Particle Gibbs with Refreshed Backward Simulation

UNIVERSITY OF CAMBRIDGE

Pete Bunch **Fredrik Lindsten** **Sumeetpal Singh**

Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, UK

{pb404, fsml2, sss40}@eng.cam.ac.uk

1. Summary

Particle Gibbs is a Markov Chain Monte Carlo algorithm which can be used for Bayesian parameter learning with Markovian state space models. It uses a particle filter at each iteration to sample a new state trajectory. The basic formulation requires large numbers of particles in order to mix well, because of particle filter degeneracy. This problem can be mitigated by including a backward simulation sweep to increase the probability of changing the particle ancestry. Here we show how a modification to the backward simulation phase, in which new states are sampled simultaneously with the ancestor indexes, can further improve mixing.

2. Preliminaries

Markovian state space model

- Sequence of latent states $x_t \in \mathcal{X} : t = 1, \dots, T$
- Sequence of observations $y_t \in \mathcal{Y} : t = 1, \dots, T$
- Model parameters $\theta \in \Theta$ with prior density $p(\theta)$
- Model densities exist and depend on the parameters:

$$x_t | x_{t-1} \sim f_{\theta,t}(x_t|x_{t-1}) \qquad y_t|x_t \sim g_{\theta,t}(y_t|x_t)$$

- The objective is to approximate the joint posterior density over all the unknown variables (with normalising constant Z):

$$p(\theta, x_{1:T}|y_{1:T}) = \frac{1}{Z} \cdot p(\theta) \prod_{t=1}^{T} g_{\theta,t}(y_t|x_t) f_{\theta,t}(x_t|x_{t-1}) \tag{1}$$

- An ideal Gibbs Sampler would target this by alternately sampling the conditionals $p(x_{1:T}|\theta, y_{1:T})$ and $p(\theta|x_{1:T}, y_{1:T})$.
- Parameter conditional can be sampled directly or targeted with Metropolis-Hastings. State conditional is generally intractable.

Particle Filter

- Recursively approximates the sequence of joint filtering densities $p(x_{1:t}|\theta, y_{1:t})$ using N particles, each one a weighted realisation of the state sequence:

$$\{x_{1:t}^{(i)},\ w_t^{(i)}\ :\ i = 1, \dots, N\}$$

- For each t, i, sample ancestor $a_t^{(i)} \in \{1, \dots, N\}$ and state $x_t^{(i)}$ from:

$$\frac{w_{t-1}^{(a_t^{(i)})}}{\sum_j w_{t-1}^{(j)}} q_t(x_t|x_{t-1}^{(a_t)})$$

- Update trajectory and weight:

$$x_{1:t}^{(i)} = x_{1:t-1}^{(a_t^{(i)})} \cup x_t^{(i)} \qquad w_t^{(i)} = \frac{f_{\theta,t}(x_t^{(i)}|x_{t-1}^{(a_t^{(i)})}) g_{\theta,t}(y_t|x_t^{(i)})}{q_t(x_t^{(i)}|x_{t-1}^{(a_t^{(i)})})}$$

- Exhibits path-space degeneracy, a significant deficiency. Not all particles are propagated to the next time instant. The number of unique states appearing in the trajectories decreases as we look back in time.

3. Particle Gibbs

- Consider all the random variable comprising a particle filter:

$$\mathbf{x}_t = \{x_t^{(i)}\}_{i=1}^N \quad , \quad \mathbf{a}_t = \{a_t^{(i)}\}_{i=1}^N \quad , \quad t = 2, \dots, T$$

- Let $K \in \{1, \dots, N\}$ be the index of one particular reference trajectory, with ancestry:

$$b_t = \begin{cases} K & t = T \\ a_{t+1}^{(b_{t+1})} & t = T-1, \dots, 1 \end{cases}$$

- Construct extended target distribution over all all these variables:

$$\pi(\theta, \mathbf{a}_{2:T}, \mathbf{x}_{1:T}, K) = \frac{1}{N^T} \cdot p(\theta, x_{1:T}^{(b_{1:T})}|y_{1:T}) \times \prod_{i \neq b_1} q_1(x_1^{(i)}) \prod_{t=2}^{T} \left[\prod_{i \neq b_t} \frac{w_{t-1}^{(a_t^{(i)})}}{\sum_j w_{t-1}^{(j)}} q_t(x_t^{(i)}|x_{t-1}^{(a_t^{(i)})}) \right] \tag{2}$$

- This may be targeted without approximation, and admits the desired posterior as a marginal. [1]
- Sample alternately from conditional distributions for:
  - θ (as before)
  - $\{\mathbf{a}_{2:T}^{(-b_{2:T})}, \mathbf{x}_{1:T}^{(-b_{1:T})}\}$ (conditional particle filter)
  - K (Sample from final set of weights)

4. Particle Gibbs with Backward Simulation

- Particle Gibbs mixing can be very slow, due to path-space degeneracy of conditional particle filter. Successive reference trajectories are likely to have a near-identical ancestry.
- Mitigated by backward simulation, an extra sampling stage to modify ancestors one at a time, instead of all at once. [2, 3]
- Sample new reference ancestor index for each t (backwards in time) from:

$$\pi(a_t^{(b_t)}|\theta, \mathbf{a}_{2:t-1}, \mathbf{x}_{1:t-1}, a_{t+1:T}^{(b_{t+1:T})}, x_{t:T}^{(b_{t:T})}, K) = \frac{w_{t-1}^{(a_t^{(b_t)})} f_{\theta,t}(x_t^{(b_t)}|x_{t-1}^{(a_t^{(b_t)})})}{\sum_j w_{t-1}^{(j)} f_{\theta,t}(x_t^{(b_t)}|x_{t-1}^{(j)})} \tag{3}$$

- This is a collapsed Gibbs move. [4] Future variables of reference trajectory are marginalised.

5. Refreshed Backward Simulation

- Backward simulation is ineffective if the transition density is informative (low variance). See Figure 1.
- Improve mixing by simultaneously sampling new state(s) along with each ancestor index:

$$\pi(a_t^{(b_t)}, x_t^{(b_t)}|\theta, \mathbf{a}_{2:t-1}, \mathbf{x}_{1:t-1}, a_{t+1:T}^{(b_{t+1:T})}, x_{t+1:T}^{(b_{t+1:T})}, K) = \frac{w_{t-1}^{(a_t^{(b_t)})} f_{\theta,t}(x_t^{(b_t)}|x_{t-1}^{(a_t^{(b_t)})}) g_{\theta,t}(y_t|x_t^{(b_t)}) f_{\theta,t+1}(x_{t+1}^{(b_{t+1})}|x_t^{(b_t)})}{\sum_j w_{t-1}^{(j)} \int f_{\theta,t}(x|x_{t-1}^{(j)}) g_{\theta,t}(y_t|x) f_{\theta,t+1}(x_{t+1}^{(b_{t+1})}|x) dx} \tag{4}$$

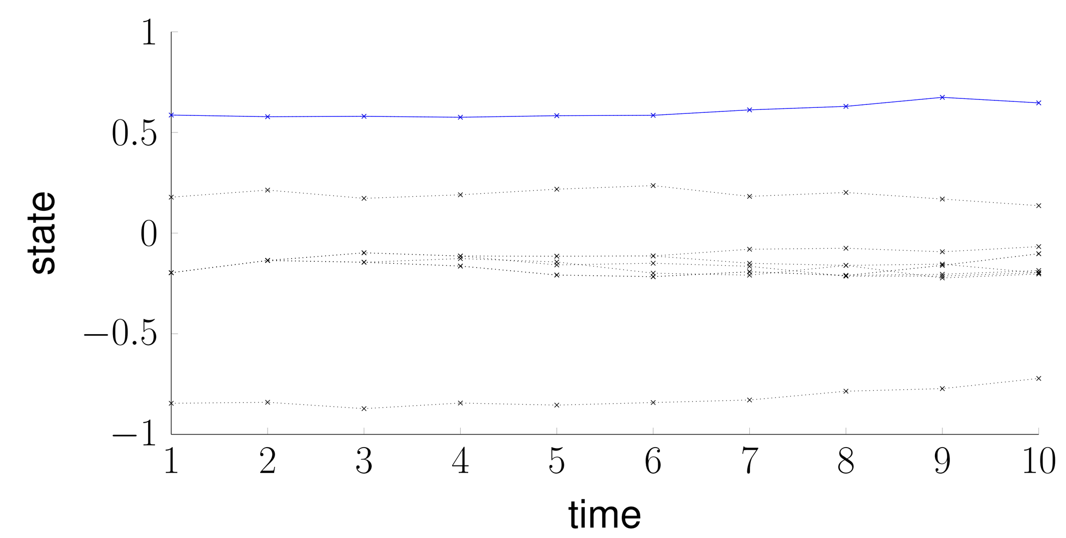


Figure 1: *A backward simulation sweep does not result in any changes to the reference particle ancestry (blue).*

- No change to extended target distribution.
- Intuitively, sample state so as to bridge the gap between discontinuous past and future.
- Easily extended to multiple time instants (i.e. sample $x_{t:t+L-1}^{(i)}$).

6. Markov Kernels for Refreshed Sampling

- Cannot sample (4) directly (continuous-discrete). Instead target with a Markov kernel.
- Write (4) in simplified form:

$$\pi(a_t, x_t|\mathbf{x}_{t-1}) = \frac{w_{t-1}^{(a_t)} \rho_t(x_t|x_{t-1}^{(a_t)})}{\sum_j w_{t-1}^{(j)} \int \rho_t(x|x_{t-1}^{(j)}) dx} \tag{5}$$

- Can be targeted with Metropolis-Hastings (as in [5] for state smoothing).
- Alternatively, use conditional importance sampling. Particle Gibbs principle applied to a single time instant using extended target:

$$\eta(\mathbf{a}_t, \mathbf{x}_t, c) = \frac{1}{N} \pi(x_t^{(c)}, a_t^{(c)}|\mathbf{x}_{t-1}) \prod_{i \neq c} \frac{v_t^{(a_t^{(i)})}}{\sum_j v_t^{(j)}} \psi_t(x_t^{(i)}|x_{t-1}^{(a_t^{(i)})}).$$

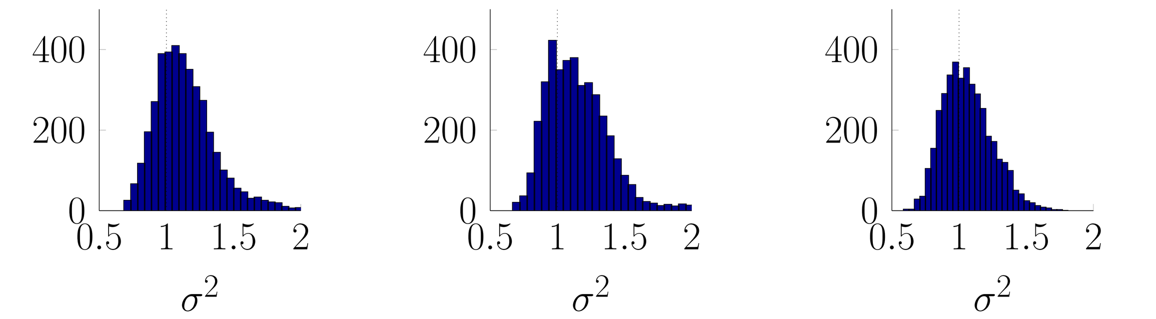
Require: Preceding particle states $\mathbf{x}_{t-1}$, current values $\{a_t^*, x_t^*\}$.

1: Sample an index uniformly $c^* \in \{1, \dots, N\}$.
2: Set $a_t^{(c^*)} = a_t^*$. Set $x_t^{(c^*)} = x_t^*$.
3: **for all** $i \in \{1, \dots, N\} \setminus c^*$ **do**
4: Sample $a_t^{(i)} \sim \frac{v_{t-1}^{(a_t)}}{\sum_j v_{t-1}^{(j)}}$. Sample $x_t^{(i)} \sim \psi_t(x_t|x_{t-1}^{(a_t^{(i)})})$.
5: **end for**
6: Sample $c' \sim \frac{u_t^{(c)}}{\sum_j u_t^{(j)}}$, where $u_t = \frac{w_{t-1}^{(a_t^{(i)})} \rho_t(x_t^{(i)}|x_{t-1}^{(a_t^{(i)})})}{v_{t-1}^{(a_t^{(i)})} \psi_t(x_t^{(i)}|x_{t-1}^{(a_t^{(i)})})}$.
7: Set $a_t' = a_t^{(c')}$. Set $x_t' = x_t^{(c')}$.
8: **return** New values $\{a_t', x_t'\}$.

Algorithm 1: *Conditional importance sampling for the joint ancestor-state conditional distributions.*

7. Simulations

- Tested on a tracking problem. Near constant velocity transition model (3D) with observations of bearing, elevation and range. Unknown parameter is scale factor on the transition covariance matrix with uninformative conjugate prior.
- 5 simulated data sets each of 100 time steps. 25000 MCMC interations (5000 burn in).
- Compared particle Gibbs (PG), with ordinary (PG-BS) and refreshed (PG-RBS) backward simulation, varying number of particles.
- PG-RBS with 100 particles takes same time as PG-BS with 200.
- PG does not work (no convergence).



(a) PG-BS ($N = 100$) (b) PG-BS ($N = 200$) (c) PG-RBS ($N = 100$)

Figure 2: *Posterior histograms*

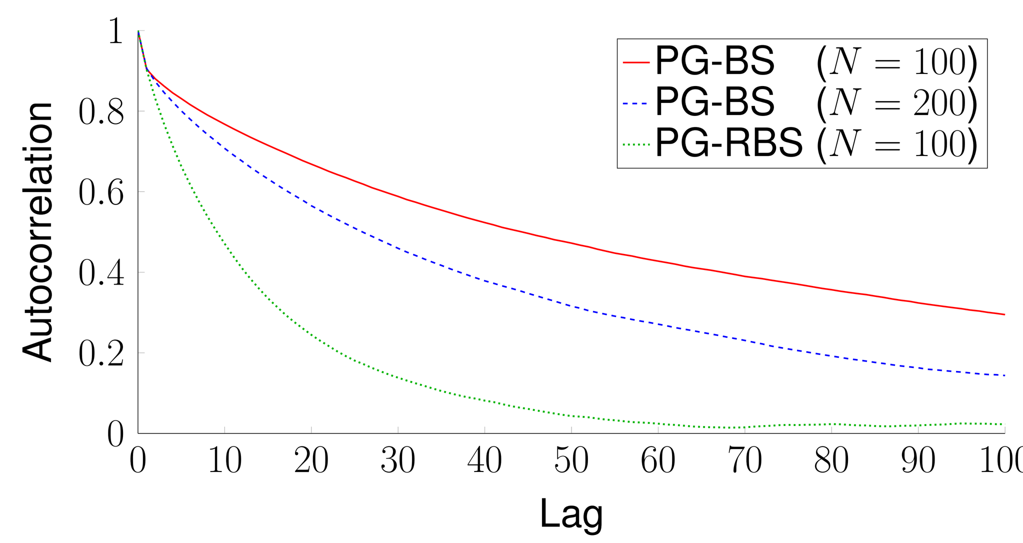


Figure 3: *Mean autocorrelation function for PG-BS and PG-RBS.*

8. Conclusions

- Simple but effective modification to standard PG-BS algorithm.
- States sampled simultaneously with ancestor indexes in backwards sweep.
- Direct sampling of conditional posterior not possible, but efficient Markov kernels exist.
- Improves mixing of Markov chain by increasing the probability of changing the ancestry.
- Can also be used with ancestor sampling [6] instead of backward simulation.

References

[1] C Andrieu, A Doucet, and R Holenstein, "Particle Markov chain Monte Carlo methods," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 72, pp. 269–342, 2010.

[2] S J Godsill, A Doucet, and M West, "Monte Carlo smoothing for nonlinear time series," *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 156–168, 2004.

[3] N Whiteley, "Discussion on "Particle Markov chain Monte Carlo methods" by Andrieu et al.," *Journal of the Royal Statistical Society: Series B*, vol. 72, no. 3, pp. 306–307, 2010.

[4] D A van Dyk and T Park, "Partially collapsed Gibbs samplers," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 790–796, 2008.

[5] P Bunch and S Godsill, "Improved particle approximations to the joint smoothing distribution using Markov chain Monte Carlo," *IEEE Transactions on Signal Processing*, vol. 61, no. 4, pp. 956–963, 2013.

[6] F Lindsten, M I Jordan, and T B Schön, "Particle Gibbs with ancestor sampling," *Journal of Machine Learning Research*, vol. 15, pp. 2145–2184, 2014.