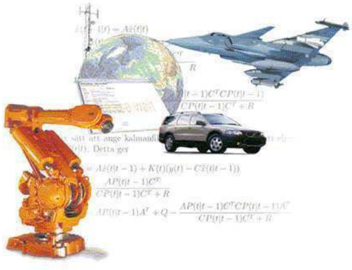
Ancestral Sampling for Particle Gibbs



Fredrik Lindsten[⋆], **Michael I. Jordan**[⋆⋆], **Thomas B. Schön**[⋆]

[⋆]Division of Automatic Control
Linköping University, Sweden

[⋆⋆]Departments of EECS and Statistics,
University of California, Berkeley, USA

This talk

Two things:

- Improve the mixing of particle Gibbs by “ancestor sampling”.
- Application to non-Markovian models.

Problem formulation

- High-dimensional target $\bar{\gamma}_T(x_{1:T}, \theta)$ on $\mathsf{X}^T \times \Theta$.
- Sample from $\bar{\gamma}_T(x_{1:T}, \theta)$ using MCMC.

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Problem formulation

- High-dimensional target $\bar{\gamma}_T(x_{1:T}, \theta)$ on $\mathsf{X}^T \times \Theta$.
- Sample from $\bar{\gamma}_T(x_{1:T}, \theta)$ using MCMC.

ex) State-space model,

$$\bar{\gamma}_T(x_{1:T}, \theta) = p(x_{1:T}, \theta \mid y_{1:T}).$$

Ideal Gibbs sampler,

1. Draw $x'_{1:T} \sim p(x_{1:T} \mid \theta, y_{1:T})$;
2. Draw $\theta' \sim p(\theta \mid x_{1:T}, y_{1:T})$.

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Particle MCMC

Particle Markov chain Monte Carlo (PMCMC),

- Use sequential Monte Carlo (SMC) to sample from $\bar{\gamma}_T(x_{1:T})$.
- Particle independent Metropolis-Hastings (PIMH)
- Particle Gibbs (PG)

N.B. Sampling θ straightforward. We drop θ to simplify notation!

Particle MCMC

Particle Markov chain Monte Carlo (PMCMC),

- Use sequential Monte Carlo (SMC) to sample from $\bar{\gamma}_T(x_{1:T})$.
- Particle independent Metropolis-Hastings (PIMH)
- **Particle Gibbs (PG)**

N.B. Sampling θ straightforward. We drop θ to simplify notation!

Particle MCMC

Particle Markov chain Monte Carlo (PMCMC),

- Use sequential Monte Carlo (SMC) to sample from $\bar{\gamma}_T(x_{1:T})$.
- Particle independent Metropolis-Hastings (PIMH)
- **Particle Gibbs (PG)**

Sequential Monte Carlo,

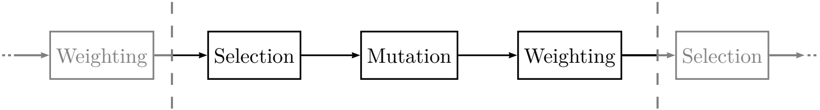
- Sequence of target densities, for $t = 1, \dots, T$,

$$\bar{\gamma}_t(x_{1:t}) = \frac{\gamma_t(x_{1:t})}{Z_t}.$$

- Approximated by collections of weighted particles.

N.B. Sampling θ straightforward. We drop θ to simplify notation!

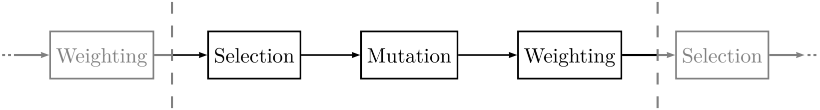
Sequential Monte Carlo – the particle filter



- Selection: $\{x_{1:t-1}^m, w_{t-1}^m\}_{m=1}^N \rightarrow \{\tilde{x}_{1:t-1}^m, 1/N\}_{m=1}^N$.
- Mutation: $x_t^m \sim R_t(dx_t \mid \tilde{x}_{1:t-1}^m)$ and $x_{1:t}^m = \{\tilde{x}_{1:t-1}^m, x_t^m\}$.
- Weighting: $w_t^m = W_t(x_{1:t}^m)$.

$$\Rightarrow \{x_{1:t}^m, w_t^m\}_{m=1}^N$$

Sequential Monte Carlo – the particle filter



- Selection + Mutation:

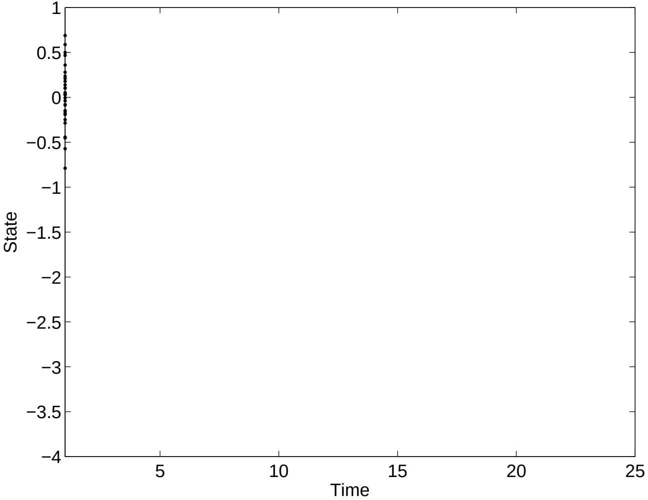
$$(a_t^m, x_t^m) \sim M_t(a_t, x_t) = \frac{w_{t-1}^{a_t}}{\sum_l w_{t-1}^l} R_t(x_t \mid x_{1:t-1}^{a_t}).$$

- Weighting: $w_t^m = W_t(x_{1:t}^m)$.

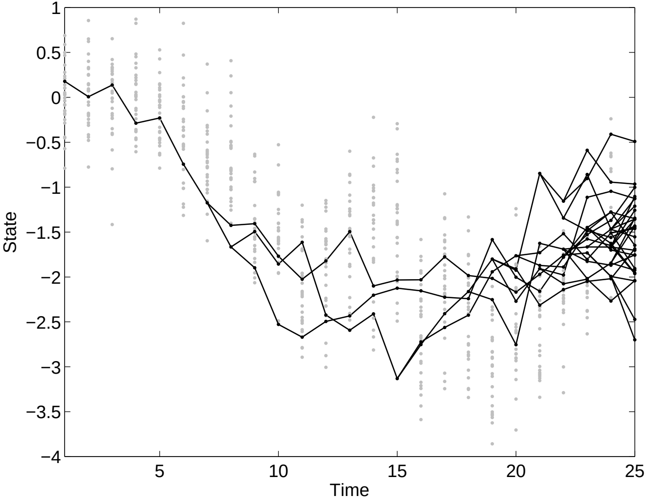
$$\Rightarrow \{x_{1:t}^m, w_t^m\}_{m=1}^N$$

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Path degeneracy



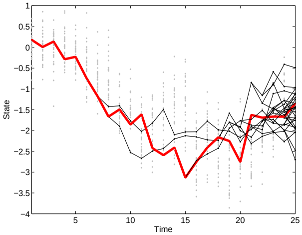
Path degeneracy



Sampling based on SMC

- With $P(x'_{1:T} = x^m_{1:T}) \propto w^m_T$ we get,

$$x'_{1:T} \overset{\text{approx.}}{\sim} \bar{\gamma}_T(x_{1:T}).$$

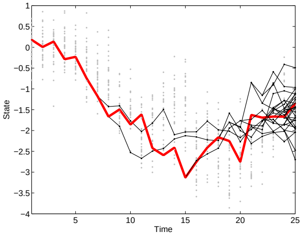


Sampling based on SMC

- With $P(x'_{1:T} = x^m_{1:T}) \propto w^m_T$ we get,

$$x'_{1:T} \overset{\text{approx.}}{\sim} \bar{\gamma}_T(x_{1:T}).$$

- Approximation can be arbitrarily bad (for small N)!

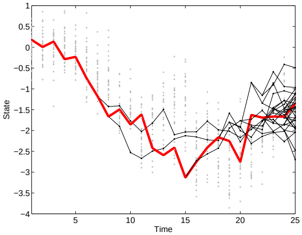


Sampling based on SMC

- With $P(x'_{1:T} = x^m_{1:T}) \propto w^m_T$ we get,

$$x'_{1:T} \overset{\text{approx.}}{\sim} \bar{\gamma}_T(x_{1:T}).$$

- Approximation can be arbitrarily bad (for small N)!
- Compensate for approximation: SMC within MCMC = PMCMC.



AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Extended target distribution

SMC generates a sample on $\mathsf{X}^{NT} \times \{1, \ldots, N\}^{N(T-1)}$ with density,

$$\psi(\mathbf{x}_{1:T}, \mathbf{a}_{2:T}) \triangleq \prod_{m=1}^{N} R_1(x_1^m) \prod_{t=2}^{T} \prod_{m=1}^{N} M_t(a_t^m, x_t^m).$$

Extended target distribution

SMC generates a sample on $\mathsf{X}^{NT} \times \{1, \ldots, N\}^{N(T-1)}$ with density,

$$\psi(\mathbf{x}_{1:T}, \mathbf{a}_{2:T}) \triangleq \prod_{m=1}^{N} R_1(x_1^m) \prod_{t=2}^{T} \prod_{m=1}^{N} M_t(a_t^m, x_t^m).$$

Introduce extended target. Let $x_{1:T}^k = x_{1:T}^{b_{1:T}} = \{x_1^{b_1}, \ldots, x_T^{b_T}\}$.

$$\phi(\mathbf{x}_{1:T}, \mathbf{a}_{2:T}, k) = \phi(x_{1:T}^{b_{1:T}}, b_{1:T}) \phi(\mathbf{x}_{1:T}^{-b_{1:T}}, \mathbf{a}_{2:T}^{-b_{2:T}} \mid x_{1:T}^{b_{1:T}}, b_{1:T})$$

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Extended target distribution

SMC generates a sample on $\mathsf{X}^{NT} \times \{1, \ldots, N\}^{N(T-1)}$ with density,

$$\psi(\mathbf{x}_{1:T}, \mathbf{a}_{2:T}) \triangleq \prod_{m=1}^{N} R_1(x_1^m) \prod_{t=2}^{T} \prod_{m=1}^{N} M_t(a_t^m, x_t^m).$$

Introduce extended target. Let $x_{1:T}^k = x_{1:T}^{b_{1:T}} = \{x_1^{b_1}, \ldots, x_T^{b_T}\}$.

$$\begin{aligned}
\phi(\mathbf{x}_{1:T}, \mathbf{a}_{2:T}, k) &= \phi(x_{1:T}^{b_{1:T}}, b_{1:T}) \phi(\mathbf{x}_{1:T}^{-b_{1:T}}, \mathbf{a}_{2:T}^{-b_{2:T}} \mid x_{1:T}^{b_{1:T}}, b_{1:T}) \\
&\triangleq \underbrace{\frac{\bar{\gamma}_T(x_{1:T}^{b_{1:T}})}{N^T}}_{\text{marginal}} \underbrace{\prod_{\substack{m=1 \\ m \neq b_1}}^{N} R_1(x_1^m) \prod_{t=2}^{T} \prod_{\substack{m=1 \\ m \neq b_t}}^{N} M_t(a_t^m, x_t^m)}_{\text{conditional}}.
\end{aligned}$$

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Particle Gibbs

Particle Gibbs

- Draw $\mathbf{x}_{1:T}^{\star,-b_{1:T}}, \mathbf{a}_{2:T}^{\star,-b_{2:T}} \sim \phi(\mathbf{x}_{1:T}^{-b_{1:T}}, \mathbf{a}_{2:T}^{-b_{2:T}} \mid x_{1:T}^{b_{1:T}}, b_{1:T})$.

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Particle Gibbs

Particle Gibbs

- Draw $\mathbf{x}_{1:T}^{\star,-b_{1:T}}, \mathbf{a}_{2:T}^{\star,-b_{2:T}} \sim \phi(\mathbf{x}_{1:T}^{-b_{1:T}}, \mathbf{a}_{2:T}^{-b_{2:T}} \mid x_{1:T}^{b_{1:T}}, b_{1:T})$.
- Draw $k^\star \sim \phi(k \mid \mathbf{x}_{1:T}^{\star,-b_{1:T}}, \mathbf{a}_{2:T}^{\star,-b_{2:T}}, x_{1:T}^{b_{1:T}}, a_{2:T}^{b_{2:T}})$.

Particle Gibbs

- Draw $\mathbf{x}_{1:T}^{\star,-b_{1:T}}, \mathbf{a}_{2:T}^{\star,-b_{2:T}} \sim \phi(\mathbf{x}_{1:T}^{-b_{1:T}}, \mathbf{a}_{2:T}^{-b_{2:T}} \mid x_{1:T}^{b_{1:T}}, b_{1:T})$.
- Draw $k^\star \sim \phi(k \mid \mathbf{x}_{1:T}^{\star,-b_{1:T}}, \mathbf{a}_{2:T}^{\star,-b_{2:T}}, x_{1:T}^{b_{1:T}}, a_{2:T}^{b_{2:T}})$.

More precisely...

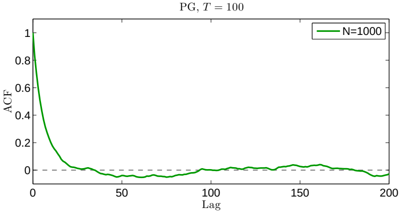
- Draw $\mathbf{x}_{1:T}^{\star,-b_{1:T}}, \mathbf{a}_{2:T}^{\star,-b_{2:T}}$ by running conditional SMC (CSMC).
- Draw $k^\star$ with $P(k^\star = m) \propto w_T^m$.

ex) Particle Gibbs

Stochastic volatility model,

$$x_{t+1} = 0.9x_t + w_t, \qquad w_t \sim \mathcal{N}(0, \theta),$$

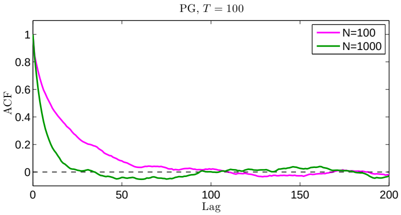
$$y_t = e_t \exp\left(\frac{1}{2}x_t\right), \qquad e_t \sim \mathcal{N}(0, 1).$$



ex) Particle Gibbs

Stochastic volatility model,

$$
\begin{aligned}
x_{t+1} &= 0.9 x_t + w_t, & w_t &\sim \mathcal{N}(0, \theta), \\
y_t &= e_t \exp\left(\frac{1}{2} x_t\right), & e_t &\sim \mathcal{N}(0, 1).
\end{aligned}
$$



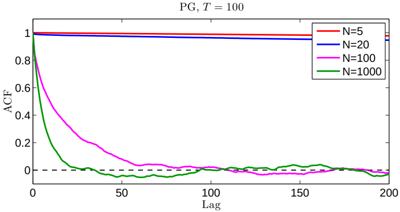
AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

ex) Particle Gibbs

Stochastic volatility model,

$$x_{t+1} = 0.9x_t + w_t, \qquad w_t \sim \mathcal{N}(0, \theta),$$

$$y_t = e_t \exp\left(\frac{1}{2}x_t\right), \qquad e_t \sim \mathcal{N}(0, 1).$$

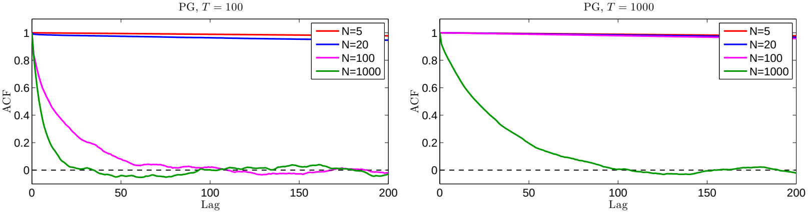


ex) Particle Gibbs

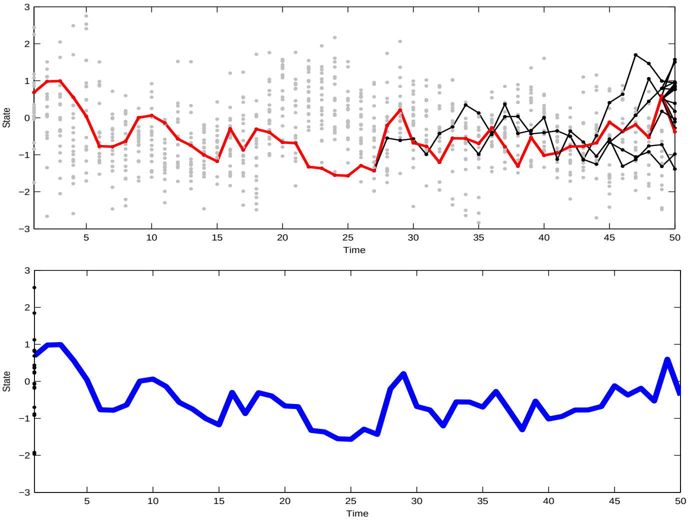
Stochastic volatility model,

$$x_{t+1} = 0.9x_t + w_t, \qquad w_t \sim \mathcal{N}(0, \theta),$$

$$y_t = e_t \exp\left(\frac{1}{2}x_t\right), \qquad e_t \sim \mathcal{N}(0, 1).$$



ex cont'd) Particle Gibbs



Poor mixing of PG – interpretation

Particle Gibbs

- Draw $\mathbf{x}_{1:T}^{\star,-b_{1:T}}, \mathbf{a}_{2:T}^{\star,-b_{2:T}} \sim \phi(\mathbf{x}_{1:T}^{-b_{1:T}}, \mathbf{a}_{2:T}^{-b_{2:T}} \mid x_{1:T}^{b_{1:T}}, b_{1:T})$.
- Draw $k^\star \sim \phi(k \mid \mathbf{x}_{1:T}^{\star,-b_{1:T}}, \mathbf{a}_{2:T}^{\star,-b_{2:T}}, x_{1:T}^{b_{1:T}}, a_{2:T}^{b_{2:T}})$.

The variables $\{x_{1:T}^{b_{1:T}}, b_{1:T-1}\}$ are never sampled!

Poor mixing of PG – interpretation

Particle Gibbs

- Draw $\mathbf{x}_{1:T}^{\star,-b_{1:T}}, \mathbf{a}_{2:T}^{\star,-b_{2:T}} \sim \phi(\mathbf{x}_{1:T}^{-b_{1:T}}, \mathbf{a}_{2:T}^{-b_{2:T}} \mid x_{1:T}^{b_{1:T}}, b_{1:T})$.
- Draw $k^\star \sim \phi(k \mid \mathbf{x}_{1:T}^{\star,-b_{1:T}}, \mathbf{a}_{2:T}^{\star,-b_{2:T}}, x_{1:T}^{b_{1:T}}, a_{2:T}^{b_{2:T}})$.

The variables $\{x_{1:T}^{b_{1:T}}, b_{1:T-1}\}$ are never sampled!

Include $b_{1:T-1}$ in the Gibbs sweep!

PG-AS

Particle Gibbs with ancestor sampling (PG-AS)

- For $t = 1, \ldots, T$, draw

$$\mathbf{x}_t^{\star,-b_t}, \mathbf{a}_t^{\star,-b_t} \sim \text{one iteration of CSMC},$$
$$(a_t^{\star,b_t} =)\ b_{t-1}^{\star} \sim \phi(b_{t-1} \mid \mathbf{x}_{1:t-1}^{\star,-b_{1:t-1}}, \mathbf{a}_{2:t-1}^{\star}, x_{1:T}^{b_{1:T}}, b_{t:T}).$$

- Draw $(k^{\star} =)\ b_T^{\star}$ with $P(b_T^{\star} = m) \propto w_T^m$.

PG-AS

Particle Gibbs with ancestor sampling (PG-AS)

- For $t = 1, \dots, T$, draw

$$\mathbf{x}_t^{\star,-b_t}, \mathbf{a}_t^{\star,-b_t} \sim \text{one iteration of CSMC},$$
$$(a_t^{\star,b_t} =)\ b_{t-1}^{\star} \sim \phi(b_{t-1} \mid \mathbf{x}_{1:t-1}^{\star,-b_{1:t-1}}, \mathbf{a}_{2:t-1}^{\star}, x_{1:T}^{b_{1:T}}, b_{t:T}).$$

- Draw $(k^\star =)\ b_T^\star$ with $P(b_T^\star = m) \propto w_T^m$.

We can show,

$$\phi(b_t \mid \mathbf{x}_{1:t}, \mathbf{a}_{2:t}, x_{t+1:T}^{b_{t+1:T}}, b_{t+1:T}) \propto w_t^{b_t} \frac{\gamma_T(x_{1:T}^k)}{\gamma_t(x_{1:t}^{b_t})}.$$

CSMC with ancestor sampling

CSMC with ancestor sampling, conditioned on $\{x'_{1:T}, b_{1:T}\}$

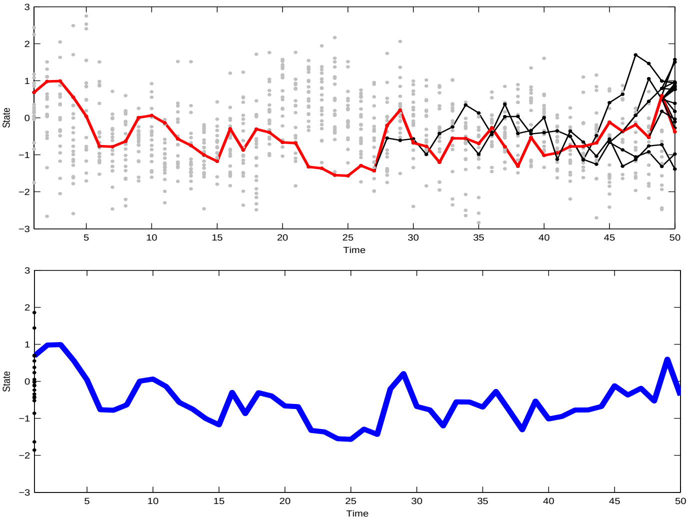
1. **Initialize ($t = 1$):**
   (a) Draw $x_1^m \sim R_1(x_1)$ for $m \neq b_1$ and set $x_1^{b_1} = x'_1$.
   (b) Set $w_1^m = W_1(x_1^m)$ for $m = 1, \dots, N$.
2. **for $t = 2, \dots, T$:**
   (a) Draw $(a_t^m, x_t^m) \sim M_t(a_t, x_t)$ for $m \neq b_t$ and set $x_t^{b_t} = x'_t$.
   (b) Draw $a_t^{b_t}$ with

$$P(a_t^{b_t} = m) \propto w_{t-1}^m \frac{\gamma_T(\{x_{1:t-1}^m, x'_{t:T}\})}{\gamma_t(x_{1:t-1}^m)}.$$

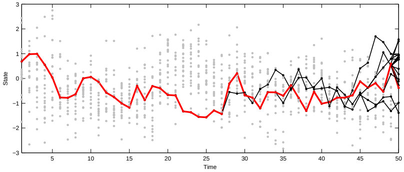
   (c) Set $x_{1:t}^m = \{x_{1:t-1}^{a_t^m}, x_t^m\}$ and $w_t^m = W_t(x_{1:t}^m)$ for $m = 1, \dots, N$.

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

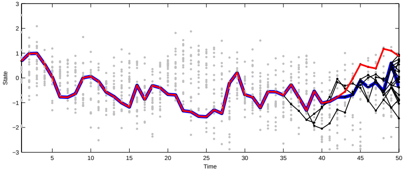
ex cont'd) PG-AS



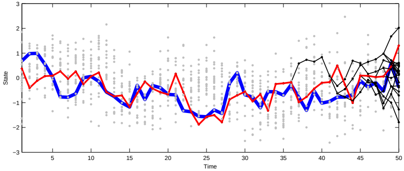
PG vs. PG-AS



PG



PG-AS

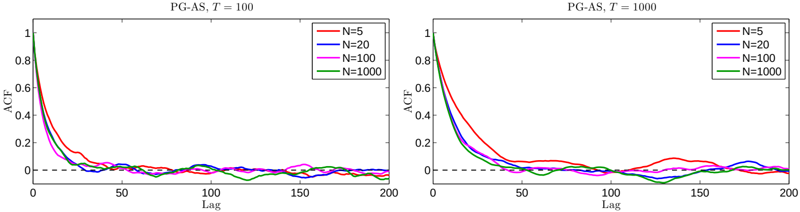


AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

ex cont'd) PG-AS

Stochastic volatility model,

$$
\begin{aligned}
x_{t+1} &= 0.9x_t + w_t, & w_t &\sim \mathcal{N}(0, \theta), \\
y_t &= e_t \exp\left(\frac{1}{2}x_t\right), & e_t &\sim \mathcal{N}(0, 1).
\end{aligned}
$$



AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Related method: PG-BS

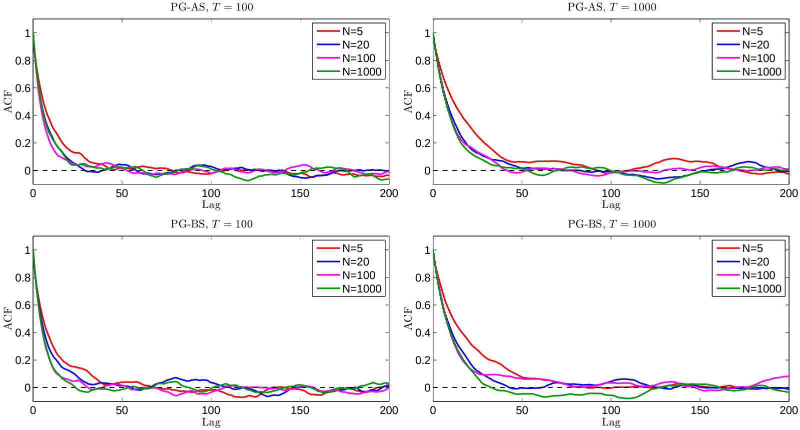
Sampling $b_{1:T-1}$ suggested by Whiteley (2010),

N. Whiteley, "Discussion on Particle Markov chain Monte Carlo methods", *Journal of the Royal Statistical Society: Series B*, 72(3):306–307, 2010.

Particle Gibbs with backward simulation (PG-BS),

- Draw $\mathbf{x}_{1:T}^{\star,-b_{1:T}}, \mathbf{a}_{2:T}^{\star,-b_{2:T}}$ by running conditional SMC.
- Draw $b_{1:T}^{\star}$ by running a backward simulator.

ex cont'd) PG-BS and PG-AS



Non-Markovian models

Main motivation for PG-AS instead of PG-BS

– **appears to be more robust to weight approximation.**

Non-Markovian models

Main motivation for PG-AS instead of PG-BS

– **appears to be more robust to weight approximation.**

Consider a non-Markovian model,

$$x_{t+1} \sim f(x_{t+1} \mid x_{1:t}),$$
$$y_t \sim g(y_t \mid x_{1:t}).$$

Backward weights depend on,

$$\frac{\gamma_T(x_{1:T})}{\gamma_t(x_{1:t})} = \frac{p(x_{1:T}, y_{1:T})}{p(x_{1:t}, y_{1:t})} = \prod_{s=t+1}^{T} g(y_s \mid x_{1:s}) f(x_s \mid x_{1:s-1}).$$

Non-Markovian models

Main motivation for PG-AS instead of PG-BS

– **appears to be more robust to weight approximation.**

Consider a non-Markovian model,

$$x_{t+1} \sim f(x_{t+1} \mid x_{1:t}),$$
$$y_t \sim g(y_t \mid x_{1:t}).$$

Backward weights depend on,

$$\frac{\gamma_T(x_{1:T})}{\gamma_t(x_{1:t})} = \frac{p(x_{1:T}, y_{1:T})}{p(x_{1:t}, y_{1:t})} \underset{\sim}{\propto} \prod_{s=t+1}^{p} g(y_s \mid x_{1:s}) f(x_s \mid x_{1:s-1}).$$

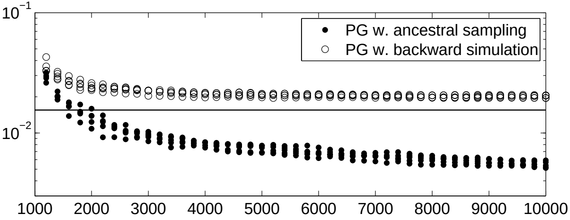
AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Rao-Blackwellized smoothing

Rao-Blackwellization,

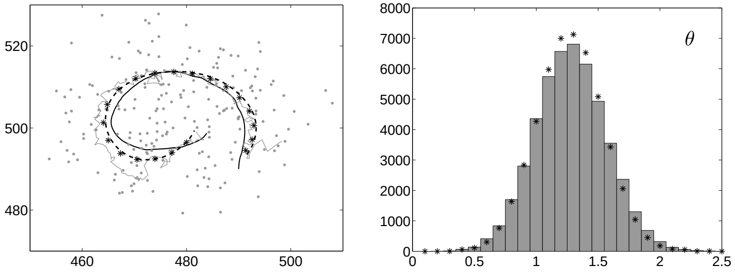
$$x_{t+1} = Ax_t + v_t, \qquad v_t \sim \mathcal{N}(0, Q),$$
$$y_t = Cx_t + e_t, \qquad e_t \sim \mathcal{N}(0, R).$$

- Marginalize 3 out of 4 states with conditional Kalman filters.
- Marginal model is non-Markovian!
- Apply PG-AS and PG-BS 1D marginal space with $N = 5$.



Target tracking

- Coordinated turn motion model with rank deficient process noise covariance.
- Noisy range-bearing measurements.
- Unknown turn rate θ.
- PG-AS with $N = 5$ and PMMH with $N = 5000$.



Conclusions

- PG-AS: A novel approach to PMCMC.
- No explicit backward pass (contrary to PG-BS).
- Easier to implement and more memory efficient than PG-BS.
- Appears to be more robust to weight approximations.
  Needs further investigation!

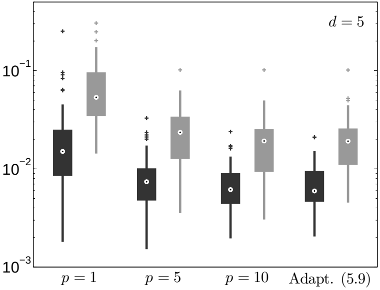
F. Lindsten, M. I. Jordan and T. B. Schön, "Ancestral Sampling for Particle Gibbs", *Accepted to the 2012 Conference on Neural Information Processing Systems (NIPS)*, Lake Tahoe, NV, USA, 2012.

Degenerate models

Degenerate state-space models,

$$x_{t+1} = Ax_t + v_t, \qquad v_t \sim \mathcal{N}(0, Q),$$
$$y_t = Cx_t + e_t, \qquad e_t \sim \mathcal{N}(0, R).$$

- $\text{rank}(Q) = 1 \Rightarrow$ degenerate model!
- Rewrite as non-degenerate non-Markovian 1D model.
- Apply PG-AS and PG-BS 1D marginal space with $N = 5$.



AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET