ITERATIVE ALGORITHMS FOR OPTIMAL STATE ESTIMATION OF JUMP MARKOV LINEAR SYSTEMS

Arnaud Doucet - Christophe Andrieu

Signal Processing Group, University of Cambridge
Department of Engineering, Trumpington Street
CB2 1PZ Cambridge, UK
Email: ad2@eng.cam.ac.uk - ca226@eng.cam.ac.uk

ABSTRACT

Jump Markov linear systems (JMLS) are linear systems whose parameters evolve with time according to a finite state Markov chain. We present three original deterministic and stochastic iterative algorithms for optimal state estimation of JMLS whose computational complexity at each iteration is linear in the data length. The first algorithm yields conditional mean estimates. The second algorithm is an algorithm that yields the marginal maximum *a posteriori* (MMAP) sequence estimate of the finite state Markov chain. The third algorithm is an algorithm that yields the MMAP sequence estimate of the continuous state of the JMLS. Convergence results for these three algorithms are obtained. Computer simulations are carried out to evaluate their performance.

1. INTRODUCTION

Jump Markov linear systems are widely used in several fields of signal processing and include, as particular cases, common models in seismic signal processing [4], [9], communications [1] and target tracking (see [8] for many examples and references). Given a set of observations, our aim is to estimate respectively the conditional mean estimate of the states and the marginal maximum *a posteriori* (MMAP) sequence estimate of the finite state Markov chain and of the continuous state of the JMLS. It is well known that exact computation of these three estimates for JMLS involves a prohibitive computational cost, exponential in the number, say T, of observations. Thus in practice, it is necessary to consider suboptimal estimation algorithms. A variety of such suboptimal algorithms has already been proposed in the literature [9], [8], [12].

In this paper, we present iterative deterministic and stochastic algorithms to solve the three above mentioned problems. They have a computational cost of $O(T)$ per iteration. Convergence results are also obtained. The range of applicability of the proposed methods is wider than the previous methods, including recent work in [5], [8]. Moreover, it can be theoretically established that they are more efficient in a specified sense.

2. MODEL AND ESTIMATION OBJECTIVES

2.1. Signal model

Throughout this paper, we will use n_z to denote the dimension of an arbitrary vector z. Let $t \in \{1, 2, ...\}$ denote discrete time and let r_t be a discrete-time, homogeneous, s-state, first-order Markov chain with transition probabilities $p_{ij} \triangleq \Pr\{r_{t+1} = j | r_t = i\}$ and initial probability distribution $p_i \triangleq \Pr\{r_1 = i\}$, $(i, j \in S)$, $S = \{1, 2, ..., s\}$. Consider the following JMLS:

$$x_{t+1} = A(r_{t+1})\, x_t + B(r_{t+1})\, v_{t+1} + F(r_{t+1})\, u_{t+1} \quad (1)$$

$$y_t = C(r_t)\, x_t + D(r_t)\, w_t + G(r_t)\, u_t \quad (2)$$

where $x_t \in \mathbb{R}^{n_x}$ is the system state, $y_t \in \mathbb{R}^{n_y}$ is the observation at time t, $u_t \in \mathbb{R}^{n_u}$ is a known deterministic input, $v_t \overset{i.i.d.}{\sim} \mathcal{N}(0, I_{n_v}) \in \mathbb{R}^{n_v}$, $w_t \overset{i.i.d.}{\sim} \mathcal{N}(0, I_{n_w}) \in \mathbb{R}^{n_w}$ and $D(i)\, D^{\mathrm{T}}(i) > 0$ $(\forall i \in S)$. $x_0 \sim \mathcal{N}(\hat{x}_0, P_0)$ where $P_0 > 0$ and x_0, v_t and w_t are mutually independent for all t. The model parameters $\lambda \triangleq \{p_i, p_{ij}, A(i), B(i), C(i), D(i), F(i), G(i), \hat{x}_0, P_0; i, j \in S\}$ are assumed known. We denote for $p, q, p < q$, $\mathbf{y}_{p:q}$ as $(y_p, ..., y_q)$, $\mathbf{x}_{p:q}$ as $(x_p, ..., x_q)$, $\mathbf{r}_{p:q}$ as $(r_p, ..., r_q)$ and finally $N(m, \Sigma)$ as $|\Sigma|^{-1/2} \exp\left(-\frac{1}{2} m^{\mathrm{T}} \Sigma^{-1/2} m\right)$.

2.2. Estimation objectives

Given $\mathbf{y}_{1:T}$ and assuming that λ is known, any required Bayesian inference for JMLS can be made using $p(\mathbf{r}_{1:T}, \mathbf{x}_{0:T} | \mathbf{y}_{1:T})$. We consider the three following optimal estimation problems:

- Conditional Mean Estimates of $\mathbf{x}_{0:T}$ and $\mathbf{r}_{1:T}$: compute optimal (in a mean square sense) estimates of $\mathbf{x}_{0:T}$ and $\mathbf{r}_{1:T}$ given by $\mathbb{E}\{\mathbf{r}_{1:T} | \mathbf{y}_{1:T}\}$ and $\mathbb{E}\{\mathbf{x}_{0:T} | \mathbf{y}_{1:T}\}$.
- MMAP Sequence Estimate of $\mathbf{r}_{1:T}$: compute optimal (in a MAP sense) state estimates of $\mathbf{r}_{1:T}$ by maximizing $p(\mathbf{r}_{1:T} | \mathbf{y}_{1:T})$.
- MMAP Sequence Estimate of $\mathbf{x}_{0:T}$: compute optimal (in a MAP sense) state estimates of $\mathbf{x}_{0:T}$ by maximizing $p(\mathbf{x}_{0:T} | \mathbf{y}_{1:T})$.

3. CONDITIONAL MEAN ESTIMATION

$\mathbb{E}\{\mathbf{r}_{1:T} | \mathbf{y}_{1:T}\}$ and $\mathbb{E}\{\mathbf{x}_{0:T} | \mathbf{y}_{1:T}\}$ are obtained by integration with respect to $p(\mathbf{r}_{1:T}, \mathbf{x}_{0:T} | \mathbf{y}_{1:T})$. If we were able to obtain N ($N \gg 1$) i.i.d. samples $\left\{\left(\mathbf{r}_{1:T}^{(k)}, \mathbf{x}_{0:T}^{(k)}\right); k = 1, ..., N\right\}$ distributed according to the joint distribution then, using the Law of Large Numbers, conditional mean estimates could be computed by averaging. Obtaining such i.i.d. samples from the posterior distribution is not feasible. This is why we propose here a Markov chain Monte Carlo (MCMC) method [11].

3.1. Algorithm

We simulate samples from the posterior distribution using the following iterative MCMC method. k denotes the iteration number.

MCMC algorithm to obtain the conditional mean estimates

1. Initialization. Set randomly $\mathbf{r}_{1:T}^{(0)}$.

Authorship based on reversed alphabetical order

2. Iteration k, $k \geq 1$

- For $t = 1, \ldots, T$, sample $r_t^{(k)} \sim p\left(r_t | \mathbf{y}_{1:T}, \mathbf{r}_{-t}^{(k)}\right)$.
- Optional step. Sample $\mathbf{x}_{0:T}^{(k)} \sim p\left(\mathbf{x}_{0:T} | \mathbf{y}_{1:T}, \mathbf{r}_{1:T}^{(k)}\right)$.

---

where $\mathbf{r}_{-t}^{(k)} \triangleq \left(r_1^{(k)}, \ldots, r_{t-1}^{(k)}, r_{t+1}^{(k-1)}, \ldots, r_T^{(k-1)}\right)$. Note that if one is not interested in estimating $\mathbf{x}_{0:T}$, then it is not necessary to sample from $p\left(\mathbf{x}_{0:T} | \mathbf{y}_{1:T}, \mathbf{r}_{1:T}^{(k)}\right)$. The obtained samples are used to compute conditional mean estimates of the states $\mathbf{r}_{1:T}$ and $\mathbf{x}_{0:T}$ using the following ergodic averages $\overline{\mathbf{r}}_{1:T} \triangleq \frac{1}{N}\sum_{k=0}^{N-1} \mathbf{r}_{1:T}^{(k)}$ and $\overline{\mathbf{x}}_{0:T} \triangleq \frac{1}{N}\sum_{k=0}^{N-1} \mathbf{x}_{0:T}^{(k)}$.

3.2. Implementation issues

This algorithm requires to be able to sample from $p\left(r_t | \mathbf{y}_{1:T}, \mathbf{r}_{-t}^{(k)}\right)$ for $t = 1, \ldots, T$ and optionally from $p\left(\mathbf{x}_{0:T} | \mathbf{y}_{1:T}, \mathbf{r}_{1:T}^{(k)}\right)$. A direct evaluation of these distributions would have a complexity $O\left(T^2\right)$. We develop here an algorithm of complexity $O(T)$, which relies on the following lemmas and proposition.

Lemma 1 *For any $t = 2, ..., T-1$*

$$\begin{aligned} &p\left(\mathbf{y}_{1:T} | \mathbf{r}_{1:T}\right) = p\left(\mathbf{y}_{1:t-1} | \mathbf{r}_{1:t-1}\right) p\left(y_t | \mathbf{y}_{1:t-1}, \mathbf{r}_{1:t}\right) \\ &\times \int p\left(\mathbf{y}_{t+1:T} | \mathbf{r}_{t+1:T}, x_t\right) p\left(x_t | \mathbf{y}_{1:t}, \mathbf{r}_{1:t}\right) dx_t \\ &p\left(\mathbf{y}_{t:T} | \mathbf{r}_{t:T}, x_{t-1}\right) = \int p\left(\mathbf{y}_{t+1:T} | \mathbf{r}_{t+1:T}, x_t\right) \\ &\times p\left(y_t, x_t | r_t, x_{t-1}\right) dx_t \end{aligned}$$

Lemma 2 *For any $t = 1, ..., T$, $p\left(\mathbf{y}_{t:T} | \mathbf{r}_{t:T}, x_{t-1}\right)$ is a Gaussian distribution of mean $M_t\left(\mathbf{r}_{t:T}\right) x_{t-1} + \mathbb{E}\left[N_t\left(\mathbf{r}_{t:T}\right)\right]$ and covariance $cov\left[L_t\left(\mathbf{r}_{t:T}\right)\right] > 0$ with $L_t\left(\mathbf{r}_{t:T}\right) \triangleq N_t\left(\mathbf{r}_{t:T}\right) N_t^{\mathsf{T}}\left(\mathbf{r}_{t:T}\right)$ where*

$$\begin{aligned} &P_{t-1|t}'^{-1}\left(\mathbf{r}_{t:T}\right) \triangleq M_t^{\mathsf{T}}\left(\mathbf{r}_{t:T}\right) L_t^{-1}\left(\mathbf{r}_{t:T}\right) M_t\left(\mathbf{r}_{t:T}\right) \\ &P_{t-1|t}'^{-1}\left(\mathbf{r}_{t:T}\right) m'_{t-1|t}\left(\mathbf{r}_{t:T}\right) \triangleq M_t^{\mathsf{T}}\left(\mathbf{r}_{t:T}\right) L_t^{-1}\left(\mathbf{r}_{t:T}\right) \mathbf{y}_{t:T} \end{aligned}$$

are given by the following backward information filter recursion.

1. *Initialization*

$$\begin{aligned} &P_{T|T}'^{-1}\left(r_T\right) = C^{\mathsf{T}}\left(r_T\right)\left(D\left(r_T\right) D^{\mathsf{T}}\left(r_T\right)\right)^{-1} C\left(r_T\right) \\ &P_{T|T}'^{-1}\left(r_T\right) m'_{T|T}\left(r_T\right) = C^{\mathsf{T}}\left(r_T\right)\left(D\left(r_T\right) D^{\mathsf{T}}\left(r_T\right)\right)^{-1} \\ &\times \left(y_T - G\left(r_T\right) u_T\right) \end{aligned} \tag{3}$$

2. *Backward recursion. For $t = T-1, \ldots, 1$,*

$$\begin{aligned} &P_{t|t+1}'^{-1}\left(\mathbf{r}_{t+1:T}\right) = A^{\mathsf{T}}\left(r_{t+1}\right) P_{t+1|t+1}'^{-1}\left(\mathbf{r}_{t+1:T}\right)\left(I_{n_x} - \right. \\ &\left. B\left(r_{t+1}\right) \mathbf{\Delta}_{t+1}\left(\mathbf{r}_{t+1:T}\right) B^{\mathsf{T}}\left(r_{t+1}\right) P_{t+1|t+1}'^{-1}\left(\mathbf{r}_{t+1:T}\right)\right) \\ &\times A^{\mathsf{T}}\left(r_{t+1}\right) \\ &P_{t|t+1}'^{-1}\left(\mathbf{r}_{t+1:T}\right) m'_{t|t+1}\left(\mathbf{r}_{t+1:T}\right) = A^{\mathsf{T}}\left(r_{t+1}\right)\left(I_{n_x} - B\left(r_{t+1}\right)\right. \\ &\left. \times \mathbf{\Delta}_{t+1}\left(\mathbf{r}_{t+1:T}\right) B^{\mathsf{T}}\left(r_{t+1}\right) P_{t+1|t+1}'^{-1}\left(\mathbf{r}_{t+1:T}\right)\right) \\ &\times P_{t+1|t+1}'^{-1}\left(\mathbf{r}_{t+1:T}\right)\left(m'_{t+1|t+1}\left(\mathbf{r}_{t+1:T}\right) - F\left(r_{t+1}\right) u_{t+1}\right) \\ &\mathbf{\Delta}_{t+1} = \left[I_{n_v} + B^{\mathsf{T}}\left(r_{t+1}\right) P_{t+1|t+1}'^{-1}\left(\mathbf{r}_{t+1:T}\right) B\left(r_{t+1}\right)\right]^{-1} \\ &P_{t|t}'^{-1}\left(\mathbf{r}_{t:T}\right) = C^{\mathsf{T}}\left(r_t\right)\left(D\left(r_t\right) D^{\mathsf{T}}\left(r_t\right)\right)^{-1} C\left(r_t\right) \\ &+ P_{t|t+1}'^{-1}\left(\mathbf{r}_{t+1:T}\right) \\ &P_{t|t}'^{-1}\left(\mathbf{r}_{t:T}\right) m'_{t|t}\left(\mathbf{r}_{t:T}\right) = C^{\mathsf{T}}\left(r_t\right)\left(D\left(r_t\right) D^{\mathsf{T}}\left(r_t\right)\right)^{-1} \\ &\times \left(y_t - G\left(r_t\right) u_t\right) + P_{t|t+1}'^{-1}\left(\mathbf{r}_{t+1:T}\right) m'_{t|t+1}\left(\mathbf{r}_{t+1:T}\right) \end{aligned} \tag{4}$$

Proposition 3 *For any $t = 2, ..., T-1$ we have*[1]

$$\begin{aligned} &p\left(r_t | \mathbf{y}_{1:T}, \mathbf{r}_{-t}\right) \propto p_{r_{t-1}r_t} p_{r_t r_{t+1}} N\left(\widetilde{y}_{t|t-1}\left(\mathbf{r}_{1:t}\right), S_t\left(\mathbf{r}_{1:t}\right)\right) \\ &\times \left|P_{t|t}\left(\mathbf{r}_{1:t}\right)\right|^{-\frac{1}{2}} \left|P_{t|t+1}'^{-1}\left(\mathbf{r}_{t+1:T}\right) + P_{t|t}^{-1}\left(\mathbf{r}_{1:t}\right)\right|^{-\frac{1}{2}} \\ &\times \exp\left(-\tfrac{1}{2}\left[m_{t|t}^{\mathsf{T}}\left(\mathbf{r}_{1:t}\right) P_{t|t+1}'^{-1}\left(\mathbf{r}_{t+1:T}\right) m_{t|t}\left(\mathbf{r}_{1:t}\right)\right.\right. \\ &-2 m_{t|t}^{\mathsf{T}}\left(\mathbf{r}_{1:t}\right) P_{t|t+1}'^{-1}\left(\mathbf{r}_{t+1:T}\right) m'_{t|t+1}\left(\mathbf{r}_{t+1:T}\right) \\ &- \left(m'_{t|t+1}\left(\mathbf{r}_{t+1:T}\right) - m_{t|t}\left(\mathbf{r}_{1:t}\right)\right)^{\mathsf{T}} P_{t|t+1}'^{-1}\left(\mathbf{r}_{t+1:T}\right) \\ &\left.\left[P_{t|t+1}'^{-1}\left(\mathbf{r}_{t+1:T}\right) + P_{t|t}^{-1}\left(\mathbf{r}_{1:t}\right)\right]^{-1} P_{t|t+1}'^{-1}\left(\mathbf{r}_{t+1:T}\right)\right. \\ &\left.\left(m'_{t|t+1}\left(\mathbf{r}_{t+1:T}\right) - m_{t|t}\left(\mathbf{r}_{1:t}\right)\right)\right]\Big) \end{aligned} \tag{5}$$

where $m_{t|t-1}\left(\mathbf{r}_{1:t}\right)$, $P_{t|t-1}\left(\mathbf{r}_{1:t}\right)$, $m_{t|t}\left(\mathbf{r}_{1:t}\right)$, $P_{t|t}\left(\mathbf{r}_{1:t}\right)$, $\widetilde{y}_{t|t-1}\left(\mathbf{r}_{1:t}\right)$ and $S_t\left(\mathbf{r}_{1:t}\right)$ are respectively the one-step ahead prediction and covariance of x_t, the filtered estimate and covariance of x_t, the innovation at time t and the covariance of this innovation, these quantities are given by the Kalman filter, (1)-(2) being linear Gaussian until t as $\mathbf{r}_{1:t}$ is known. To sum up, the algorithm proceeds as follows at iteration k.

Backward-Forward procedure

For $t = T, \ldots, 1$ compute and store $P_{t|t+1}'^{-1}\left(\mathbf{r}_{t+1:T}^{(k-1)}\right)$ and $P_{t|t+1}'^{-1}\left(\mathbf{r}_{t+1:T}^{(k-1)}\right) m'_{t|t+1}\left(\mathbf{r}_{t+1:T}^{(k-1)}\right)$ using (3)-(4).

For $t = 1, \ldots, T$

- For $i = 1, \ldots, s$, run one step-ahead the Kalman filter with $r_t = i$, store $m_{t|t}\left(\mathbf{r}_{1:t-1}^{(k)}, r_t = i\right)$ and $P_{t|t}\left(\mathbf{r}_{1:t-1}^{(k)}, r_t = i\right)$, compute up to a normalizing constant $p\left(r_t = i | \mathbf{y}_{1:T}, \mathbf{r}_{-t}^{(k)}\right)$ using (5).
- Sample $r_t^{(k)} \sim p\left(r_t = i | \mathbf{y}_{1:T}, \mathbf{r}_{-t}^{(k)}\right)$ and store only $m_{t|t}\left(\mathbf{r}_{1:t-1}^{(k)}, r_t^{(k)}\right), P_{t|t}\left(\mathbf{r}_{1:t-1}, r_t^{(k)}\right)$.

Sampling from $p\left(\mathbf{x}_{0:T} | \mathbf{y}_{1:T}, \mathbf{r}_{1:T}^{(k)}\right)$ is realized using the efficient forward filtering-backward sampling recursion described in [2]. The filtering step being realized when sampling $\mathbf{r}_{1:T}^{(k)}$, it only remains to perform the backward sampling step.

3.3. Convergence issues

It can be proved that the simulated Markov chain is uniformly geometrically ergodic [11] so a central limit theorem holds for the estimates $\overline{\mathbf{r}}_{1:T}$ and $\overline{\mathbf{x}}_{0:T}$.

Proposition 4 *Let $\varphi : S^T \times \mathbb{R}^{(T+1)n_x} \to \mathbb{R}$ and the estimate $\overline{\varphi}_N \triangleq \frac{1}{N}\sum_{k=0}^{N-1} \varphi(\mathbf{r}_{1:T}^{(k)}, \mathbf{x}_{0:T}^{(k)})$. If $\mathbb{E}_{p(\cdot|\mathbf{y}_{1:T})}\left[|\varphi(\mathbf{r}_{1:T}, \mathbf{x}_{0:T})|\right] < \infty$ then*

$$\overline{\varphi}_N \overset{a.s.}{\to} \mathbb{E}_{p(\cdot|\mathbf{y}_{1:T})}\left[\varphi(\mathbf{r}_{1:T}, \mathbf{x}_{0:T})\right]$$

If $\mathbb{E}_{p(\cdot|\mathbf{y}_{1:T})}\left[|\varphi(\mathbf{r}_{1:T}, \mathbf{x}_{0:T})|^2\right] < +\infty$, then there exists $\sigma(\varphi)$ such that

$$\sqrt{N}\left(\overline{\varphi}_N - \mathbb{E}_{p(\cdot|\mathbf{y}_{1:T})}\left[\varphi(\mathbf{r}_{1:T}, \mathbf{x}_{0:T})\right]\right) \Rightarrow \mathcal{N}\left(0, \sigma^2(\varphi)\right)$$

It is possible to easily improve the asymptotic variance of the estimates $\overline{\mathbf{r}}_{1:T}$ and $\overline{\mathbf{x}}_{0:T}$ using a Rao-Blackwellization method [7].

[1] If $P'_{t|t+1}\left(\mathbf{r}_{t+1:T}\right)$ exists then one can obtain a much simpler expression for $p\left(r_t | \mathbf{y}_{1:T}, \mathbf{r}_{-t}\right)$.

4. MMAP ESTIMATION OF $\mathbf{R}_{1:T}$

4.1. Algorithm

Let us introduce a positive sequence $\{\gamma_k; k \in \mathbb{N}\}$ satisfying $\gamma_{k+1} \geq \gamma_k$, $\lim_{k\to+\infty} \gamma_k = +\infty$. The algorithm to obtain the MMAP estimate of $\mathbf{r}_{1:T}$ is a simple modification of the previous algorithm. Thus it has a computational complexity $O(T)$ at each iteration.

Stochastic algorithm to estimate the MMAP sequence of $\mathbf{r}_{1:T}$

1. Initialization. Set randomly $\mathbf{r}_{1:T}^{(0)}$.
2. Iteration k, $k \geq 1$
   - For $t = 1, \ldots, T$, sample $r_t^{(k)} \sim \overline{p}^{\gamma_k}\left(r_t | \mathbf{y}_{1:T}, \mathbf{r}_{-t}^{(k)}\right)$.

---

where $\overline{p}^{\gamma_k}\left(r_t | \mathbf{y}_{1:T}, \mathbf{r}_{-t}^{(k)}\right) \propto \left[p\left(r_t | \mathbf{y}_{1:T}, \mathbf{r}_{-t}^{(k)}\right)\right]^{\gamma_k}$.

Remark 1 *A deterministic version of this algorithm consists of selecting* $r_t^{(k)} = \arg\max_{\{1,\ldots,s\}} p\left(r_t | \mathbf{y}_{1:T}, \mathbf{r}_{-t}^{(k)}\right)$.

4.2. Convergence issues

The deterministic version of this algorithm converges towards (a discrete equivalent of) a stationary point of $p(\mathbf{r}_{1:T} | \mathbf{y}_{1:T})$. The stochastic version of it is nothing but a stochastic annealing algorithm in discrete state space. The following result can be easily obtained.

Proposition 5 *There exist $C > 0$ and $\varepsilon > 0$ such that for $\gamma_k = C \ln(k + \varepsilon)$ and for any initial sequence $\mathbf{r}_{1:T}^{(0)}$*

$$\lim_{i\to+\infty} \Pr\left(\mathbf{r}_{1:T}^{(i)} \in \mathcal{M}(\mathbf{r}_{1:T})\right) = 1$$

where $\mathcal{M}(\mathbf{r}_{1:T})$ is the set of MMAP estimates.

5. MMAP ESTIMATION OF $\mathbf{X}_{0:T}$

The algorithm presented in this part requires the following additional assumption that $B(i) B^{\mathrm{T}}(i) > 0$ for all $i \in \{1, \ldots, s\}$[2]. We introduce a positive sequence $\{\gamma_k; k \in \mathbb{N}\}$ satisfying $\gamma_{k+1} \geq \gamma_k$, $\lim_{k\to+\infty} \gamma_k = +\infty$. The proposed algorithm proceeds as follows.

Stochastic algorithm to estimate the MMAP sequence of $\mathbf{x}_{0:T}$

1. Initialization. Set randomly $\mathbf{x}_{0:T}^{(0)}$.
2. Iteration k, $k \geq 1$
   - For $t = 0, \ldots, T$, sample $x_t^{(k)} \sim \overline{p}^{\gamma_k}\left(x_t | \mathbf{y}_{1:T}, \mathbf{x}_{-t}^{(k)}\right)$.

---

where $\overline{p}^{\gamma_k}\left(x_t | \mathbf{y}_{1:T}, \mathbf{x}_{-t}^{(k)}\right) \propto \left[p\left(x_t | \mathbf{y}_{1:T}, \mathbf{x}_{-t}^{(k)}\right)\right]^{\gamma_k}$ and $\mathbf{x}_{-t}^{(k)} \triangleq \left(x_0^{(k)}, \ldots, x_{t-1}^{(k)}, x_{t+1}^{(k-1)}, \ldots, x_T^{(k-1)}\right)$.

Remark 2 *A deterministic version of this algorithm consists of selecting* $x_t^{(k)} = \arg\max_{x_t} p\left(x_t | \mathbf{y}_{1:T}, \mathbf{x}_{-t}^{(k)}\right)$.

[2] If $B(i) B^{\mathrm{T}}(i) \neq 0$, then the JMLS can be transformed to a new system where the noise covariance matrix is positive definite.

5.1. Implementation issues

It is necessary to sample from $\overline{p}^{\gamma_k}\left(x_t | \mathbf{y}_{1:T}, \mathbf{x}_{-t}^{(k)}\right)$ for any t. A direct evaluation of these distributions has a complexity $O(T^2)$. We develop here an algorithm whose complexity is $O(T)$. Conditional upon $\mathbf{x}_{0:T}$, the system (1)-(2) is a standard hidden Markov model (HMM) [10].

Lemma 6 *For any $t = 2, ..., T-2$*

$$\begin{array}{l} p(\mathbf{y}_{1:T}, \mathbf{x}_{0:T}) = p(\mathbf{y}_{1:t-1}, \mathbf{x}_{0:t-1}) p(y_t, x_t | \mathbf{y}_{1:t-1}, \mathbf{x}_{0:t-1}) \times \\ \sum_{i=1}^{s} p(\mathbf{y}_{t+1:T}, \mathbf{x}_{t+1:T} | x_t, r_t = i) p(r_t = i | \mathbf{y}_{1:t}, \mathbf{x}_{0:t}) \\ p(\mathbf{y}_{t:T}, \mathbf{x}_{t:T} | x_{t-1}, r_{t-1} = i) = \sum_{j=1}^{s} p_{ij} \times \\ p(\mathbf{y}_{t+1:T}, \mathbf{x}_{t+1:T} | y_t, x_t, r_t = j) p(y_t, x_t | r_t = j, x_{t-1}) \end{array} \quad (6)$$

Proposition 7 *For any $t = 2, ..., T-2$ we have*

$$x_t | (\mathbf{y}_{1:T}, \mathbf{x}_{-T}) \sim \sum_{i=1}^{s} \sum_{j=1}^{s} \alpha_t(i, j) \mathcal{N}(m_t(i, j), P_t(i, j)) \quad (7)$$

where

$$\begin{array}{l} P_t^{-1}(i, j) = \left(B(i) B^{\mathrm{T}}(i)\right)^{-1} + C^{\mathrm{T}}(i) \left(D(i) D^{\mathrm{T}}(i)\right)^{-1} C(i) \\ + A^{\mathrm{T}}(j) \left(B(j) B^{\mathrm{T}}(j)\right)^{-1} A(j) \\ m_t(i, j) = P_t(i, j) \left[\left(B(i) B^{\mathrm{T}}(i)\right)^{-1} (A(i) x_{t-1} + F(i) u_t) \right. \\ + C^{\mathrm{T}}(i) \left(D(i) D^{\mathrm{T}}(i)\right)^{-1} (y_t - G(i) u_t) \\ \left. + A^{\mathrm{T}}(j) \left(B(j) B^{\mathrm{T}}(j)\right)^{-1} (x_{t+1} - F(j) u_{t+1})\right] \end{array} \quad (8)$$

$\alpha_t(i, j) = \left[\sum_{i,j} \widetilde{\alpha}_t(i, j)\right]^{-1} \widetilde{\alpha}_t(i, j)$ *where* $\widetilde{\alpha}_t(i, j)$ *is given by*

$$\begin{array}{l} \widetilde{\alpha}_t(i, j) = p(\mathbf{y}_{t+2:T}, \mathbf{x}_{t+2:T} | y_{t+1}, x_{t+1}, r_{t+1} = j) p_{ij} \\ \times p(r_t = i | \mathbf{y}_{1:t-1}, \mathbf{x}_{0:t-1}) N^{-1}(m_t(i, j), P_t(i, j)) \\ \times N\left(y_t - G(i) u_t, D(i) D^{\mathrm{T}}(i)\right) \\ \times N\left(A(i) x_{t-1} + F(i) u_t, B(i) B^{\mathrm{T}}(i)\right) \\ \times N\left(x_{t+1} - F(j) u_{t+1}, B(i) B^{\mathrm{T}}(i)\right) \\ \times N\left(y_{t+1} - C(j) x_{t+1} - G(j) u_{t+1}, D(j) D^{\mathrm{T}}(j)\right) \end{array} \quad (9)$$

To sample from $\overline{p}^{\gamma_k}(x_t | \mathbf{y}_{1:T}, \mathbf{x}_{-T})$ for $t = 1, \ldots, T$, the algorithm proceeds as follows at iteration k:

Backward-Forward procedure

For $t = T, \ldots, 1$ for any $r_{t-1} = 1, \ldots, s$ compute and store $p(\mathbf{y}_{t:T}, \mathbf{x}_{t:T} | y_{t-1}, x_{t-1}, r_{t-1})$ using the backward recursion of HMM models [10], see (6).

For $t = 1, \ldots, T$

- For $i = 1, \ldots, s$, run one step-ahead the HMM one-step ahead predictor to obtain $p(r_t = i | \mathbf{y}_{1:t-1}, \mathbf{x}_{0:t-1})$, compute $p\left(x_t | \mathbf{y}_{1:T}, \mathbf{x}_{-t}^{(k)}\right)$ using (7) to (9).
- Sample $x_t^{(k)} \sim \overline{p}^{\gamma_k}\left(x_t | \mathbf{y}_{1:T}, \mathbf{x}_{-t}^{(k)}\right)$.

5.2. Convergence issues

The deterministic version of this algorithm converges towards a stationary point of $p(\mathbf{x}_{0:T} | \mathbf{y}_{1:T})$. Convergence of the stochastic annealing process in this continuous unbounded space framework has not yet been proved and is under study.

6. APPLICATION TO DIGITAL COMMUNICATIONS

Code Division Multiple Access (CDMA) provides a means of separating the signals of multiple users transmitting simultaneously and occupying the same bandwidth. The system performance is greatly enhanced if the receiver employs some means of suppressing narrowband interference prior to signal "despreading" [8]. A widely used model for the sampled received signal y_t consists of the spread spectrum signal r_t from N users, the narrowband interference i_t and observation noise w_t, the interfering signal being obtained by passing white noise through an all pole filter, with both poles at $z = 0.99$ that is

$$y_t = r_t + i_t + \sigma_w w_t \quad (10)$$

$$i_t = 1.98 i_{t-1} - 0.980 i_{t-2} + \sigma_e e_t \quad (11)$$

where $w_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ and $e_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ are mutually independent. The power of the received spread sprectrum signal for each user was held constant with amplitudes ± 1, randomly selected and r_t was binomially distributed. The CDMA spread spectrum model (10) and (11) can be easily re-expressed as the JMLS of (1) and (2), for the state vector $x_t = \begin{pmatrix} i_t & i_{t-1} \end{pmatrix}^{\mathrm{T}}$ and $u_t = 1$ for all t. In our simulation studies we considered a single user, with $\sigma_e = 0.03$ and we applied the five algorithms (conditional mean, D(deterministic)/S(stochastic) MMAP of x_t and r_t) for increasing observation noise. We present the Bit Error Rate (BER) for these algorithms. To evaluate the BER from $\mathbb{E}[\, r_t | \, \mathbf{y}_{1:T}]$, we set $\widehat{r}_t = 1$ if $\mathbb{E}[\, r_t | \, \mathbf{y}_{1:T}] > 0$ and $\widehat{r}_t = -1$ otherwise. To evaluate the BER from $\mathbf{x}_{0:T}^{MMAP} \triangleq \arg\max p(\, \mathbf{x}_{0:T} | \, \mathbf{y}_{1:T})$, we substract $\mathbf{x}_{0:T}^{MMAP}$ from the observations then set $\widehat{r}_t = 1$ if the residual signal is positive and $\widehat{r}_t = -1$ otherwise. The algorithms were run on 400 points, and averaged over 100 independent runs. In all cases, the algorithms were initialized randomly. In the example, we discard the first $N_0 = 20$ samples simulated by MCMC algorithm to estimate $\mathbb{E}[\, r_t | \, \mathbf{y}_{1:T}]$[11]. Then taking into account the following $N = 50$ iterations of the MCMC algorithm has appear to be sufficient, the estimates being stabilized. The deterministic optimization algorithms are iterated until convergence. Convergence occurs after no more than 5 iterations in all experiments. To use the stochastic optimization algorithms, we implement a linear cooling schedule $\gamma_k = \alpha k + \beta$ as it is usually done in practice for a number $N = 50$ of iterations with $\gamma_1 = 1$ and $\gamma_N = 10$. Then the final iterations are performed using the deterministic algorithms. The results are displayed in the following table. The stochastic algorithms appear of course much less sensitive to initialization than their deterministic counterparts. They are of great interest at low signal to noise ratio.

σ_w	$\mathbb{E}[\, \mathbf{r}_{1:T} \, \vert \, \mathbf{y}_{1:T}]$	D/S MMAP $\mathbf{r}_{1:T}$	D/S MMAP $\mathbf{x}_{0:T}$
0.5	4.02	4.61/4.37	5.36/4.82
0.6	5.67	6.82/6.01	7.37/6.39
0.7	9.27	12.23/9.44	13.68/9.85
0.8	12.06	17.32/12.32	19.05/13.01
0.9	15.68	22.11/15.94	24.03/16.06
1.0	18.42	27.32/19.01	30.32/21.08

7. DISCUSSION

The deterministic strategy to obtain the MMAP estimate of $\mathbf{r}_{1:T}$ has been developed in [9] but the proposed popular SMLR algorithm has a complexity $O(T^2)$. In the case of MA models [1] and in a state-space framework [3], algorithms of complexity $O(T)$ have been proposed. However, the algorithm in [3] is based on an approximate initialization of a backward recursion and assumes that $A(i)$ is regular for any i. Our algorithm has a similar complexity but it does not rely on any approximation and makes no assumption on $A(i)$. Furthermore, its range of applicability is wider than the algorithms presented in [5], [8] as the assumption $B(i)\, B^{\mathrm{T}}(i) > 0$ is not necessary. In the case where r_t is an independent sequence, the stochastic version of this algorithm is ensured to have a lower maximum correlation than the algorithm described in [5] according to [7, Th. 5.1]. The deterministic version is ensured to have a better asymptotic convergence rate than the EM algorithm in [8]. Indeed it is a simple coordinate ascent method that limits the amount of missing data [6]. The same remark holds for the deterministic version of the proposed algorithm to maximize $p(\, \mathbf{x}_{0:T} | \, \mathbf{y}_{1:T})$ compared to the EM algorithm presented in [8]. To sum up, the algorithms proposed here are of great interest as they have a computational complexity similar to the most recent algorithms but better theoretical and practical properties.

8. REFERENCES

[1] O. Cappé, A. Doucet, É. Moulines and M. Lavielle, "Simulation-Based Methods for Blind Maximum-Likelihood Filter Identification", to appear *Signal Processing*, 1998.

[2] C.K. Carter and R. Kohn, "On Gibbs Sampling for State Space Models", *Biometrika*, vol. 81, no. 3, pp. 541-553, 1994.

[3] C.K. Carter and R. Kohn, "Markov Chain Monte Carlo Methods in Conditionally Gaussian State Space Models", *Biometrika*, vol. 83, pp. 589-601, 1996.

[4] F. Champagnat, Y. Goussard and J. Idier, "Unsupervised Deconvolution of Sparse Spike Trains using Stochastic Approximation", *IEEE Trans. Signal Processing*, vol. 44, pp. 2988-2998, 1996.

[5] A. Doucet, A.S.L. Logothetis and V. Krishnamurthy, "State Estimation of Jump Markov Linear Systems via Stochastic Sampling Algorithms", in *Proc. IEEE Conf. CDC'98*.

[6] J.A. Fessler and A.O. Hero, "Space-alternating Generalized Expectation-Maximization Algorithm", *IEEE Trans. Signal Processing*, vol. 42, no. 10, pp. 2664-2677, 1994.

[7] J.S. Liu, W.H. Wong and A. Kong, "Covariance Structure of the Gibbs Sampler with Applications to the Comparisons of Estimators and Augmentation Schemes", *Biometrika*, vol. 81, pp. 27-40, 1994.

[8] A.S.L. Logothetis and V. Krishnamurthy, "A Bayesian EM Framework for Estimating Jump Markov Linear Systems", to appear *IEEE Trans. Signal Processing*, 1998.

[9] J.M. Mendel, *Maximum-Likelihood Deconvolution*, New York : Springer-Verlag, 1990.

[10] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proc. IEEE*, vol. 77, no. 2, pp. 257-285, 1989.

[11] L. Tierney, "Markov Chains for Exploring Posterior Distributions", *Ann. Stat.*, vol. 22, pp. 1701-1762, 1994.

[12] J.K. Tugnait, "Adaptive Estimation and Identification for Discrete Systems with Markov Jump Parameters", *IEEE Trans. Automatic Control*, vol. 25, pp. 1054-1065, 1982.