Identification of jump Markov linear models using particle filters

UPPSALA UNIVERSITET

Andreas Svensson[†], Thomas B. Schön[†] and Fredrik Lindsten[‡]

[†]Department of Information Technology, Uppsala University and [‡]Department of Engineering, University of Cambridge

UNIVERSITY OF CAMBRIDGE

Summary

- Identification of parameters in jump Markov linear models.
- Theoretically established convergence properties (even for a finite number of particles $N \geq 2$ in the involved particle filter!).
- Rao-Blackwellization for decreased variance.
- Computationally more efficient than previous methods that combine particle smoothers and EM.
- An estimate of the state space smoothing distribution explicitly obtained as a 'by-product'.

Contribution: Maximum Likelihood identification of jump Markov linear models

Problem formulation

Maximum Likelihood identification⋆ *of parameters*⋆⋆⋆ *in jump Markov linear models*⋆⋆*:*

⋆ Maximum Likelihood identification:

$$\widehat{\theta} = \underset{\theta}{\text{argmax}}\ p_\theta(y_{1:T}) \quad (1)$$

⋆⋆ Jump Markov linear model:

$$\begin{aligned} s_{t+1} &\sim p(s_{t+1}|s_t) \triangleq \pi_{s_t,s_{t+1}} \\ z_{t+1} &= A_{s_{t+1}} z_t + B_{s_{t+1}} u_t + w_t \\ y_t &= C_{s_t} z_t + D_{s_t} u_t + v_t \end{aligned} \quad (2)$$

with modes $s_t \in \{1, \dots, K\}$, linear states $z_t \in \mathbb{R}^{n_z}$ and covariances $\mathbb{E}\left[w_t^T w_t\right] = Q_{s_{t+1}}$, $\mathbb{E}\left[v_t^T v_t\right] = R_{s_{t+1}}$.

⋆⋆⋆ Parameters to identify:

$$\theta = \{A_n, B_n, C_n, D_n, Q_n, R_n, \{\pi_{n,m}\}_{m=1}^K\}_{n=1}^K \quad (3)$$

(K and n_z are assumed to be known)

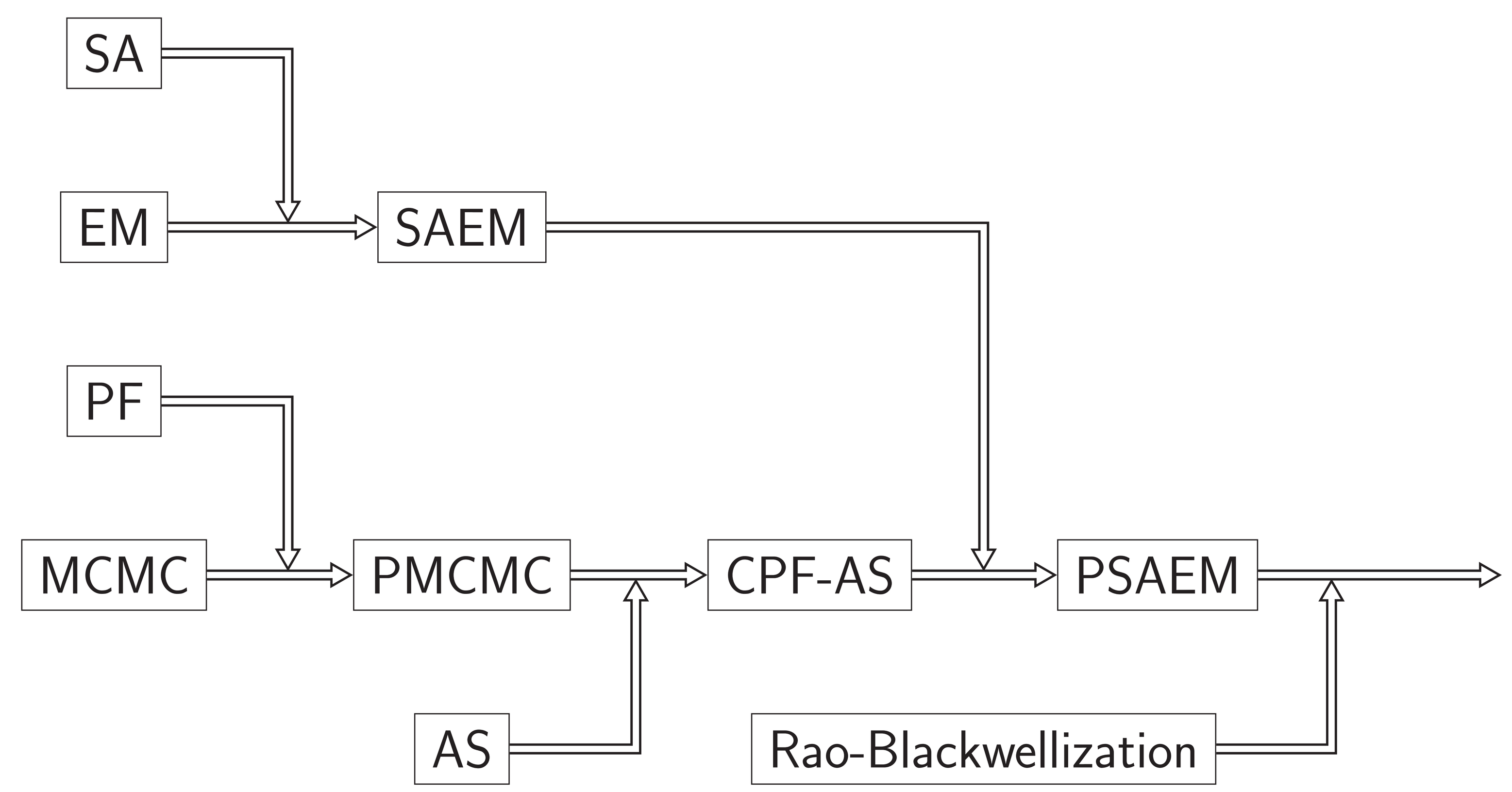
Background

Identification of nonlinear systems using Expectation Maximization (EM) and particle smoothing has previously been proposed[1], as well as a more efficient related methodology involving stochastic approximation and a Markov chain Monte Carlo construction, PSAEM[2]. The present contribution is a Rao-Blackwellized version of PSAEM, adopted particularly to identify jump Markov linear models.

1. Thomas B. Schön, Adrian Wills and Brett Ninness. **System identification of nonlinear state-space models.** *Automatica*, 47(1):39-49, January 2011.
2. Fredrik Lindsten **An efficient stochastic approximation EM algorithm using conditional particle filters.** *Proceedings of the 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013.

Key methods

- **EM** *(Expectation Maximization)* Framework for learning parameters using latent variables (e.g. state space variables).
- **PF** *(Particle Filter)* A sequential Monte Carlo method for finding the posterior distribution of state space variables.
- **SAEM** *(Stochastic Approximation EM)* Extension of EM based on stochastic E-step (e.g. when using methods such as PF).
- **MCMC** *(Markov Chain Monte Carlo)* Framework for exploration of (complicated) probability densities, based on the construction of a Markov kernel related to the distribution of interest.
- **PMCMC** *(Particle MCMC)* An MCMC construction relying on a repeated use of the PF to construct the Markov kernel.
- **CPF-AS** *(Conditional PF with Ancestor Sampling)* A particular PMCMC method using Ancestor Sampling to increase mixing properties.
- **Rao-Blackwellization** To treat the linear substructure analytically.



Algorithm strategy

[0] Intitialize $\widehat{\theta}_0$, $s'_{1:T}[0]$ and $\widehat{\mathcal{Q}}_k(\theta) \equiv 0$
for $k \geq 1$ **do**
[1] Given $s'_{1:T}[k-1]$ and θ_{k-1}, run RB CPF-AS to get $\{s^i_{1:T}, w_T\}_{i=1}^N$
[2] Compute $\widehat{\mathcal{Q}}_k(\theta) = (1-\gamma_k)\widehat{\mathcal{Q}}_{k-1}(\theta) + \gamma_k \sum_i w_T^i \log p_\theta(y_{1:T}, s^i_{1:T})$
[3] Compute $\widehat{\theta}_k = \text{argmax}\ \widehat{\mathcal{Q}}_k(\theta)$
[4] Set $s'_{1:T}[k] = s^J_{1:T}$ with $\mathbb{P}(J = i) = w^i_T$
end for

RB CPF-AS: Rao-Blackwellized Conditional Particle Filter with Ancestor Sampling. See reference for algorithmic statement.

Theoretical foundation

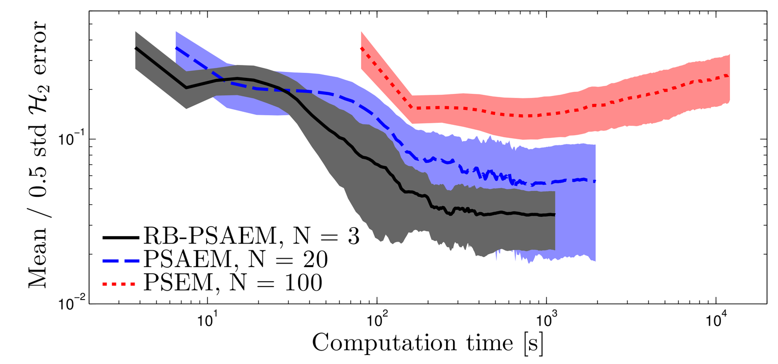
The construction of PSAEM is well supported by MCMC theory; the CPF-AS defines an ergodic Markov kernel with invariant distribution $p_\theta(s_{1:T}, z_{1:T}|y_{1:T})$. This implies convergence to a (not necessarily global) maximizer $\widehat{\theta}$ of $p_\theta(y_{1:T})$ even for a finite number of particles N in the particle filter.

This result is also well reflected in practice, as only $N = 3$ particles are used in the numerical examples.

Numerical example

Example 1 Identification of a one-dimensional jump Markov linear model with 2 modes on simulated data. The following methods are compared:

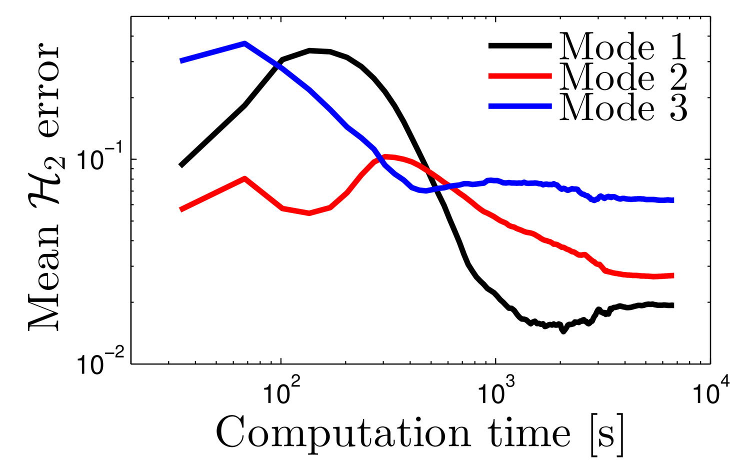
1. Rao-Blackwellized PSAEM with $N = 3$ particles,
2. Original PSAEM with $N = 20$,
3. PS+EM with $N = 100$ forward particles and $M = 20$ backward trajectories.



Mean (solid lines) and 0.5 standard deviation (coloured areas) $\mathcal{H}_2$ error for 7 runs of Rao-Blackwellized PSAEM (black), PSAEM (blue) and PS+EM (red).

From the figure (note the log-log scale) it is clear that our new Rao-Blackwellized PSAEM algorithm has a significantly better performance, both in terms of mean and in variance between different runs, compared to the previous algorithms.

Example 2 Identification of a two-dimensional system with 3 modes. Still, N = 3. The figure shows the mean (over 10 runs) $\mathcal{H}_2$ error for each mode. Thus, the Rao-Blackwellized PSAEM algorithm has the ability to catch the system dynamics fairly well even of multi-dimensional systems.



Mean $\mathcal{H}_2$ error for each mode.

Reference

Andreas Svensson, Thomas B. Schön and Fredrik Lindsten. **Identification of jump Markov linear models using particle filters.** In *Proceedings of the 53rd IEEE Conference on Decision and Control (CDC)*, Los Angeles, CA, USA, December, 2014. (accepted for publication)

http://www.it.uu.se/katalog/andsv164/Research/RB-PSAEM

andreas.svensson@it.uu.se, thomas.schon@it.uu.se, fredrik.lindsten@eng.cam.ac.uk