Non-Linear Filtering in General State-Space Models

Sergio Hernandez
PhD Student
MSCS Victoria University of Wellington

Agenda

- General State Space Models
- Bayesian Implementation
- Extended Kalman Filters
- Particle Filters.
- Discussion

General State Space Models

- Graphical model for representing conditional dependencies between a set of random variables.
- The goal is to construct the PDF of the state based on all available information. This PDF constitutes the complete solution of the filtering problem.
- Optimal and sub-optimal solutions for the non-linear filtering problem uses linear approximations.

General State Space Models

- Non Linear State Space Model Definition

$x_{t+1} = f_{\theta}(x_t) + u_t$ State Equation

$y_t = g_{\theta}(x_t) + v_t$ Observation Equation

$\theta = \{u, v, f, g, x_0\}$ Model Parameters

Bayesian Implementation

- The Bayesian inference model defines

$$Prior = p(x_0, \theta)$$

$$p(x_{0:t}, \theta) = p(x_{0,} \theta) \prod_{t=1}^{t} p(x_t | x_{t-1}, \theta)$$

$$p(x_{0:t} | y_{1:t}, \theta) = \frac{p(x_{0:t}, \theta)\, p(y_{1:t} | x_{0:t}, \theta)}{\int p(x_{0:t}, \theta)\, p(y_{1:t} | x_{0:t}, \theta)\, d\, x_{0:t}}$$

Bayesian Implementation

- Prediction (Chapman-Kolmogorov):

$$p(x_t | y_{1:t-1}) = \int p(x_t | x_{t-1})\, p(x_{t-1} | y_{1:t-1})\, d\, x_{t-1}$$

- Update (Filtering):

$$p(x_t | y_{1:t}) = \frac{p(y_t | x_t)\, p(x_t | y_{1:t-1})}{\int p(y_t | x_t)\, p(x_t | y_{1:t-1})\, dx_t}$$

Extended Kalman Filter

- A linear filter means that all the parameters of the model are linear and all the noise comes from a Gaussian distribution.
- The EKF approximates the state and observation equations with the first-order Taylor series evaluated at the current state.
- If the “true” density is multimodal, then a single Gaussian may not work.

Extended Kalman Filter

$$p(x_{t-1}|y_{1:t-1}) \approx N(x_{t-1}; m_{t-1|t-1}, P_{t-1|t-1})$$

$$p(x_t|y_{1:t-1}) \approx N(x_t; m_{t|t-1}, P_{t|t-1})$$

$$p(x_t|y_{1:t}) \approx N(x_t; m_{t|t}, P_{t|t})$$

Extended Kalman Filter

$$m_{t|t-1} = f_t(m_{t-1|t-1})$$

$$m_{t|t} = m_{t|t-1} + K_t(y_t - g_t(m_{t|t-1}))$$

$$P_{t|t-1} = Q_{t-1} + \hat{F}_t P_{t-1|t-1} \hat{F}_t^T$$

$$P_{t|t} = P_{t|t-1} - K_t \hat{G}_t P_{t|t-1}$$

$$\hat{F}_t = \frac{\partial f_t(x)}{\partial x}$$

$$\hat{G}_t = \frac{\partial g_t(x)}{\partial x}$$

$$S_t = \hat{G}_t P_{t|t-1} \hat{G}_t^T + R_t$$

$$K_t = P_{t|t-1} \hat{G}_t^T S_t^{-1}$$

Sequential Monte Carlo (SMC)

- Importance sampling can replace the expectation step by an unbiased estimate.

$$p(x_{0:t}|y_{1:T},\theta)\simeq\sum_{i=1}^{T}\omega_t^i\,\delta(x_{0:t}-x_{0:t}^i) \qquad \sum_i \omega_t^i=1$$

- A small effective sample size implies large variance of the weights and is usually called weight degeneracy.

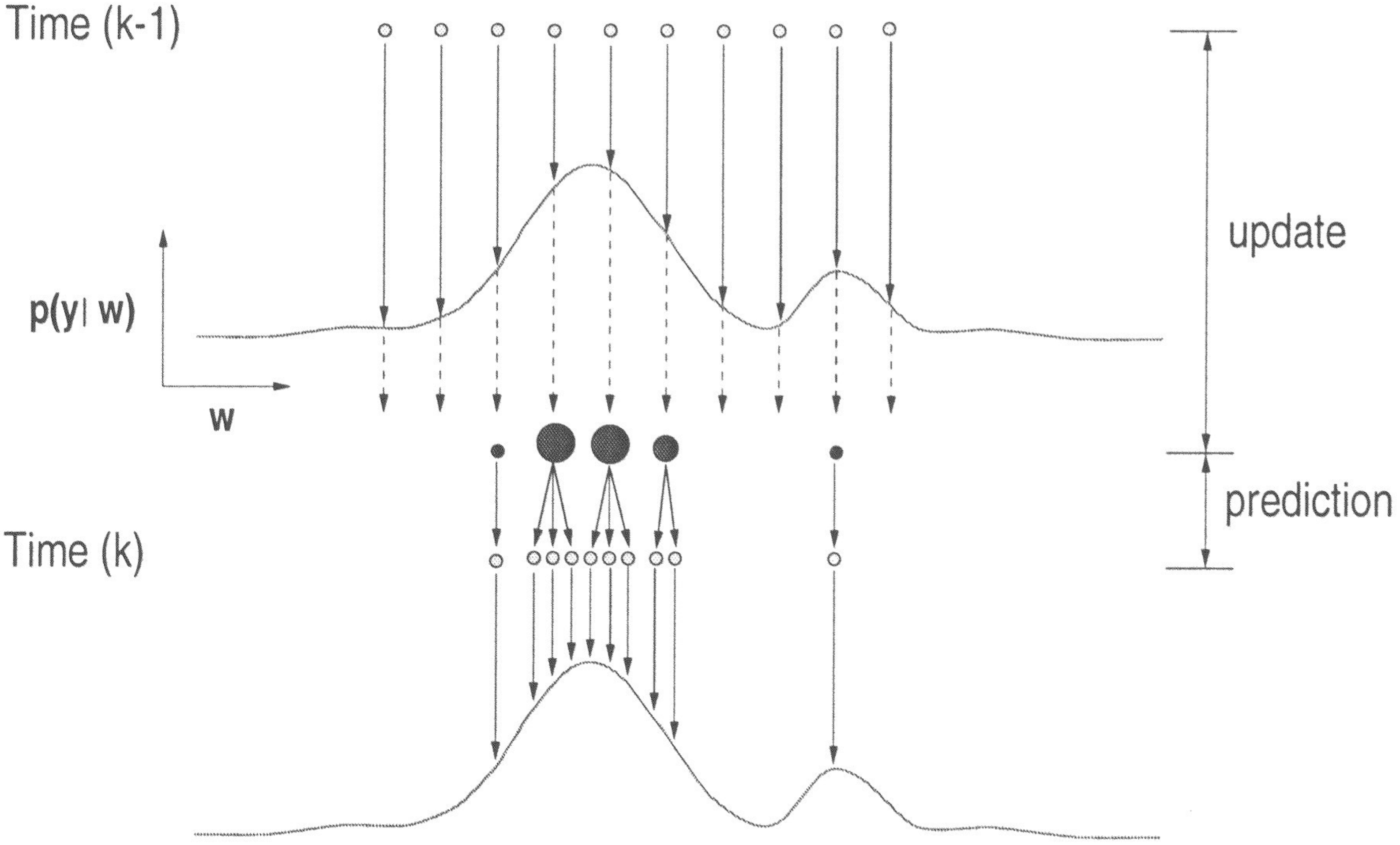
$$N_{eff}=\frac{N_s}{1+Var(\omega_t^i)}\simeq\frac{1}{\sum_{i=1}^{N_s}(\omega_t^i)}\leq N_s$$

Sequential Monte Carlo (SMC)

- SMC performs sequential approximations of probability distributions using importance sampling.
- Avoids the degeneracy problem by re-sampling at each iteration the points of the samples according to their weights.

$$p(x_t|y_{1:t},\theta)\simeq\sum_{i=1}^{N_t}\omega_T^i\,\delta(x_t-x_t^i)$$

Sequential Monte Carlo (SMC)



de Freitas, J. F. G., Niranjan, M., Gee, A. H., Doucet, A. Sequential Monte Carlo Methods to Train Neural Network Models Neural Comp. 2000 12: 955-993

Sequential Monte Carlo (SMC)

- Importance sample from transition prior

$$\tilde{x}_t \simeq p(x_t | x_{0:t-1}, y_{1:t})$$

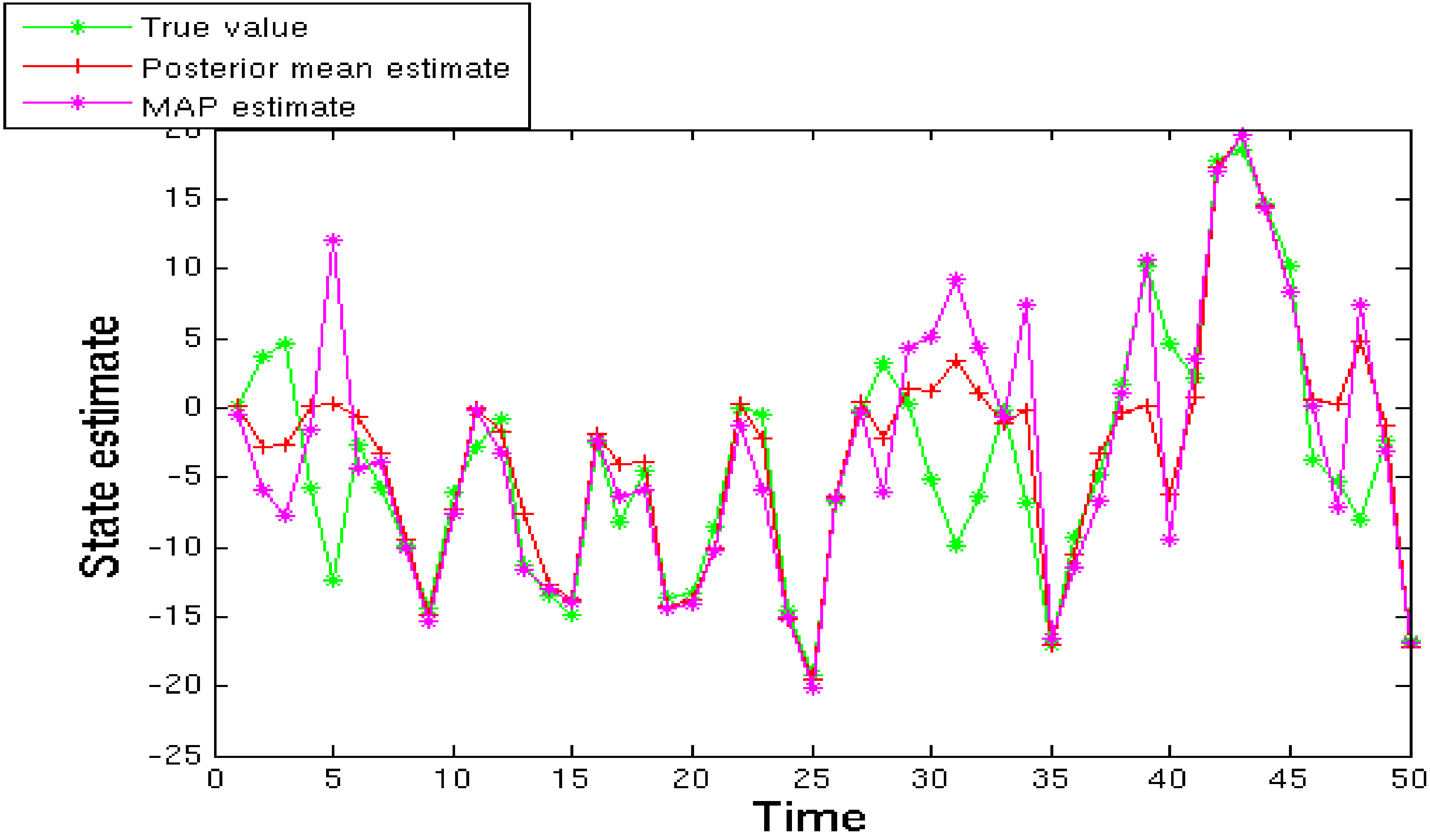
- Evaluate and normalize importance weights

$$\tilde{\omega}_t^i \simeq p(y_t | \tilde{x}_t^i)$$

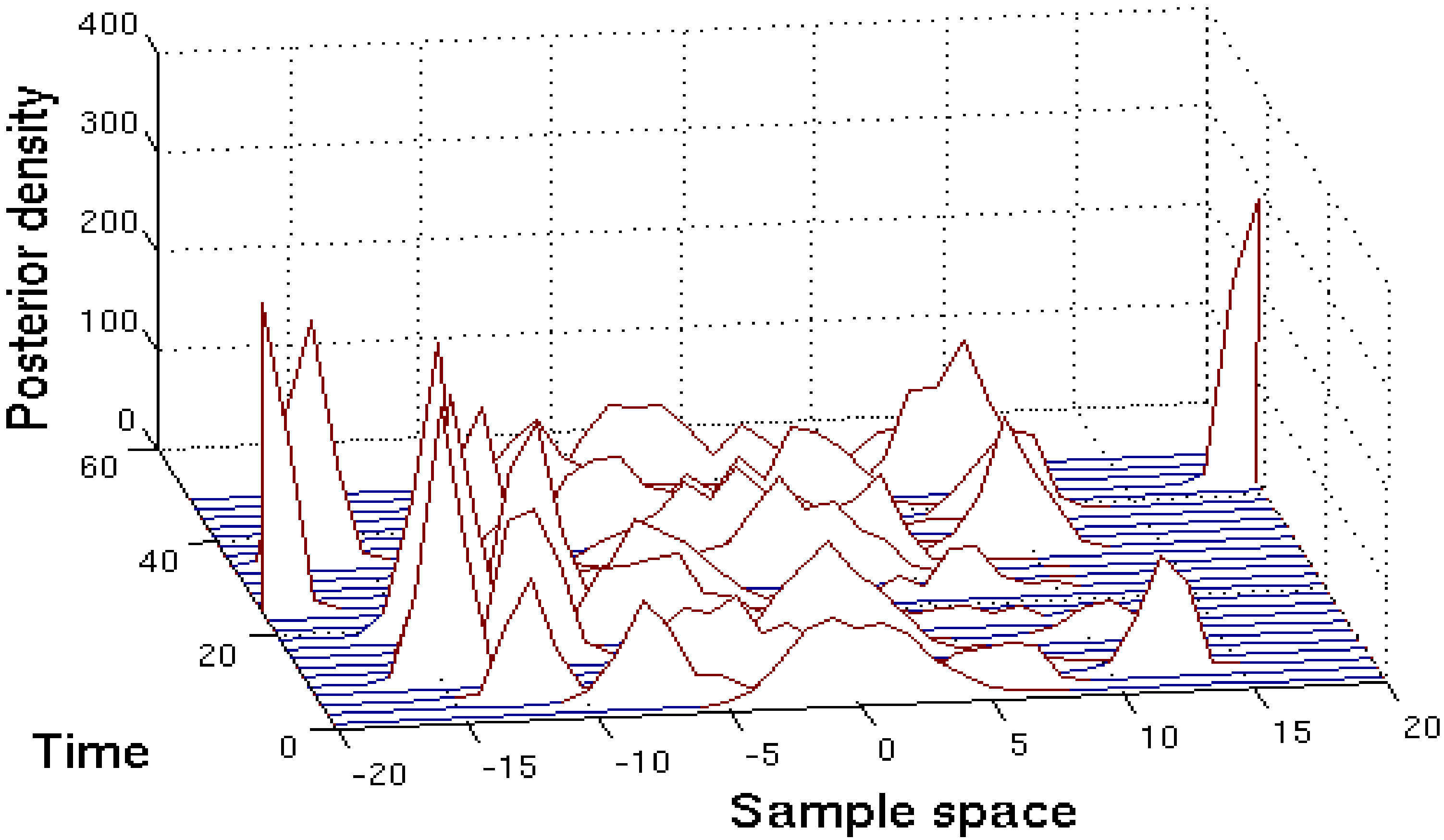
- Resample N particles from the importance sampling distribution.

$$\widehat{P}(x_{0:T} | y_{1:T}) = \sum \tilde{\omega}_T^i \delta(x_{0:T} - x_{0:T}^i)$$

Sequential Monte Carlo (SMC)



Sequential Monte Carlo (SMC)



Discussion

- SMC methods beats the “curse of dimensionality” and make tractable an un-tractable problem.
- Weight degeneracy and variance controlling is still a open discussion. The combination of deterministic and stochastic approaches has shown good results.
- It's essential to sample from the “right” importance distribution, so you can use new information to build a better proposal distribution.

Discussion

- SMC methods are almost an standard in tracking and computer vision problems requiring on-line processing.
- Some applications that can be explored using GSS and SMC:
  - Temporal clustering and classification.
  - Sequential patterns discovery.
  - Long term anomaly detection.
  - Reinforcement learning and causality analysis.

Some References

- Daum, F., "Nonlinear filters: beyond the Kalman filter," Aerospace and Electronic Systems Magazine, IEEE , vol.20, no.8pp. 57- 69, Aug. 2005.
- Arulampalam, M.S.; Maskell, S.; Gordon, N.; Clapp, T., "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on] , vol.50, no.2pp.174-188, Feb 2002
- Doucet, Arnaud; Freitas, Nando de; Gordon, Neil (Eds.) ,”Sequential Monte Carlo Methods in Practice”, Springer Verlag, New York, 2001.

Acknowledge

- Thanks to the Chilean Commission of Science and Technology (Conicyt) with the loan number 7172-CH of the World Bank and the Victoria University of Wellington who are funding my PhD.
- http://www.conicyt.cl
- http://www.pbct.cl

http://www.eso.org/projects/alma/