

Un algorithme d'estimation non supervisée dans des modèles Markoviens à sauts

Yohan PETETIN¹, François DESBOUVRIES²

¹CEA/LIST, 91191 Gif-sur-Yvette, France

²Télécom SudParis/CITI & CNRS UMR 5157, 91000 Evry, France

yohan.petetin@cea.fr, francois.desbouvries@telecom-sudparis.eu

Résumé – Cette communication traite du problème du calcul de l'estimateur bayésien optimal (au sens de l'erreur quadratique moyenne) dans des modèles markoviens à sauts. Nous utilisons pour ce faire une technique récente de calcul rapide et exact de cet estimateur dans une classe de modèles de Markov triplet particuliers. Le problème se ramène alors au problème d'estimation des paramètres de ces modèles; nous proposons une méthode d'estimation basée d'une part sur la minimisation d'une divergence de Kullback-Leibler et d'autre part sur une étape d'estimation non supervisée.

Abstract – This paper addresses the problem of computing the optimal bayesian estimator (in the mean square sense) in some jump Markov models. We make use of a recent algorithm which computes this estimator fastly and exactly in some particular Triplet Markov models. The problem then reduces to a parameter estimation problem; our two-step estimator is based on a Kullback-Leibler divergence minimization and an unsupervised parameter estimation step.

1 Introduction

Nous nous intéressons au problème de filtrage statistique dans des modèles de Markov à sauts. $\{\mathbf{x}_k\}_{k \geq 0} \in \mathbf{R}^p$ désigne une suite de vecteurs aléatoires (v.a.) cachés et $\{\mathbf{y}_k\}_{k \geq 0} \in \mathbf{R}^q$ une suite de v.a. observés. Nous considérons également un troisième processus discret $\{r_k\}_{k \geq 0}$, $r_k \in \{1, \dots, K\}$ visant à modéliser des changements de régime. Un modèle classique dans ce contexte de changements de régimes est le modèle de Markov caché à sauts markoviens (Jump Markov state space system ou JMSS) qui s'écrit

$$p^1(\mathbf{x}_{0:k}, \mathbf{y}_{0:k}, \mathbf{r}_{0:k}) = p^1(r_0) \underbrace{\prod_{i=1}^k p^1(r_i | r_{i-1})}_{p^1(\mathbf{r}_{0:k})} \times \underbrace{p^1(\mathbf{x}_0 | r_0) \prod_{i=1}^k f_{i|i-1}(\mathbf{x}_i | \mathbf{x}_{i-1}, r_i)}_{p^1(\mathbf{x}_{0:k} | \mathbf{r}_{0:k})} \underbrace{\prod_{i=0}^k g_i(\mathbf{y}_i | \mathbf{x}_i, r_i)}_{p^1(\mathbf{y}_{0:k} | \mathbf{x}_{0:k}, \mathbf{r}_{0:k})}. \quad (1)$$

Dans ce modèle, la suite des états discrets $\{r_k\}_{k \geq 0}$ est une chaîne de Markov dont la loi de transition est donnée par $p^1(r_i | r_{i-1})$, et conditionnellement aux sauts $\mathbf{r}_{0:k}$, la suite $\{\mathbf{x}_k, \mathbf{y}_k\}_{k \geq 0}$ forme une chaîne de Markov cachée. La transition de l'état caché est donnée par la densité de probabilité $f_{k|k-1}(\mathbf{x}_k | \mathbf{x}_{k-1}, r_k)$ et la vraisemblance par $g_k(\mathbf{y}_k | \mathbf{x}_k, r_k)$. Dans ce papier, nous nous plaçons dans des modèles linéaires et gaussiens ($\mathcal{N}(\mathbf{x}; \mathbf{m}; \mathbf{P})$ désigne la densité de probabilité de la loi multi-normale de

moyenne \mathbf{m} et de matrice de covariance \mathbf{P} prise au point \mathbf{x}) :

$$p(\mathbf{x}_0 | r_0) = \mathcal{N}(\mathbf{x}_0; \mathbf{m}_0; \mathbf{P}_0), \quad (2)$$

$$f_{k|k-1}(\mathbf{x}_k | \mathbf{x}_{k-1}, r_k) = \mathcal{N}(\mathbf{x}_k; \mathbf{F}_k(r_k) \mathbf{x}_{k-1}; \mathbf{Q}_k(r_k)), \quad (3)$$

$$g_k(\mathbf{y}_k | \mathbf{x}_k, r_k) = \mathcal{N}(\mathbf{y}_k; \mathbf{H}_k(r_k) \mathbf{x}_k; \mathbf{R}_k(r_k)). \quad (4)$$

Il est bien connu que dans ce modèle le calcul de l'estimateur bayésien optimal (au sens de la minimisation de l'erreur quadratique moyenne) de \mathbf{x}_k ,

$$\Theta_k = \mathbb{E}(\mathbf{x}_k | \mathbf{y}_{0:k}), \quad (5)$$

est un problème NP-difficile. Deux grandes classes d'approximations ont été proposées :

- les méthodes d'approximations numériques de type "Interacting Multiple Models" (IMM) [2] consistent à propager plusieurs filtres de Kalman (FK) dont la sortie est pondérée de façon à approcher l'estimateur (5) ;
- la seconde classe d'approximations repose sur l'utilisation des méthodes de Monte Carlo séquentielles. Plus précisément, Θ_k se réécrit

$$\Theta_k = \sum_{\mathbf{r}_{0:k}} p^1(\mathbf{r}_{0:k} | \mathbf{y}_{0:k}) \mathbb{E}(\mathbf{x}_k | \mathbf{y}_{0:k}, \mathbf{r}_{0:k}); \quad (6)$$

puis $p^1(\mathbf{r}_{0:k} | \mathbf{y}_{0:k})$ est approché par échantillonnage d'importance séquentiel tandis que $\mathbb{E}(\mathbf{x}_k | \mathbf{y}_{0:k}, \mathbf{r}_{0:k})$ est calculé par FK [5].

En pratique l'utilisation de ces modèles nécessite la connaissance de la loi des sauts $p^1(r_k | r_{k-1})$; si ce n'est pas le cas $p^1(r_k | r_{k-1})$ peut être estimé par filtrage particulière (FP).

De nouvelles techniques d'estimation dans un modèle à sauts ont été proposées récemment, basées sur le résultat suivant. La loi de probabilité (1) fait partie d'un ensemble de lois qui toutes s'écrivent sous la forme

$$p(\mathbf{x}_{0:k}, \mathbf{y}_{0:k}, \mathbf{r}_{0:k}) = p(\mathbf{x}_0, \mathbf{y}_0, r_0) \times \prod_{i=1}^k p(\mathbf{x}_i, \mathbf{y}_i, r_i | \mathbf{x}_{i-1}, \mathbf{y}_{i-1}, r_{i-1}); \quad (7)$$

dit d'une autre façon, le modèle de Markov à sauts (1) est le modèle de Markov triplet (7) particulier qui satisfait

$$p^1(\mathbf{x}_k, \mathbf{y}_k, r_k | \mathbf{x}_{k-1}, \mathbf{y}_{k-1}, r_{k-1}) = p^1(r_k | r_{k-1}) \times f_{k|k-1}(\mathbf{x}_k | \mathbf{x}_{k-1}, r_k) g_k(\mathbf{y}_k | \mathbf{x}_k, r_k).$$

Or il a été prouvé récemment le résultat suivant [8] [1]. Dans le modèle (7), si

- i) $p(\mathbf{y}_k, r_k | \mathbf{x}_{k-1}, \mathbf{y}_{k-1}, r_{k-1}) = p(\mathbf{y}_k, r_k | \mathbf{y}_{k-1}, r_{k-1})$,
- ii) $E(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_{k-1:k}, \mathbf{r}_{k-1:k})$ est calculable,

alors l'estimateur bayésien optimal (5) est calculable exactement et récursivement avec un coût de calcul linéaire en nombre d'observations (les formules de filtrage ne sont pas rappelées ici faute de place).

Bien que le modèle de Markov à sauts (1) ne vérifie pas les conditions i) et ii), ce résultat ouvre la voie à de nouvelles méthodes de filtrage statistique dans un JMSS, dans la mesure où il est possible de définir un modèle satisfaisant i) et ii) (- et donc dans lequel le problème de filtrage est résolu de façon très rapide) "proche" (selon un sens à préciser) du modèle p^1 originel. Différentes contributions ont été proposées dans ce sens [9] [4] [6]. En particulier dans [6] la démarche est la suivante. La classe des modèles de Markov Triplet satisfaisant les deux contraintes physiques (3) (4) est précisée; il s'agit d'une famille de lois paramétrée qui dépend de deux paramètres \mathbf{F}^2 et \mathbf{H}^2 (les indices sont omis pour simplicité de notation). Puis dans cette classe nous identifions les lois telles que le calcul exact rapide soit possible (c'est-à-dire satisfaisant i) et ii)), ce qui fixe \mathbf{H}^2 , et enfin dans cette sous-classe la loi (c'est-à-dire implicitement le \mathbf{F}^2) la plus proche du modèle (1).

Dans cet article nous proposons une démarche qui s'affranchit des contraintes physiques proposées (cf. [6, §III-A, Proposition 2]) pour le cas où dans le problème considéré celles-ci ne s'avèreraient pas pertinentes; par ailleurs la méthode proposée dans cet article s'affranchit de la résolution de l'équation matricielle que doit satisfaire \mathbf{H}^2 (cf. [6, eq. (54) p. 3647]), équation qui n'admet pas toujours de solution. Notre démarche est la suivante. Nous considérons d'emblée une famille de modèles de Markov triplet $p^2(\cdot)$ pour lesquels le filtrage exact est faisable rapidement :

$$p^2(\mathbf{x}_k, \mathbf{y}_k, r_k | \mathbf{x}_{k-1}, \mathbf{y}_{k-1}, r_{k-1}) = p^2(\mathbf{y}_k, r_k | \mathbf{y}_{k-1}, r_{k-1}) \times p^2(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_{k-1:k}, \mathbf{r}_{k-1:k}) \quad (8)$$

avec

$$p^2(\mathbf{y}_k | \mathbf{y}_{k-1}, \mathbf{r}_{k-1:k}) = \mathcal{N}(\mathbf{y}_k; \mathbf{H}^2(\cdot) \mathbf{y}_{k-1}; \Sigma^y(\cdot)), \quad (9)$$

$$p^2(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_{k-1:k}, \mathbf{r}_{k-1:k}) = \mathcal{N}(\mathbf{x}_k; \mathbf{B}(\cdot) \begin{pmatrix} \mathbf{x}_{k-1} \\ \mathbf{y}_{k-1} \\ \mathbf{y}_k \end{pmatrix}; \Sigma^x(\cdot)) \quad (10)$$

(la notation abrégée $\mathbf{B}(\cdot)$ désigne une matrice $\mathbf{B}(\mathbf{r}_{k-1:k})$ dépendant de r_{k-1} et r_k). Cette famille de lois dépend alors des paramètres $p^2(r_k | \mathbf{y}_{k-1}, r_{k-1})$ (nous fixons dès à présent $p^2(r_k | \mathbf{y}_{k-1}, r_{k-1}) = p^2(r_k | r_{k-1}) = p^1(r_k | r_{k-1})$), $\mathbf{B}(\mathbf{r}_{k-1:k})$, $\Sigma^x(\mathbf{r}_{k-1:k})$, $\mathbf{H}^2(\mathbf{r}_{k-1:k})$ et $\Sigma^y(\mathbf{r}_{k-1:k})$. Le problème devient alors un problème d'estimation de paramètres qui, à notre connaissance, n'a pas été abordé dans ce cadre (les paramètres de la loi initiale ne sont pas estimés et nous prenons $p^2(\mathbf{x}_0, \mathbf{y}_0, r_0) = p^1(\mathbf{x}_0, \mathbf{y}_0, r_0)$). Nous abordons ce problème en deux temps. Dans un premier temps, $\mathbf{H}^2(\mathbf{r}_{k-1:k})$, $\Sigma^y(\mathbf{r}_{k-1:k})$ et $p^2(r_k | r_{k-1})$ sont fixés tandis que $\mathbf{B}(\mathbf{r}_{k-1:k})$ et $\Sigma^x(\mathbf{r}_{k-1:k})$ sont obtenus par minimisation de la divergence de Kullback-Leibler (DKL) entre $p^1(\mathbf{x}_{0:k}, \mathbf{y}_{0:k}, \mathbf{r}_{0:k})$ et $p^2(\mathbf{x}_{0:k}, \mathbf{y}_{0:k}, \mathbf{r}_{0:k})$. Dans un second temps, $\mathbf{H}^2(\mathbf{r}_{k-1:k})$, $\Sigma^y(\mathbf{r}_{k-1:k})$ et $p^2(r_k | r_{k-1})$ sont estimés de manière non supervisée via l'algorithme EM [3].

2 Estimation paramétrique

2.1 Estimation par minimisation de DKL

Posons $\mathbf{t}_k = (\mathbf{x}_k, \mathbf{y}_k, r_k)$. Nous cherchons à minimiser la DKL entre $p^1(\cdot)$ et $p^2(\cdot)$,

$$\text{DKL}(p^1(\mathbf{t}_{0:k}), p^2(\mathbf{t}_{0:k})) = \sum_{\mathbf{r}_{0:k}} p^1(\mathbf{r}_{0:k}) \times \text{DKL}(p^1(\mathbf{x}_{0:k}, \mathbf{y}_{0:k} | \mathbf{r}_{0:k}), p^2(\mathbf{x}_{0:k}, \mathbf{y}_{0:k} | \mathbf{r}_{0:k})), \quad (11)$$

en fonction de $\mathbf{B}(\mathbf{r}_{k-1:k})$ et de $\Sigma^x(\mathbf{r}_{k-1:k})$. Puisque $p^1(\mathbf{r}_{0:k})$ ne dépend pas de ces paramètres et que

$$p^1(\mathbf{x}_{0:k}, \mathbf{y}_{0:k} | \mathbf{r}_{0:k}) = p^1(\mathbf{x}_0, \mathbf{y}_0 | r_0) \prod_{i=1}^k p^1(\mathbf{y}_i | \mathbf{x}_{i-1}, r_i) \times p^1(\mathbf{x}_i | \mathbf{x}_{i-1}, \mathbf{y}_i, r_i), \quad (12)$$

$$p^2(\mathbf{x}_{0:k}, \mathbf{y}_{0:k} | \mathbf{r}_{0:k}) = p^1(\mathbf{x}_0, \mathbf{y}_0 | r_0) \prod_{i=1}^k p^2(\mathbf{y}_i | \mathbf{y}_{i-1}, \mathbf{r}_{i-1:i}) \times p^2(\mathbf{x}_i | \mathbf{x}_{i-1}, \mathbf{y}_i, \mathbf{r}_{i-1:i}), \quad (13)$$

la minimisation de (11) par rapport à $\mathbf{B}(\mathbf{r}_{k-1:k})$ et à $\Sigma^x(\mathbf{r}_{k-1:k})$ se ramène au problème de la minimisation de

$$\text{DKL}(p^1(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_k, r_k), p^2(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_{k-1:k}, r_{k-1:k})).$$

Cette DKL est minimum lorsque

$$p^1(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_k, r_k) = p^2(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_{k-1:k}, \mathbf{r}_{k-1:k}),$$

et il reste à choisir $\mathbf{B}(\mathbf{r}_{k-1:k})$ et $\Sigma^x(\mathbf{r}_{k-1:k})$ tels que cette condition soit vérifiée. Finalement les paramètres $\mathbf{B}(\mathbf{r}_{k-1:k})$ et

$\Sigma^x(\mathbf{r}_{k-1:k})$ issus de la minimisation de la DKL entre le modèle $p^1(\cdot)$ et $p^2(\cdot)$ s'écrivent

$$\mathbf{B}(\mathbf{r}_{k-1:k}) = \begin{bmatrix} \mathbf{F}(\mathbf{r}_{k-1:k}) - \tilde{\mathbf{K}}(\mathbf{r}_{k-1:k})\mathbf{H}(\mathbf{r}_{k-1:k})\mathbf{F}(\mathbf{r}_{k-1:k}) \\ \mathbf{0} \\ \tilde{\mathbf{K}}(\mathbf{r}_{k-1:k}) \end{bmatrix} \quad (14)$$

$$\Sigma^x(\mathbf{r}_{k-1:k}) = (\mathbf{I} - \tilde{\mathbf{K}}(\cdot)\mathbf{H}(\cdot))\mathbf{Q}(\cdot), \quad (15)$$

$$\text{où } \tilde{\mathbf{K}}(\mathbf{r}_{k-1:k}) = (\mathbf{H}(\cdot)\mathbf{Q}(\cdot))^T [\mathbf{R}(\cdot) + \mathbf{H}(\cdot)\mathbf{Q}(\cdot)\mathbf{H}(\cdot)^T]^{-1}.$$

2.2 Estimation non supervisée

Il reste maintenant à estimer $p^2(r_k = l | r_{k-1} = m)$, $\mathbf{H}^2(r_{k-1} = l, r_k = m)$ et $\Sigma^x(r_{k-1} = l, r_k = m)$ avec $(l, m) \in \{1, \dots, K\}^2$. Ces paramètres sont estimés de manière supervisée, à partir d'un jeu d'observations $\mathbf{y}_{0:n}$ et de l'algorithme EM. Cet algorithme requiert d'abord le calcul de l'espérance de la log-vraisemblance des données complètes

$$\begin{aligned} \mathbb{E}(\log(p^2(\mathbf{r}_{0:n}, \mathbf{y}_{0:n}) | \mathbf{y}_{0:n})) &= \mathbb{E}(\log(p^2(\mathbf{r}_{0:n})) | \mathbf{y}_{0:n}) + \\ \mathbb{E}(\log(p^2(\mathbf{y}_{0:n} | \mathbf{r}_{0:n})) | \mathbf{y}_{0:n}). \end{aligned} \quad (16)$$

D'après (8),

$$\begin{aligned} \mathbb{E}(\log(p^2(\mathbf{r}_{0:n})) | \mathbf{y}_{0:n}) &= \mathbb{E}(\log(p^2(r_0)) | \mathbf{y}_{0:n}) + \\ \sum_{k=1}^n \mathbb{E}(\log(p^2(r_k | r_{k-1})) | \mathbf{y}_{0:n}), \end{aligned} \quad (17)$$

$$\begin{aligned} \mathbb{E}(\log(p^2(\mathbf{y}_{0:n} | \mathbf{r}_{0:n})) | \mathbf{y}_{0:n}) &= \mathbb{E}(\log(p^2(\mathbf{y}_0 | r_0)) | \mathbf{y}_{0:n}) + \\ \sum_{k=1}^n \mathbb{E}(\log(p^2(\mathbf{y}_k | \mathbf{y}_{k-1}, \mathbf{r}_{k-1:k})) | \mathbf{y}_{0:n}). \end{aligned} \quad (18)$$

Dans le modèle de Markov triplet vérifiant (8), le couple (\mathbf{y}_k, r_k) est markovien. Par conséquent, et contrairement au modèle $p^1(\cdot)$, $p^2(r_{k-1}, r_k | \mathbf{y}_{0:n})$ est calculable rapidement (voir par exemple [7]). Finalement, les seconds termes du membre de droite de (17) et (18) sont calculables exactement et s'expriment respectivement en fonction de $p^2(r_k = l | r_{k-1} = m)$, $\mathbf{H}^2(l, m)$ et de $\Sigma^x(l, m)$ pour $(l, m) \in \{1, \dots, K\}^2$.

Il reste à maximiser (17) et (18) en fonction de ces paramètres d'intérêt. La réactualisation de $p^2(r_k | r_{k-1})$ est donnée par les formules de Baum-Welch adaptée aux modèles de Markov couple [7]. Pour la réactualisation de $\mathbf{H}^2(l, m)$, $\Sigma^x(l, m)$, nous notons

$$A^1(l, m) = \sum_{k=1}^n p^2(r_{k-1} = l, r_k = m | \mathbf{y}_{0:n}) \mathbf{y}_k \mathbf{y}_k^T, \quad (19)$$

$$A^2(l, m) = \sum_{k=1}^n p^2(r_{k-1} = l, r_k = m | \mathbf{y}_{0:n}) \mathbf{y}_k \mathbf{y}_{k-1}^T, \quad (20)$$

$$A^3(l, m) = \sum_{k=1}^n p^2(r_{k-1} = l, r_k = m | \mathbf{y}_{0:n}) \mathbf{y}_{k-1} \mathbf{y}_{k-1}^T. \quad (21)$$

On montre alors que les paramètres $\mathbf{H}^2(l, m)$ et $\Sigma^x(l, m)$ maximisant (18) sont données par

$$\begin{aligned} \mathbf{H}^2(l, m) &= A^2(l, m) \times A^3(l, m)^{-1}, \quad (22) \\ \Sigma^x(l, m) &= \frac{A^1(l, m) - A^2(l, m) \times A^3(l, m)^{-1} A^2(l, m)^T}{\sum_{k=1}^n p^2(r_{k-1} = l, r_k = m | \mathbf{y}_{0:n})}. \end{aligned} \quad (23)$$

3 Simulations

3.1 Scénario

Nous considérons un modèle à sauts $p^1(\cdot)$ avec 3 régimes possibles : $r_k \in \{1, 2, 3\}$, $p^1(r_k | r_{k-1}) = 0.8$ si $r_{k-1} = r_k$ et 0.1 sinon. Nous prenons $\mathbf{x}_k = [p_x, \dot{p}_x, p_y, \dot{p}_y]^T$,

$$\begin{aligned} \mathbf{F}(r) &= \begin{bmatrix} 1 & \frac{\sin(\omega_r T_e)}{\omega_r} & 0 & -\frac{1 - \cos(\omega_r T_e)}{\omega_r} \\ 0 & \cos(\omega_r T_e) & 0 & -\sin(\omega_r T_e) \\ 0 & \frac{1 - \cos(\omega_r T_e)}{\omega_r} & 1 & \frac{\sin(\omega_r T_e)}{\omega_r} \\ 0 & \sin(\omega_r T_e) & 0 & \cos(\omega_r T_e) \end{bmatrix}, \\ \mathbf{Q}(r) &= \sigma_v^2(r) \begin{bmatrix} \frac{T_e^3}{3} & \frac{T_e^2}{2} & 0 & 0 \\ \frac{T_e^2}{2} & T_e & 0 & 0 \\ 0 & 0 & \frac{T_e^3}{3} & \frac{T_e^2}{2} \\ 0 & 0 & \frac{T_e^2}{2} & T_e \end{bmatrix}, \end{aligned}$$

avec $\omega_r = [0, 6\pi/180, -6\pi/180]$, $T_e = 2$, $\sigma_v^2(r) = [7, 10, 10]$, $\mathbf{H}(r_k) = \mathbf{I}_4$ et $\mathbf{R}(r_k) = \mathbf{I}_4$ (\mathbf{H}_k and \mathbf{R}_k ne dépendent pas de r_k). Une première trajectoire de longueur $T = 500$ est générée suivant le modèle (1) afin d'estimer les paramètres des lois $p^2(r_k | r_{k-1})$ et $p^2(\mathbf{y}_k | \mathbf{y}_{k-1}, \mathbf{r}_{k-1:k})$. Nous utilisons 25 itérations de l'algorithme EM.

Ensuite, nous générons $P = 200$ trajectoires de longueur $T = 200$ et nous restaurons les données avec 3 estimateurs : un algorithme de FP basé sur la loi d'importance $p^1(r_k | r_{k-1})$ et qui utilise $N = 100$ particules ; l'estimateur basé sur l'IMM ; et notre estimateur basé sur l'algorithme de filtrage exact [8] qui utilise les paramètres estimés à l'étape précédente.

Le critère de performance utilisé est l'EQM par rapport à l'estimateur issu du FK $\hat{\mathbf{x}}_{k,p}^{\text{KF}}$ utilisant les vrais sauts. Plus précisément, pour une trajectoire p donnée et un estimateur $\hat{\mathbf{x}}_{k,p}$ à l'instant k , nous calculons

$$\text{EQM}(k) = \frac{1}{P} \sum_{p=1}^P (\hat{\mathbf{x}}_{k,p} - \hat{\mathbf{x}}_{k,p}^{\text{KF}})^2. \quad (24)$$

3.2 Résultats

Pour l'estimation de $p^2(r_k = l | r_{k-1} = m)$, l'algorithme EM a été initialisé avec $p^{2,\text{init}}(r_k = l | r_{k-1} = m) = 1/3$, quels que soient (l, m) . Par ailleurs, nous initialisons $\mathbf{H}^{2,\text{init}}(\mathbf{r}_{k-1:k}) = 0.3\mathbf{F}(r_k)$ et $\Sigma^{y,\text{init}}(\mathbf{r}_{k-1:k}) = 20 \times 10^3(\mathbf{R} + \mathbf{H}\mathbf{Q}(r_k)\mathbf{H}^T)$. Après application de l'algorithme EM, la matrice associée à la

loi de transition $p^2(r_k = l | r_{k-1} = m)$ estimée est donnée par

$$\pi(r_{k-1}, r_k) = \begin{bmatrix} 0.7926 & 0.0971 & 0.1104 \\ 0.0710 & 0.8351 & 0.0938 \\ 0.0983 & 0.1023 & 0.7994 \end{bmatrix}. \quad (25)$$

Sur la figure 1, nous avons tracé l'EQM des trois estimateurs utilisés pour l'estimation des données cachées. Les estimateurs basés sur le FP dans le modèle $p^1(\cdot)$ et le filtrage exact dans le modèle $p^2(\cdot)$ se comportent de la même manière et améliorent l'estimateur basé sur l'IMM dans le modèle $p^1(\cdot)$.

Les avantages de l'algorithme proposé sont les suivants : la loi des sauts n'a pas besoin d'être connue et l'estimation des paramètres à partir de l'algorithme EM ne nécessite pas d'approximations (contrairement au modèle $p^1(\cdot)$, la quantité (17) est calculable rapidement); d'autre part, l'estimation des données cachées se fait en coût de calcul linéaire en nombre d'observations et ne requiert pas l'utilisation de tirages Monte-Carlo. Cet algorithme se révèle, en moyenne, 18 fois plus rapide que l'algorithme de FP lorsque le rééchantillonnage est exécuté à chaque instant, et 10 fois plus rapide si les instants de rééchantillonnage sont espacés selon un critère de nombre de particules efficaces.

En revanche, lorsque les données suivent effectivement le modèle $p^1(\cdot)$ les performances de l'algorithme exact dans le modèle $p^2(\cdot)$ dépendent des paramètres du modèle de référence $p^1(\cdot)$. En effet, une partie des paramètres du modèle $p^2(\cdot)$ est obtenue par contrainte de minimisation de la DKL (11), qui elle-même dépend de $\mathbf{F}(r_k)$, $\mathbf{H}(r_k)$, $\mathbf{Q}(r_k)$ et $\mathbf{R}(r_k)$. L'étude des performances de cet algorithme sur des données ne suivant pas nécessairement le modèle $p^1(\cdot)$ fera l'objet d'un travail ultérieur.

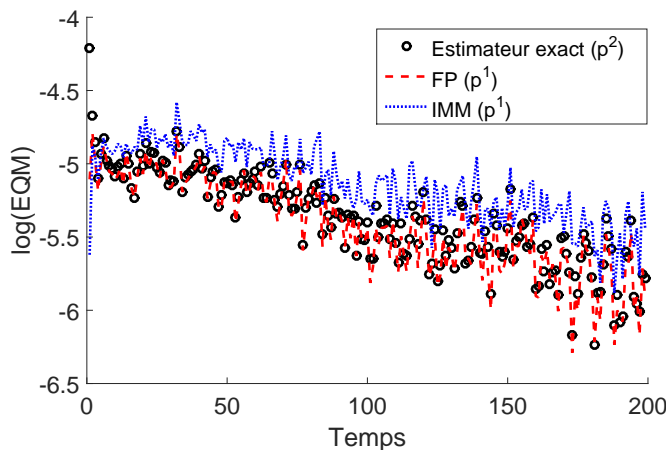


FIGURE 1 – Log EQM pour l'estimateur exact dans le modèle $p^2(\cdot)$ estimé, l'estimateur basé sur le filtrage particulaire dans $p^1(\cdot)$ et sur l'IMM dans $p^1(\cdot)$.

4 Conclusion

Dans cette communication, nous avons abordé le problème d'estimation de paramètres de modèles Markov triplet dans lesquels l'estimateur bayésien optimal (au sens de la minimisation de l'EQM) est calculable exactement, en un coût linéaire en nombre d'observations, et qui ne repose pas sur l'utilisation de tirages Monte Carlo. Cette procédure d'estimation se fait en deux étapes : la première étape consiste à estimer une partie des paramètres en prenant le modèle à sauts classique comme modèle de référence ; la seconde étape estime les autres paramètres, en particulier la loi des sauts, de manière non supervisée. La restauration des données, dans le modèle obtenu, donne des résultats similaires au filtrage particulaire mais avec un coût de calcul moindre.

Références

- [1] N. Abbassi, D. Benboudjema, and W. Pieczynski. Kalman filtering approximations in triplet Markov Gaussian switching models. In *IEEE Workshop on Statistical Signal Processing*, Nice, France, June 28-30 2011.
- [2] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*. John Wiley and sons, New-York, 2001.
- [3] A. P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society (B)*, 39(1) :1–38, 1977.
- [4] S. Derrode and W. Pieczynski. Exact fast computation of optimal filter in gaussian switching linear systems. *IEEE Signal Processing Letters*, 20(7) :701–04, 2013.
- [5] A. Doucet, N. J. Gordon, and V. Krishnamurthy. Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, 49(3) :613–24, March 2001.
- [6] Y. Petetin and F. Desbouvries. A class of fast exact bayesian filters in dynamical models with jumps. *IEEE Transactions on Signal Processing*, 62(4) :3643–53, 2014.
- [7] W. Pieczynski. Pairwise Markov chains. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5) :634–39, May 2003.
- [8] W. Pieczynski. Exact filtering in conditionally Markov switching hidden linear models. *Comptes Rendus Mathématiques*, 349(9-10) :587–590, May 2011.
- [9] W. Pieczynski, S. Derrode, N. Abbassi, Y. Petetin, and F. Desbouvries. Exact optimal filtering in an approximating switching system. In *Actes Ateliers Traitement et Analyse de l'Information - Méthodes et Applications (TAIMA)*, Hammamet, Tunisie, 2013.