Exact Calculation of Optimal Filters in Hidden Markov Switching Long-Memory Chain

Draft version: June 2009

Wojciech Pieczynski
Telecom SudParis, Dept CITI,
8, rue Charles Fourier, 91 000 Evry, France
E-mail: Wojciech.Pieczynski@it-sudparis.eu
http://www-public.it-sudparis.eu/~pieczyn/

Abstract

This paper considers the problem of continuous state estimation in the presence of random switches. There are three random chains $X_1^N = (X_1, ..., X_N)$, $R_1^N = (R_1, ..., R_N)$, and $Y_1^N = (Y_1, ..., Y_N)$. The random variables X_i, Y_i, and R_i take their values from $\mathbb{R}^q$, $\mathbb{R}^m$, and $S = \{1, ..., s\}$, respectively. Y_1^N is observed, X_1^N and R_1^N are hidden, and the problem is to estimate (R_1^N, X_1^N) from Y_1^N. In the classical probabilistic models R_1^N is Markovian, X_1^N is linear Markovian conditionally on R_1^N, and $(Y_1, ..., Y_N)$ are independent conditionally on (R_1^N, X_1^N). Neither exact filtering nor smoothing is possible with polynomial complexity in time in such models and the different research works mainly concern different approximate algorithms. More recently, another class of models, in which exact filtering and smoothing with polynomial complexity in time are feasible, has been proposed. In the latter models the distribution of the triplet (X_1^N, R_1^N, Y_1^N) is defined by a Markov distribution of (R_1^N, X_1^N), and then X_1^N is assumed to be linear Markovian conditionally on (R_1^N, Y_1^N). Subsequently, two extensions of these models have been specified. In the first one, the Markovianity of (R_1^N, Y_1^N) is extended to the « partial » Markovianity, in which (R_1^N, Y_1^N) is Markovian with respect to R_1^N, but may not be with respect to Y_1^N. In the second one, (R_1^N, Y_1^N) remains Markovian and X_1^N is assumed to be linear conditionally on (R_1^N, Y_1^N), but is not assumed Markovian. The aim of this paper is to propose a family of models admitting both these extensions simultaneously. In the new models proposed the distribution of Y_1^N conditional on R_1^N can be of the « long dependence » kind, and it is the same for the distribution of X_1^N conditional on (R_1^N, Y_1^N). We show that the Kalman-like exact filtering remains feasible with polynomial complexity in time in the new models models.

1. Introduction

Let us consider $X_1^N = (X_1, ..., X_N)$ and $Y_1^N = (Y_1, ..., Y_N)$ two sequences of random vectors, and let $R_1^N = (R_1, ..., R_N)$ be a finite-values random chain. Each X_n takes its values from R^q, while Y_n takes its values from R^m. The sequences X_1^N and R_1^N are hidden and the sequence Y_1^N is observed. For each $n = 1, \ldots, N$, we will set $x_1^n = (x_1, ..., x_n)$, $y_1^n = (y_1, ..., y_n)$, and $r_1^n = (r_1, ..., r_n)$. We deal with the problem of filtering, which consists of the computation, for each $n = 1, \ldots, N$, of the conditional expectation $E[X_n | Y_1^n = y_1^n]$. To simplify, we will set $E[X_n | Y_1^n = y_1^n] = E[X_n | y_1^n]$. As is well known, this conditional expectation is the optimal estimation of X_n from Y_1^n, when the squared error is concerned. This expectation can be considered – which will be done in this paper - as given by the distribution $p(r_n | y_1^n)$, which is the distribution of R_n conditional on $Y_1^n = y_1^n$, and by the conditional expectation $E[X_n | R_n = r_n, Y_1^n = y_1^n]$, denoted by $E[X_n | r_n, y_1^n]$. We have

$$E[X_n | y_1^n] = \sum_{r_n} E[X_n | r_n, y_1^n] p(r_n | y_1^n) \tag{1.1}$$

Finally, the problem considered is to compute $p(r_{n+1} | y_1^{n+1})$ and $E[X_{n+1} | r_{n+1}, y_1^{n+1}]$ from $p(r_n | y_1^n)$ and $E[X_n | r_n, y_1^n]$. The most classical model to define the distribution of the triplet $T_1^N = (X_1^N, R_1^N, Y_1^N)$, in use for about thirty years, is the so-called "conditionally Gaussian state-space linear model" (CGSSLM), which consists of considering that R_1^N is a Markov chain and, roughly speaking, (X_1^N, Y_1^N) is the classical linear system conditionally on R_1^N. This is summarized in the following:

$$R_1^N \text{ is a Markov chain;} \tag{1.2}$$

$$X_{n+1} = F_n(R_n) X_n + G_n(R_n) W_n; \tag{1.3}$$

$$Y_n = H_n(R_n) X_n + J_n(R_n) Z_n, \tag{1.4}$$

where X_1, W_1, …, W_N are independent (conditionally on R_1^N) Gaussian vectors in R^q, Z_1, …, Z_N are independent (conditionally on R_1^N) Gaussian vectors in R^m, $F_1(R_1)$, …, $F_N(R_N)$, $G_1(R_1)$, …, $G_N(R_N)$ are matrices of size $q \times q$ depending on switches, and $H_1(R_1)$, …, $H_N(R_N)$, $J_1(R_1)$, …, $J_N(R_N)$ are matrices of size $q \times m$ also depending on switches. Therefore the classical

Kalman filter can be used when $R_1^N = r_1^N$ is known; however, it has been well known since the publication of (Tugnait, 1982) that the exact computation of neither $E[X_n | r_n, y_1^n]$ nor $E[X_n | r_n, y_1^N]$ is feasible with linear - or even polynomial - complexity in time in such models when R_1^N is not known. The difficulty comes from the fact that conditional probabilities $p(y_{n+1} | y_1^n)$ are not computable with a reasonable complexity. This is a constant problem in all the classical models and the very reason for this is the fact that in the classical model (1.2)-(1.4) the couple (R_1^N, Y_1^N) is not Markovian. Then different approximations have to be used and a rich bibliography on the classical methods concerning the subject can be seen in recent books (Costa *et al.* 2005, Ristic *et al.* 2004, Cappe *et al.* 2005,), among others. Roughly speaking, there are two families of approximating methods: the stochastic ones, based on the Monte Carlo Markov Chains (MCMC) principle (Doucet *et al.* 2001, Andrieu *et al.* 2003, Cappe *et al.* 2005, Giordani *et al.* 2007), among others, and deterministic ones (Costa *et al.* 2005, Zoeter *et al.* 2006), among others. Further recent results concerning different applications of these models and related approximation methods can be seen in recent works (Germani *et al.*, 2006; Ho & Chen, 2006; Kim *et al.*, 2007; Lee & Dullerud 2007; Zhou and Shumway 2008; Johnson and Sakoulis 2008; Orguner & Demirekler 2008), among others.

To remedy this impossibility of exact computation, different models have been proposed since 2008. Two of them, proposed in (Abbassi and Pieczynski 2008, Pieczynski 2008), are based on the following two general assumptions: (i) R_1^N is a Markov – or a semi-Markov chain, the difference being of little importance here ; (ii) X_1^N and Y_1^N are independent conditionally on R_1^N. As (R_1^N, Y_1^N) is Markovian in the proposed models, the conditional ditributions $p(y_{n+1} | y_1^n)$ are computable, which implies that the exact filtering and smoothing are also. Further, prediction is workable (Bardel and Desbouvries 2009). More sophisticated models, in which the hypothesis (ii) is relaxed but the possibility of exact filtering remains were proposed in (Pieczynski 2009a; Pieczynski and Desbouvries 2009). In the latter models, the Markovianity of (R_1^N, Y_1^N) is kept, which still allows exact filtering and exact smoothing with complexity linear in time to be performed. Subsequently, based on the recent model proposed in (Lanchantin *et al.* 2008), two extensions to “partially” Markov models, which can include the “long-memory” ones (Beran and Taqqu 1994; Doukhan *et al.* 2003), have been introduced. In the first one the Markovianity of (R_1^N, Y_1^N) has been relaxed and replaced by the “partial” Markovianity, in which (R_1^N, Y_1^N) is Markovian with respect to R_1^N but is not necessarily Markovian with respect to Y_1^N (Pieczynski *et al.* 2009). In the second one, the distribution of the state chain X_1^N conditional on (R_1^N, Y_1^N) remains linear but is no longer necessarily Markovian (Pieczynski 2009b).

The aim of the present paper is to consider both the latter extensions simultaneously. Roughly speaking, we propose a general model in which

although neither $p(y_1^N | r_1^N, x_1^N)$ nor $p(x_1^N | r_1^N, y_1^N)$ are Markovian, the filtering can be performed with complexity polynomial in time.
The new model is proposed and discussed in the next section, and the exact computation of smoothing is described in the third one. The fourth section contains some conclusions and perspectives.

2. Conditionally Markov switching linear chain (CMSLC)

Let (X_1^N, R_1^N, Y_1^N) be the triplet of random sequences as specified above. The distribution of the couple (R_1^N, Y_1^N) will be assumed to be a "pairwise partially Markov chain" (PPMC) distribution recently introduced in (Lanchantin *et al.* 2008). The distribution $p(r_1^N, y_1^N)$ of a PPMC (R_1^N, Y_1^N) can be defined by $p(r_1, y_1)$ and the transitions $p(r_{n+1}, y_{n+1} | r_1^n, y_1^n)$ verifying

$$p(r_{n+1}, y_{n+1} | r_1^n, y_1^n) = p(r_{n+1}, y_{n+1} | r_n, y_1^n) . \tag{2.1}$$

Such a law is called "partially" Markovian as it can be seen as being Markovian with respect to the variables R_1^N, but being not necessarily Markovian with respect to the variables Y_1^N.

Definition 1

A triplet (X_1^N, R_1^N, Y_1^N) will be said to be a "conditionally Markov switching linear chain" (CMSLC) if it verifies

(R_1^N, Y_1^N) is a PPMC ; (2.2)

for $n = 1, \ldots, N-1$, $X_{n+1} = F^{n+1}(R_{n+1}, Y_{n+1}) X_1^n + G_{n+1}(R_{n+1}, Y_{n+1}) W_{n+1}$, (2.3)

with $F^{n+1}(r_{n+1}, y_{n+1}) = [F_1^{n+1}(r_{n+1}, y_{n+1}), F_2^{n+1}(r_{n+1}, y_{n+1}), \ldots, F_n^{n+1}(r_{n+1}, y_{n+1})]$, where for each $i = 1, \ldots, n$, $F_i^{n+1}(r_{n+1}, y_{n+1})$ is a matrix of size $q \times q$ depending on (r_{n+1}, y_{n+1}), $G_{n+1}(r_{n+1}, y_{n+1})$ is a matrix of size $q \times q$ depending on (r_{n+1}, y_{n+1}), and X_1, W_1, …, W_N are independent centred vectors in R^q such that each W_n is independent of (R_1^N, Y_1^N).

Let us point out the following aspects of the model (2.2)-(2.3), underlying its differences with the classical ones:

(a) the model (2.2)-(2.3) is said to be "conditionally Markov switching" because the switching process R_1^N is Markovian conditionally on Y_1^N ; however, it does not need to be Markovian according to its own distribution (without conditioning upon Y_1^N). We may recall that such models are richer and more efficient that the classical hidden Markov chains in which R_1^N is Markovian (Derrode and Pieczynski 2004, Pieczynski 2007);

(b) similarly, the model is said to be "conditionally linear" because X_1^N is linear conditionally on (R_1^N, Y_1^N); however, contrary to the classical models, it is not necessarily linear according to its own distribution (without conditioning upon (R_1^N, Y_1^N));

(c) the distribution of Y_1^N conditional on (X_1^N, R_1^N) is a very complex one, while it is, in general, very simple in the classical models. However, this additional complexity enriches the model and does not interfere in the computations of interest;

(d) the Gaussianity is not needed, either at the X_1^N distribution level or at the Y_1^N one.

We see that in "CMSLC" the word "conditionally" concerns the Markovianity of R_1^N as well as the linearity of X_1^N.

The oriented dependence graphs of the classical models, the long-memory models proposed in (Pieczynski *et al.*, 2009), and the CMSLC proposed in the present paper are presented in Figure 1.

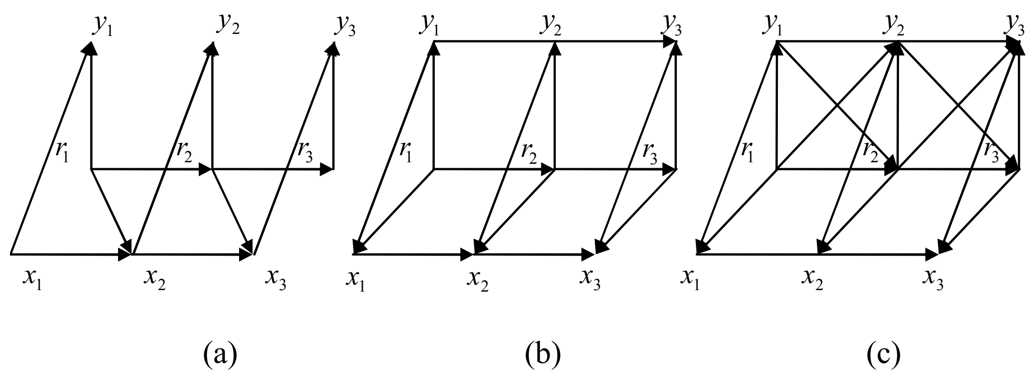

Figure 1: Dependence oriented graphs of: (a) classical model; (b) recent long-memory model; (c) new conditionally Markov switching linear chain (CMSLC) model.

3. Filtering withCMSLC

In the following, we assume that $p(r_{n+1}, y_{n+1} | r_n, y_1^n)$ are given in a closed form. The main property of the CMSLC model is that $p(y_{n+1} | y_1^n)$ is linked to $p(r_n | y_1^n)$ by

$$p(y_{n+1} | y_1^n) = \sum_{r_{n+1}} \sum_{r_n} p(r_n | y_1^n) p(r_{n+1}, y_{n+1} | r_n, y_1^n), \tag{2.4}$$

which comes from the fact that (R_1^N, Y_1^N) is a PPMC. Thus $p(y_2 | y_1^1)$, …, $p(y_{n+1} | y_1^n)$ are computable with complexity linear in time. This is the key point because the lack of the computability of $p(r_n | y_1^n)$ with complexity linear in times is the very reason for the impossibility of exact filtering in classical models.
We can state the following result:

Lemma

Let us consider a CMSLC (X_1^N, R_1^N, Y_1^N). Then we have:

(i) $p(r_{n+1} | y_1^{n+1})$ is given from $p(r_n | y_1^n)$ by

$$p(r_{n+1} | y_1^{n+1}) = \frac{1}{p(y_{n+1} | y_1^n)} \sum_{r_n} p(r_n | y_1^n) p(r_{n+1}, y_{n+1} | r_n, y_1^n); \tag{2.5}$$

(ii) for $n = 1, \ldots, N-1$, and $i = 1, \ldots, n$, the distribution $p(x_i | r_{n+1}, y_1^{n+1})$ is given from the distribution $p(x_i | r_n, y_1^n)$ by

$$p(x_i | r_{n+1}, y_1^{n+1}) = \frac{\sum_{r_n} p(r_n | y_1^n) p(r_{n+1}, y_{n+1} | r_n, y_1^n) p(x_i | r_n, y_1^n)}{p(y_{n+1} | y_1^n) p(r_{n+1} | y_1^{n+1})}, \tag{2.7}$$

where $p(r_{n+1} | y_1^{n+1})$ is computable with (2.5) and $p(r_{n+1}, y_{n+1} | r_n, y_1^n)$ are given.

Proof

(i) is given by the following classical computation: $p(r_{n+1} | y_1^{n+1}) = \sum_{r_n} p(r_{n+1}, r_n | y_1^{n+1}) = \frac{1}{p(y_{n+1} | y_1^n)} \sum_{r_n} p(r_{n+1}, r_n, y_{n+1} | y_1^n)$, which leads to the results knowing that

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

to show (ii), we can write:

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

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

Knowing that according to the model we have $p(r_{n+1}, y_{n+1} | x_i, r_n, y_1^n) = p(r_{n+1}, y_{n+1} | r_n, y_1^n)$, it gives

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

which is (2.7) and ends the proof.

Proposition

Let us consider a CMSLC (X_1^N, R_1^N, Y_1^N). Then for $n = 1, \ldots, N-1$, and $i = 1, \ldots, n$, $E[X_i | r_{n+1}, y_1^{n+1}]$ is given from $E[X_i | r_n, y_1^n]$ by

$$E[X_i | r_{n+1}, y_1^{n+1}] = \frac{\sum_{r_n} p(r_n | y_1^n) p(r_{n+1}, y_{n+1} | r_n, y_1^n) E[X_i | r_n, y_1^n]}{p(y_{n+1} | y_1^n) p(r_{n+1} | y_1^{n+1})} , \tag{2.8}$$

and $E[X_{n+1} | r_{n+1}, y_1^{n+1}]$ is given from $E[X_1 | r_{n+1}, y_1^{n+1}]$, …, $E[X_n | r_{n+1}, y_1^{n+1}]$ by

$$E[X_{n+1} | r_{n+1}, y_1^{n+1}] = \sum_{i=1}^{n} F_i^{n+1}(r_{n+1}, y_{n+1}) E[X_i | r_{n+1}, y_1^{n+1}] \tag{2.9}$$

Proof

(2.8) is a direct consequence of (2.7). To show (2.9), let us take the expectation of (2.3) conditional on $(R_{n+1}, Y_{n+1}) = (r_{n+1}, y_{n+1})$. As the randomn variables G_1, …, G_{N-1} are centred, we have $E[X_{n+1} | r_{n+1}, y_1^{n+1}] = \sum_{i=1}^{n} F_i^{n+1}(r_{n+1}, y_{n+1}) E[X_i | r_{n+1}, y_1^{n+1}]$, which ends the proof.

4. Conclusions and perspectives

We presented a "Conditionally Markov switching linear chain" (CMSLC) model (X_1^N, R_1^N, Y_1^N), in which both hidden switches process R_1^N and hidden states process X_1^N can be recovered from the observed process Y_1^N by a Kalman-like filtering with complexity polynomial in time. None of the distributions $p(x_1^N | r_1^N, y_1^N)$, $p(y_1^N | r_1^N, x_1^N)$ needs to be Markovian, and can be, in particular, of the "long-memory" kind.

Tackling the parameter problem in such models, using the general "Expectation-Maximization" (EM) principle (McLachlan and Khrishnan 1996) or the general "Iterative Conditional Estimation" (ICE) principle (Derrode and Pieczynski 2004), is undoubtedly among the most important perspectives.

References

Abbassi, N. and Pieczynski, W. (2008). "Exact filtering in semi-Markov jumping system." Paper presented at the Sixth International Conference of Computational Methods in Sciences and Engineering, September 25-30, Hersonissos, Crete, Greece.

Andrieu, C., Davy, C. M., and Doucet, A. (2003). "Efficient particle filtering for jump Markov systems. Application to time-varying autoregressions." *IEEE Trans on Signal Processing*, 51(7): 1762-1770.

Bardel, N. and Desbouvries, F. (2009). "Exact Bayesian Prediction in non-Gaussian Markov-Switching Model." Paper presented at the XIIIth International Conference on Applied Stochastic Models and Data Analysis (ASDMDA), Vilnius, Lithuania, June 30-July 3.

Beran J. and Taqqu M. S. (1994). "*Statistics for Long-Memory processes*." Monographs on Statistics and Applied Probability, Chapman and Hall, New York.

Cappé, O., Moulines E., and Ryden T. 2005. *Inference in hidden Markov models*, Springer.

Costa, O. L. V., Fragoso, M. D., and Marques, R. P. (2005). "*Discrete time Markov jump linear systems*." New York, Springer-Verlag.

Derrode S. and Pieczynski W. (2004). "Signal and Image Segmentation using Pairwise Markov Chains." *IEEE Trans. on Signal Processing*, 52(9): 2477-2489.

Doucet, A., Gordon, N. J. and Krishnamurthy, V. (2001). "Particle filters for state estimation of Jump Markov Linear Systems." *IEEE Trans. on Signal Processing*, 49(3): 613-624.

Doukhan P., Oppenheim G., and Taqqu M. S. (2003). "*Long-Range Dependence*." Birkhauser.

Germani, A. Manes, C., and Palumbo, P. (2006). "Filtering for bimodal systems: the case of unknown switching statistics." *IEEE Trans. on Circuits and Systems,* 53(6): 393-404.

Giordani P., Kohn R., and van Dijk, D. (2007). "A unified approach to nonlinearity, structural change, and outliers." *Journal of Econometrics*, 137: 112-133.

Ho, T.-J. and Chen, B.-S. (2006). "Novel extended Viterbi-based multiple-model algorithms for state estimation of discrete-time systems with Markov jump parameters." *IEEE Trans. on Signal Processing,* 54(2): 393-404.

Johnson, L. D. and Sakoulis, G. (2008). "Maximizing equity market sector predictability in a Bayesian time-varying parameter model." *Computational Statistics & Data Analysis*, 52(6): 3083-3106.

Kim, C.-J., Piger, J. and Starz, R. (2007). "Estimation of Markov regime-switching regression models with endogenous switching." *Journal of Econometrics,* 143: 263-273.

Lanchantin, P., Lapuyade-Lahorgue, J., and Pieczynski, W. (2008). "Unsupervised segmentation of triplet Markov chains hidden with long-memory noise." *Signal Processing,* 88(5): 1134-1151.

Lee, J.-W. and Dullerud, G. E. (2007). "A stability and contractiveness analysis of discrete-time Markovian jump linear systems." *Automatica,* 43: 168-173.

McLachlan, G. J. and Khrishnan, T. (1996). *"The EM algorithm and extensions".* Wiley series in Probability and Statistics, Wiley Interscience.

Orguner, U. and Demirekler, M. (2008). "Risk-sensitive filtering for jump Markov linear systems." *Automatica,* 44: 109-118.

Pieczynski, W. (2007). "Multisensor triplet Markov chains and theory of evidence." *International Journal of Approximate Reasoning*, 45(1): 1-16.

Pieczynski, W. (2008). "Exact calculation of optimal filter in semi-Markov switching model." Paper presented at the Fourth World Conference of the International Association for Statistical Computing (IASC 2008), December 5-8, Yokohama, Japan.

Pieczynski, W., Abbassi, N., and Ben Mabrouk, M. (2009) "Exact filtering and smoothing of Markov switching linear system hidden with Gaussian long-memory Noise." Paper presented at the XIII International Conference Applied

Stochastic Models and Data Analysis, (ASMDA 2009), June 30- July 3, Vilnius, Lithuania.

Pieczynski, W. (2009a). "Exact filtering in Markov marginal switching hidden models." Submitted to *Comptes Rendus Mathématique*.

Pieczynski, W. (2009b). "Exact smoothing in hidden conditionally Markov switching chains." Paper presented at the XIII International Conference Applied Stochastic Models and Data Analysis, (ASMDA 2009), June 30- July 3, Vilnius, Lithuania.

Pieczynski, W. and Desbouvries, F. (2009). "Exact Bayesian smoothing in triplet switching Markov chains." Paper presented at the conference "Complex data modeling and computationally intensive statistical methods for estimation and prediction" (S. Co 2009), September 14-16, Milan, Italy.

Ristic, B. Arulampalam, S. and Gordon, N. (2004). "*Beyond the Kalman Filter - Particle filters for tracking applications.*" Artech House, Boston, USA.

Tugnait, J. K. (1982). "Adaptive estimation and identification for discrete systems with Markov jump parameters." *IEEE Trans. on Automatic Control*, AC-25: 1054-1065.

Zhou, T. N. and Shumway, R. (2008). "One-step approximations for detecting regime changes in the state space model with application to the influenza data." *Computational Statistics & Data Analysis*, 52(5): 2277-2291.

Zoeter, O. and Heskes, T. (2006). "Deterministic approximate inference techniques for conditionally Gaussian state space models." *Statistical Computation*, 16: 279-292.