



**HAL**  
open science

# Exact Bayesian estimation in constrained Triplet Markov Chains

Yohan Petetin, François Desbouvries

► **To cite this version:**

Yohan Petetin, François Desbouvries. Exact Bayesian estimation in constrained Triplet Markov Chains. 2014 MLSP : IEEE International Workshop on Machine Learning for Signal Processing, Sep 2014, Reims, France. 10.1109/MLSP.2014.6958847 . hal-01262438

**HAL Id: hal-01262438**

**<https://hal.science/hal-01262438>**

Submitted on 2 Apr 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# EXACT BAYESIAN ESTIMATION IN CONSTRAINED TRIPLET MARKOV CHAINS

*Yohan Petetin*

CEA, LIST  
Gif-sur-Yvette, 91191 CEDEX, France  
yohan.petetin@cea.fr

*François Desbouvries*

Telecom SudParis, CITI Department  
Evry, 91000 France  
francois.desbouvries@it-sudparis.eu

## ABSTRACT

The Jump Markov state-space system (JMSS) is a well known model for representing dynamical models with jumps. However inference in a JMSS model is NP-hard, even in the conditionally linear and Gaussian case. Suboptimal solutions include Sequential Monte Carlo (SMC) and Interacting Multiple Models (IMM) methods. In this paper, we build a constrained Triplet Markov Chain (TMC) model which is close to the given JMSS model, and in which moments of interest can be computed exactly (without resorting to numerical nor Monte Carlo approximations) and at a computational cost which is linear in the number of observations. Additionally, a side advantage of our technique is that it can be used easily in a partially known model context.

**Index Terms**— Jump Markov state-space systems, Triplet Markov chains, Bayesian estimation, Expectation Maximization

## 1. INTRODUCTION

Dynamical systems with jumps are ubiquitous in signal processing. They are usually modeled as follows. Let  $\{\mathbf{y}_k\}_{k \geq 0} \in \mathbf{R}^p$  be the sequence of observations and  $\{\mathbf{x}_k\}_{k \geq 0} \in \mathbf{R}^m$  the sequence of hidden states. In many practical problems we are given the transition probability density function (pdf)  $f_{i|i-1}(\mathbf{x}_i|\mathbf{x}_{i-1}, r_i)$ , which locally models the evolution of the hidden state; as well as the likelihood pdfs  $g_i(\mathbf{y}_i|\mathbf{x}_i, r_i)$ , which describes how observation  $\mathbf{y}_i$  is related to state  $\mathbf{x}_i$ . Both pdfs depend on the realization of a third discrete variable  $r_i$  (with  $r_i \in \{1, \dots, K\}$ ) which models the so-called "jumps" or regime switchings.

It remains to model the joint pdf of variables  $(\mathbf{x}_{0:k}, \mathbf{y}_{0:k}, \mathbf{r}_{0:k})$ , in which  $\mathbf{x}_{0:k}$ , say, denotes  $\{\mathbf{x}_i\}_{i=0}^k$ . A popular model is the

so-called JMSS [1] [2] [3]:

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

In this model both the discrete process  $\{r_k\}_{k \geq 0}$  and the joint process  $\{(r_k, \mathbf{x}_k)\}_{k \geq 0}$  are markovian; given  $\mathbf{r}_{0:k}$ ,  $(\mathbf{x}_k, \mathbf{y}_k)$  is a Hidden Markov chain (HMC) with transition pdfs  $f_{i|i-1}(\mathbf{x}_i|\mathbf{x}_{i-1}, r_i)$  and likelihood pdfs  $g_i(\mathbf{y}_i|\mathbf{x}_i, r_i)$ ; and

$$p^1(\mathbf{x}_i|\mathbf{x}_{i-1}, r_i) = f_{i|i-1}(\mathbf{x}_i|\mathbf{x}_{i-1}, r_i), \quad (2)$$

$$p^1(\mathbf{y}_i|\mathbf{x}_i, r_i) = g_i(\mathbf{y}_i|\mathbf{x}_i, r_i). \quad (3)$$

In this paper we focus on the recursive computation of the first and second order moments of  $p(\mathbf{x}_k|\mathbf{y}_{0:k})$ ,

$$\Phi_k = \mathbb{E}(f(\mathbf{x}_k)|\mathbf{y}_{0:k}) = \int f(\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{0:k})d\mathbf{x}_k, \quad (4)$$

where  $f(\mathbf{x}) = \mathbf{x}$  or  $f(\mathbf{x}) = \mathbf{x}\mathbf{x}^T$ , in a partially known JMSS model, i.e. in model (1) where  $f_{i|i-1}(\mathbf{x}_i|\mathbf{x}_{i-1}, r_i)$  and  $g_i(\mathbf{y}_i|\mathbf{x}_i, r_i)$  are known, but not  $p^1(r_i|r_{i-1})$ .

Let us first assume that the model is completely known. Even if model (1) is a simple extension of the classical HMC model, inference in a JMSS is a difficult problem. The exact computation of  $\Phi_k$  is impossible in general, and is still NP-hard in the case where

$$f_{i|i-1}(\mathbf{x}_i|\mathbf{x}_{i-1}, r_i) = \mathcal{N}(\mathbf{x}_i; \mathbf{F}_i(r_i)\mathbf{x}_{i-1}; \mathbf{Q}_i(r_i)), \quad (5)$$

$$g_i(\mathbf{y}_i|\mathbf{x}_i, r_i) = \mathcal{N}(\mathbf{y}_i; \mathbf{H}_i(r_i)\mathbf{x}_i; \mathbf{R}_i(r_i)) \quad (6)$$

( $\mathcal{N}(\mathbf{x}; \mathbf{m}; \mathbf{P})$  denotes the Gaussian pdf with mean  $\mathbf{m}$  and covariance matrix  $\mathbf{P}$  taken at point  $\mathbf{x}$ ). So computing  $\Phi_k$  requires suboptimal approximation techniques, even in the linear and Gaussian case. Available methods include the IMM

[4] [5] [6] which runs several Kalman filters (KF), the outputs of which are combined according to the parameters of the model and to the available observations. On the other hand, SMC techniques are powerful methods to approach  $\Phi_k$  [7] [8] [9] [10]. In this last set of methods, a set of weighted random samples  $\{\mathbf{r}_{0:k}^i, w_k^i\}_{i=1}^N$  approximates  $p^1(\mathbf{r}_{0:k}|\mathbf{y}_{0:k})$ , while given  $\mathbf{r}_{0:k}$ ,  $p^1(\mathbf{x}_{0:k}|\mathbf{r}_{0:k}, \mathbf{y}_{0:k})$  is a Gaussian pdf computable via KF, which leads to the following approximation of  $p^1(\mathbf{x}_{0:k}|\mathbf{y}_{0:k})$ :

$$p^1(\mathbf{x}_{0:k}|\mathbf{y}_{0:k}) \approx \sum_{i=1}^N w_k(\mathbf{r}_{0:k}^i) \mathcal{N}(\mathbf{x}_{0:k}; \mathbf{m}_k(\mathbf{r}_{0:k}^i); \mathbf{P}_k(\mathbf{r}_{0:k}^i)). \quad (7)$$

Let us finally briefly discuss the case where  $p^1(r_k|r_{k-1})$  is not available. In that case the Expectation-Maximization (EM) algorithm can only be implemented approximately (as is done in [11]) because this algorithm relies on the computation of probabilities  $p^1(r_{i-1}, r_i|\mathbf{y}_{0:k})$  which, again, is NP-hard in a JMSS model.

The contents of this paper is as follows. We do not focus on the current approximations in the JMSS (1), (5)-(6). We rather propose a class of statistical models  $p^2(\mathbf{x}_{0:k}, \mathbf{y}_{0:k}, \mathbf{r}_{0:k})$  which are "close enough" to  $p^1(\mathbf{x}_{0:k}, \mathbf{y}_{0:k}, \mathbf{r}_{0:k})$ , and which satisfies the following properties:

- i) the unknown transition probability matrix  $p^2(r_k|r_{k-1})$  can be estimated via the EM algorithm;
- ii) once  $p^2(r_k|r_{k-1})$  is known or estimated,  $\Phi_k$  in (4) can be computed exactly (i.e., without resorting to any numerical nor Monte Carlo approximations), and at a computational cost which is linear in the number of observations.

By "close enough", we indeed mean that our model  $p^2$  should satisfy:

- iii)  $p^2(\mathbf{r}_{0:k}) = p^1(\mathbf{r}_{0:k})$ ;
- iv)  $p^2(\mathbf{x}_k|\mathbf{x}_{k-1}, r_k)$  and  $p^2(\mathbf{y}_k|\mathbf{x}_k, r_k)$  respectively coincide with pdfs  $f_{i|i-1}(\mathbf{x}_i|\mathbf{x}_{i-1}, r_i)$  in (5) and  $g_i(\mathbf{y}_i|\mathbf{x}_i, r_i)$  in (6);
- v)  $p^2(\cdot)$  minimizes the Kullback-Leibler divergence (KLD) with the JMSS root model  $p^1(\cdot)$ .

Let us now briefly describe our methodology. Our statistical models  $p^2(\cdot)$  are built from TMC [12], which are models which take into account regime switchings and which generalize the JMSS model (1). It has been shown in some recent contributions that the exact and fast computation of  $\Phi_k$  is possible in some TMC models [13] [14]. In this paper, we exploit these particular TMC in order to build models which satisfy constraints i) to v) above.

The rest of this paper is organized as follows. In section 2 we briefly recall TMC models and we particularly focus

on conditionnally linear and Gaussian ones. In section 3 we assume that the model is completely known, and we build a class of constrained TMC models  $p^2(\cdot)$  which are close to the JMSS model  $p^1(\cdot)$ , and in which  $\Phi_k$  can be computed exactly and efficiently. In section 4 we next assume that the model is partially known, and we address parameter estimation via the EM algorithm which, by contrast with model  $p^1(\cdot)$ , can be directly implemented in model  $p^2(\cdot)$ . Section 5 is devoted to simulations, and we end the paper with a conclusion.

## 2. TRIPLET MARKOV CHAINS

### 2.1. JMSS as a particular TMC

As we already recalled, in the JMSS model (1),  $\{r_k\}_{k \geq 0}$  is a Markov chain (MC), and given  $\mathbf{r}_{0:k}$ ,  $(\mathbf{x}_k, \mathbf{y}_k)$  is an HMC, i.e. given  $\mathbf{r}_{0:k}$ ,  $\{\mathbf{x}_k\}_{k \geq 0}$  is an MC and observations  $\{\mathbf{y}_i\}$  are conditionnally independent with  $p^1(\mathbf{y}_i|\mathbf{x}_{0:k}, \mathbf{r}_{0:k}) = p^1(\mathbf{y}_i|\mathbf{x}_i, r_i) = g_i(\mathbf{y}_i|\mathbf{x}_i, r_i)$ . On the other hand, let  $\mathbf{t}_k = (\mathbf{x}_k, \mathbf{y}_k, r_k)$ . From (1) we see that the JMSS model also satisfies

$$\begin{aligned} p^1(\mathbf{t}_i|\mathbf{t}_{0:i-1}) &= p^1(r_i|r_{i-1})f_{i|i-1}(\mathbf{x}_i|\mathbf{x}_{i-1}, r_i)g_i(\mathbf{y}_i|\mathbf{x}_i, r_i) \\ &= p^1(\mathbf{t}_i|\mathbf{t}_{i-1}) \end{aligned} \quad (8)$$

so the *triplet* chain  $\{\mathbf{t}_k = (\mathbf{x}_k, \mathbf{y}_k, r_k)\}_{k \geq 0}$  is an MC, and as such the JMSS model is one particular TMC; by TMC, we mean any process  $\{\mathbf{t}_k\}_{k \geq 0}$  such that  $\mathbf{t}_k$  is an MC, i.e. such that the joint pdf  $p^2(\mathbf{t}_{0:k})$  factorizes as

$$p^2(\mathbf{t}_{0:k}) = p^2(\mathbf{t}_0) \prod_{i=1}^k p(\mathbf{t}_i|\mathbf{t}_{i-1}). \quad (9)$$

### 2.2. Filtering in linear and Gaussian TMC

In this section we focus on the subclass of TMC models (9) in which  $\{r_k\}_{k \geq 0}$  is an MC and the transition pdf of  $(\mathbf{x}_k, \mathbf{y}_k)$  given  $(\mathbf{x}_{k-1}, \mathbf{y}_{k-1}, \mathbf{r}_{k-1:k})$  is a Gaussian, i.e. on those TMC models which satisfy

$$p^2(\mathbf{t}_k|\mathbf{t}_{k-1}) = p^2(r_k|r_{k-1})p^2(\mathbf{x}_k, \mathbf{y}_k|\mathbf{t}_{k-1}, r_k), \quad (10)$$

where

$$p^2(\mathbf{x}_k, \mathbf{y}_k|\mathbf{t}_{k-1}, r_k) = \mathcal{N}(\mathbf{x}_k, \mathbf{y}_k; \mathbf{B}_k(\mathbf{r}_{k-1:k}) \begin{bmatrix} \mathbf{x}_{k-1} \\ \mathbf{y}_{k-1} \end{bmatrix}; \Sigma_k(\mathbf{r}_{k-1:k})), \quad (11)$$

$$\mathbf{B}_k(\mathbf{r}_{k-1:k}) = \begin{bmatrix} \mathbf{F}_k^1(\mathbf{r}_{k-1:k}) & \mathbf{F}_k^2(\mathbf{r}_{k-1:k}) \\ \mathbf{H}_k^1(\mathbf{r}_{k-1:k}) & \mathbf{H}_k^2(\mathbf{r}_{k-1:k}) \end{bmatrix}, \quad (12)$$

$$\Sigma_k(\mathbf{r}_{k-1:k}) = \begin{bmatrix} \Sigma_k^{11}(\mathbf{r}_{k-1:k}) & \Sigma_k^{21}(\mathbf{r}_{k-1:k})^T \\ \Sigma_k^{21}(\mathbf{r}_{k-1:k}) & \Sigma_k^{22}(\mathbf{r}_{k-1:k}) \end{bmatrix}. \quad (13)$$

The linear and Gaussian JMSS model (1), (5)-(6) is of course one such model; it corresponds to the particular setting  $\mathbf{F}_k^1(\mathbf{r}_{k-1:k}) = \mathbf{F}_k(r_k)$ ,  $\mathbf{F}_k^2(\mathbf{r}_{k-1:k}) = \mathbf{0}$ ,  $\mathbf{H}_k^1(\mathbf{r}_{k-1:k}) =$

$\mathbf{H}_k(r_k)\mathbf{F}_k(r_k)$ ,  $\mathbf{H}_k^2(\mathbf{r}_{k-1:k}) = \mathbf{0}$ ,  $\Sigma_k^{11}(\mathbf{r}_{k-1:k}) = \mathbf{Q}_k(r_k)$ ,  $\Sigma_k^{21}(\mathbf{r}_{k-1:k}) = \mathbf{H}_k(r_k)\mathbf{Q}_k(r_k)$  and  $\Sigma_k^{22}(\mathbf{r}_{k-1:k}) = \mathbf{R}_k(r_k) + \mathbf{H}_k(r_k)\mathbf{Q}_k(r_k)\mathbf{H}_k(r_k)^T$ .

We now address the computation of moment  $\Phi_k$  in (4) in the class of models (10)-(13). Since the linear and Gaussian JMSS model is one element of this class, it is clear from section 1 above that the exact computation of  $\Phi_k$  will not be possible for all models of this class. However, it has been proved [13] [15] [14] that in the particular case where

$$\mathbf{H}_k^1(\mathbf{r}_{k-1:k}) = \mathbf{0}, \quad (14)$$

$\Phi_k$  can be computed exactly at a computational cost which is linear in the number of observations (the formulas for computing  $\Phi_k$  are not recalled here for want of space, but can be found in [13] and [14]).

### 3. EXACT BAYESIAN ESTIMATION IN CONSTRAINED TMC MODELS

In all this section we assume that  $p^2(r_k|r_{k-1})$  is known (the estimation problem of  $p^2(r_k|r_{k-1})$  will be addressed in §4). We use the result recalled in section 2.2 in order to build a class of statistical models which all share the transition and likelihood pdfs (5)-(6) of the linear and Gaussian JMSS, and in which one can compute  $\Phi_k$  exactly.

#### 3.1. Constrained TMC models

We begin with building a set of TMC models  $p^2(\cdot)$  such that conditions iii) and iv) in Section 1 hold. We have the following result (the proof of Propositions 1 to 3 below is omitted due to lack of space).

**Proposition 1** *Let  $p^1$  be given by (1), (5)-(6). The linear and Gaussian TMC model (9)-(13) defined by  $p^2(r_0) = p^1(r_0)$ ,  $p^2(r_k|r_{k-1}) = p^1(r_k|r_{k-1})$ , and*

$$\mathbf{B}_k(\mathbf{r}_{k-1:k}) = \begin{bmatrix} \mathbf{F}_k(r_k) - \mathbf{F}_k^2(\mathbf{r}_{k-1:k})\mathbf{H}_{k-1}(r_{k-1}) & \mathbf{F}_k^2(\mathbf{r}_{k-1:k}) \\ \mathbf{H}_k(r_k)\mathbf{F}_k(r_k) - \mathbf{H}_k^2(\mathbf{r}_{k-1:k})\mathbf{H}_{k-1}(r_{k-1}) & \mathbf{H}_k^2(\mathbf{r}_{k-1:k}) \end{bmatrix}, \quad (15)$$

$$\Sigma_k(\mathbf{r}_{k-1:k}) = \begin{bmatrix} \Sigma_k^{11}(\mathbf{r}_{k-1:k}) & \Sigma_k^{21}(\mathbf{r}_{k-1:k})^T \\ \Sigma_k^{21}(\mathbf{r}_{k-1:k}) & \Sigma_k^{22}(\mathbf{r}_{k-1:k}) \end{bmatrix}, \quad (16)$$

$$\Sigma_k^{11}(\mathbf{r}_{k-1:k}) = \mathbf{Q}_k(r_k) - \mathbf{F}_k^2(\mathbf{r}_{k-1:k})\mathbf{R}_{k-1}(r_{k-1})\mathbf{F}_k^2(\mathbf{r}_{k-1:k})^T \quad (17)$$

$$\Sigma_k^{21}(\mathbf{r}_{k-1:k}) = \mathbf{H}_k(r_k)\mathbf{Q}_k(r_k) - \mathbf{H}_k^2(\mathbf{r}_{k-1:k})\mathbf{R}_{k-1}(r_{k-1})\mathbf{F}_k^2(\mathbf{r}_{k-1:k})^T, \quad (18)$$

$$\Sigma_k^{22}(\mathbf{r}_{k-1:k}) = \mathbf{R}_k(r_k) - \mathbf{H}_k^2(\mathbf{r}_{k-1:k})\mathbf{R}_{k-1}(r_{k-1})\mathbf{H}_k^2(\mathbf{r}_{k-1:k})^T + \mathbf{H}_k(r_k)\mathbf{Q}_k(r_k)\mathbf{H}_k(r_k)^T, \quad (19)$$

and where parameters  $\mathbf{F}_k^2(\mathbf{r}_{k-1:k})$  and  $\mathbf{H}_k^2(\mathbf{r}_{k-1:k})$  can be arbitrarily chosen, provided  $\Sigma_k(\mathbf{r}_{k-1:k})$  is a positive definite covariance matrix for all  $k$ , satisfy the constraints

$$p^2(\mathbf{r}_{0:k}) = p^1(\mathbf{r}_{0:k}), \quad (20)$$

$$p^2(\mathbf{x}_k|\mathbf{x}_{k-1}, r_k) = f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}, r_k), \quad (21)$$

$$p^2(\mathbf{y}_k|\mathbf{x}_k, r_k) = g_k(\mathbf{y}_k|\mathbf{x}_k, r_k). \quad (22)$$

**Remark 1** *In the particular case where  $\mathbf{F}_k^2(\mathbf{r}_{k-1:k}) = \mathbf{0}$  and  $\mathbf{H}_k^2(\mathbf{r}_{k-1:k}) = \mathbf{0}$  the model reduces to the linear and Gaussian JMSS (1), (5)-(6).*

#### 3.2. Exact filtering in constrained TMC models

We now focus on the computation of  $\Phi_k$  in our constrained TMC models parametrized by  $\mathbf{F}_k^2(\mathbf{r}_{k-1:k})$  and  $\mathbf{H}_k^2(\mathbf{r}_{k-1:k})$ . The result of [13], recalled in §2.2, enables us to propose models in which  $\Phi_k$  can be computed efficiently:

**Proposition 2** *Let  $p^2(\cdot)$  be a constrained linear and Gaussian TMC model described Proposition 1. If  $\mathbf{H}_k^2(\mathbf{r}_{k-1:k})$  satisfies*

$$\mathbf{H}_k(r_k)\mathbf{F}_k(r_k) - \mathbf{H}_k^2(\mathbf{r}_{k-1:k})\mathbf{H}_{k-1}(r_{k-1}) = \mathbf{0} \quad (23)$$

then  $p^2(r_k|\mathbf{y}_{0:k})$  and  $E(f(\mathbf{x}_k)|\mathbf{y}_{0:k}, r_k)$  can be computed exactly and at a computational cost which is linear in the number of observations. Finally  $\Phi_k$  is computed as

$$\Phi_k = \sum_{r_k} p(r_k|\mathbf{y}_{0:k})E(f(\mathbf{x}_k)|\mathbf{y}_{0:k}, r_k). \quad (24)$$

#### 3.3. KLD minimization

Note that the exact computation of  $\Phi_k$  relies on the value of  $\mathbf{H}_k^2(\mathbf{r}_{k-1:k})$  but not on that of  $\mathbf{F}_k^2(\mathbf{r}_{k-1:k})$ . So the set of constrained models in which  $\Phi_k$  can be computed exactly, described by Proposition 1 plus condition (23), remains parametrized by  $\mathbf{F}_k^2(\mathbf{r}_{k-1:k})$ ; among this set we now look for pdf  $p^2$  closest to  $p^1$  in KLD sense, i.e. we take into account condition v) of section 1. We have the following result.

**Proposition 3** *Let  $p^1(\cdot)$  be the linear and Gaussian JMSS model (1), (5)-(6), and let  $p^2(\cdot)$  be the class of models described by Proposition 1, in which condition (23) holds, and thus a set of models where  $\Phi_k$  can be computed exactly. Parameters  $\mathbf{F}_k^2(\mathbf{r}_{k-1:k})$  which minimize the KLD between  $p^2(\mathbf{t}_{0:k})$  and  $p^1(\mathbf{t}_{0:k})$  are given by*

$$\mathbf{F}_k^{2,\text{opt}}(\mathbf{r}_{k-1:k}) = \mathbf{Q}_k(r_k)\mathbf{H}_k(r_k)^T \times \left[ \mathbf{R}_k(r_k) + \mathbf{H}_k(r_k)\mathbf{Q}_k(r_k)\mathbf{H}_k(r_k)^T \right]^{-1} \mathbf{H}_k^2(\mathbf{r}_{k-1:k}). \quad (25)$$

#### 4. EM ALGORITHM IN A PARTIALLY KNOWN CONSTRAINED TMC MODEL

Until now, we assumed that the probability transitions  $p^2(r_k|r_{k-1}) = p^1(r_k|r_{k-1})$  of MC  $\mathbf{r}_{0:k}$  were known. This is not necessarily the case in practice, so in this section we address the estimation problem of the probability distribution  $p^2(r_k|r_{k-1})$ . It turns out that an advantage of the technique described in section 3 is that the EM algorithm [16] can easily be implemented in model  $p^2(\cdot)$ , which is not the case in model  $p^1(\cdot)$ , as we now recall.

So let us first assume the JMSS model (1), (5)-(6). The EM algorithm is based on the computation of expectation

$$E(\log(p^1(\mathbf{t}_{0:k})|\mathbf{y}_{0:k})) \quad (26)$$

and on its maximization w.r.t.  $p^1(r_k|r_{k-1})$ . Developing (26) we see that we need to compute

$$E(\log(p^1(r_i|r_{i-1})|\mathbf{y}_{0:k})) = \sum_{r_{i-1:i}} \log(p^1(r_i|r_{i-1}))p^1(\mathbf{r}_{i-1:i}|\mathbf{y}_{0:k}) \quad (27)$$

for all  $i$ ,  $1 \leq i \leq k$ . However, it is well known that in the linear and Gaussian JMSS (1), (5)-(6), the computation of  $p^1(r_{i-1}, r_i|\mathbf{y}_{0:k})$  is an NP-hard problem [7].

Here we propose to implement the EM algorithm, not in  $p^1$ , but in a model  $p^2$  described by propositions 1, 2 and 3, and so which is close to model  $p^1$ . As above, we need to compute the probabilities  $p^2(\mathbf{r}_{i-1:i}|\mathbf{y}_{0:k})$ ; however now (23) holds, and one can show easily that under this condition  $p^2(\mathbf{y}_k, r_k|\mathbf{t}_{k-1}) = p^2(\mathbf{y}_k, r_k|\mathbf{y}_{k-1}, r_{k-1})$ , and next that  $(\mathbf{y}_k, r_k)$  is an MC with transition

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

in which

$$p^2(\mathbf{y}_k|\mathbf{y}_{k-1}, \mathbf{r}_{k-1:k}) = \mathcal{N}(\mathbf{y}_k; \mathbf{H}_k^2(\mathbf{r}_{k-1:k})\mathbf{y}_{k-1}; \Sigma_k^{22}(\mathbf{r}_{k-1:k})),$$

where  $\mathbf{H}_k^2(\mathbf{r}_{k-1:k})$  satisfies (23) and  $\Sigma_k^{22}(\mathbf{r}_{k-1:k})$  is defined in (19). Consequently, the computation of  $p^2(r_{i-1}, r_i|\mathbf{y}_{0:k})$ , can be performed efficiently by the algorithm described in [14], which is an adaptation to pairwise Markov chains (PMC) (i.e., models in which  $\{(r_k, \mathbf{y}_k)\}_{k \geq 0}$  is an MC) of the Forward-Backward algorithm, and the maximization of  $E(\log(p^2(\mathbf{t}_{0:k})|\mathbf{y}_{0:k}))$  as a function of  $p^2(r_k|r_{k-1})$  can be performed efficiently by an adaptation to PMC of the Baum-Welch algorithm [17].

#### 5. SIMULATIONS

In this final section we validate our technique via simulations. We consider the linear and Gaussian JMSS (1), (5)-(6) where

$$\mathbf{F}_k(r) = \begin{bmatrix} 1 & \frac{\sin(\omega_r T)}{\omega_r} & 0 & -\frac{1-\cos(\omega_r T)}{\omega_r} \\ 0 & \cos(\omega_r T) & 0 & -\sin(\omega_r T) \\ 0 & \frac{1-\cos(\omega_r T)}{\omega_r} & 1 & \frac{\sin(\omega_r T)}{\omega_r} \\ 0 & \sin(\omega_r T) & 0 & \cos(\omega_r T) \end{bmatrix},$$

$$\mathbf{Q}_k(r) = \sigma_v^2(r) \begin{bmatrix} \frac{T^3}{3} & \frac{T^2}{2} & 0 & 0 \\ \frac{T^2}{2} & T & 0 & 0 \\ 0 & 0 & \frac{T^3}{3} & \frac{T^2}{2} \\ 0 & 0 & \frac{T^2}{2} & T \end{bmatrix},$$

$\mathbf{H}_k = \mathbf{I}_4$  and  $\mathbf{R}_k = \mathbf{I}_4$ . We set  $T = 2$ ,  $r_k \in \{1, 2, 3\}$  represents the behavior of a target: straight, left turn and right turn. So we set  $w_r \in \{0, 6\pi/180, -6\pi/180\}$  and  $\sigma_v(r) \in \{7, 10, 10\}$ . The true transition probabilities of MC  $\{r_k\}$  are defined by  $p^1(r_k|r_{k-1}) = 0.8$  if  $r_k = r_{k-1}$  and  $p^1(r_k|r_{k-1}) = 0.1$  if  $r_k \neq r_{k-1}$ , and we assume them unknown. So our goal is twofold :

1. we look for estimating  $p^1(r_k|r_{k-1})$  via the PMC Baum-Welch algorithm in our constrained TMC model  $p^2(\cdot)$  which satisfies (23) and (25);
2. we compute  $\Phi_k$  in model  $p^2(\cdot)$  and we compare our estimate with estimates based on IMM and particle filtering (PF).

##### 5.1. Parameter estimation step

We first generate a data set of length  $K = 500$  according to the linear and Gaussian JMSS model (1), (5)-(6), and we initialize the transition probabilities as  $p^2(r_k|r_{k-1}) = 1/3$ . We use 10 iterations of the PMC Baum-Welch algorithm. The estimated transition probability matrix  $\pi(r_{k-1}, r_k) = p^2(r_k|r_{k-1})$  is given by

$$\pi(r_{k-1}, r_k) = \begin{bmatrix} 0.7972 & 0.1012 & 0.1016 \\ 0.1031 & 0.8074 & 0.0895 \\ 0.1119 & 0.1030 & 0.7851 \end{bmatrix}. \quad (28)$$

**Remark 2** Matrix (28) is obtained from a unique set of data. If we average the transition probability matrix over  $P = 100$  sets of data we get

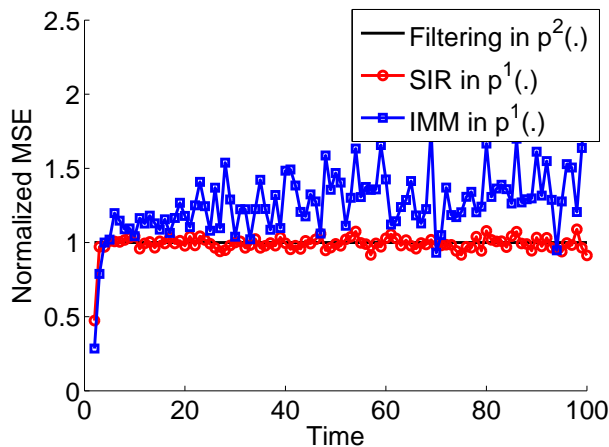
$$\pi^{\text{mean}}(r_{k-1}, r_k) = \begin{bmatrix} 0.7946 & 0.0995 & 0.1059 \\ 0.0951 & 0.8024 & 0.1025 \\ 0.1020 & 0.1059 & 0.7921 \end{bmatrix}.$$

##### 5.2. Computation of $\Phi_k$

Now, our model  $p^2(\cdot)$  is completely defined. We generate  $P = 200$  sets of data with length  $K = 100$ . For each trajectory  $p$ ,  $1 \leq p \leq 200$ , we compute the exact estimate

$\hat{x}_{k,p,\text{TMC}}$  in our model  $p^2(\cdot)$  (with transition probabilities (28) estimated at the previous step); an estimate  $\hat{x}_{k,p,\text{SIR}}$  based on the sampling importance resampling (SIR) algorithm with importance distribution  $p^1(r_k|r_{k-1})$  and with  $N = 150$  particles [7]; an estimate  $\hat{x}_{k,p,\text{IMM}}$  based on the IMM algorithm [4]; and the estimate  $\hat{x}_{k,p,\text{KF}}$  based on a KF which uses the true jumps and which is our benchmark solution. Note that the PF and the IMM algorithms use the true transition probabilities  $p^1(r_k|r_{k-1})$ . For each estimate, we compute the averaged mean square error (MSE)  $\text{MSE}(k) = \frac{1}{P} \sum_{p=1}^P (\hat{x}_{k,p,\cdot} - \hat{x}_{k,p,\text{KF}})^2$ .

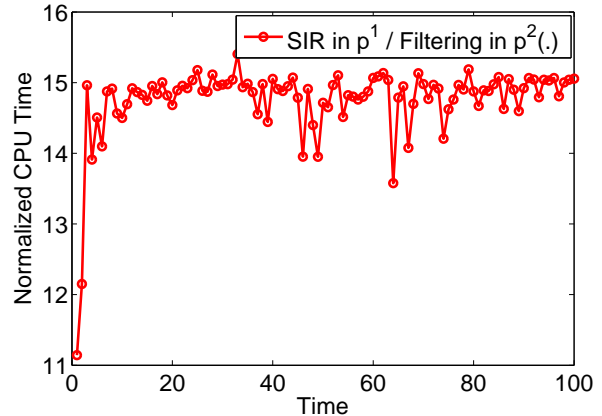
In Fig. 1, we have displayed the averaged MSEs, normalized w.r.t. our estimate  $\hat{x}_{k,p,\text{TMC}}$ . Note that our estimator, though computed in model  $p^2(\cdot)$  with estimated parameters  $p(r_k|r_{k-1})$ , performs similarly to the SIR based solution in model  $p^1(\cdot)$  with known parameters. However we do not use Monte Carlo samples; as a result, our algorithm is approximately fifteen times faster than the SIR based solution as can be seen from Fig. 2.



**Fig. 1.** Ratio of the exact filtering technique in model  $p^2(\cdot)$  (black line), the SIR based one in model  $p^1(\cdot)$  (red circles) and the IMM based one in model  $p^1(\cdot)$  (blue squares).

## 6. CONCLUSION

In this paper we proposed a new inference technique in a partially known JMSS system  $p^1(\cdot)$ . We built a set of constrained TMC models  $p^2(\cdot)$  which include  $p^1(\cdot)$ , and which all share the state transition and likelihood pdfs of  $p^1(\cdot)$ . Among this set we identified a subclass of models in which a conditional moment of interest can be computed exactly at a computational cost which is linear in the number of observations, and we identified within this set the model which minimizes the KLD to  $p^1(\cdot)$ . Finally we addressed the parameter estimation problem in the case where the pdf of the jumps is unknown. Simulations show that our inference technique with unknown pa-



**Fig. 2.** Ratio of the CPU time of the exact filtering technique in model  $p^2(\cdot)$  to that of the SIR based one in model  $p^1(\cdot)$ .

rameters performs similarly to an SMC based solution with known parameters but is fifteen times faster.

## 7. REFERENCES

- [1] G. Ackerson and K.S. Fu, "On state estimation in switching environments," *IEEE Transactions on Automatic Control*, vol. 4, pp. 429–434, February 1970.
- [2] H. Akashia and H. Kumamoto, "Random sampling approach to state estimation in switching environments," *Automatica*, vol. 13, no. 4, pp. 429–434, July 1977.
- [3] C-J. Kim, "Dynamic linear models with Markov switching," *J. of Econometrics*, vol. 60, pp. 1–22, 1994.
- [4] H. A. P. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Transactions on Automatic Control*, vol. 33, no. 8, pp. 780–783, 1988.
- [5] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting multiple model methods in target tracking: a survey," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 1, pp. 103–123, 1998.
- [6] X.R. Li and V.P. Jilkov, "Survey of maneuvering target tracking. Part V. Multiple-model methods," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 4, pp. 1255–1321, 2005.
- [7] A. Doucet, N. J. Gordon, and V. Krishnamurthy, "Particle filters for state estimation of jump Markov linear systems," *IEEE Transactions on Signal Processing*, vol. 49, no. 3, pp. 613–24, March 2001.
- [8] C. Andrieu, M. Davy, and A. Doucet, "Efficient particle filtering for jump Markov systems," *IEEE Transactions on Signal Processing*, vol. 51, pp. 1762–1770, 2002.

- [9] A. Doucet, S. J. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197–208, 2000.
- [10] T. Schön, F. Gustafsson, and P.-J. Nordlund, "Marginalized particle filters for mixed linear nonlinear state-space models," *IEEE Transactions on Signal Processing*, vol. 53, pp. 2279–2289, 2005.
- [11] V.P. Jilkov and X. R. Li, "Online bayesian estimation of transition probabilities for Markovian jump systems," *IEEE Transactions on Signal Processing*, vol. 52, pp. 1620–1630, 2004.
- [12] W. Pieczynski, "Chaînes de Markov triplet," *Comptes Rendus de l'Académie des Sciences - Mathématiques*, vol. 335, pp. 275–278, 2002, in French.
- [13] W. Pieczynski, "Exact filtering in conditionally Markov switching hidden linear models," *Comptes Rendus Mathématiques*, vol. 349, no. 9-10, pp. 587–590, May 2011.
- [14] W. Pieczynski, "Exact smoothing in hidden conditionally Markov switching linear models," *Communications in Statistics - Theory and Methods*, vol. 40, no. 16, pp. 2823–2829, may 2011.
- [15] 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.
- [16] 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)*, vol. 39, no. 1, pp. 1–38, 1977.
- [17] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Stat.*, vol. 37, pp. 1554–63, 1966.