Exact Approximation of Rao-Blackwellized Particle Filters

Adam M. Johansen[⋆], Nick Whiteley[†] and Arnaud Doucet[‡]

[⋆]University of Warwick Department of Statistics,
[†]University of Bristol School of Mathematics,
[‡]University of Oxford Department of Statistics

a.m.johansen@warwick.ac.uk
http://go.warwick.ac.uk/amjohansen/talks

July 11th, 2012
16th IFAC SysID

Context

Filtering in State-Space Models:

- SIR Particle Filters [GSS93]
- Rao-Blackwellized Particle Filters [AD02, CL00]

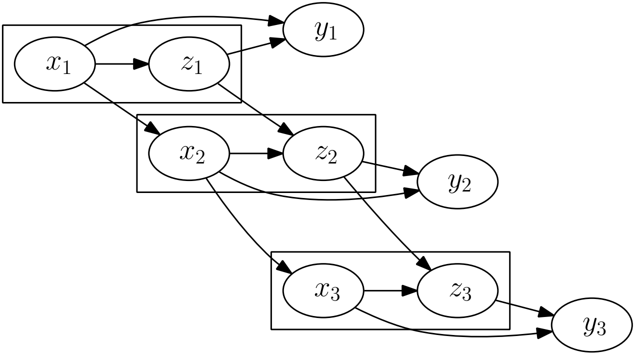
Exact Approximation of Monte Carlo Algorithms:

- Particle MCMC [ADH10]
- Local SMC [JD12]

Approximating the RBPF

- Approximated Rao-Blackwellized Particle Filters [CSOL11]

A (Rather Broad) Class of Hidden Markov Models



- Unobserved Markov chain $\{(X_n, Z_n)\}$ transition f.
- Observed process $\{Y_n\}$ conditional density g.
- Density:

$$p(x_{1:n}, z_{1:n}, y_{1:n}) = f_1(x_1, z_1) g(y_1 | x_1, z_1) \prod_{i=2}^{n} f(x_i, z_i | x_{i-1}, z_{i-1}) g(y_i | x_i, z_i).$$

Formal Solutions

- Filering and Prediction Recursions:

$$p(x_n, z_n | y_{1:n}) = \frac{p(x_n, z_n | y_{1:n-1}) g(y_n | x_n, z_n)}{\int p(x'_n, z'_n | y_{1:n-1}) g(y_n | x'_n, z'_n) d(x'_n, z'_n)}$$

$$p(x_{n+1}, z_{n+1} | y_{1:n}) = \int p(x_n, z_n | y_{1:n}) f(x_{n+1}, z_{n+1} | x_n, z_n) d(x_n, z_n)$$

- Smoothing:

$$p((x,z)_{1:n} | y_{1:n}) \propto p((x,z)_{1:n-1} | y_{1:n-1}) f((x,z)_n | (x,z)_{n-1}) g(y_n | (x,z)_n)$$

A Simple SIR Filter

Algorithmically, at iteration n:

- Given $\{W_{n-1}^i, (X,Z)_{1:n-1}^i\}$ for $i = 1, \ldots, N$:
- **Resample**, obtaining $\{\frac{1}{N}, (\widetilde{X}, \widetilde{Z})_{1:n-1}^i\}$.
  - Sample $(X,Z)_n^i \sim q_n(\cdot|(\widetilde{X}, \widetilde{Z})_{n-1}^i)$
  - Weight $W_n^i \propto \frac{f((X,Z)_n^i|(\widetilde{X},\widetilde{Z})_{n-1}^i)g(y_n|(X,Z)_n^i)}{q_n((X,Z)_n^i|(\widetilde{X},\widetilde{Z})_{n-1}^i)}$

Actually:

- Resample efficiently.
- Only resample when necessary.
- ...

A Rao-Blackwellized SIR Filter

Algorithmically, at iteration n:

- Given $\{W_{n-1}^{X,i}, (X_{1:n-1}^i, p(z_{1:n-1}|X_{1:n-1}^i, y_{1:n-1})\}$
- **Resample**, obtaining $\{\frac{1}{N}, (\widetilde{X}_{1:n-1}^i, p(z_{1:n-1}|\widetilde{X}_{1:n-1}^i, y_{1:n-1}))\}$.
- For $i = 1, \ldots, N$:
  - Sample $X_n^i \sim q_n(\cdot|\widetilde{X}_{n-1}^i)$
  - Set $X_{1:n}^i \leftarrow (\widetilde{X}_{1:n-1}^i, X_n^i)$.
  - Weight $W_n^{X,i} \propto \frac{p(X_n^i, y_n|\widetilde{X}_{n-1}^i)}{q_n(X_n^i|\widetilde{X}_{n-1}^i)}$
  - Compute $p(z_{1:n}|y_{1:n}, X_{1:n}^i)$.

Requires analytically tractable substructure.

An Approximate Rao-Blackwellized SIR Filter

Algorithmically, at iteration n:

- Given $\{W_{n-1}^{X,i}, (X_{1:n-1}^{i}, \widehat{p}(z_{1:n-1}|X_{1:n-1}^{i}, y_{1:n-1})\}$
- **Resample**, obtaining $\{\frac{1}{N}, (\widetilde{X}_{1:n-1}^{i}, \widehat{p}(z_{1:n-1}|\widetilde{X}_{1:n-1}^{i}, y_{1:n-1}))\}$.
- For $i = 1, \ldots, N$:
  - Sample $X_n^i \sim q_n(\cdot|\widetilde{X}_{n-1}^{i})$
  - Set $X_{1:n}^{i} \leftarrow (\widetilde{X}_{1:n-1}^{i}, X_n^i)$.
  - Weight $W_n^{X,i} \propto \frac{\widehat{p}(X_n^i, y_n|\widetilde{X}_{n-1}^{i})}{q_n(X_n^i|\widetilde{X}_{n-1}^{i})}$
  - Compute $\widehat{p}(z_{1:n}|y_{1:n}, X_{1:n}^{i})$.

Is approximate; how does error accumulate?

Exactly Approximated Rao-Blackwellized SIR Filter

At time $n = 1$

- Sample, $X_1^i \sim q^x\left(\cdot \mid y_1\right)$.
- Sample, $Z_1^{i,j} \sim q^z\left(\cdot \mid X_1^i, y_1\right)$.
- Compute and normalise the local weights

$$w_1^z\left(X_1^i, Z_1^{i,j}\right) := \frac{p(X_1^i, y_1, Z_1^{i,j})}{q^z\left(Z_1^{i,j} \middle| X_1^i, y_1\right)}, W_1^{z,i,j} := \frac{w_1^z\left(X_1^i, Z_1^{i,j}\right)}{\sum_{k=1}^M w_1^z\left(X_1^i, Z_1^{i,k}\right)},$$

$$\text{define } \widehat{p}(X_1^i, y_1) := \frac{1}{M}\sum_{j=1}^M w_1^z\left(X_1^i, Z_1^{i,j}\right).$$

- Compute and normalise the top-level weights

$$w_1^x\left(X_1^i\right) := \frac{\widehat{p}(X_1^i, y_1)}{q^x\left(X_1^i | y_1\right)}, \; W_1^{x,i} := \frac{w_1^x\left(X_1^i\right)}{\sum_{k=1}^N w_1^x\left(X_1^k\right)}.$$

At times $n \geq 2$

- Resample

$$\left\{ W_{n-1}^{x,i}, \left(X_{1:n-1}^{i}, \left\{ W_{n-1}^{z,i,j}, Z_{1:n-1}^{i,j} \right\}_{j} \right) \right\}_{i}$$

to obtain

$$\left\{ \frac{1}{N}, \left(\widetilde{X}_{1:n-1}^{i}, \left\{ \overline{W}_{n-1}^{z,i,j}, \overline{Z}_{1:n-1}^{i,j} \right\}_{j} \right) \right\}_{i}.$$

- Resample $\{\overline{W}_{n-1}^{z,i,j}, \overline{Z}_{1:n-1}^{i,j}\}_j$ to obtain $\{\frac{1}{M}, \widetilde{Z}_{1:n-1}^{i,j}\}_j$.
- Sample $X_n^i \sim q^x(\cdot|\widetilde{X}_{1:n-1}^i, y_{1:n})$; set $X_{1:n}^i := (\widetilde{X}_{1:n-1}^i, X_n^i)$.
- Sample $Z_n^{i,j} \sim q^z \left(\cdot |\, X_{1:n}^i, y_{1:n}, \widetilde{Z}_{1:n-1}^{i,j} \right)$; set $Z_{1:n}^{i,j} := (\widetilde{Z}_{1:n-1}^{i,j}, Z_n^{i,j})$.

- Compute and normalise the local weights

$$w_n^z\left(X_{1:n}^i, Z_{1:n}^{i,j}\right) := \frac{p\left(X_n^i, y_n, Z_n^{i,j} \,\middle|\, \widetilde{X}_{n-1}^i, \widetilde{Z}_{n-1}^{i,j}\right)}{q^z\left(Z_n^{i,j} \,\middle|\, X_{1:n}^i, y_{1:n}, \widetilde{Z}_{1:n-1}^{i,j}\right)},$$

$$\begin{aligned} \widehat{p}(X_n^i, y_n | \widetilde{X}_{1:n-1}^i, y_{1:n-1}) : &= \frac{1}{M}\sum_{j=1}^{M} w_n^z\left(X_{1:n}^i, Z_{1:n}^{i,j}\right), \\ W_n^{z,i,j} : &= \frac{w_n^z\left(X_{1:n}^i, Z_{1:n}^{i,j}\right)}{\sum_{k=1}^{M} w_n^z\left(X_{1:n}^i, Z_{1:n}^{i,k}\right)}. \end{aligned}$$

- Compute and normalise the top-level weights

$$w_n^x\left(X_{1:n}^i\right) := \frac{\widehat{p}(X_n^i, y_n | \widetilde{X}_{1:n-1}^i, y_{1:n-1})}{q^x(X_n^i | \widetilde{X}_{1:n-1}^i, y_{1:n})},$$

$$W_n^{x,i} := \frac{w_n^x\left(X_{1:n}^i\right)}{\sum_{k=1}^{N} w_n^x\left(X_{1:n}^k\right)}.$$

How can this be justified?

- As an extended space SIR algorithm.
- Via unbiased estimation arguments.

Note also the $M = 1$ and $M \to \infty$ cases.

How does this differ from CSOL11?

Principally in the *local* weights, benefits including:

- Valid (N-consistent) for all $M \geq 1$ rather than (M, N)-consistent.
- Computational cost $\mathcal{O}(MN)$ rather than $\mathcal{O}(M^2N)$.
- Only requires knowledge of joint behaviour of x or z; doesn't require say $p(x_n|x_{n-1}, z_{n-1})$.

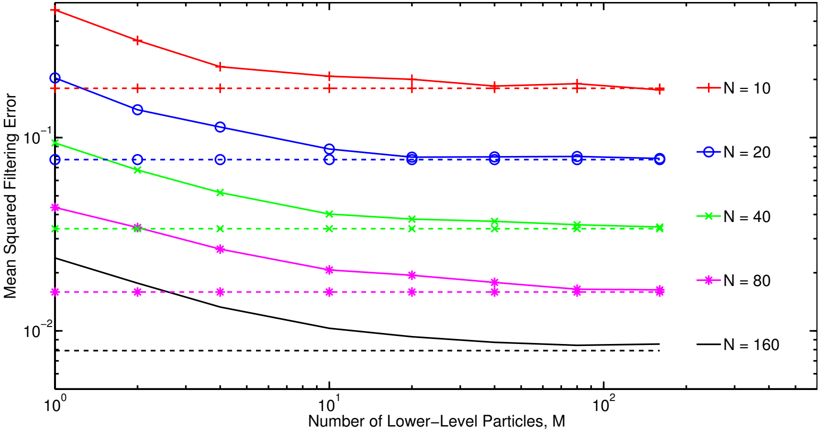
Toy Example: Model

We use a simulated sequence of 100 observations from the model defined by the densities:

$$
\begin{aligned}
\mu(x_1, z_1) =& \mathcal{N}\left(\left(\begin{array}{c} x_1 \\ z_1 \end{array}\right); \left(\begin{array}{c} 0 \\ 0 \end{array}\right) \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right]\right) \\
f(x_n, z_n | x_{n-1}, z_{n-1}) =& \mathcal{N}\left(\left(\begin{array}{c} x_n \\ z_n \end{array}\right); \left(\begin{array}{c} x_{n-1} \\ z_{n-1} \end{array}\right), \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right]\right) \\
g(y_n | x_n, z_n) =& \mathcal{N}\left(y_n; \left(\begin{array}{c} x_n \\ z_n \end{array}\right), \left[\begin{array}{cc} \sigma_x^2 & 0 \\ 0 & \sigma_z^2 \end{array}\right]\right)
\end{aligned}
$$

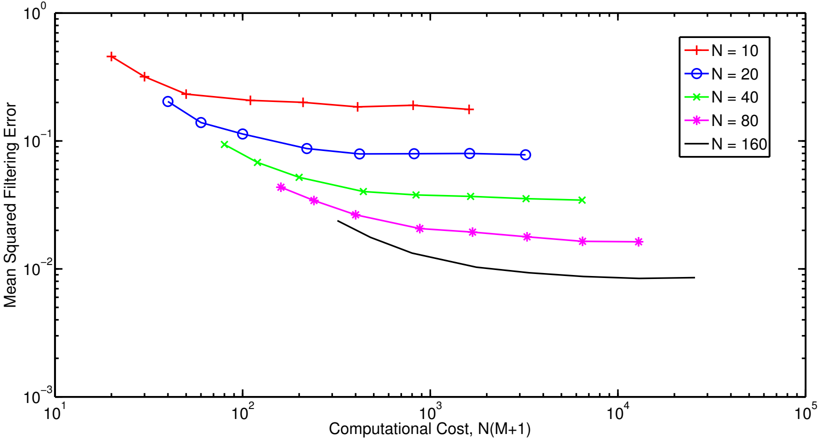
Consider IMSE (relative to optimal filter) of filtering estimate of first coordinate marginals.

Approximation of the RBPF



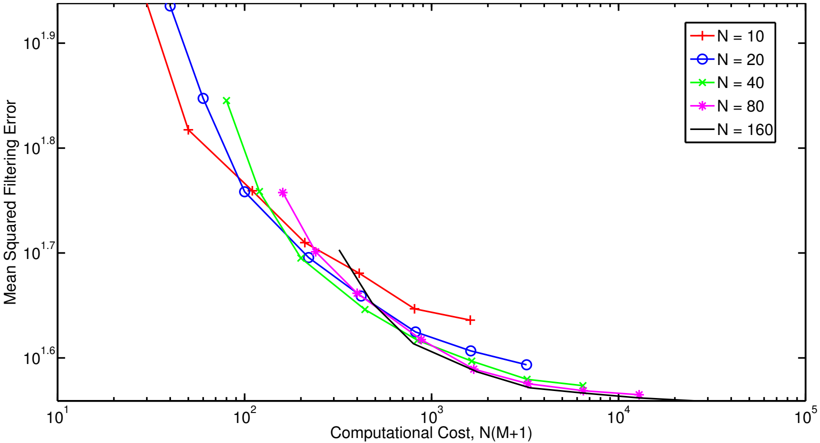
For $\sigma_x^2 = \sigma_z^2 = 1$.

Computational Performance



For $\sigma_x^2 = \sigma_z^2 = 1$.

Computational Performance



For $\sigma_x^2 = 10^2$ and $\sigma_z^2 = 0.1^2$.

In Conclusion

- SMC can be used hierarchically.
- Software implementation is not difficult [Joh09].
- The Rao-Blackwellized particle filter can be approximated *exactly*
  - Can reduce estimator variance at fixed cost.
  - Attractive for distributed/parallel implementation.
  - Allows combination of different sorts of particle filter.
  - Can be combined with other techniques for parameter estimation etc..

References

C. Andrieu and A. Doucet.
Particle filtering for partially observed Gaussian state space models.
Journal of the Royal Statistical Society B, 64(4):827–836, 2002.

C. Andrieu, A. Doucet, and R. Holenstein.
Particle Markov chain Monte Carlo.
Journal of the Royal Statistical Society B, 72(3):269–342, 2010.

R. Chen and J. S. Liu.
Mixture Kalman filters.
Journal of the Royal Statistical Society B, 62(3):493–508, 2000.

T. Chen, T. Schön, H. Ohlsson, and L. Ljung.
Decentralized particle filter with arbitrary state decomposition.
IEEE Transactions on Signal Processing, 59(2):465–478, February 2011.

N. J. Gordon, S. J. Salmond, and A. F. M. Smith.
Novel approach to nonlinear/non-Gaussian Bayesian state estimation.
IEE Proceedings-F, 140(2):107–113, April 1993.

A. M. Johansen and A. Doucet.
Hierarchical particle sampling for intractable state-space models.
CRiSM working paper, University of Warwick, 2012. In preparation.

A. M. Johansen.
SMCTC: Sequential Monte Carlo in C++.
Journal of Statistical Software, 30(6):1–41, April 2009.