Supplementary material for: Recursive parameter estimation for jump Markov linear systems

Mehmet Çetinkaya[a], Emre Özkan[a]

[a]*Department of Electrical and Electronics Engineering, Middle East Technical University, Turkey*

Abstract

This supplementary material contains further examples to "Recursive parameter estimation for jump Markov linear systems" [1] as well as the derivation of the sufficient statistics for the state transition and measurement matrices of a jump Markov linear system.

1. Simulation Results

In this section, the performance of the proposed online EM algorithm is evaluated on two simulation examples adapted from the literature [2] and [3]. We first investigate the transfer function identification problem for a 2^{nd} order systems. The second simulation demonstrates identifying the measurement noise characteristics of a 4^{th} order system. The performance of the proposed algorithm is evaluated against a state-of-the-art offline EM algorithm presented in [2].

1.1. Second Order System

Consider the system given in [2], whose transfer functions and transition probability matrix are

$$H_1(z) = \frac{0.7406z^{-1} + 0.004861z^{-1}}{1 + 0.6178z^{-1} + 0.4385z^{-2}}, \tag{1a}$$

$$H_2(z) = \frac{-1.461z^{-1} + 1.98z^{-1}}{1 - 1.1814z^{-1} + 0.2715z^{-2}}, \tag{1b}$$

$$\Pi = \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{bmatrix}. \tag{1c}$$

These transfer function representations can be converted to state-space models as follows. Given

$$H_k(z) = \frac{b_1^k z^{-1} + b_2^k z^{-2}}{1 + a_1^k z^{-1} + a_2^k z^{-2}} \tag{2}$$

the state transition, measurement and input matrices of the corresponding model in controllable canonical form are

$$F^k = \begin{bmatrix} 0 & 1 \\ -a_2^k & -a_1^k \end{bmatrix}, H^k = \begin{bmatrix} b_2^k & b_1^k \end{bmatrix}, G^k = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{3}$$

respectively. We run the online EM algorithm on the JMLS switching between these state-space models, where $k = 1, 2$. The measurement noise variances are $R^1 = 0.002$, $R^2 = 0.005$ and process noise covariances are $Q^1 = \text{Diag}(0.045, 0.03)$, $Q^2 = \text{Diag}(0.032, 0.02)$.

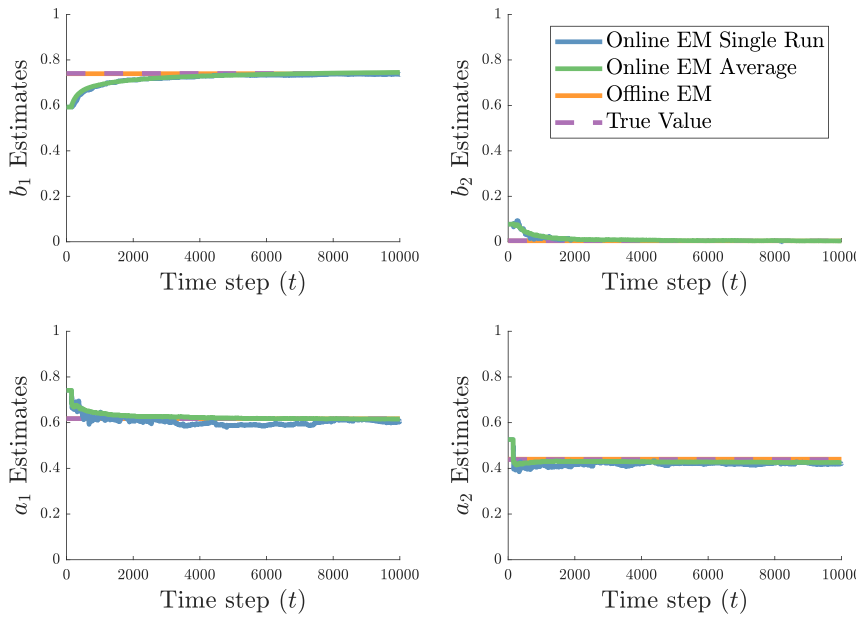

Figure 1: Single run parameter estimates of online and offline EM algorithms and average parameter estimates of online EM algorithm over 100 independent runs for the second order system given in section 1.1, for the first mode.

The system is excited with the input $u_t \sim \mathcal{N}(0,1)$, and the total simulation time is $t = 10,000$. The unknown parameters of the models are initialized with a 20% error as $\hat{F}^i = 1.2F^i$ and $\hat{H}^i = 0.8H^i$. Step sizes used in the simulation are chosen as $\eta_t^1 = t^{-0.7}$ for the sufficient statistics related to the state transition matrix, and $\eta_t^2 = t^{-0.6}$ for those related to the measurement matrix. The burn-in period is set to $t_b = 150$.

The estimated parameters $a_1^k, a_2^k, b_1^k, b_2^k$ for $k = 1, 2$, averaged over 100 MC runs, are given in Figures 1 and 2 where the transfer function is parameterized as in (2). Table 1 summarizes the parameter estimates, percent errors, and simulation times. Both EM algorithms yield parameter estimates that are close to the true values. The online EM algorithm processed 10,000 measurements in 38 seconds, whereas the batch method required 150 iterations for parameter convergence, taking approximately $19,000$ seconds to complete the iterations.

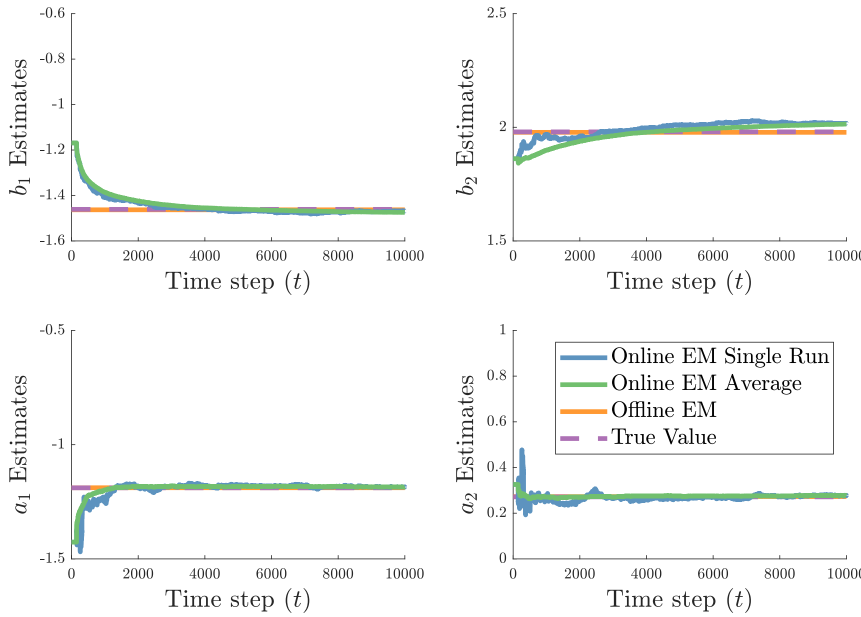

Figure 2: Single run parameter estimates of online and offline EM algorithms and average parameter estimates of online EM algorithm over 100 independent runs for the second order system given in section 1.1, for the second mode.

Table 1: Parameter estimates, absolute percent errors with respect to true values and simulation times of the second order system given in section 1.1

Parameter	b_1^1	b_2^1	a_1^1	a_2^1
True Values	0.741	0.005	0.618	0.439
Online EM Single Run Estimate	0.737	0.007	0.607	0.421
Percent Error	0.5%	40%	1.8%	4.1%
Online EM Average Estimate	0.745	0.004	0.616	0.426
Percent Error	0.6%	20%	0.3%	3.0%
Offline EM Estimate	0.739	0.004	0.617	0.440
Percent Error	0.3%	20%	0.2%	0.2%
Parameter	b_1^2	b_2^2	a_1^2	a_2^2
True Values	-1.461	1.980	-1.181	0.271
Online EM Single Run Estimate	-1.468	2.016	-1.185	0.278
Percent Error	0.5%	1.8%	0.3%	2.6%
Online EM Average Estimate	-1.475	2.014	-1.185	0.277
Percent Error	1.0%	1.7%	0.3%	2.2%
Offline EM Estimate	-1.463	1.977	-1.186	0.273
Percent Error	0.1%	0.2%	0.4%	0.7%
Online EM Computation Time	38 seconds			
Offline EM Computation Time	19000 seconds			

1.2. Fourth Order System with Unknown Noise Characteristics

Consider the nearly constant velocity model with position measurements [4]. The state-space model is

$$F^k = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \otimes I_2, H^k = \begin{bmatrix} I_2 & 0 \end{bmatrix}, \tag{4a}$$

$$Q^k = q \begin{bmatrix} \frac{T^3}{3} & \frac{T^2}{2} \\ \frac{T^2}{2} & T \end{bmatrix} \otimes I_2, \tag{4b}$$

where $\otimes$ is the Kronecker product, $q = 1$, $T = 1s$, and the measurements are obtained either with nominal measurement noise $R^1 = 100I_2$, or outlier measurement noise $R^2 = 100R^1$. The TPM is adapted to

$$\Pi = \begin{bmatrix} 0.9 & 0.1 \\ 0.9 & 0.1 \end{bmatrix} \tag{4c}$$

so that measurements corrupted with Gaussians of severely increased covariances are obtained approximately 10 percent of the time [3].

The initial noise covariance estimates consist of a lower value for the nominal noise covariance and a slightly larger value for the outlier mode's covariance. The initial estimates are chosen as $\hat{R}^1 = 0.2R^1$ and $\hat{R}^2 = 5R^1$. To ensure an adequate exploration of the outlier mode before any maximization step, a burn-in period of $t_b = 150$ is chosen. The step size for the stochastic approximation is set to $\eta_t = t^{-0.8}$.

Figures 3 and 4 illustrate the estimated measurement noise covariances. Table 2 presents the final estimates and corresponding percent errors. Both the online EM and the offline EM algorithms have similar accuracy in noise parameter estimates. The online EM algorithm took 41 seconds to process the whole sequence, while the offline EM took 1268 seconds to perform 10 iterations.

1.3. Impact of Parameter Estimates on State Estimation Accuracy

Finally, to assess the quality of the parameter estimates obtained by the algorithms, we run a GPB2 filter using three sets of parameters: the true parameters θ_{True}, the offline EM estimates $\hat{\theta}_{\text{EM}}$, and the averaged estimates from the online EM algorithm $\hat{\theta}_{\text{OEM}}^{\text{Avg.}}$. We perform 100 Monte Carlo simulations and compare the RMSEs for each case. The results, presented in Table 3, show that the filter using the true parameters achieves the smallest RMSE, as expected. It is closely followed by the filter using estimates from the offline EM algorithm, and then by the online EM algorithm. The RMSEs obtained using the parameters of both the offline and online EM algorithms are comparable to the RMSE of the filter that uses the true parameters.

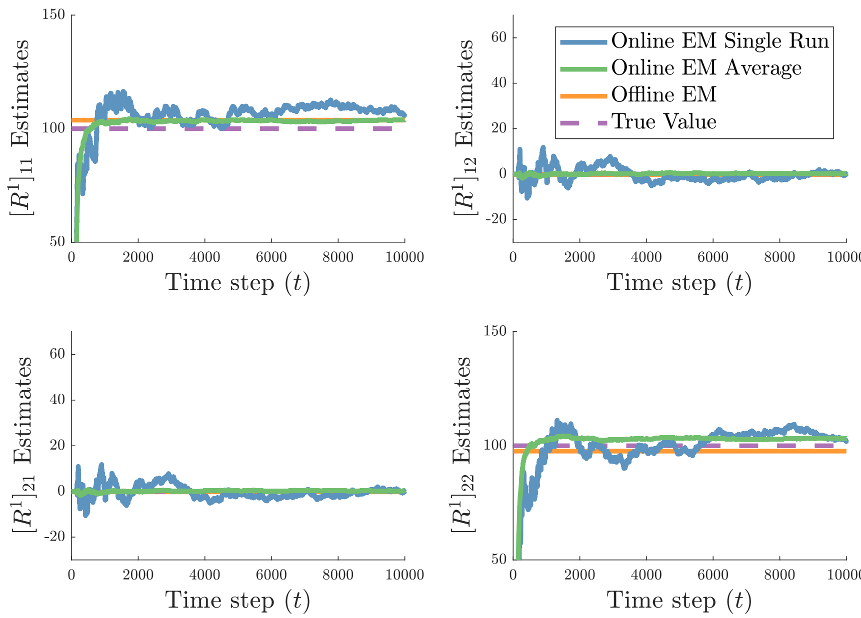

Figure 3: Single run noise covariance matrix estimates of online and offline EM algorithms and average parameter estimates of online EM algorithm over 100 independent runs for the fourth order system given in section 1.2, for the first (nominal) mode.

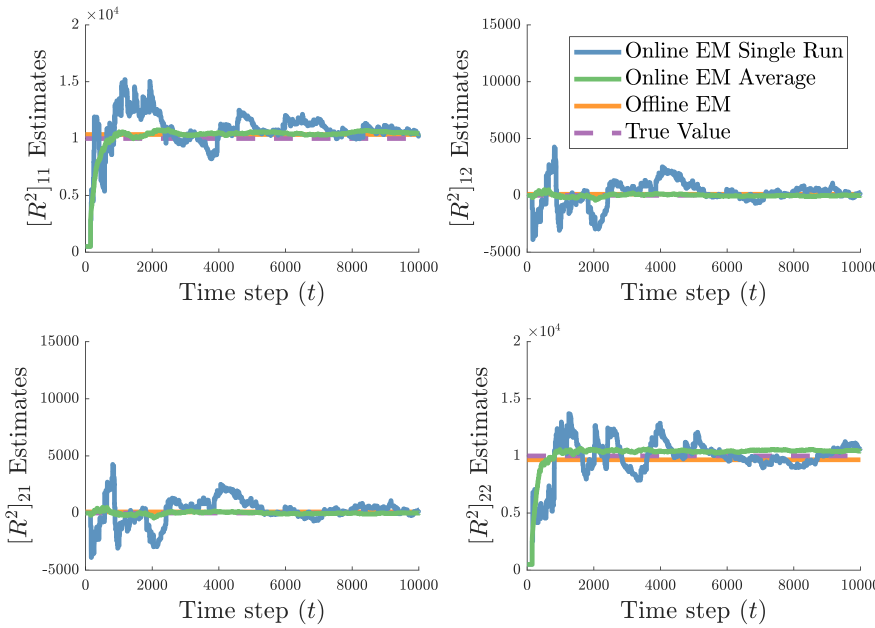

Figure 4: Single run noise covariance matrix estimates of online and offline EM algorithms and average parameter estimates of online EM algorithm over 100 independent runs for the fourth order system given in section 1.2, for the second (outlier) mode.

Table 2: Parameter estimates, absolute percent errors with respect to true values and simulation times of the fourth order system given in section 1.2

Parameter	$[R^1]^1_{11}$	$[R^1]^1_{12}$	$[R^1]^1_{22}$
True Values	100	0	100
Online EM Single Run Estimate	105	0	102
Percent Error	5%	NA	2%
Online EM Average	104	0	103
Percent Error	4%	NA	3%
Offline EM Estimate	104	0	98
Percent Error	4%	NA	2%
Parameter	$[R^2]^1_{11}$	$[R^2]^1_{12}$	$[R^2]^1_{22}$
True Values	10000	0	10000
Online EM Single Run Estimate	10263	270	10640
Percent Error	2.6%	NA	6.4%
Online EM Average Estimate	10448	6	10410
Percent Error	4.5%	NA	4.1%
Offline EM Estimate	10355	96	9662
Percent Error	3.5%	NA	3.4%
Online EM Computation Time	41 seconds		
Offline EM Computation Time	1268 seconds		

Table 3: RMSEs of GPB2 filter run with averaged parameter estimates by the online EM, parameter estimates by the offline EM, and the true parameters

Example Number	$\hat{\theta}^{Avg.}_{OEM}$	$\hat{\theta}_{EM}$	θ_{True}
Example 1 RMSE	98.03	96.15	93.56
Example 2 RMSE	964.1	964.0	964.0

2. Derivation of the Sufficient Statistics

In this section, we derive the sufficient statistics for the state transition and measurement matrices, F^k and H^k. These are obtained by differentiating the expected complete data log-likelihood with respect to the model parameters. The complete data log-likelihood is given by

$$
\begin{aligned}
\log p(y_{0:t}, x_{0:t}, r_{0:t}; \theta) &\stackrel{\pm}{=} \sum_{i=0}^{t} \log \left(p_{r_i}(y_i|x_i; \theta_{r_i})\, p_{r_i}(x_i|x_{i-1}; \theta_{r_i}) \right) \\
&\stackrel{\pm}{=} -0.5 \sum_{i=0}^{t} \sum_{k=1}^{K} \mathbb{1}(r_i = k)(y_i - H^k x_i)^T (R^k)^{-1} (y_i - H^k x_i) \\
&\quad - 0.5 \sum_{i=1}^{t} \sum_{k=1}^{K} \mathbb{1}(r_i = k)(x_i - F^k x_{i-1} - G^k u_{i-1})^T (Q^k)^{-1} (x_i - F^k x_{i-1} - G^k u_{i-1})
\end{aligned}
\tag{5}
$$

where $\stackrel{\pm}{=}$ denotes equality up to a constant independent of H^k and F^k. By taking the trace of this scalar expression and applying the cyclic property of the trace operator, it can be rewritten as

$$
\begin{aligned}
\log p(y_{0:t}, x_{0:t}, r_{0:t}; \theta) &\stackrel{\pm}{=} -0.5 \sum_{i=0}^{t} \sum_{k=1}^{K} \mathbb{1}(r_i = k)\, \mathrm{Tr}([-2H^k x_i y_i^T + H^k x_i x_i^T (H^k)^T](R^k)^{-1}) \\
&\quad -0.5 \sum_{i=1}^{t} \sum_{k=1}^{K} \mathbb{1}(r_i = k)\, \mathrm{Tr}([-2F^k x_{i-1}(x_i - G^k u_{i-1})^T + F^k x_{i-1} x_{i-1}^T (F^k)^T](Q^k)^{-1})
\end{aligned}
\tag{6}
$$

Taking the expectation of the above expression and employing results from [5, Eq.s (100), (118)]

$$
\frac{\partial\, \mathrm{Tr}(XA)}{\partial X} = A^T, \tag{7}
$$

$$
\frac{\partial\, \mathrm{Tr}(XBX^T C)}{\partial X} = C^T X B^T + CXB = 2CXB \textbf{ for } B = B^T, C = C^T, \tag{8}
$$

we can then maximize the auxiliary function

$$
\begin{aligned}
\frac{\partial Q(\theta, \theta')}{\partial H^k} &= \frac{\partial \mathbb{E}_{\theta'}\{\log p(y_{0:t}, x_{0:t}, r_{0:t}; \theta)\}}{\partial H^k} \\
&= -0.5 (R^k)^{-1} \mathbb{E}_{\theta'} \left\{ \sum_{i=0}^{t} \mathbb{1}(r_i = k)(-2 y_i x_i^T + 2 H^k x_i x_i^T) \right\} = 0.
\end{aligned}
\tag{9}
$$

Solving this equation yields the maximizer for H^k as

$$
\hat{H}^k = \mathbb{E}_{\theta'} \left\{ \sum_{i=0}^{t} \mathbb{1}(r_i = k) y_i x_i^T \right\} \left(\mathbb{E}_{\theta'} \left\{ \sum_{i=0}^{t} \mathbb{1}(r_i = k) x_i x_i^T \right\} \right)^{-1} \tag{10}
$$

$$
= \mathbb{E}_{\theta'} \left\{ S_{k,t}^{4T}(z_{0:t}) \right\} \left(\mathbb{E}_{\theta'} \left\{ S_{k,t}^{0}(z_{0:t}) \right\} \right)^{-1}, \tag{11}
$$

where the sufficient statistics for H^k are given by

$$S^4_{k,t}\left(z_{0:t}\right) = \sum_{i=0}^{t} \mathbb{1}\left(r_i = k\right) x_i y_i^T, \text{ and} \tag{12}$$

$$S^0_{k,t}\left(z_{0:t}\right) = \sum_{i=0}^{t} \mathbb{1}\left(r_i = k\right) x_i x_i^T. \tag{13}$$

We now differentiate the auxiliary function with respect to F^k

$$\begin{aligned}\frac{\partial Q(\theta,\theta')}{\partial F^k} &= \frac{\partial \mathbb{E}_{\theta'}\{\log p(y_{0:t}, x_{0:t}, r_{0:t};\theta)\}}{\partial F^k} \\ &= -0.5(Q^k)^{-1}\mathbb{E}_{\theta'}\left\{\sum_{i=1}^{t}\mathbb{1}(r_i = k)(-2(x_i - G^k u_{i-1})x_{i-1}^T + 2F^k x_{i-1}x_{i-1}^T)\right\} = 0. \end{aligned} \tag{14}$$

Solving this equation yields the maximizer for F^k as

$$\hat{F}^k = \mathbb{E}_{\theta'}\left\{\sum_{i=1}^{t}\mathbb{1}(r_i = k)(x_i - G^k u_{i-1})x_{i-1}^T\right\}\left(\mathbb{E}_{\theta'}\left\{\sum_{i=1}^{t}\mathbb{1}(r_i = k)x_{i-1}x_{i-1}^T\right\}\right)^{-1} \tag{15}$$

$$= \left(\mathbb{E}_{\theta'}\left\{S^{2T}_{k,t}(z_{0:t})\right\} - \mathbb{E}_{\theta'}\left\{S^{3T}_{k,t}(z_{0:t})\right\}\right)\left(\mathbb{E}_{\theta'}\left\{S^1_{k,t}(z_{0:t})\right\}\right)^{-1}, \tag{16}$$

where the sufficient statistics for F^k are given by

$$S^1_{k,t}\left(z_{0:t}\right) = \sum_{i=1}^{t} \mathbb{1}\left(r_i = k\right) x_{i-1} x_{i-1}^T, \tag{17}$$

$$S^2_{k,t}\left(z_{0:t}\right) = \sum_{i=1}^{t} \mathbb{1}\left(r_i = k\right) x_{i-1} x_i^T, \text{ and} \tag{18}$$

$$S^3_{k,t}\left(z_{0:t}\right) = \sum_{i=1}^{t} \mathbb{1}(r_i = k) x_{i-1}(G^k u_{i-1})^T. \tag{19}$$

References

[1] M. Çetinkaya, E. Özkan, Recursive parameter estimation for jump markov linear systems, manuscript under revision (2025).

[2] M. P. Balenzuela, A. G. Wills, C. Renton, B. Ninness, Parameter estimation for jump Markov linear systems, Automatica 135 (2022) 109949. `doi:https://doi.org/10.1016/j.automatica.2021.109949`.

[3] M. Roth, E. Özkan, F. Gustafsson, A student's t filter for heavy tailed process and measurement noise, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 5770–5774. `doi:10.1109/ICASSP.2013.6638770`.

[4] Y. Bar-Shalom, X. R. Li, T. Kirubarajan, Estimation with applications to tracking and navigation: theory algorithms and software, John Wiley & Sons, 2004.

[5] K. B. Petersen, M. S. Pedersen, et al., The matrix cookbook, Technical University of Denmark 7 (15) (2008) 510.