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

Wojciech Pieczynski, Telecom SudParis and CNRS, France

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 \( \mathbb{R}^q \), while \( Y_n \) takes its values from \( \mathbb{R}^m \). The sequences \( X_1^N \) and \( R_1^N \) are hidden and the sequence \( Y_1^N \) is observed. We deal with the problem of filtering, which consists of the computation, for each \( n = 1, ..., N \), 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

\[
E[X_n|y_1^n] = \sum_{r_n} E[X_n|r_n, y_1^n] p(r_n|y_1^n)
\]  \hspace{1cm} (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:
\[ R_1^N \text{ is a Markov chain; } \]  
\[ X_{n+1} = F_n(R_n)X_n + G_n(R_n)W_n; \]  
\[ Y_n = H_n(R_n)X_n + J_n(R_n)Z_n, \]

where \( X_1, W_1, \ldots, W_N \) are independent (conditionally on \( R_1^N \)) Gaussian vectors in \( R^q \), \( Z_1, \ldots, Z_N \) are independent (conditionally on \( R_1^N \)) Gaussian vectors in \( R^m \), \( F_1(R_1), \ldots, F_N(R_N), G_1(R_1), \ldots, G_N(R_N) \) are matrices of size \( q \times q \) depending on switches, and \( H_1(R_1), \ldots, H_N(R_N), J_1(R_1), \ldots, J_N(R_N) \) are matrices of size \( q \times m \) also depending on switches. Therefore the classical Kalman filter can be used when \( R_1^N = \pi_1^N \) is known; however, it has been well known since the publication of (Tugnait, 1982) that the exact computation of neither \( E[X_n | P_{n-1}, \pi_1^N] \) nor \( E[X_n | P_{n-1}, Y_1^N] \) 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 deep 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 probabilities \( p(y_{n+1} | y_1^n) \) are computable and the exact filtering and smoothing are also. More sophisticated models, in which the hypothesis (ii) is relaxed but the possibility of exact filtering remains were proposed in (Pieczynski 2009a; Pieczynski and Desbovriers 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 \((R^N_i, Y^N_i)\) has been relaxed and replaced by the “partial” Markovianity, in which \((R^N_i, Y^N_i)\) is Markovian with respect to \(R^N_i\) but is not necessarily Markovian with respect to \(Y^N_i\) (Pieczyński et al. 2009). In the second one, the distribution of the state chain \(X^N_i\) conditional on \((R^N_i, Y^N_i)\) remains linear but is no longer necessarily Markovian (Pieczyński 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 \(p(y^N_i | x^N_i, x^N_i)\) nor \(p(x^N_i | y^N_i, y^N_i)\) 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 chains (CMSLC)

Let \((X^N_i, R^N_i, Y^N_i)\) be the triplet of random sequences as above. The distribution of the couple \((R^N_i, Y^N_i)\) will be assumed to be a “pairwise partially Markov chain” (PPMC) distribution recently introduced in (Lanchantin et al. 2008). The distribution \(p(r^N_i, y^N_i)\) of a PPCMC \((R^N_i, Y^N_i)\) can be defined by \(p(r_i, y_i)\) and the transitions \(p(r_n, y_n | r_{i_{n+1}}, y_{i_{n+1}})\) verifying

\[
p(r_n, y_n | r_{n+1}, y_{n+1}) = p(r_n, y_n | r_{n+1}, y_{n+1})
\]  

(2.1)

Such a law is called “partially” Markovian as it can be seen as being Markovian with respect to the variables \(R^N_i\), but is not necessarily Markovian with respect to the variables \(Y^N_i\).

Definition 1

A triplet \((X^N_i, R^N_i, Y^N_i)\) will be said to be a “conditionally Markov switching linear chain” (CMSLC) if it verifies

\[
(R^N_i, Y^N_i) \text{ is a PPCMC} ;
\]  

(2.2)

for \(n = 1, \ldots, N-1,

\]

(2.3)
with $F_{n+1}(r_{n+1}, y_{n+1}) = [F_1(r_{n+1}, y_{n+1}), F_2(r_{n+1}, y_{n+1}), \ldots, F_n(r_{n+1}, y_{n+1})]$, where each $F_i(r_{n+1}, y_{n+1})$ 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 $(r_{n+1}, y_{n+1})$, $H_{n+1}(r_{n+1}, y_{n+1})$ is a vector of size $q$ depending on $(r_{n+1}, y_{n+1})$, and $X_1, W_1, \ldots, W_N$ 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 in the general case;

(b) similarly, the model is said to be “conditionally linear” because $X_i^N$ is linear conditionally on $(R_i^N, Y_i^N)$; however, contrary to the classical models, it is not necessarily linear according to its distribution conditional $R_i^N$;

(c) the distribution of $Y_i^N$ conditional on $(X_i^N, R_i^N)$ 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 $X_i^N$ distribution level or at the $Y_i^N$ one.

We see that in “CMSLC” the word “conditionally” concerns the Markovianity of $R_i^N$ as well as the linearity of $X_i^N$.

3. Filtering with CMSLC

In the following, we assume that $p(r_{n+1}, y_{n+1} | r_n, y_n^N)$ are given in a closed form.

The main property of the model is that $p(y_{n+1} | y_n^N)$ is linked to $p(r_n | y_n^N)$ by

$$p(y_{n+1} | y_n^N) = \sum_{r_{n+1}} \sum_{y_{n+1}} p(r_{n+1} | y_{n+1}) p(r_n | y_n^N) p(y_{n+1} | r_{n+1}, y_{n+1})$$

(2.4)

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

**Lemma**

Let us consider n CMSLC \((X_i^n, R_i^n, Y_i^n)\). Then we have:

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

\[
p(r_{n+1}|y_1^{n+1}) = \frac{1}{p(y_n^1|y_1^n)} \sum r_n p(r_n|y_n) p(r_{n+1}, y_{n+1}|r_n, y_n^n); \tag{2.5}
\]

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

\[
p(x_i|r_{n+1}, y_i^{n+1}) = \frac{\sum r_n p(x_i|r_n, y_i^n) p(r_{n+1}, y_{n+1}|r_n, y_n^n)}{p(y_n^1|y_1^n)p(r_{n+1}|y_1^{n+1})}, \tag{2.7}
\]

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

**Proof**

(i) is given by the following classical computation:

\[
p(r_{n+1}|y_1^{n+1}) = \sum r_n p(r_{n+1}, r_n|y_1^{n+1}) = \frac{1}{p(y_n^1|y_1^n)} \sum r_n p(r_{n+1}, r_n, y_{n+1}|y_1^n),
\]

leads to the results knowing that

\[
p(r_{n+1}, y_{n+1}|r_n, y_n^n) = p(r_n|y_n^n) p(r_{n+1}, y_{n+1}|r_n, y_n^n);
\]

(ii), we have

\[
p(x_i|r_{n+1}, y_i^{n+1}) = \frac{p(x_i, r_{n+1}, y_{n+1}|y_1^n)}{p(r_{n+1}, y_{n+1}|y_1^n)} = \frac{p(x_i, r_{n+1}, y_{n+1}|y_1^n)}{p(y_n^1|y_1^n)p(r_{n+1}|y_1^{n+1})} = \sum r_n \frac{p(x_i, r_n, r_{n+1}, y_{n+1}|y_1^n)}{p(y_n^1|y_1^n)p(r_{n+1}|y_1^{n+1})}.
\]

Knowing that according to the model we have
\[ p(r_{n+1}, y_{n+1} \mid x_n, r_n, y_n) = p(r_{n+1}, y_{n+1} \mid f_n, y_n), \quad \text{it gives} \]
\[ \sum_{r_n} \frac{p(r_n \mid y_n \mid x_n, r_{n+1}, y_{n+1}) p(r_{n+1}, y_{n+1} \mid r_n, y_n) p(x_n \mid r_n, y_n)}{p(y_{n+1} \mid y_n) p(r_{n+1} \mid y_{n+1})}, \] which is (2.7) and ends the proof.

Proposition

Let us consider a CMSLC \((X_1^N, R_1^N, Y_1^N)\). Let \( n \in \{1, \ldots, N-1\} \). Then for \( i = 1, \ldots, n \), \( E[X_i \mid r_{n+1}, y_{n+1}] \) is given from \( E[X_i \mid r_n, y_n] \) by
\[
E[X_i \mid r_{n+1}, y_{n+1}] = \sum_{r_n} \frac{p(r_n \mid y_n \mid x_n, r_{n+1}, y_{n+1}) p(r_{n+1}, y_{n+1} \mid r_n, y_n) E[X_i \mid r_n, y_n]}{p(y_{n+1} \mid y_n) p(r_{n+1} \mid y_{n+1})}, \tag{2.8}
\]
and \( E[X_{n+1} \mid r_{n+1}, y_{n+1}] \) is given from \( E[X_1 \mid r_n, y_n], \ldots, E[X_n \mid r_n, y_n] \) by
\[
E[X_{n+1} \mid r_{n+1}, y_{n+1}] = H_{n+1} (r_{n+1}, y_{n+1}) + \sum_{i=1}^{n} F_i^{n+1} (r_{n+1}, y_{n+1}) \sum_{r_n} \frac{p(r_n \mid y_n \mid x_n, r_{n+1}, y_{n+1}) p(r_{n+1}, y_{n+1} \mid r_n, y_n) E[X_i \mid r_n, y_n]}{p(y_{n+1} \mid y_n) p(r_{n+1} \mid y_{n+1})}, \tag{2.9}
\]

Proof

(2.8) is a direct consequence of (2.7). To show (2.9), let us take the expectation of (2.3) conditional on \((R_{n+1}, Y_{n+1}) = (r_{n+1}, y_{n+1})\). We have
\[
E[X_{n+1} \mid r_{n+1}, y_{n+1}] = H_{n+1} (r_{n+1}, y_{n+1}) + \sum_{i=1}^{n} F_i^{n+1} (r_{n+1}, y_{n+1}) E[X_i \mid r_{n+1}, y_{n+1}] = H^{n+1} (r_{n+1}, y_{n+1}) + \sum_{i=1}^{n} F_i^{n+1} (r_{n+1}, y_{n+1}) \sum_{r_n} \frac{p(r_n \mid y_n \mid x_n, r_{n+1}, y_{n+1}) p(r_{n+1}, y_{n+1} \mid r_n, y_n) E[X_i \mid r_n, y_n]}{p(y_{n+1} \mid y_n) p(r_{n+1} \mid y_{n+1})},
\]
which is obtained using (2.8) and ends the proof.

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

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

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

References


