

Rapid System Identification for Jump Markov Non-Linear Systems

André R. Braga, Carsten Fritsche, Marcelo G. S. Bruno and Fredrik Gustafsson

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-144023>

N.B.: When citing this work, cite the original publication.

Braga, A. R., Fritsche, C., Bruno, M. G. S., Gustafsson, F., (2017), Rapid System Identification for Jump Markov Non-Linear Systems, *Proc. 2017 IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, , 714-718.
<https://doi.org/10.1109/CAMSAP.2017.8313089>

Original publication available at:

<https://doi.org/10.1109/CAMSAP.2017.8313089>

Copyright: IEEE

<http://www.ieee.org/>

©2017 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.



Rapid System Identification for Jump Markov Non-Linear Systems

André R. Braga
Campus Quixadá
Federal University of Ceará
63902-580 Quixadá CE, Brazil

Carsten Fritsche, Fredrik Gustafsson
Department of Electrical Engineering
Linköping University
SE-581 83 Linköping, Sweden

Marcelo G.S. Bruno
Division of Electronics Engineering
Aeronautics Institute of Technology
12228-900 São José dos Campos SP, Brazil

Abstract—This work evaluates a previously introduced algorithm called Particle-Based Rapid Incremental Smoother within the framework of state inference and parameter identification in Jump Markov Non-Linear System. It is applied to the recursive form of two well-known Maximum Likelihood based algorithms who face the common challenge of online computation of smoothed additive functionals in order to accomplish the task of model parameter estimation. This work extends our previous contributions on identification of Markovian switching systems with the goal to reduce the computational complexity. A benchmark problem is used to illustrate the results.

I. INTRODUCTION

A lot of estimation tasks are confronted with a category of system models that can operate in multiple regimes. Usually, the change is governed by a Markov chain and, additionally to this discrete mode variable, there can also be continuous state-space variables leading to a so-called hybrid system. This type of dynamical system has attracted growing attention from the filtering and identification community in a large variety of applications, see e.g. [1]–[3]. Particularly non-linear problems lack an exact solution for the estimation of unknown discrete and continuous variables and it is one of the reasons recursive approaches using Sequential Monte Carlo (SMC) methods, namely Particle Filter (PF), emerged as the most successful methods because they are easy to implement and parallelize [4], [5].

Moreover, in most practical situations, the problem of system identification, i.e. the estimation of unknown model parameters, arises as an increase in difficulty. The most popular solving tools are based on the Maximum Likelihood (ML) concept since it has been studied for almost a century. Among famous algorithms, we can highlight the Expectation Maximization (EM) [6] and the Recursive Maximum Likelihood (RML), based on the gradient ascent optimization method [7]. However, when we are interested in their online implementation, the recursive computation of smoothed expectations for additive functionals is needed in order to obtain either the sufficient statistics for EM or the score function for RML. A natural and straight forward choice for a solution would be to use the PF, since it provides an approximation of the joint smoothing distribution by means of the particle trajectories and its corresponding weights [8]. But the effect known as degeneracy turns it to be unfeasible. This is due to the PF's resampling step, that leads to the depletion of state trajectories (for more details refer to [9], [10]) and a poor approximation of the smoothing density. Although it can be partially solved using fixed lag smoothing [8], [11], a better approach uses the

principle that, conditionally on the observations, the latent (hybrid) process still satisfies the Markov property when evolving backward in time [9]. This supported the development of the Forward Filter Backward Smoother (FFBSm) [12], where a backward pass was included to better approximate the joint smoothing distribution but at a cost of significantly increased computational complexity. Furthermore, due to the backward pass, the complexity also increases with the arrival of new measurements, making it impractical for online problems. To overcome this limitation, an online approach was proposed based on the forward smoothing approach [13]. One of the limitations was solved, but there is still a quadratic complexity in the number of particles. A good candidate for a way to cope with such large computational burden is the path-based solution [14], based on a coarse approximation of the backward kernel. This approach could indeed reduce the complexity but lead to some performance degradation as exposed in [15].

Our main interest lies in the so-called Particle-Based Rapid Incremental Smoother (PaRIS) algorithm, proposed and extensively evaluated in [16]–[18]. It uses a modified Forward Filter Backward Simulator (FFBSi) [19] (employing rejection-sampling) as a way of reducing complexity. This paper introduces a PaRIS based adaptation to recursively solve the hard problem of state estimation in Jump Markov Non-Linear System (JMNLs) under parameter uncertainty.

The remainder of this work is organized as follows. The modeling for JMNLs is described in Section II as well as the considered system identification algorithms. In Section III, the required computation for the online solution is described and the PaRIS version for models with switching regimes is also presented. The performance is assessed in Section IV while the final conclusions are drawn in Section V.

II. PROBLEM FORMULATION

Consider the hybrid discrete-time system model as follows

$$r_t \sim \Pi(r_t|r_{t-1}), \quad (1a)$$

$$x_t \sim f_{r_t}(x_t|x_{t-1}; \theta_{r_t}), \quad (1b)$$

$$y_t \sim g_{r_t}(y_t|x_t; \theta_{r_t}), \quad (1c)$$

where the discrete mode $r_t \in \{1, \dots, K\}$ and the continuous variable $x_t \in \mathbb{R}^{n_x}$ are both latent but indirectly observed through the measurements $y_t \in \mathbb{R}^{n_y}$. The mode change is ruled by a finite state Markov process (1a) and we assume it to define both the continuous state transition density and measurements likelihood in (1b) and (1c), respectively. They are parametrized

by the unknown vector θ_{r_t} and the transition kernel for the discrete state is defined as $\pi_{k\ell} = \mathbb{P}(\ell|k) = \mathbb{P}(r_t = \ell | r_{t-1} = k)$. For notational brevity, we also consider Π as the $K \times K$ Transition Probability Matrix (TPM), with each element being $\pi_{k\ell}$.

We are concerned with the task of jointly estimating recursively the latent states (x_t, r_t) and the model parameters $\theta = (\{\theta_k\}_{k=1}^K, \Pi)$ from the measurements $y_{1:t} = [y_1, \dots, y_t]^T$ available up to time t .

Following the frequentist philosophy, the ML approach considers the estimate $\hat{\theta}^{\text{ML}}$ as the outcome of a direct optimization on the log-likelihood function of the observed data, $\log p(y_{1:n}; \theta)$. According to [20], there are two main strategies for solving the problem: marginalization and data augmentation, which are briefly presented in Sections II-A and II-B, respectively.

A. Recursive Maximum Likelihood

For the recursive gradient ascent, at discrete time $n + 1$, the parameter estimate can be updated according to

$$\theta_{n+1} = \theta_n + \gamma_{n+1} \nabla_{\theta} \log p(y_n | y_{1:n-1}; \theta_{0:n}). \quad (2)$$

where the score of the predictive likelihood $p(y_n | y_{1:n-1}; \theta_{0:n})$ given *all* parameter estimates $\theta_{0:n}$ approximates the Fisher's identity [5] and can be expressed as

$$\begin{aligned} \nabla_{\theta} \log p(y_n | y_{1:n-1}; \theta_{0:n}) \\ = \nabla_{\theta} \log p(y_{1:n}; \theta_{0:n}) - \nabla_{\theta} \log p(y_{1:n-1}; \theta_{0:n-1}), \end{aligned} \quad (3)$$

and $\{\gamma_n\}_{n \geq 1}$ is a non-negative scalar that needs to be adjusted at each iteration to ensure, a minima, that the computed likelihood sequence is non-decreasing [21].

Based on the Fisher's identity [5] computed for JMNLS, the score can then be written as

$$\nabla_{\theta} \log p(y_{1:n}; \theta) = \sum_{k=1}^K \sum_{\ell=1}^K \mathcal{S}_{k\ell,n}^{(1)} + \sum_{k=1}^K \mathcal{S}_{k,n}^{(2)}, \quad (4)$$

where, $\mathcal{S}_{k\ell,n}^{(1)}$ and $\mathcal{S}_{k,n}^{(2)}$ are additive functionals computed as in [22].

B. Expectation Maximization

Usually more stable than gradient techniques [5], the EM algorithm is a popular and robust alternative for maximizing the log-likelihood. It approximates the joint log-likelihood by a function, $\mathcal{Q}(\theta, \theta')$, described as a projection onto space defined by available observations and a current estimate θ' of the likelihood maximizer [23],

$$\begin{aligned} \mathcal{Q}(\theta, \theta') &= \mathbb{E}_{\theta'} \{ \log p(y_{1:n}, z_{1:n} | \theta) | y_{1:n} \} \\ &= \int \log p(y_{1:n}, z_{1:n} | \theta) p(z_{1:n} | y_{1:n}, \theta') dz_{1:n}. \end{aligned} \quad (5)$$

Assuming the densities involved in the probabilistic model belong to the exponential family, the auxiliary quantity, or projection function for JMNLS can be written as

$$\begin{aligned} \mathcal{Q}(\theta, \theta') &= \sum_{k=1}^K \sum_{\ell=1}^K \mathcal{S}_{k\ell,n}^{(1)} \log(\pi_{k\ell}) \\ &+ \sum_{k=1}^K \left(\left\langle \psi_k(\theta_k), \mathcal{S}_{k,n}^{(3)} \right\rangle - \mathcal{A}_k(\theta_k) \mathcal{S}_{k,n}^{(2)} \right), \end{aligned} \quad (6)$$

where the additive functionals $\mathcal{S}_{k\ell,n}^{(1)}$, $\mathcal{S}_{k,n}^{(2)}$ and $\mathcal{S}_{k,n}^{(3)}$ are also called sufficient statistics [13] and are computed as in [15].

III. ONLINE COMPUTATION OF ADDITIVE FUNCTIONALS

In order to solve the required expectations, we first define the joint state $z_t = [x_t \ r_t]$ and then pack them into $\mathcal{S}_n = \mathbb{E}_{\theta} \{ \mathcal{S}_n(z_{0:n}) | y_{1:n} \}$ where $\mathcal{S}_n(z_{0:n})$ is the set of additive functionals. By defining the auxiliary function $T_t(z_t) \triangleq \mathbb{E}_{\theta} \{ \mathcal{S}_t(z_{0:t}) | z_t, y_{1:t} \}$, the smoothed additive functional can then be explained as a filtered estimate of $T_t(z_t)$, i.e.

$$\mathcal{S}_t = \mathbb{E}_{\theta} \{ T_t(z_t) | y_{1:t} \} = \int T_t(z_t) p(z_t | y_{1:t}) dz_t. \quad (7)$$

If we additionally count on the additive structure of the recursion, it is even possible to compute $T_t(z_t)$ using

$$\begin{aligned} T_t(z_t) &= \int [T_{t-1}(z_{t-1}) + s_t(z_{t-1}, z_t)] \\ &\quad \times p(z_{t-1} | z_t, y_{1:t-1}; \theta) dz_{t-1}, \end{aligned} \quad (8)$$

where the recursion is initialized with $T_0(z_0) = 0$ and we end up with an online solution by computing the integral in (7).

A. SMC Implementation

In general, due to the intractability of the involved integrals, an exact solution for the expectations is unfeasible. An appropriate solution is obtained with the following decomposition

$$p(x_{1:t}, r_t | y_{1:t}) = p(r_t | x_{1:t}, y_{1:t}) p(x_{1:t} | y_{1:t}). \quad (9)$$

In order to simplify the notation, the dependence on θ is here omitted. SMC techniques are employed in order to approximate $p(x_{1:t} | y_{1:t})$ and a conditional Hidden Markov Model (HMM) filter for $p(r_t | x_{1:t}, y_{1:t})$. This method was proposed in [15] and is also called Rao-Blackwellized Particle Filter (RBPf). The posterior approximation is the set $\{x_{1:t}^{(i)}, \{\alpha_{t|t}^{(i)}(\ell)\}_{\ell=1}^K, w_t^{(i)}\}_{i=1}^N$ where $\alpha_{t|t}^{(i)}(\ell) \triangleq \mathbb{P}(r_t = \ell | x_{1:t}^{(i)}, y_{1:t})$ and $\ell \in \{1, \dots, K\}$.

For the smoothed additive functionals computation in an online fashion, the forward smoother is employed and the backward density can be computed as

$$\begin{aligned} p(x_{1:t-1}, r_{t-1} = k | x_t^{(i)}, r_t = \ell, y_{1:t-1}; \theta) \\ \approx \sum_{j=1}^N \frac{\tilde{w}_t^{(i,j)}(k, \ell)}{\sum_{u=1}^N \sum_{m=1}^K \tilde{w}_t^{(i,u)}(m, \ell)} \delta(x_{1:t-1} - x_{1:t-1}^{(j)}), \end{aligned} \quad (10)$$

with

$$\tilde{w}_t^{(i,j)}(k, \ell) = f_{\ell}(x_t^{(i)} | x_{t-1}^{(j)}; \theta_{\ell}) \pi_{k\ell} \alpha_{t-1|t-1}^{(j)}(k) w_{t-1}^{(j)}. \quad (11)$$

The auxiliary function $\hat{T}_t^{(i)}(\ell) \approx T(x_t^{(i)}, r_t = \ell)$ can then be updated recursively according to the approximation of (8) as follows

$$\begin{aligned} \hat{T}_t^{(i)}(\ell) &= \sum_{j=1}^N \sum_{k=1}^K \left(\frac{\tilde{w}_t^{(i,j)}(k, \ell)}{\sum_{u=1}^N \sum_{m=1}^K \tilde{w}_t^{(i,u)}(m, \ell)} \right. \\ &\quad \times \left. \left[\hat{T}_{t-1}^{(j)}(k) + s_t(x_{t-1}^{(j)}, r_{t-1} = k, x_t^{(i)}, r_t = \ell) \right] \right). \end{aligned} \quad (12)$$

B. PaRIS for JMNLs

The approach described before has a computational complexity of $\mathcal{O}(K^2 N^2)$ and, in order to make it linear with N , we employ a modified version of the PaRIS algorithm. It was firstly proposed in [16], theoretically explained in [17] and extended to the RML framework in [18]. It is supported by the FFBSi concept [19] where, conditioned on each particle $x_t^{(i)}$ and $r_t = \ell$, candidate backward sample indexes $\{J_\ell^{(i,s)}\}_{s=1}^M$ are drawn according to the Markov transition kernel, i.e.

$$\mathbb{P}(J_\ell^{(i,s)} = j) = \frac{w_{t-1}^{(j)} f_\ell(x_t^{(i)} | x_{t-1}^{(j)}; \theta_\ell)}{\sum_{u=1}^N w_{t-1}^{(u)} f_\ell(x_t^{(i)} | x_{t-1}^{(u)}; \theta_\ell)}. \quad (13)$$

In order to compute the normalizing constant in the denominator of (13), we still end up with an algorithm having a quadratic complexity with N . This can be resolved by introducing the mild assumption that the backward density is uniformly bounded by a constant $\bar{\varepsilon}$ [24]. It is then possible to apply an accept-reject sampling scheme by drawing a candidate $J_\ell^{(i,s)}$ with probability proportional to $\{w_{t-1}^{(j)}\}_{j=1}^N$ and then accepting this candidate with probability $\frac{f_\ell(x_t^{(i)} | x_{t-1}^{(J_\ell^{(i,s)})}; \theta_\ell)}{\bar{\varepsilon}}$. Then, instead of computing each subsequent auxiliary statistics as the expected sum of all the previous statistics and the incremental term under the retrospective dynamics induced by the backward kernel, we draw $2 \leq M \ll N$ indexes and the modified version of (12) can be written as

$$\begin{aligned} \hat{T}_t^{(i)}(\ell) &= \frac{1}{M} \sum_{s=1}^M \sum_{k=1}^K \left(\frac{\tilde{w}_t^{(i, J_\ell^{(i,s)})}(k, \ell)}{\sum_{u=1}^M \sum_{m=1}^K \tilde{w}_t^{(i, J_\ell^{(i,u)})}(m, \ell)} \right. \\ &\quad \left. \times \left[\hat{T}_{t-1}^{(J_\ell^{(i,s)})}(k) + s_t(x_{t-1}^{(J_\ell^{(i,s)})}, r_{t-1} = k, x_t^{(i)}, r_t = \ell) \right] \right), \end{aligned} \quad (14)$$

where

$$\tilde{w}_t^{(i, J_\ell^{(i,s)})}(k, \ell) = \pi_{k\ell} \alpha_{t-1|t-1}^{(J_\ell^{(i,s)})}(k). \quad (15)$$

As N approaches infinity, the amount of draws needed for the accept-reject scheme will tend to a constant [16]. Hence, the computational complexity of the algorithm will be reduced to $\mathcal{O}(K^2 NM)$.

IV. SIMULATION RESULTS

The performance of the proposed joint state inference and parameter identification algorithm is assessed on the following benchmark model

$$x_t = \frac{1}{2}x_{t-1} + 25 \frac{x_{t-1}}{1 + x_{t-1}^2} + 8 \cos[1.2(t-1)] + v_t, \quad (16a)$$

$$y_t = \frac{x_t^2}{20} + e_{r_t}, \quad (16b)$$

where the initial state x_0 and process noise v_t are distributed according to $x_0 \sim \mathcal{N}(0, 1)$ and $v_t \sim \mathcal{N}(0, 1)$, respectively. It is further assumed that the measurement noise e_{r_t} is mode dependent, and switches between $K = 2$ modes. The parameters $\theta_{r_t} = \{\mu_{r_t}, \sigma_{r_t}\}$ of the corresponding measurement noise probability density function (pdf) are assumed to be unknown,

and are estimated together with the TPM and the states. The true parameters were chosen as follows: $\pi_{11} = 0.95$, $\pi_{22} = 0.8$, $\theta_1 = \{0, 1\}$ and $\theta_2 = \{3, 2\}$, whereas the filter has been initialized with the following parameters: $\hat{\pi}_{11}^0 = 0.5$, $\hat{\pi}_{22}^0 = 0.5$, $\hat{\theta}_1^0 = \{-1, 5\}$ and $\hat{\theta}_2^0 = \{4, 5\}$.

We additionally consider different choices of the step-size γ_t for each identification approach, namely $\gamma_t = t^{-0.5}$ (RML) and $\gamma_t = t^{-0.7}$ (EM). For a more detailed study of its influence, refer to [22].

From the model given in (16), we simulate a single state and measurement trajectory of length 10000, and perform 100 Monte Carlo runs of the RBPF on this data. Although the algorithm jointly estimates the states and the parameters, we will only show results for the parameter identification.

In Fig. 1, we provide comparative results for $N = 150$ and $M = 3$. It is clear that the PaRIS-based approach delivers minimal impact on estimation performance for noise mean and standard deviation while a slight bias is perceived for the TPM estimation.

We now compare the methods relative to their computational complexity by addressing the average Root Mean Square (RMS) of the estimation error over the iterations/time (discarding the first 50 iterations) and discrete modes with respect to 50, 100, 150, 200, 250 and 300 particles (additionally 900 for PaRIS). The complexity is computed as $\mathcal{O}(K^2 N^2)$ and $\mathcal{O}(K^2 NM)$ for the original forward smoother and PaRIS, respectively. The results are depicted in Fig. 2 for each parameter separately. The advantage of employing the rapid smoother in terms of computational complexity for a desired average RMS error in both μ and σ estimation tasks is easily perceived. We can further notice that for $M = 3$, estimating the TPM with PaRIS does not yield a comparable performance even when increasing in the number of particles. We thus simulate a few additional points (with 50, 150 and 300 particles) with $M = 15$, which are also provided in Fig. 2. These results highlight a higher dependency of TPM estimation performance on the number of backward samples. A possible explanation of this is the fact that the unknown parameter (II) is the one that directly acts on the state transition.

V. CONCLUSION

This paper, we proposed a new version of the PaRIS algorithm suited for systems with switching regimes. The algorithm aims at solving the problem of unknown parameter estimation using particle based methods with linear complexity (in the number of particles). The experiments were performed with both EM and RML for a highly non-linear benchmark problem. The results confirmed the efficiency of the algorithm in solving the problem with minimal impact on performance but there was a larger sensitivity on the number of backward samples needed for a proper estimation of the discrete state transition matrix.

ACKNOWLEDGMENT

André R. Braga was supported by CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico, CISB - Centro de Pesquisa e Inovação Suelco-Brasileiro and Saab AB. Carsten Fritsche was supported by the Vinnova Industry Excellence Center ELLIIT at Linköping University.

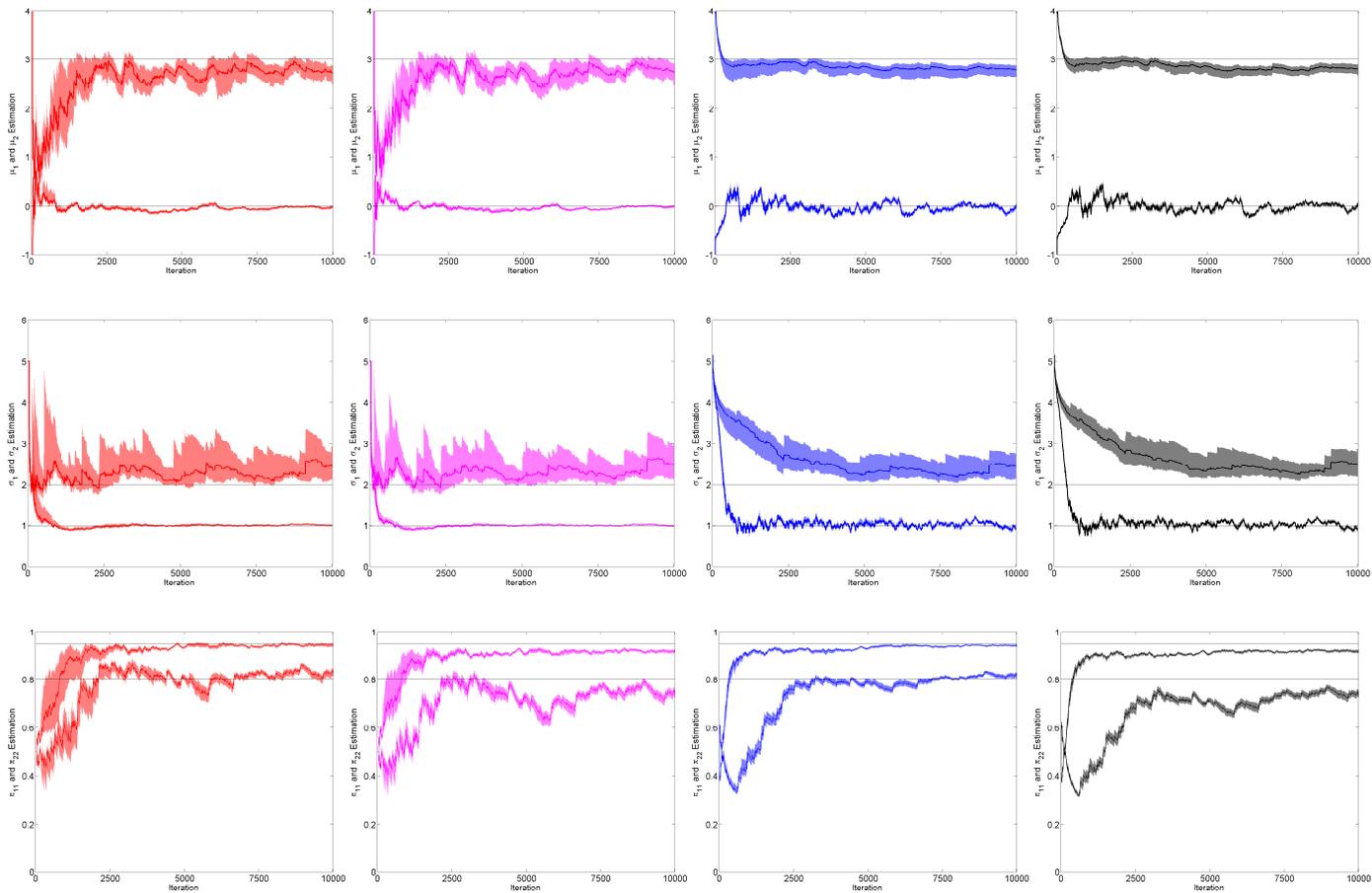


Figure 1. Parameter identification results. From left to right: EM, EM PaRIS, RML and RML PaRIS. From top to bottom: (μ_1, μ_2) , (σ_1, σ_2) , and (π_{11}, π_{22}) . The lines show the averages and the shaded areas show the upper and lower bounds.

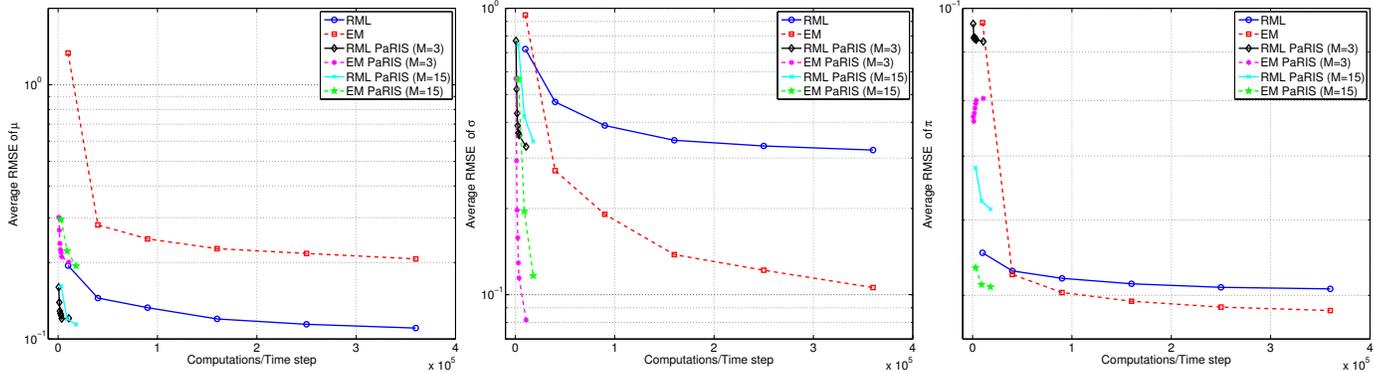


Figure 2. Average RMS error of measurement noise mean (left), measurement noise standard deviation (center) and transition matrix elements (right).

REFERENCES

- [1] L. E. Svensson and N. Williams, "Optimal Monetary Policy under Uncertainty in DSGE Models: A Markov Jump-Linear-Quadratic Approach," National Bureau of Economic Research, Working Paper 13892, March 2008.
- [2] K. A. Loparo and F. Abdel-Malek, "A Probabilistic Approach to Dynamic Power System Security," *IEEE Transactions on Circuits and Systems*, vol. 37, no. 6, pp. 787–798, Jun 1990.
- [3] J.-F. Liao and B.-S. Chen, "Robust Mobile Location Estimator with NLOS Mitigation using Interacting Multiple Model Algorithm," *IEEE Transactions on Wireless Communications*, vol. 5, no. 11, pp. 3002–3006, November 2006.
- [4] C. Fritsche, E. Özkan, and F. Gustafsson, "Online EM Algorithm for Jump Markov Systems," in *2012 15th International Conference on Information Fusion*, July 2012, pp. 1941–1946.
- [5] N. Kantas, A. Doucet, S. S. Singh, J. Maciejowski, and N. Chopin, "On Particle Methods for Parameter Estimation in State-Space Models," *Statist. Sci.*, vol. 30, no. 3, pp. 328–351, 08 2015.
- [6] 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, Series B*, vol. 39, no. 1, pp. 1–38, 1977.
- [7] J. Nocedal and S. J. Wright, *Numerical Optimization*, ser. Springer Series in Operations Research and Financial Engineering. Berlin: Springer, 2006.
- [8] G. Kitagawa, "Monte Carlo Filter and Smoother for Non-Gaussian Non-linear State Space Models," *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. pp. 1–25, 1996.
- [9] A. Doucet, S. Godsill, and C. Andrieu, "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, Jul. 2000.
- [10] P. E. Jacob, L. M. Murray, and S. Rubenthaler, "Path Storage in the Particle Filter," *Statistics and Computing*, vol. 25, no. 2, pp. 487–496, 2015.
- [11] J. Olsson, O. Cappe, R. Douc, and E. Moulines, "Sequential Monte Carlo Smoothing with Application to Parameter Estimation in Nonlinear State Space Models," *Bernoulli*, vol. 14, no. 1, pp. 155–179, 02 2008.
- [12] R. Douc, A. Garivier, E. Moulines, and J. Olsson, "On the Forward Filtering Backward Smoothing Particle Approximations of the Smoothing Distribution in General State Spaces Models," *ArXiv e-prints*, Apr. 2009.
- [13] P. Del Moral, A. Doucet, and S. Singh, "Forward Smoothing using Sequential Monte Carlo," Cambridge University Engineering Department, Tech. Rep. CUED/F-INFENG/TR638, 2010.
- [14] O. Cappe, "Online Sequential Monte Carlo EM Algorithm," in *Statistical Signal Processing, 2009. SSP '09. IEEE/SP 15th Workshop on*, Aug 2009, pp. 37–40.
- [15] E. Özkan, F. Lindsten, C. Fritsche, and F. Gustafsson, "Recursive Maximum Likelihood Identification of Jump Markov Nonlinear Systems," *Signal Processing, IEEE Transactions on*, vol. 63, no. 3, pp. 754–765, Feb 2015.
- [16] J. Westerborn and J. Olsson, "Efficient Particle-Based Online Smoothing in General Hidden Markov Models," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 8003–8007.
- [17] J. Olsson and J. Westerborn, "Efficient Particle-Based Online Smoothing in General Hidden Markov Models: The PaRIS Algorithm," *Bernoulli*, vol. 23, no. 3, pp. 1951–1996, 08 2017.
- [18] —, "Efficient Parameter Inference in General Hidden Markov Models Using the Filter Derivatives," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 3984–3988.
- [19] S. J. Godsill, A. Doucet, and M. West, "Monte Carlo Smoothing for Non-Linear Time Series," *Journal of the American Statistical Association*, vol. 99, pp. 156–168, 2004.
- [20] T. B. Schön, F. Lindsten, J. Dahlin, J. Wägberg, C. A. Naesseth, A. Svensson, and L. Dai, "Sequential Monte Carlo methods for system identification," in *Proceedings of the 17th IFAC Symposium on System Identification (SYSID)*, Beijing, China, October 2015.
- [21] O. Cappé, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models (Springer Series in Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005.
- [22] A. R. Braga, C. Fritsche, F. Gustafsson, and M. Bruno, "Gradient-Based Recursive Maximum Likelihood Identification of Jump Markov Non-Linear Systems," to appear in 2017 20th International Conference on Information Fusion (FUSION) (FUSION 2017), Xi'an, P.R. China, Jul. 2017.
- [23] S. Gibson and B. Ninness, "Robust Maximum-Likelihood Estimation of Multivariable Dynamic Systems," *Automatica*, vol. 41, no. 10, pp. 1667 – 1682, 2005.
- [24] R. Douc, A. Garivier, E. Moulines, and J. Olsson, "Sequential Monte Carlo Smoothing for General State Space Hidden Markov Models," *The Annals of Applied Probability*, vol. 21, no. 6, pp. 2109–2145, 12 2011.