UPPSALA UNIVERSITET

Improving the particle filter in high dimensions using conjugate artificial process noise

Anna Wigren
joint work with Fredrik Lindsten and Lawrence Murray

Department of Information Technology, Uppsala University

July 4, 2018

anna.wigren@it.uu.se

Introduction

High-dimensional state space models arise in:

- oceanography
- numerical weather prediction
- epidemiology

anna.wigren@it.uu.se

Introduction

High-dimensional state space models arise in:

- oceanography
- numerical weather prediction
- epidemiology

Consider models of the form:

$$x_t = f(x_{t-1}, v_t)$$
$$y_t = Cx_t + e_t$$

anna.wigren@it.uu.se

UPPSALA UNIVERSITET

Introduction

High-dimensional state space models arise in:

- oceanography
- numerical weather prediction
- epidemiology

Consider models of the form:

$$x_t = f(x_{t-1}, v_t)$$
$$y_t = Cx_t + e_t$$

Current techniques solving the HD filtering problem include:

- the ensemble Kalman filter
- various versions of the particle filter

anna.wigren@it.uu.se

Some background on particle filters

Approximates the filtering distribution as:

$$\hat{p}^N(x_t|y_{1:t}) = \sum_{i=1}^{N} w_t^i \delta_{x_t^i}(x_t)$$

Often impossible to sample from true distribution $\rightarrow$ use a proposal q.

Two common choices:

- Standard proposal $p(x_t|x_{t-1})$
- Locally optimal proposal $p(x_t|x_{t-1}, y_t)$

[1] C. Snyder, T. Bengtsson, and M. Morzfeld. Performance bounds for particle filters using the optimal proposal. *Monthly Weather Review*, 143(11):4750–4761, 2015

anna.wigren@it.uu.se

Some background on particle filters

Approximates the filtering distribution as:

$$\hat{p}^N(x_t|y_{1:t}) = \sum_{i=1}^{N} w_t^i \delta_{x_t^i}(x_t)$$

Often impossible to sample from true distribution $\rightarrow$ use a proposal q.

Two common choices:

- Standard proposal $p(x_t|x_{t-1})$
- Locally optimal proposal $p(x_t|x_{t-1}, y_t)$

The optimal proposal has a minimal degeneracy [1].

BUT: Intractable unless $p(y_t|x_t)$ is conjugate to $p(x_t|x_{t-1})$.

---

[1] C. Snyder, T. Bengtsson, and M. Morzfeld. Performance bounds for particle filters using the optimal proposal. *Monthly Weather Review*, 143(11):4750–4761, 2015

anna.wigren@it.uu.se

The particle filter in HD

Suffers from degeneracy:
One particle weight is close to one, all others are close to zero — poor approximation!

Can be avoided if the number of particles increases exponentially with the state dimension [1].

[1] C. Snyder, T. Bengtsson, and M. Morzfeld. Performance bounds for particle filters using the optimal proposal. *Monthly Weather Review*, 143(11):4750–4761, 2015

anna.wigren@it.uu.se

The particle filter in HD

Suffers from degeneracy:
One particle weight is close to one, all others are close to zero — poor approximation!

Can be avoided if the number of particles increases exponentially with the state dimension [1].

Previous particle filter adaptions for HD-problems:

- Approximate the particle filter

We suggest:

- Approximate the model (and use regular particle filters)

[1] C. Snyder, T. Bengtsson, and M. Morzfeld. Performance bounds for particle filters using the optimal proposal. *Monthly Weather Review*, 143(11):4750–4761, 2015

anna.wigren@it.uu.se

The approximate model

Add artificial process noise in an extra state update:

$$
\begin{aligned}
x'_t &= f(x_{t-1}, v_t) \\
x_t &= x'_t + \varepsilon \xi_t \\
y_t &= C x_t + e_t
\end{aligned}
$$

where

- ε adjusts the magnitude of the artificial noise
- $\xi_t \sim \mathcal{N}(0, S)$ where S is a covariance matrix

Note:
The second and third rows correspond to a linear-Gaussian model.

anna.wigren@it.uu.se

A particle filter for the approximate model

Take x_t' and x_t as states, use the combined proposal:

$$q(x_t, x_t' | x_{t-1}, y_t) = p(x_t' | x_{t-1}) p(x_t | x_t', y_t) = p(x_t' | x_{t-1}) \mathcal{N}(x_t | \mu, \Sigma)$$

anna.wigren@it.uu.se

UPPSALA UNIVERSITET

A particle filter for the approximate model

Take x'_t and x_t as states, use the combined proposal:

$$q(x_t, x'_t | x_{t-1}, y_t) = p(x'_t | x_{t-1}) p(x_t | x'_t, y_t) = p(x'_t | x_{t-1}) \mathcal{N}(x_t | \mu, \Sigma)$$

Step-by-step:

1) Propagate standard
2) Propagate optimal
3) Update weight optimal

$$x'_t = f(x_{t-1}, v_t)$$
$$x_t = x'_t + \varepsilon \xi_t$$
$$y_t = C x_t + e_t$$

anna.wigren@it.uu.se

A particle filter for the approximate model

Take x'_t and x_t as states, use the combined proposal:

$$q(x_t, x'_t | x_{t-1}, y_t) = p(x'_t | x_{t-1}) p(x_t | x'_t, y_t) = p(x'_t | x_{t-1}) \mathcal{N}(x_t | \mu, \Sigma)$$

Step-by-step:

1) Propagate standard
2) Propagate optimal
3) Update weight optimal

$$x'_t = f(x_{t-1}, v_t)$$
$$x_t = x'_t + \varepsilon \xi_t$$
$$y_t = C x_t + e_t$$

Intuition:

Adding noise + optimal proposal shifts particles towards observation.

anna.wigren@it.uu.se

UPPSALA UNIVERSITET

A particle filter for the approximate model

Take x_t' and x_t as states, use the combined proposal:

$$q(x_t, x_t' | x_{t-1}, y_t) = p(x_t' | x_{t-1}) p(x_t | x_t', y_t) = p(x_t' | x_{t-1}) \mathcal{N}(x_t | \mu, \Sigma)$$

Step-by-step:

1) Propagate standard
2) Propagate optimal
3) Update weight optimal

$$x_t' = f(x_{t-1}, v_t)$$
$$x_t = x_t' + \varepsilon \xi_t$$
$$y_t = C x_t + e_t$$

Intuition:

Adding noise + optimal proposal shifts particles towards observation.

Alternative view:

Bias-variance trade-off

anna.wigren@it.uu.se

Numerical results

Evaluate the method using two different models:

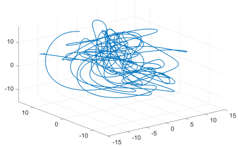
- Linear Gaussian:

$$x_t = Ax_{t-1} + v_t,$$

- Lorenz'96:

$$dx_k = \big((x_{k+1} - x_{k-2})x_{k-1} - x_k + F\big)dt + b(x_k)dW_k,$$

Observations: $y_t = Cx_t + e_t$



anna.wigren@it.uu.se

UPPSALA UNIVERSITET

Numerical results

Evaluate the method using two different models:

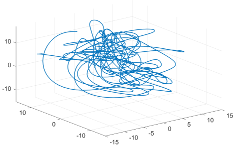
- Linear Gaussian:

$$x_t = Ax_{t-1} + v_t,$$

- Lorenz'96:

$$dx_k = \big((x_{k+1} - x_{k-2})x_{k-1} - x_k + F\big)dt + b(x_k)dW_k,$$

Observations: $y_t = Cx_t + e_t$



Performance measures:

- Marginal log-likelihood, $\log Z = \log p(y_{1:T})$
- MSE of state estimate
- Effective sample size, $N_{eff} = 1/\sum_{i=1}^{N}(w_{t-1}^i)^2$

anna.wigren@it.uu.se

Choosing covariance matrix

Conjugate artificial noise

$$\xi_t \sim \mathcal{N}(0, S)$$

anna.wigren@it.uu.se

Choosing covariance matrix

Conjugate artificial noise

$$\xi_t \sim \mathcal{N}(0, S)$$

The identity matrix:

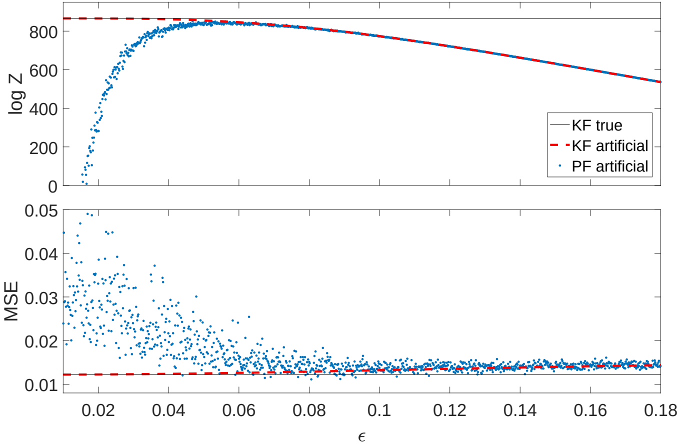
- Noise of the same magnitude added to all states
- No correlation between states

The weighted sample covariance matrix:

- Allows for correlation between states

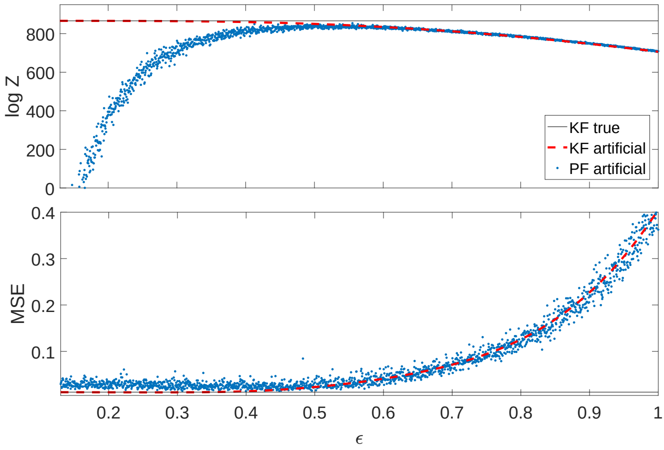
anna.wigren@it.uu.se

Linear Gaussian (identity)



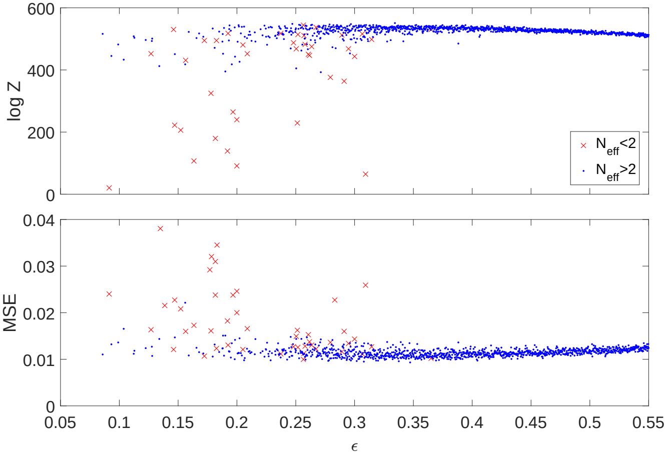
anna.wigren@it.uu.se

Linear Gaussian (sample covariance)



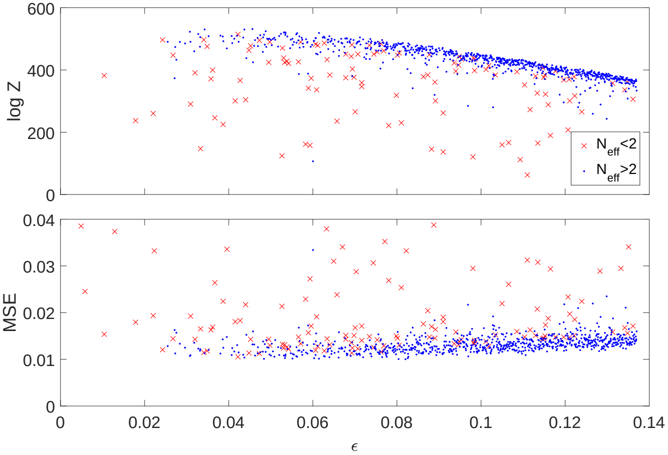
anna.wigren@it.uu.se

Lorenz'96 (sample covariance)



anna.wigren@it.uu.se

Lorenz’96 (identity)



anna.wigren@it.uu.se

Conclusion

- Adding artificial process noise can significantly improve the performance
- Considered filtering problem, but can be used for parameter estimation

anna.wigren@it.uu.se

Conclusion

- Adding artificial process noise can significantly improve the performance
- Considered filtering problem, but can be used for parameter estimation

Future work:

- If HD enough this method also breaks down $\rightarrow$ combine with other methods
- How to choose ε and the covariance matrix?
- Use the method on real application

anna.wigren@it.uu.se

Conclusion

- Adding artificial process noise can significantly improve the performance
- Considered filtering problem, but can be used for parameter estimation

Future work:

- If HD enough this method also breaks down $\rightarrow$ combine with other methods
- How to choose ε and the covariance matrix?
- Use the method on real application

More details:
https://arxiv.org/abs/1801.07000

anna.wigren@it.uu.se