C. R. Acad. Sci. Paris, Ser. I 349 (2011) 587–590

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Statistics/Theory of Signals

Exact filtering in conditionally Markov switching hidden linear models

Filtrage exact dans les modèles conditionnellement linéaires à sauts markoviens

Wojciech Pieczynski

Institut Telecom, Telecom SudParis, Dept. CITI, CNRS UMR 5157, 91000 Evry, France

ARTICLE INFO

Article history:
Received 3 September 2010
Accepted after revision 7 February 2011
Available online 12 April 2011

Presented by Paul Deheuvels

ABSTRACT

In the classical setup of Markov switching hidden linear systems, the exact evaluation of optimal filters requires computations with complexity increasing exponentially with respect to the number of observations. In the present article, we propose a new family of models which overcome this difficulty, and render practically feasible these calculations.

RÉSUMÉ

Dans les modèles classiques linéaires cachés à sauts markoviens, le calcul exact des filtres optimaux est d'une complexité exponentielle par rapport au nombre d'observations, ce qui nécessite l'utilisation de techniques d'approximation. Dans cet article, nous introduisons une nouvelle famille de modèles qui rendent ces opérations possibles en pratique.

Version française abrégée

On considère trois processus stochastiques à temps discret $X = \{X_i\colon\ i \geqslant 1\}$, $R = \{R_i\colon\ i \geqslant 1\}$, et $Y = \{Y_i\colon\ i \geqslant 1\}$. On suppose que, pour $i \geqslant 1$, $X_i \in \mathbb{R}^m$, $Y_i \in \mathbb{R}^q$, alors que $R_i \in S = \{1, \dots, s\}$ prend ses valeurs dans un ensemble S fini. Dans l'intervalle de temps $\{1, \dots, n\}$, $X_1^n = \{X_1, \dots, X_n\}$ et $R_1^n = \{R_1, \dots, R_n\}$ sont cachés, et $Y_n^1 = \{Y_1, \dots, Y_n\}$ est observé. On désigne par $p(x_1^n, r_1^n, y_1^n)$ la densité de probabilité jointe de (X_1^n, R_1^n, Y_1^n), supposée exister et définie sur $\mathbb{R}^{nm} \times S^n \times \mathbb{R}^{nq}$. Dans ce contexte, nous cherchons à évaluer les quantités $p(r_n|y_1^n) = p(r_n|Y_1^n = y_1^n)$, $\mathbb{E}(X_n|y_1^n) = \mathbb{E}(X_n|Y_1^n = y_1^n)$ et $\mathrm{Var}(X_n|y_1^n) = \mathrm{Var}(X_n|Y_1^n = y_1^n)$, avec des notations évidentes. Considérons le modèle classique suivant, dit *modèle conditionnellement gaussien linéaire* : (i) R est markovien ; (ii) $X_{n+1} = F_{n+1}(R_{n+1})X_n + G_{n+1}(R_{n+1})W_{n+1}$; et $Y_n = H_n(R_n)X_n + J_n(R_n)Z_n$, où X_0 est un vecteur gaussien, $W_1, \dots, W_n, \dots, Z_1, \dots, Z_n, \dots$ sont gaussiens indépendants et indépendents de X_0, et $F_1(R_1), G_1(R_1), H_1(R_1), J_1(R_1), \dots, F_n(R_n), G_n(R_n), H_n(R_n), J_n(R_n), \dots$, sont des matrices de dimensions correspondantes. Les modèles de ce type ont été appliqués à un nombre important de problèmes. Lorsque l'on utilise ce modèle dans le problème de filtrage que nous considérons dans cet article, le calcul des expressions d'intérêt requiert un nombre d'opérations croissant de manière exponentielle en fonction de la durée n d'observation. Ceci est en particulier dû au calcul de $p(y_{n+1}|y_1^n)$, expression que l'on se sait pas évaluer par un nombre d'opérations proportionnel à n. Pour résoudre le problème, on est alors amené à faire appel à des approximations, soit déterministes [4,10], soit fondées sur des méthodes markoviennes de Monte Carlo (MCMC [2,3,6]).

L'objet de cette Note est d'introduire une nouvelle famille de modèles, dits *modèles cachés conditionnellement linéaires à sauts markoviens* [MCCLSM], pour lesquels $p(r_n|y_1^n)$, $\mathbb{E}(X_n|y_1^n)$ et $\mathrm{Var}(X_n|y_1^n)$, sont calculables avec une complexité propor-

E-mail address: Wojciech.Pieczynski@it-sudparis.eu.

1631-073X/$ – see front matter
doi:10.1016/j.crma.2011.02.007

tionnelle à n. Ceci est rendu possible notamment par le fait que, dans un modèle MCCLSM, le couple (R, Y) est markovien, ce qui rend facilement calculable $p(y_{n+1}|y_1^n)$, et pallie la principale difficulté inhérente aux modèles classiques. Par contre, dans ce même modèle, le couple (X, R) n'est, en général, pas markovien, ce qui le distingue du cas des modèles classiques où cette propriété est requise. Plus précisémment, un MCCLSM est un modèle dans lequel : (i) $T = (X, R, Y)$ est markovien avec les transitions vérifiant $p(x_{n+1}, r_{n+1}, y_{n+1} \mid x_n, r_n, y_n) = p(r_{n+1}, y_{n+1} \mid r_n, y_n)p(x_{n+1} \mid x_n, r_n, y_n, r_{n+1}, y_{n+1})$, $\forall n \geqslant 1$ (le couple (R, Y) est alors markovien); (ii) il existe une suite $W_2, \ldots, W_n, \ldots$, de vecteurs centrés indépendants, avec chaque W_{n+1} indépendant de (R_1^{n+1}, Y_1^{n+1}), telle que $X_{n+1} = F_{n+1}(R_{n+1}, Y_{n+1})X_n + G_{n+1}(R_{n+1}, Y_{n+1})W_{n+1} + H_{n+1}(R_{n+1}, Y_{n+1})$, où $F_2, G_2, \ldots, F_n, G_n, \ldots$, sont des fonctions de $S \times \mathbb{R}^q$ dans l'ensemble des matrices de taille $m \times m$, et $H_2, \ldots, H_n, \ldots$, sont des fonctions de $S \times \mathbb{R}^q$ dans $\mathbb{R}^m$.

1. Introduction

Let $X = \{X_i\colon i \geqslant 1\}$ be a *hidden* random sequence, $Y = \{Y_i\colon i \geqslant 1\}$ be an *observed* random sequence, and $R = \{R_i\colon i \geqslant 1\}$ be a *hidden* discrete random sequence, where for $i \geqslant 1$, we let $X_i \in \mathbb{R}^m$, $Y_i \in \mathbb{R}^q$ and $R_i \in S = \{1, \ldots, s\}$. For each choice of the observation time $n \geqslant 1$, we assume that $X_1^n := \{X_1, \ldots, X_n\}$, $R_1^n := \{R_1, \ldots, R_n\}$, and $Y_1^n := \{Y_1, \ldots, Y_n\}$, have a joint probability density $p(x_1^n, r_1^n, y_1^n)$ on $\mathbb{R}^{mn} \times S^n \times \mathbb{R}^{qn}$. In this paper we are concerned with the computation of $p(r_n|y_1^n) = p(r_n|Y_1^n = y_1^n)$, $\mathbb{E}[X_n|y_1^n] = \mathbb{E}[X_n|Y_1^n = y_1^n]$, and $\mathrm{Var}[X_n|y_1^n] = \mathrm{Var}[X_n|Y_1^n = y_1^n]$, with self-explanatory notation. This problem is often referred to as that of *switching filters*, because of the fact that the indexes of the discrete process R govern the *changes of regime* (or *switches*, or *jumps*) in the distribution of the random pair (X_n, Y_n). Models of this kind have been repeatedly used during the last decades, and give rise to a large number of contributions (refer to [6], and to the recently published volumes [3] and [4] which cite an extended bibliography on the subject).

In the present article, we introduce a family of distributions $p(x_1^n, r_1^n, y_1^n)$, $n = 1, 2, \ldots$, which allow the quantities of interest at time $n + 1$, namely $p(r_{n+1}|y_1^{n+1})$, $\mathbb{E}[X_{n+1}|y_1^{n+1}]$, and $\mathrm{Var}[X_{n+1}|y_1^{n+1}]$, to be easily computed from $p(r_n|y_1^n)$, $\mathbb{E}[X_n|y_1^n]$, $\mathrm{Var}[X_n|y_1^n]$, and y_{n+1}. Because of this, the *switching filter* can be computed with complexity linear in the time observation n, which is of obvious importance for practical use. By contrast, the models currently used in the literature do not afford this computational efficiency, and thus require approximations to keep the computational complexity in check.

The most commonly used model in the literature is the *Conditionally Gaussian Linear State-Space Model* [CGLSSM], where one assumes that R is a Markov chain, with (X, Y) being some version of a linear system, conditionally on R. We have, namely,

$$R \text{ is a Markov chain;} \tag{1}$$

$$X_{n+1} = F_{n+1}(R_{n+1})X_n + G_{n+1}(R_{n+1})W_{n+1}; \tag{2}$$

$$Y_n = H_n(R_n)X_n + J_n(R_n)Z_n; \tag{3}$$

where X_0 is Gaussian, $W_1, \ldots, W_n, \ldots$, are independent (and independent from X_0) Gaussian vectors in $\mathbb{R}^m$, $Z_1, \ldots, Z_n, \ldots$, are independent (and independent from X_0) Gaussian vectors in $\mathbb{R}^q$, and $F_1(R_1), G_1(R_1), H_1(R_1), J_1(R_1), \ldots, F_n(R_n)$, $G_n(R_n), H_n(R_n), J_n(R_n), \ldots$, are matrices of appropriate dimensions which depend on the switches (see the dependence graph in Fig. 1(a)). For fixed $R_1 = r_1, \ldots, R_n = r_n, \ldots$, the computation of $\mathbb{E}[X_{n+1}|y_1^{n+1}]$ from $\mathbb{E}[X_n|y_1^n]$ and y_{n+1} is obtained by the well-known Kalman filter, with complexity linear with respect to the number n of observations. On the other hand (see, e.g., [9]) such calculations cannot be achieved when R is Markov, so that different approximation-based methods have to be used. Among the latter, we should mention stochastic methods (see, e.g., [2] and [3]), and deterministic-type techniques (refer to [4] and [10]). In fact, setting $V = (X, R)$, we have classically

$$p\big(v_{n+1}\big|y_1^{n+1}\big) = \frac{p(y_{n+1}|v_{n+1}) \int_{\mathbb{R}^m \times S} p(v_n|y_1^n)p(v_{n+1}|v_n)\,\mathrm{d}v_n}{p(y_{n+1}|y_1^n)}, \tag{4}$$

but there is no known closed-form solution of this recursion.

As previously mentioned, the aim of the present Note is to introduce a family of joint probability densities $p(x_1^n, r_1^n, y_1^n)$ for which the exact computation of the filter is achievable with a complexity proportional to n. These distributions differ from that of the classical models in the literature (see, e.g., [2–4,6,9,10]). Our models are inspired, in part, from that proposed recently in [1] and [8], which assume independence of X_1^n and Y_1^n conditionally on R_1^n. We will extend largely these assumptions, making use of the dependence graphs displayed in Fig. 1, (b) and (c). We see that model (b) seems to be quite close to the classical model (a); however, in contrast to the latter, in the former the filter can be computed exactly.

Definition 1.1. Let $X = \{X_i\colon i \geqslant 1\}$, $Y = \{Y_i\colon i \geqslant 1\}$, and $R = \{R_i\colon i \geqslant 1\}$ be three random sequences, where for $i \geqslant 1$, we let $X_i \in \mathbb{R}^m$, $Y_i \in \mathbb{R}^q$ and $R_i \in S = \{1, \ldots, s\}$. The triplet $T = (X, R, Y)$ will be called "conditionally Markov switching hidden linear model" (CMSHLM) if:

(i) $T = (X, R, Y)$ is a Markov chain with transitions verifying

$$p(x_{n+1}, r_{n+1}, y_{n+1} \mid x_n, r_n, y_n) = p(r_{n+1}, y_{n+1} \mid r_n, y_n)p(x_{n+1} \mid x_n, r_n, y_n, r_{n+1}, y_{n+1}), \quad \forall n \geqslant 1 \tag{5}$$

(the chain (R, Y) is then Markov);

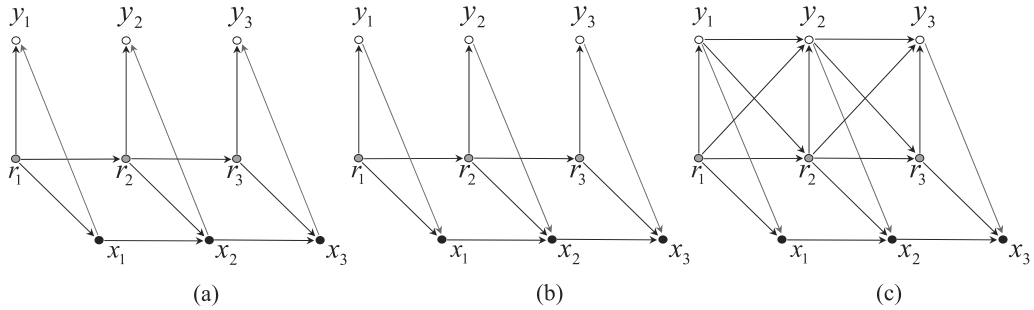

Fig. 1. Dependence oriented graphs of: (a) classical model; (b) new simplified model; (c) general new model.

(ii) There exist independent zero-mean random vectors $W_2, \ldots, W_n$ in R^m, where W_{n+1} is independent of (R_1^{n+1}, Y_1^{n+1}) for each $n \geqslant 1$, such that

$$X_{n+1} = F_{n+1}(R_{n+1}, Y_{n+1})X_n + G_{n+1}(R_{n+1}, Y_{n+1})W_{n+1} + H_{n+1}(R_{n+1}, Y_{n+1}), \quad \forall n \geqslant 1, \tag{6}$$

where $F_2, G_2, \ldots, F_n, G_n, \ldots,$ are functions from $S \times \mathbb{R}^q$ to the set of $m \times m$ matrices, and $H_2, \ldots, H_n, \ldots,$ are functions from $S \times \mathbb{R}^q$ to $\mathbb{R}^m$.

Let us notice that according to (5), in CMSHLM $p(r_{n+1}, y_{n+1} \mid x_n, r_n, y_n) = p(r_{n+1}, y_{n+1} \mid r_n, y_n)$, which means that X_n and (R_{n+1}, Y_{n+1}) are independent given (R_n, Y_n). This fact will be of importance in the proof of Proposition 2.1 below.

Let us note the generality of the CMSHLM model: transitions $p(r_{n+1}, y_{n+1} \mid r_n, y_n)$ of the Markov chain (R, Y) can be of any form and it is the same for the functions $F_2, G_2, H_2, \ldots, F_n, G_n, H_n, \ldots$. In particular, the chain R need not be Markovian; this is why the model is called "conditionally Markov switching" and not "Markov switching". As studied in [5], such a Markov chain (R, Y) can be much more efficient in unsupervised segmentation than the classical hidden Markov chain (HMC), in which R is Markovian. Some conditions on a Markov chain (R, Y) under which R is Markovian are provided in [7]. However, (R, Y) can be the classical HMC, which gives the new "simplified" model, presented in Fig. 1(b).

Let us remark that in the CMSHLM the hidden chain (X_1^n, R_1^n) has a very complex structure and is, in general, neither linear nor Markovian. This is very different from the classical model, in which (X_1^n, R_1^n) is Markovian. However, an important common point is that (X_1^n, R_1^n) is Markovian conditionally on Y_1^n in both of them.

Finally, in the classical model (1)–(3), the chain (X, R) is Markovian and (R, Y) is not, while (X, R) is not Markovian in CMSHLM and (R, Y) is. It is not clear that either of these properties should be inherently better suited to real situations; however, the possibility of exact calculations with complexity linear with respect to the number n of observations is a clear advantage of CMSHLM over the classical CGLSSM (1)–(3).

Let us also remark that, upon removing the arrows "from y_n to x_n" in (b) and (c), we obtain the models proposed in [8].

2. Exact filtering with complexity linear in the number of observations

Let $X = \{X_i\colon i \geqslant 1\}$, $Y = \{Y_i\colon i \geqslant 1\}$, and $R = \{R_i\colon i \geqslant 1\}$ be a CMSHLM as defined in Definition 1.1. The problem we address is to compute $p(r_{n+1}|y_1^{n+1})$, $\mathbb{E}[X_{n+1} \mid r_{n+1}, y_1^{n+1}]$, and $\mathbb{E}[(X_{n+1})(X_{n+1})^T \mid r_{n+1}, y_1^{n+1}]$ from $p(r_n|y_1^n)$, $\mathbb{E}[X_n \mid r_n, y_1^n]$, $\mathbb{E}[(X_n)(X_n)^T \mid r_n, y_1^n]$, y_{n+1}, and $p(r_{n+1}, y_{n+1} \mid r_n, y_n)$.

Let us observe that, according to the model, we have

$$p\big(r_n, r_{n+1}, y_{n+1} \bigm| y_1^n\big) = p(r_{n+1}, y_{n+1} \mid r_n, y_n)p\big(r_n \big| y_1^n\big). \tag{7}$$

We shall establish the following proposition:

Proposition 2.1. *Consider a "conditionally Markov switching hidden linear model" (X, R, Y) as defined in Definition* 1.1. *For fixed y_1^{n+1}, let $a(r_n, r_{n+1})$ be the distribution* (7)*.*

Then

$$p\big(r_{n+1} \big| y_1^{n+1}\big) = \frac{1}{p(y_{n+1} | y_1^n)} \sum_{r_n} a(r_n, r_{n+1}), \tag{8}$$

with

$$p\big(y_{n+1} \big| y_1^n\big) = \sum_{r_n, r_{n+1}} a(r_n, r_{n+1}), \tag{9}$$

and

$$\mathbb{E}\big[X_{n+1}\,\big|\,r_{n+1}, y_1^{n+1}\big] = F_{n+1}(r_{n+1}, y_{n+1})\left(\sum_{r_n} p\big(r_n\,\big|\,r_{n+1}, y_1^{n+1}\big)\mathbb{E}\big[X_n\,\big|\,r_n, y_1^n\big]\right) + H_{n+1}(r_{n+1}, y_{n+1}), \tag{10}$$

with

$$p\big(r_n\,\big|\,r_{n+1}, y_1^{n+1}\big) = \frac{p(r_{n+1}, y_{n+1}\mid r_n, y_n)p(r_n|y_1^n)}{\sum_{r_n} p(r_{n+1}, y_{n+1}\mid r_n, y_n)p(r_n|y_1^n)}. \tag{11}$$

In addition, if $\Gamma_n = \operatorname{Var}[W_n]$ exists for each $n = 2, 3, \ldots,$ then $\operatorname{Var}[X_n \mid r_n, y_1^n]$ also exists for each $n = 1, 2, \ldots,$ and we have

$$\begin{aligned}
&\mathbb{E}\big[(X_{n+1})(X_{n+1})^T\,\big|\,r_{n+1}, y_1^{n+1}\big]\\
&\quad= F_{n+1}(r_{n+1}, y_{n+1})\left(\sum_{r_n} p\big(r_n\,\big|\,r_{n+1}, y_1^{n+1}\big)\mathbb{E}\big[(X_n)(X_n)^T\,\big|\,r_n, y_1^n\big]\right)\big(F_{n+1}(r_{n+1}, y_{n+1})\big)^T\\
&\qquad+ G_{n+1}(r_{n+1}, y_{n+1})(\Gamma_{n+1})^T\big(G_{n+1}(r_{n+1}, y_{n+1})\big)^T\\
&\qquad+ F_{n+1}(r_{n+1}, y_{n+1})\left(\sum_{r_n} p\big(r_n\,\big|\,r_{n+1}, y_1^{n+1}\big)\mathbb{E}\big[X_n\,\big|\,r_n, y_1^n\big]\right)\big(H_{n+1}(r_{n+1}, y_{n+1})\big)^T\\
&\qquad+ H_{n+1}(r_{n+1}, y_{n+1})\left(\sum_{r_n} p\big(r_n\,\big|\,r_{n+1}, y_1^{n+1}\big)\mathbb{E}\big[X_n\,\big|\,r_n, y_1^n\big]\right)^T\big(F_{n+1}(r_{n+1}, y_{n+1})\big)^T\\
&\qquad+ H_{n+1}(r_{n+1}, y_{n+1})\big(H_{n+1}(r_{n+1}, y_{n+1})\big)^T,
\end{aligned} \tag{12}$$

which implies

$$\begin{aligned}
&\operatorname{Var}\big[X_{n+1}\,\big|\,r_{n+1}, y_1^{n+1}\big]\\
&\quad= \mathbb{E}\big[(X_{n+1})(X_{n+1})^T\,\big|\,r_{n+1}, y_1^{n+1}\big] - \mathbb{E}\big[X_{n+1}\,\big|\,r_{n+1}, y_1^{n+1}\big]\big(\mathbb{E}\big[X_{n+1}\,\big|\,r_{n+1}, y_1^{n+1}\big]\big)^T.
\end{aligned} \tag{13}$$

Proof. As $p(r_{n+1}|y_1^{n+1}) = \frac{1}{p(y_{n+1}|y_1^n)}\sum_{r_n} p(r_n, r_{n+1}, y_{n+1}\mid y_1^n) = \frac{1}{p(y_{n+1}|y_1^n)}\sum_{r_n} a(r_n, r_{n+1})$, we have (8). The relation (9) is immediate from $a(r_n, r_{n+1}) = p(r_n, r_{n+1}, y_{n+1}\mid y_1^n)$. To obtain (10), we use (6) and compute successively

$$\begin{aligned}
\mathbb{E}\big[X_{n+1}\,\big|\,r_{n+1}, y_1^{n+1}\big] &= F_{n+1}(r_{n+1}, y_{n+1})\mathbb{E}\big[X_n\,\big|\,r_{n+1}, y_1^{n+1}\big] + H_{n+1}(r_{n+1}, y_{n+1})\\
&= F_{n+1}(r_{n+1}, y_{n+1})\sum_{r_n}\mathbb{E}\big[r_n, X_n\,\big|\,r_{n+1}, y_1^{n+1}\big] + H_{n+1}(r_{n+1}, y_{n+1})\\
&= F_{n+1}(r_{n+1}, y_{n+1})\left(\sum_{r_n} p\big(r_n\,\big|\,r_{n+1}, y_1^{n+1}\big)\mathbb{E}\big[X_n\,\big|\,r_n, r_{n+1}, y_1^{n+1}\big]\right) + H_{n+1}(r_{n+1}, y_{n+1}).
\end{aligned}$$

Now, according to (5), $p(r_{n+1}, y_{n+1}\mid x_n, r_n, y_n) = p(r_{n+1}, y_{n+1}\mid r_n, y_n)$, which means that X_n and (R_{n+1}, Y_{n+1}) are independent given (R_n, Y_n). This implies $\mathbb{E}[X_n\mid r_n, r_{n+1}, y_1^{n+1}] = \mathbb{E}[X_n\mid r_n, y_1^n]$, which leads to (10). The relation (12) is obtained by classical calculus, applying $\mathbb{E}[(X_{n+1})(X_{n+1})^T\mid r_{n+1}, y_1^{n+1}]$ to X_{n+1} given with (6) and using the fact that

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

which follows from the equality $\mathbb{E}[(X_n)(X_n)^T\mid r_n, r_{n+1}, y_1^{n+1}] = \mathbb{E}[(X_n)(X_n)^T\mid r_n, y_1^n]$. Finally, (11) is immediate from (7). □

References

[1] N. Abbassi, W. Pieczynski, Exact filtering in semi-Markov jumping system, in: Sixth International Conference of Computational Methods in Sciences and Engineering, Hersonissos, Crete, Greece, September 25–30, 2008.
[2] C. Andrieu, C.M. Davy, A. Doucet, Efficient particle filtering for jump Markov systems. Application to time-varying autoregressions, IEEE Trans. Signal Process. 51 (7) (2003) 1762–1770.
[3] O. Cappé, E. Moulines, T. Ryden, Inference in Hidden Markov Models, Springer, 2005.
[4] O.L.V. Costa, M.D. Fragoso, R.P. Marques, Discrete Time Markov Jump Linear Systems, Springer-Verlag, New York, 2005.
[5] S. Derrode, W. Pieczynski, Signal and image segmentation using pairwise Markov chains, IEEE Trans. Signal Process. 52 (9) (2004) 2477–2489.
[6] P. Giordani, R. Kohn, D. van Dijk, A unified approach to nonlinearity, structural change, and outliers, J. Econometrics 137 (2007) 112–133.
[7] W. Pieczynski, Multisensor triplet Markov chains and theory of evidence, Internat. J. Approx. Reason. 45 (1) (2007) 1–16.
[8] W. Pieczynski, Exact calculation of optimal filter in semi-Markov switching model, in: Fourth World Conference of the International Association for Statistical Computing (IASC 2008), Yokohama, Japan, December 5–8, 2008.
[9] J.K. Tugnait, Adaptive estimation and identification for discrete systems with Markov jump parameters, IEEE Trans. Automat. Control AC-25 (1982) 1054–1065.
[10] O. Zoeter, T. Heskes, Deterministic approximate inference techniques for conditionally Gaussian state space models, Statist. Comput. 16 (2006) 279–292.