The Extended Parameter Filter

Yusuf B. Erol
Lei Li
Bharath Ramsundar
Stuart J. Russell

June 16, 2013
Model

Goal: To have an online Bayesian algorithm that can track $p(\theta \mid y_{0:T})$ and $p(x_t \mid y_{0:T})$. 

$x_t = f_\theta(x_{t-1}) + \nu_t$

$y_t = g(x_t) + \omega_t$
Particle Filter

- Particle filter is the standard tool for nonlinear and non-Gaussian state estimation problems.
- Particle filters fail under static parameters or deterministic behavior.
- Existing solutions are either biased or computationally expensive or restricted to specific model classes.
Storvik’s Filter

• Samples from $p(\theta \mid x_{0:t})$ at each time step for each particle path.
• Requires fixed-dimensional sufficient statistics, $p(\theta \mid x_{0:t}) = p(\theta \mid T_t)$, for constant-time online filtering.
• Restricted to a specific model class.

\[ x_t = F^T(x_{t-1})\theta + \epsilon_t, \quad \epsilon_t \sim N(0, Q) \]
Parameter Estimation

Separability

The Extended Parameter Filter

Separable Transition Models

\[ x_t = f_\theta(x_{t-1}) + v_t \]
\[ y_t = g(x_t) + w_t \]

Definition
A system is separable if the transition function \( f_\theta(x_{t-1}) \) can be written as \( f_\theta(x_{t-1}) = \ell(x_{t-1})^T h(\theta) \) for some arbitrary \( \ell(\cdot) \) and \( h(\cdot) \) and if the stochastic i.i.d. noise \( v_t \) has log-polynomial density.

Theorem
For a separable system, there exist fixed-dimensional sufficient statistics for the Gibbs density, \( p(\theta \mid x_0:T) \).

Main difference is \( h(\theta) \), which is an arbitrary mapping from \( \theta \rightarrow h(\theta) \).
Main Idea

- Generalizes Storvik’s filter to non-separable models via polynomial approximation.
- Given $f_\theta(\cdot)$, approximate it using polynomials to $\hat{f}_\theta(x) = \ell^T(x)h(\theta)$.
- Propagate-Weight-Resample in particle filtering fashion.
- Sample from $p(\theta \mid x_{0:t}) \approx \hat{p}(\theta \mid x_{0:t}) = \hat{p}(\theta \mid S_t)$ at each time step for each particle path.
The SIN model

\[ x_t = \sin(\theta x_{t-1}) + \nu_t, \quad \nu_t \sim N(0, \sigma^2) \]

\[ y_t = x_t + w_t, \quad w_t \sim N(0, \sigma_{obs}^2) \] (1)

\[ \sin(\theta x_{t-1}) \approx x_{t-1} \theta - \frac{1}{3!} x_{t-1}^3 \theta^3 + \frac{1}{5!} x_{t-1}^5 \theta^5 \ldots \]

\[ \approx \begin{bmatrix} x_{t-1} & -\frac{1}{3!} x_{t-1}^3 & \frac{1}{5!} x_{t-1}^5 & \ldots \end{bmatrix} \begin{bmatrix} \theta \\ \theta^3 \\ \theta^5 \\ \vdots \end{bmatrix} \]

(2)
Results

(a) Identifiability

(b) Approximating Gibbs
Results

(c) Particle filter (SIR), (d) Liu–West filter, $N = 50000$

(e) EPF, $N = 1000$
The Extended Parameter Filter

Fast online algorithm for both state and parameter estimation, with constant time per update;

Provable error bounds: Error reduces with number of particles and the order of approximation;

Introducing *separability*, an adequate condition for the existence of the sufficient statistic $T_t$ for the Gibbs density $p(\theta \mid x_{0:t})$;

Handling a general class of nonlinear/non-Gaussian models: polynomial approximation so as to generate approximate sufficient statistics for non-separable models.
Future Work

• Automating the polynomial approximation is necessary for high-dimensional parameter spaces.
• Sampling from **multivariate log-polynomial** densities can be done via Metropolis-Hastings or slice sampling. An efficient black-box sampler for such densities is needed.