1^{n+1})$ et $E[X_{n+1}|r_{n+1}, y_1^{n+1}]$ à partir de $p(r_n|y_1^n)$ et $E[X_n|r_n, y_1^n]$. Nous pouvons écrire classiquement

$$p(x_{n+1}|r_{n+1}, y_1^n) = \int_{\mathcal{R}} p(x_{n+1}|x_n, r_{n+1})p(x_n|r_{n+1}, y_1^n)dx_n \quad (4)$$

$$p(x_{n+1}|r_{n+1}, y_1^{n+1}) = \frac{p(x_{n+1}, y_{n+1}|r_{n+1}, Y_1^n)}{p(y_{n+1}|r_{n+1}, y_1^n)} =$$

$$\frac{p(x_{n+1}|r_{n+1}, y_1^n)p(y_{n+1}|r_{n+1}, x_{n+1})}{p(y_{n+1}|r_{n+1}, y_1^n)} \quad (5)$$

Le calcul de (4) ne pose pas de problème et l'on obtient une solution explicite. Le problème vient du fait que le dénominateur $p(y_{n+1}|r_{n+1}, y_1^n)$ figurant dans (5) n'est pas calculable avec une complexité linéaire en N . Cependant, cette quantité serait calculable récursivement si le couple (R_1^N, Y_1^N) était une chaîne de Markov cachée classique, dont le graphe de dépendance est présenté en Figure 1, (b). En effet, on aurait

$$p(y_{n+1}|r_{n+1}, Y_1^n) = \frac{p(y_{n+1}, r_{n+1}|y_1^n)}{p(r_{n+1}|y_1^n)} = \frac{p(y_{n+1}|y_1^n)p(r_{n+1}|y_1^{n+1})}{p(r_{n+1}|y_1^n)},$$

avec les quantités $p(y_{n+1}|y_1^n)$, $p(r_{n+1}|y_1^{n+1})$, et $p(r_{n+1}|y_1^n)$ calculable classiquement. Notre première méthode consiste en utilisation de (5) avec $p(y_{n+1}|r_{n+1}, Y_1^n)$ calculé comme dans la chaîne de Markov cachée classique. Afin d'effectuer ces calculs on doit disposer des paramètres de la chaîne cachée en question ; nous les estimons par la méthode « Iterative Conditional Estimation » (ICE [4]), qui est ici modifiée pour correspondre au caractère adaptatif du filtrage. Nous obtenons ainsi une méthode de filtrage « partiellement » non supervisée dans la mesure où une partie des paramètres nécessaires est estimée uniquement à partir de la chaîne observée Y_1^N .

Cette démarche peut être étendue aux modèles plus généraux, comme les chaînes de Markov couples [4], ou chaînes couples partiellement de Markov (CCPM [8]),

dont le graphe de dépendance orienté est présenté à la Figure 1, (c) (on notera que la flèche reliant les y va directement de y_1 à y_3). Ce dernier modèle admet pour loi

$$p(r_1^N, y_1^N) = p(r_1, y_1) \prod_{n=1}^{N-1} p(r_{n+1}, y_{n+1}|r_n, y_1^n), \quad (6)$$

l'appellation « partiellement » de Markov venant du fait que le couple (R_1^N, Y_1^N) est markovien par rapport à R_1^N , mais ne l'est pas nécessairement par rapport à Y_1^N . Dans un tel modèle R_1^N est markovien conditionnellement à Y_1^N et dans le cas gaussien considéré ici la loi $p(y_{n+1}|r_{n+1}, y_1^n)$, qui est gaussienne, est calculable par des formules analogues à celles utilisées dans le cas des chaînes de Markov cachées classiques [8]. En calculant $p(y_{n+1}|r_{n+1}, y_1^n)$ dans (5) par cette démarche définit notre deuxième méthode de calcul approximative de (5).

Notons que d'autres modèles, comme les chaînes semi-markoviennes, pourraient être utilisées de manière analogue en permettant de concevoir autant de méthodes de filtrage approximatif originales.

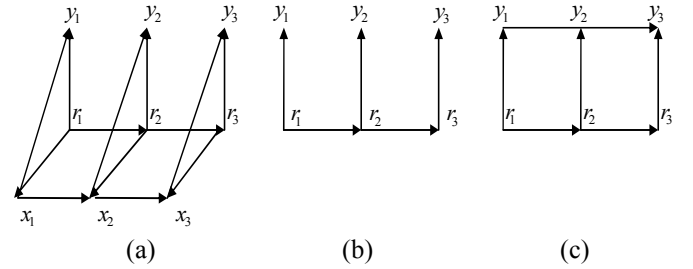


Figure 1 : graphes orientés de dépendance. (a) modèle classique, (b) CMC classique, (c) CCPM.

Nous comparons nos deux méthodes avec la méthode approchée utilisant partiellement le filtrage particulière décrite dans [3, 6], Présenté ici succinctement, son principe est le suivant. Considérons

$$p(x_n, r_1^n|y_1^n) = p(r_1^n|y_1^n)p(x_n|r_1^n, y_1^n) \quad (7)$$

Le problème est alors de calculer $p(x_n|r_1^n, y_1^n)$ et $p(r_1^n|y_1^n)$. Le filtrage de Kalman permet de calculer les paramètres des densités gaussiennes $p(x_n|r_1^n, y_1^n)$, et on utilise le filtrage particulière pour approximer la probabilité jointe a posteriori $p(r_1^n|y_1^n)$.

3 Simulations

Soit R_1^N une chaîne de Markov stationnaire à deux états ($m = 2$), avec $\Omega = \{\lambda_1, \lambda_2\}$, on considère qu'elle est stationnaire au sens strict, avec pour loi initiale $p_1 = (0.5, 0.5)$ pour matrice des transitions symétrique : $p(r_2|r_1 = r_2) = 1 - \rho$, $p(r_2|r_1 \neq r_2) = \rho$. Le paramètre ρ définit ainsi la « fréquence moyenne des sauts » dans un intervalle de temps donné.

Les fonctions F, G, H et K sont définies par $F(\lambda_1) = -a$, $F(\lambda_2) = a$, $G(\lambda_1) = -b$, $G(\lambda_2) = b$, et, pour $i = 1, 2$, $H(\lambda_i) = \Delta_i$, $K(\lambda_i) = S_i$. Nous considérons $b > 0$, $a > 0$, $\Delta_1 \neq \Delta_2$ et $S_1 \neq S_2$.

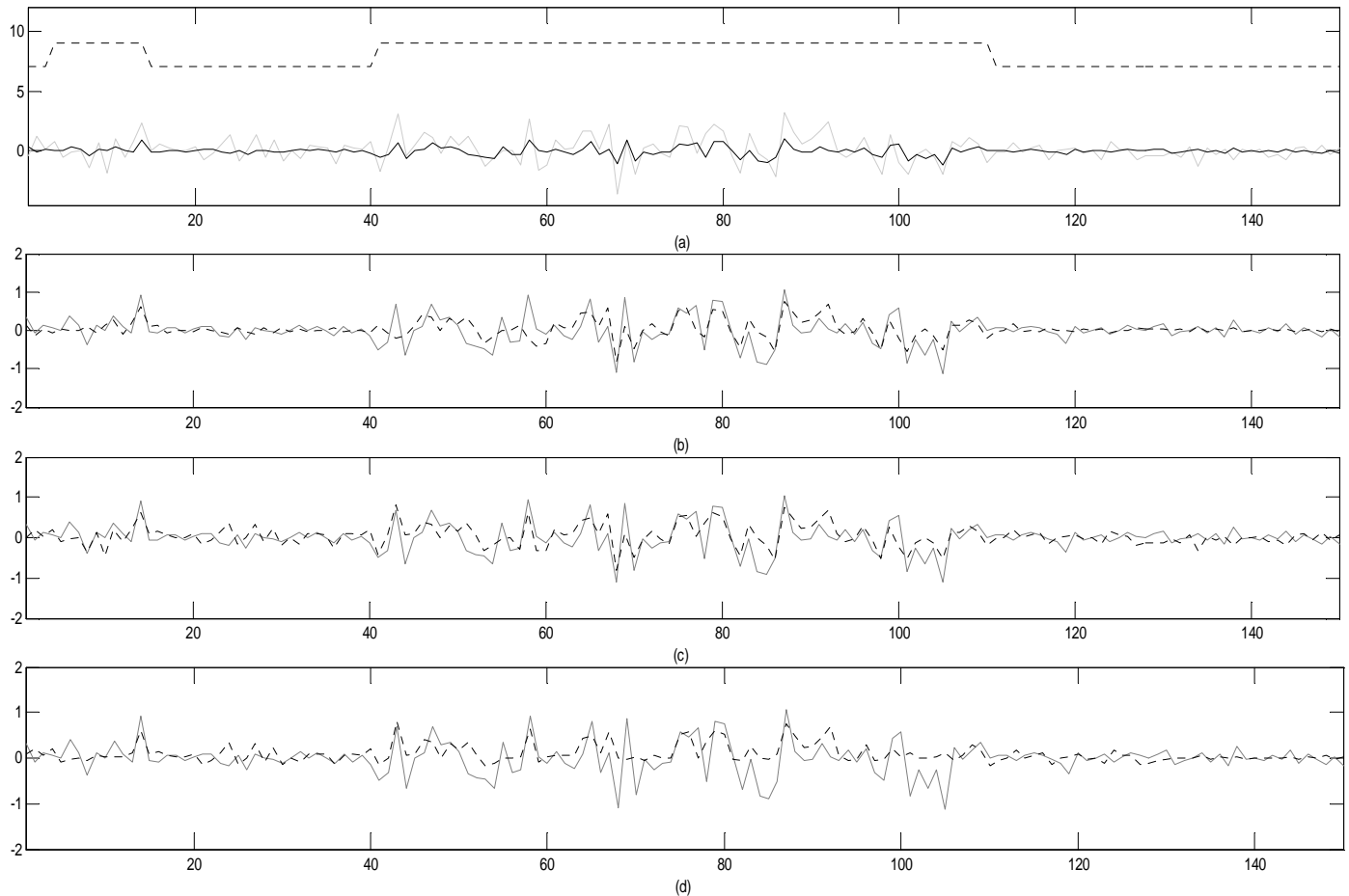
Nous présentons ci-après les résultats des simulations correspondant aux paramètres $b = 2$, $a = 0.25$, $\Delta_1 = 0.1$, $\Delta_2 = 0.5$, et $S_1 = 0.5$, $S_2 = 1$. Le paramètre ρ prend plusieurs valeurs possibles, et la différence entre le vrai signal caché $x = x_1^N$ et le signal estimé

$\hat{x} = \hat{x}_1^N$ est mesurée par les quantités $\frac{1}{N} \sum_{n=1}^N (x_n - \hat{x}_n)^2$, présentées dans le Tableau 1. Nous avons utilisé le tirage de 250 particules dans l'algorithme FP, leur augmentation n'apportant pas d'amélioration significative des résultats.

Nous constatons que les résultats obtenus avec les nouvelles méthodes sont presque toujours de meilleure qualité que ceux obtenus avec la méthode FP. Cette dernière donne des résultats comparables lorsque ρ est petit, ce qui signifie que les sauts sont rares. Dans les situations contraires, lorsque les sauts deviennent plus fréquents la différence des résultats au profit des nouvelles méthodes devient plus significative. Certains résultats significatifs sont présentés dans le Tableau 1 et à la Figure 2, d'autres simulations effectuées confirment ces tendances générales. Ces différents résultats tendent à montrer que la méthode FP éprouve plus de difficultés à gérer les sauts que les méthodes proposées.

Tab 1 : les écarts entre le signal simulé et les restaurations par le filtrage particulaire (FP), la méthode utilisant les chaînes couples partiellement de Markov (CCPM), et celle utilisant les chaînes de Markov cachées (CMC)

ρ	0.02	0.40	0.80
FP	0.0836	0.1647	0.1873
CCPM	0.0784	0.0874	0.1055
CMC	0.0966	0.1237	0.1263



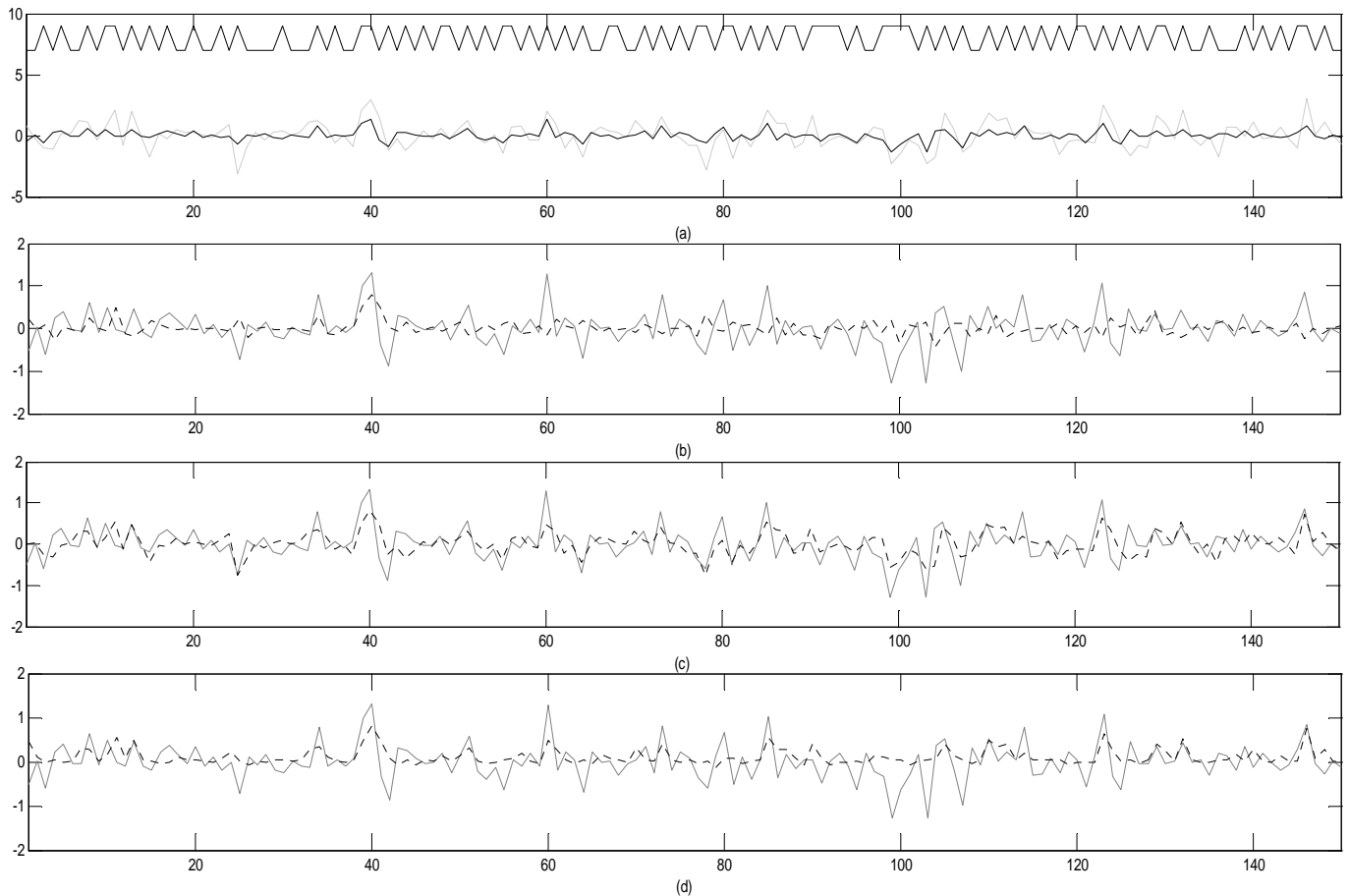


Figure 2 : lignes 1-4 correspondent à $\rho = 0.02$, et les lignes 5-8 correspondent à $\rho = 0.80$. (a) : trajectoires simulées des trois processus ; (b) : restauration par le filtrage particulaire (FP), (c) : restauration par la méthode utilisant les chaînes couples partiellement de Markov (CCPM), et (d) : restauration par la méthode utilisant les chaînes de Markov cachées (CMC). La vraie trajectoire est en continu et les restaurations en pointillés

4 Conclusions

Nous avons proposé deux méthodes originales - dont une partiellement non supervisée - de filtrage approché, présentant des formules analytiques et une complexité linéaire en nombre d'observations, dans les systèmes linéaires gaussiens à sauts markoviens. Les résultats de simulations présentés montrent l'existence de situations dans lesquelles ces méthodes sont compétitives par rapport à celles utilisant partiellement le filtrage particulaire.

5 Références

1. C. Andrieu, M. Davy, and A. Doucet, Efficient particle filtering for jump Markov systems. Application to time-varying autoregressions, *IEEE Trans. on Signal Processing*, Vol. 51, No. 7, pp. 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. S. Derrode and W. Pieczynski, Signal and Image Segmentation using Pairwise Markov Chains, *IEEE Trans. on Signal Processing*, Vol. 52, No. 9, pp. 2477-2489, 2004.
5. A. Doucet, N. J. Gordon, and V. Krishnamurthy, Particle filters for state estimation of Jump Markov Linear Systems, *IEEE Trans. on Signal Processing*, Vol. 49, pp. 613-624, 2001.
6. Y. Ephraim and N. Merhaw, Hidden Markov processes, *IEEE Trans. on Information Theory*, Vol. 48, No. 6, pp. 1518-1569, 2002.
7. A. H. Jazwinski, *Stochastic processes and filtering theory*, Vol. 64, Mathematics in Science and Engineering, Academic Press, San Diego, 1970.
8. P. Lanchantin, J. Lapuyade-Lahorgue, and W. Pieczynski, Unsupervised segmentation of triplet Markov chains with long-memory noise, *Signal Processing*, Vol. 88, No. 5, pp 1134-1151, May 2008.
9. W. Pieczynski and F. Desbouvries, On triplet Markov chains, International Symposium on Applied Stochastic Models and Data Analysis, (ASMDA 2005), Brest, France, May 2005.
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.