Maximum likelihood identification in nonlinear state space models

"The particle filter provides a systematic way of exploring the state space"

UPPSALA UNIVERSITET

Thomas Schön
Division of Systems and Control
Department of Information Technology
Uppsala University.

Email: thomas.schon@it.uu.se,
www: `user.it.uu.se/~thosc112`

Joint work with: **Fredrik Lindsten** (University of Cambridge and University of Oxford), **Lennart Ljung** (Linköping University), **Christian A. Naesseth** (Linköping University), **Brett Ninness** (Newcastle University) and **Adrian Wills** (MRA, Australia).

The nonlinear SSM

A state space model (SSM) consists of a Markov process $\{x_t\}_{t\geq 1}$ that is indirectly observed via a measurement process $\{y_t\}_{t\geq 1}$,

$$\begin{aligned} x_{t+1} \mid x_t &\sim f_\theta(x_{t+1} \mid x_t, u_t), \\ y_t \mid x_t &\sim g_\theta(y_t \mid x_t, u_t), \\ x_1 &\sim \mu_\theta(x_1), \\ (\theta &\sim \pi(\theta)). \end{aligned}$$

$$\begin{aligned} x_{t+1} &= a_\theta(x_t, u_t) + v_{\theta,t}, \\ y_t &= c_\theta(x_t, u_t) + e_{\theta,t}, \\ x_1 &\sim \mu_\theta(x_1), \\ (\theta &\sim \pi(\theta)). \end{aligned}$$

We observe

$$y_{1:T} \triangleq \{y_1, \dots, y_T\}, \quad \text{and possibly } u_{1:T} \triangleq \{u_1, \dots, u_T\}.$$

(leaving the latent variables $x_{1:T}$ unobserved).

Identification problem: Find θ based on $y_{1:T}$ (and $u_{1:T}$).

Two commonly used problem formulations

Maximum likelihood (ML) formulation – model the unknown parameters as a deterministic variable and solve

$$\widehat{\theta}_{\mathsf{ML}} = \arg\max_{\theta \in \Theta} p_\theta(y_{1:T}).$$

Bayesian formulation – model the unknown parameters as a random variable $\theta \sim \pi(\theta)$ and compute

$$p(\theta \mid y_{1:T}) = \frac{p(y_{1:T} \mid \theta)\pi(\theta)}{p(y_{1:T})}.$$

UPPSALA UNIVERSITET

Central object – the likelihood

The likelihood is computed by marginalizing the joint density

$$p_\theta(x_{1:T}, y_{1:T}) = \mu_\theta(x_1) \prod_{t=1}^{T} g_\theta(y_t \mid x_t) \prod_{t=1}^{T-1} f_\theta(x_{t+1} \mid x_t).$$

w.r.t. the state sequence $x_{1:T}$,

$$p_\theta(y_{1:T}) = \int p_\theta(x_{1:T}, y_{1:T}) \mathrm{d}x_{1:T}.$$

We are averaging $p_\theta(x_{1:T}, y_{1:T})$ over all possible state sequences.

Equivalently we have

$$p_\theta(y_{1:T}) = \prod_{t=1}^{T} p_\theta(y_t \mid y_{1:t-1}) = \prod_{t=1}^{T} \int g_\theta(y_t \mid x_t) \underbrace{p_\theta(x_t \mid y_{1:t-1})}_{\text{key challenge}} \mathrm{d}x_t.$$

UPPSALA UNIVERSITET

Identification strategy – data augmentation

Motivation: If we had access to the complete likelihood

$$p_\theta(x_{1:T}, y_{1:T}) = \mu_\theta(x_1) \prod_{t=1}^{T} g_\theta(y_t \mid x_t) \prod_{t=1}^{T-1} f_\theta(x_{t+1} \mid x_t)$$

the problem would be **much** easier.

Key idea: Treat the state sequence $x_{1:T}$ as an *auxiliary variable* that is estimated together with θ.

The data augmentation strategy breaks the original problem into two new and closely linked problems.

Augmentation strategy via EM

Expectation maximization (EM) employs the complete likelihood $p_\theta(x_{1:T}, y_{1:T})$ as a **substitute** for the observed likelihood $p_\theta(y_{1:T})$.

They are of course related,

$$p_\theta(x_{1:T}, y_{1:T}) = p_\theta(x_{1:T} \mid y_{1:T}) p_\theta(y_{1:T}).$$

It can be shown that by iteratively maximizing

$$\mathcal{Q}(\theta, \theta_k) \triangleq \int \log p_\theta(x_{1:T}, y_{1:T}) p_{\theta_k}(x_{1:T} \mid y_{1:T}) \mathrm{d}x_{1:T} \\ = \mathrm{E}_{\theta_k}\left[\log p_\theta(x_{1:T}, y_{1:T}) \mid y_{1:T}\right],$$

w.r.t. θ, the obtained sequence $\{\theta_k\}_{k\geq 1}$ will result in a monotonical increase in likelihood values.

Sequential Monte Carlo – particle filter

SMC approximates a **sequence** of probability distributions on a sequence of probability spaces of **increasing dimension**.

Specifically (for nonlinear SSMs): SMC offers numerical approximations to the state estimation problems.

Ex: The filtering problem for the nonlinear SSM amounts to

$$p_\theta(x_t \mid y_{1:t}) = \frac{g_\theta(y_t \mid x_t) \int f_\theta(x_t \mid x_{t-1}) p_\theta(x_{t-1} \mid y_{1:t-1}) \mathrm{d}x_{t-1}}{p_\theta(y_t \mid y_{1:t-1})}$$

Here, the particle filter maintains an empirical approximation

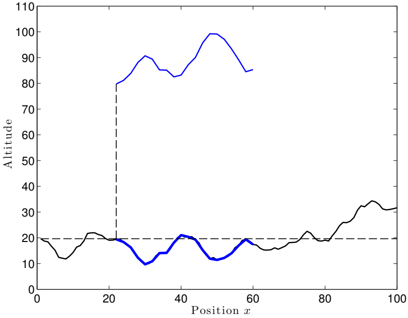
$$\widehat{p}_\theta(x_t \mid y_{1:t}) = \sum_{i=1}^{N} w_t^i \delta_{x_t^i}(x_t),$$

that converge to the true filtering distribution as $N \longrightarrow \infty$.

UPPSALA UNIVERSITET

The particle filter – toy problem

Consider a toy 1D localization problem.



Dynamic model:

$$x_{t+1} = x_t + u_t + v_t,$$

where x_t denotes position, u_t denotes velocity (known), $v_t \sim \mathcal{N}(0, 5)$ denotes an unknown disturbance.

Measurements:

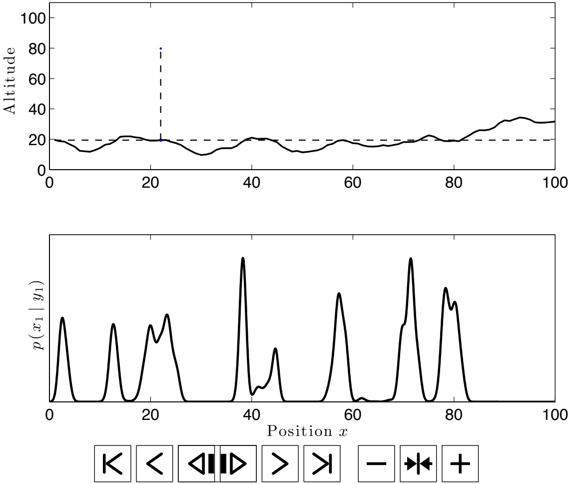
$$y_t = h(x_t) + e_t.$$

where $h(\cdot)$ denotes the world model (here the terrain height) and $e_t \sim \mathcal{N}(0, 1)$ denotes an unknown disturbance.

The same idea has been used in many applications, see e.g.

Thomas Schön, Fredrik Gustafsson, and Per-Johan Nordlund. **Marginalized particle filters for mixed linear/nonlinear state-space models**. *IEEE Transactions on Signal Processing*, 53(7):2279-2289, July 2005.

The particle filter – toy problem



Highlights two **key capabilities** of the PF:

1. Automatically handles an unknown and dynamically changing number of hypotheses.
2. Work with nonlinear/non-Gaussian models.

Using EM and particle smoothing together

Algorithm 1 EM for identifying nonlinear dynamical systems

1. **Initialise:** Set $k = 1$ and choose an initial θ_1.
2. **While** not converged **do:**

(a) **Expectation (E) step:** Compute

$$\mathcal{Q}(\theta, \theta_k) = \int \log p_\theta(x_{1:T}, y_{1:T}) \underbrace{p_{\theta_k}(x_{1:T} \mid y_{1:T})} \mathrm{d}x_{1:T}$$

using **sequential Monte Carlo** (particle smoother).

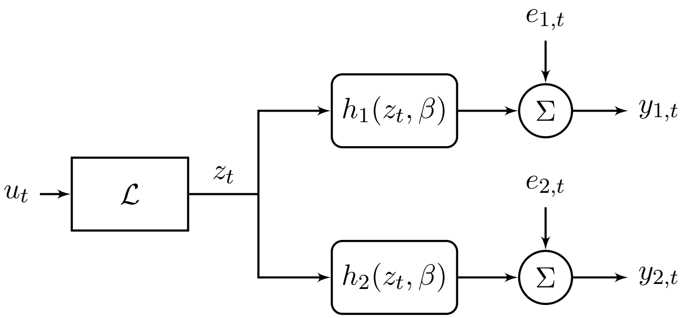
(b) **Maximization (M) step:** Compute $\theta_{k+1} = \arg\max_{\theta \in \Theta} \mathcal{Q}(\theta, \theta_k)$

(c) $k \leftarrow k + 1$

Thomas B. Schön, Adrian Wills and Brett Ninness. **System Identification of Nonlinear State-Space Models**. *Automatica*, 47(1):39-49, January 2011.

UPPSALA UNIVERSITET

Ex – blind Wiener identification (I/III)



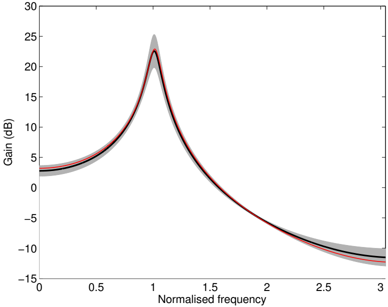
$$x_{t+1} = \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} x_t \\ u_t \end{pmatrix}, \qquad u_t \sim \mathcal{N}(0, Q),$$

$$z_t = Cx_t, \qquad y_t = h(z_t, \beta) + e_t, \qquad e_t \sim \mathcal{N}(0, R).$$

Id. problem: Find $\mathcal{L}$, β, r_1, and r_2 based on $\{y_{1,1:T}, y_{2,1:T}\}$.

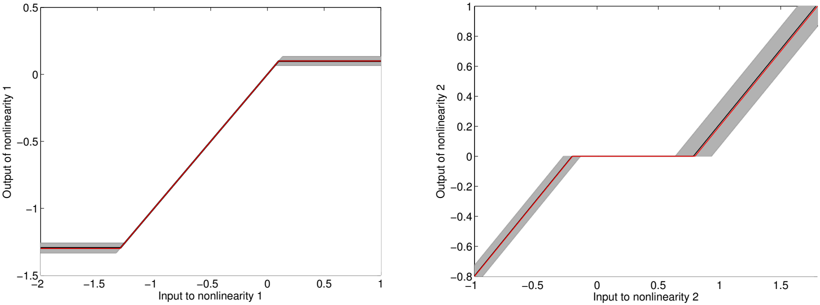
Ex – blind Wiener identification (II/III)

- Second order LGSS model with complex poles.
- Results obtained using $T = 1\,000$ samples.
- Employ the EM-PS with $N = 100$ particles.
- The plots are based on 100 realisations of data.
- Nonlinearities (dead zone and saturation) shown on next slide.



Bode plot of estimated mean (black), true system (red) and the result for all 100 realisations (gray).

Ex – blind Wiener identification (III/III)



Estimated mean (black), true static nonlinearity (red) and the result for all 100 realisations (gray).

Adrian Wills, Thomas B. Schön, Lennart Ljung and Brett Ninness. **Identification of Hammerstein-Wiener Models**. *Automatica*, 49(1): 70-81, January 2013.

Bayesian system identification

Problem formulation: Compute the posterior distribution

$$p(\theta \mid y_{1:T}) = \frac{p(y_{1:T} \mid \theta)\pi(\theta)}{p(y_{1:T})},$$

for the nonlinear SSM.

Key challenge: That there is no closed form expression available.

Solution: Construct a Markov chain with $p(\theta, x_{1:t} \mid y_{1:t})$ as its stationary distribution.

SMC is used to build an MCMC kernel with $p(\theta, x_{1:t} \mid y_{1:t})$ as its stationary distribution **without** introducing any systematic errors!

If there is time have a look at a Wiener example.

Conclusion

1. EM = state smoothing + optimization problem.
2. Conveyed the intuition underlying SMC.
3. Used SMC within EM to compute ML estimates of nonlinear state space models.

A lot of interesting research that remains to be done!!

Tutorial paper available on arXiv (1503.06058) this morning.

Thomas B. Schön, Fredrik Lindsten, Johan Dahlin, Johan Wågberg, Christian A. Naesseth, Andreas Svensson and Liang Dai. **Sequential Monte Carlo methods for system identification**. *Submitted to the 17th IFAC Symposium on System Identification (SYSID)*, Beijing, China, October 2015.

PhD course on the topic is available here

`user.it.uu.se/~thosc112/CIDS.html`

Some references to our work

Maximum likelihood system identification

Thomas B. Schön, Adrian Wills and Brett Ninness. **System Identification of Nonlinear State-Space Models**. *Automatica*, 47(1):39-49, January 2011.

Adrian Wills, Thomas B. Schön, Lennart Ljung and Brett Ninness. **Identification of Hammerstein-Wiener models**. *Automatica*, 49(1):70-81, January 2013.

Bayesian system identification

Fredrik Lindsten, Michael I. Jordan and Thomas B. Schön. **Particle Gibbs with ancestor sampling.** *Journal of Machine Learning Research (JMLR)*, 15:2145-2184, June 2014.

Fredrik Lindsten and Thomas B. Schön. **Backward simulation methods for Monte Carlo statistical inference.** *Foundations and Trends in Machine Learning*, 6(1):1-143, 2013.

Fredrik Lindsten, Thomas B. Schön and Michael I. Jordan. **Bayesian semiparametric Wiener system identification.** *Automatica*, 49(7):2053-2063, July 2013.

SMC in high dimensions

Christian A. Naesseth, Fredrik Lindsten and Thomas B. Schön. **Nested sequential Monte Carlo Methods**. *arXiv preprint*, February 2015.

SMC convergence in one slide...

Let $\varphi : \mathsf{X} \mapsto \mathbb{R}$ be some test function of interest. The expectation

$$\mathrm{E}_{\theta}\left[\varphi(x_t) \mid y_{1:t}\right] = \int \varphi(x_t) p_{\theta}(x_t \mid y_{1:t}) \mathrm{d}x_t,$$

can be estimated by the particle filter

$$\widehat{\varphi}_t^N \triangleq \sum_{i=1}^{N} w_t^i \varphi(x_t^i).$$

The **CLT** governing the convergence of this estimator states

$$\sqrt{N}\left(\widehat{\varphi}_t^N - \mathrm{E}_{\theta}\left[\varphi(x_t) \mid y_{1:t}\right]\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma_t^2(\varphi)).$$

The **likelihood estimate** $\widehat{p}_{\theta}(y_{1:t}) = \prod_{s=1}^{t} \left\{ \frac{1}{N} \sum_{i=1}^{N} \bar{w}_s^i \right\}$ from the PF is **unbiased**, $\mathrm{E}_{\psi_\theta}\left[\widehat{p}_{\theta}(y_{1:t})\right] = p_{\theta}(y_{1:t})$ for any value of N and there are **CLTs available** as well.