Iterations of Filtering

http://dx.doi.org/10.1080/01621459.2016.1222291
(arxiv 1511.06286)

P. Guarniero, **A. M. Johansen** and A. Lee

University of Warwick
a.m.johansen@warwick.ac.uk
www2.warwick.ac.uk/fac/sci/statistics/staff/academic/johansen/talks/

Bristol, September 13th, 2016

SuSTaIn and Me

MSci	Cambridge, Physics October 1998 – June 2002
PhD	Cambridge, Engineering October 2002 – September 2006
PDRF	**Brunel Fellowship, Bristol** October 2006 – August 2008
Afterwards	Warwick, Statistics September 2008–present

What I did in Bristol. . .

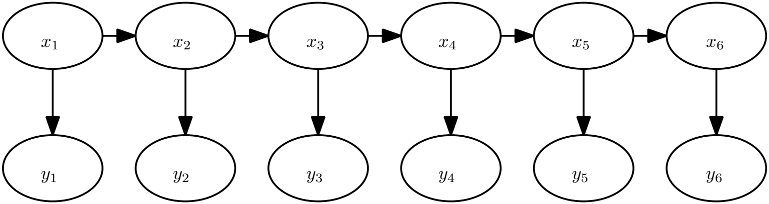
- Monte Carlo filtering of piecewise-deterministic processes N. Whiteley, J., and S. Godsill. *Journal of Computational and Graphical Statistics*, 20(1):119–139 March 2011.
- On Solving Integral Equations Using Markov Chain Monte Carlo A. Doucet, J. and V. B. Tadić. *Applied Mathematics and Computation* 216:2869–2880, 2010.
- SMCTC: Sequential Monte Carlo in C++, *Journal of Statistical Software* 30(6):1–41, April 2009.
- A note on auxiliary particle filters J. and A. Doucet, *Statistics and Probability Letters* 72(12):1498–1504, September 2008.
- Particle methods for maximum likelihood estimation in latent variable models J., A. Doucet, and M. Davy, *Statistics and Computing*, 18(1):47-57, March 2008.
- Single molecule-level analysis of the subunit composition of the T-cell receptor on live T cells J. R. James, S. S. White, R. W. Clarke, J., et al., *Proceedings of the National Academy of Science, USA* 104(45):17662-17667, November 2007.
- Simulation of the annual loss distribution in operational risk via Panjer recursions and Volterra integral equations for value at risk and expected shortfall estimation G. W. Peters, J., and A. Doucet, *Journal of Operational Risk* 2(3):29–58, Fall 2007.

Outline

- Background: SMC and PMCMC
- Iterative Lookahead Methods
  - Motivation
  - Methodology
  - Applications: linear Gaussian and stochastic volatility
  - Ongoing work: diffusion bridges
- Conclusions

Discrete Time Filtering

Online inference for Hidden Markov Models:



- Given *transition* $f_\theta(x_{n-1}, x_n)$,
- and *likelihood* $g_\theta(x_n, y_n)$,
- use $p_\theta(x_n|y_{1:n})$ to characterize latent state, but,

$$p_\theta(x_n|y_{1:n}) = \frac{\int p_\theta(x_{n-1}|y_{1:n-1}) f_\theta(x_{n-1}, x_n) dx_{n-1} g_\theta(x_n, y_n)}{\int \int p_\theta(x_{n-1}|y_{1:n-1}) f_\theta(x_{n-1}, x_n') dx_{n-1} g_\theta(x_n', y_n) dx_n'}$$

isn't often tractable.

Particle Filtering

A (sequential) Monte Carlo (SMC) scheme to approximate the filtering distributions.

A Simple Particle Filter [4]

At $n = 1$:

- Sample $X_1^1, \ldots, X_1^N \sim \mu_\theta$.

For $n > 1$:

- Sample

$$X_n^1, \ldots, X_n^N \sim \frac{\sum_{j=1}^N g_\theta(X_{n-1}^j, y_{n-1}) f_\theta(X_{n-1}^j, \cdot)}{\sum_{k=1}^n g_\theta(X_{n-1}^k, y_{n-1})}$$

- Approximate $p_\theta(dx_n|y_{1:n}), p_\theta(y_{1:n})$ with

$$\widehat{p_\theta}(\cdot|y_{1:n}) = \frac{\sum_{j=1}^N g_\theta(X_n^j, y_n)\delta_{X_n^j}}{\sum_{k=1}^N g_\theta(X_n^k, y_n)}, \frac{\widehat{p_\theta}(y_{1:n})}{\widehat{p_\theta}(y_{1:n-1})} = \frac{1}{n}\sum_{j=1}^N g_\theta(X_n^k, y_n)$$

Online Particle Filters for Offline Parameter Estimation

Particle Markov chain Monte Carlo (PMCMC) [2]

- Embed SMC within MCMC,
- justified via explicit auxiliary variable construction,
- or in some cases by a pseudomarginal [1] argument.
- Very widely applicable,
- but prone to poor mixing when SMC performs poorly for some θ [7, Section 4.2.1].
- Is valid for *very* general SMC algorithms.

Twisting the HMM (a complement to [8])

Given (μ, f, g) and $y_{1:T}$, introducing
$\boldsymbol{\psi} := (\psi_1, \psi_2, \dots, \psi_T), \psi_t \in \mathcal{C}_b(\mathsf{X}, (0, \infty))$ and

$$\tilde{\psi}_0 := \int_{\mathsf{X}} \mu(x_1)\, \psi_1(x_1)\, dx_1 \quad \tilde{\psi}_t(x_t) := \int_{\mathsf{X}} f(x_t, x_{t+1})\, \psi_{t+1}(x_{t+1})\, dx_{t+1}$$

we obtain $(\mu_1^{\boldsymbol{\psi}}, \{f_t^{\boldsymbol{\psi}}\}, \{g_t^{\boldsymbol{\psi}}\})$, with

$$\mu_1^{\boldsymbol{\psi}}(x_1) := \frac{\mu(x_1)\psi_1(x_1)}{\tilde{\psi}_0}, \qquad f_t^{\boldsymbol{\psi}}(x_{t-1}, x_t) := \frac{f(x_{t-1}, x_t)\, \psi_t(x_t)}{\tilde{\psi}_{t-1}(x_{t-1})}$$

and the sequence of non-negative functions

$$g_{T1}^{\boldsymbol{\psi}}(x_1) := g(x_1, y_1)\, \frac{\tilde{\psi}_1(x_1)}{\psi_1(x_1)} \tilde{\psi}_0, \qquad g_t^{\boldsymbol{\psi}}(x_t) := g(x_t, y_t)\, \frac{\tilde{\psi}_t(x_t)}{\psi_t(x_t)}.$$

Proposition

For any sequence of bounded, continuous and positive functions $\boldsymbol{\psi}$*, let*

$$Z_{\boldsymbol{\psi}} := \int_{\mathsf{X}^T} \mu_1^{\boldsymbol{\psi}}(x_1)\, g_1^{\boldsymbol{\psi}}(x_1) \prod_{t=2}^{T} f_t^{\boldsymbol{\psi}}(x_{t-1}, x_t)\, g_t^{\boldsymbol{\psi}}(x_t)\, dx_{1:T}.$$

Then, $Z_{\boldsymbol{\psi}} = p_\theta(y_{1:T})$ *for any such* $\boldsymbol{\psi}$*.*

The optimal choice is:

$$\psi_t^*(x_t) := g(x_t, y_t)\, \mathbb{E}\left[\prod_{p=t+1}^{T} g(X_p, y_p) \,\middle|\, \{X_t = x_t\} \right], \quad x_t \in \mathsf{X},$$

for $t \in \{1, \ldots, T-1\}$. Then, $Z_{\boldsymbol{\psi}^*}^N = p(y_{1:T})$ with probability 1.

Towards Iterative Auxiliary Particle Filters [5]

ψ-Auxiliary Particle Filter

1. Sample $\xi_1^i \sim \mu^{\psi}$ independently for $i \in \{1, \ldots, N\}$.
2. For $t = 2, \ldots, T$, sample independently

$$\xi_t^i \sim \frac{\sum_{j=1}^N g_{t-1}^{\psi}(\xi_{t-1}^j) f_t^{\psi}(\xi_{t-1}^j, \cdot)}{\sum_{j=1}^N g_{t-1}^{\psi}(\xi_{t-1}^j)}, \qquad i \in \{1, \ldots, N\}.$$

Necessary features of ψ

1. It is possible to sample from f_t^{ψ}.
2. It is possible to evaluate g_t^{ψ}.
3. To be useful: $\mathbb{V}(\widehat{Z}_{\psi}^N)$ must be small.

A Recursive Approximtion

Proposition

The sequence ψ^* *satisfies* $\psi^*_T(x_T) = g(x_T, y_T)$, $x_T \in \mathsf{X}$ *and*

$$\psi^*_t(x_t) = g(x_t, y_t)\, f\left(x_t, \psi^*_{t+1}\right), \quad x_t \in \mathsf{X}, \quad t \in \{1, \ldots, T-1\}.$$

Algorithm 1 Recursive function approximations

For $t = T, \ldots, 1$:

1. Set $\psi^i_t \leftarrow g\left(\xi^i_t, y_t\right) f\left(\xi^i_t, \psi_{t+1}\right)$ for $i \in \{1, \ldots, N\}$.
2. Choose ψ_t as a member of Ψ on the basis of $\xi^{1:N}_t$ and $\psi^{1:N}_t$.

Iterated Auxiliary Particle Filters

Algorithm 2 An iterated auxiliary particle filter with parameters (N_0, k, τ)

1. Initialize: set ${\psi_t}^0 = \mathbf{1}$. $l \leftarrow 0$.
2. Repeat:
   2.1 Run a $\boldsymbol{\psi}^l$-APF with N_l particles; set $\hat{Z}_l \leftarrow Z_{\boldsymbol{\psi}^l}^{N_l}$.
   2.2 If $l > k$ and $\mathrm{sd}(\hat{Z}_{l-k:l})/\mathrm{mean}(\hat{Z}_{l-k:l}) < \tau$, go to 3.
   2.3 Compute $\boldsymbol{\psi}^{l+1}$ using Algorithm 1.
   2.4 If $N_{l-k} = N_l$ and the sequence $\hat{Z}_{l-k:l}$ is not monotonically increasing, set $N_{l+1} \leftarrow 2N_l$.
   Otherwise, set $N_{l+1} \leftarrow N_l$.
   2.5 Set $l \leftarrow l + 1$. Go to 2a.
3. Run a $\boldsymbol{\psi}^l$-APF. Return $\hat{Z} := Z_{\boldsymbol{\psi}}^{N_l}$.

An Elementary Implementation

Function Approximation

- Numerically obtain:

$$(m_t^*, \Sigma_t^*, \lambda_t^*) = \arg\min_{(m,\Sigma,\lambda)} \sum_{i=1}^{N} (\mathcal{N}(x_t^i, m, \Sigma) - \lambda \psi_t^i)^2$$

- Set:

$$\psi_t(x_t) := \mathcal{N}(x_t; m_t^*, \Sigma_t^*) + c(N, m_t^*, \Sigma_t^*).$$

Stopping Rule

- $k = 3$ or $k = 5$ in the following examples
- $\tau = 0.5$

Resampling

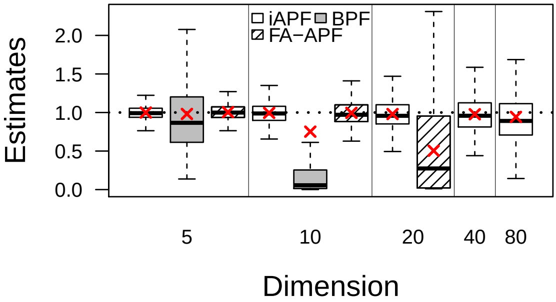
- Multinomial when ESS $< N/2$.

A Linear Gaussian Model: Behaviour with Dimension

$$\mu = \mathcal{N}(\cdot; \mathbf{0}, I_d) \qquad f(x, \cdot) = \mathcal{N}(\cdot; Ax, I_d)$$

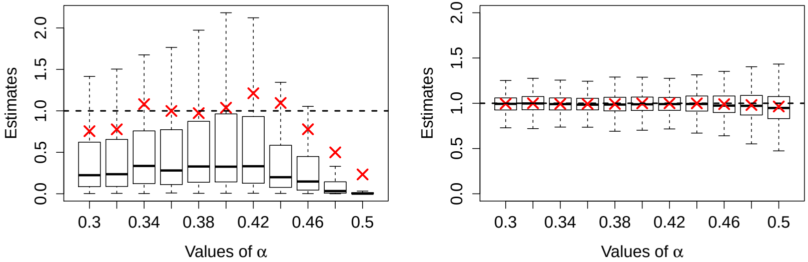
$$\text{and } g(x, \cdot) = \mathcal{N}(\cdot; x, I_d) \qquad \text{where } A_{ij} = 0.42^{|i-j|+1},$$

Box plots of $\hat{Z}/Z$ for different $|\mathsf{X}|$ (1000 replicates; $T = 100$).



Linear Gaussian Model: Sensitivity to Parameters

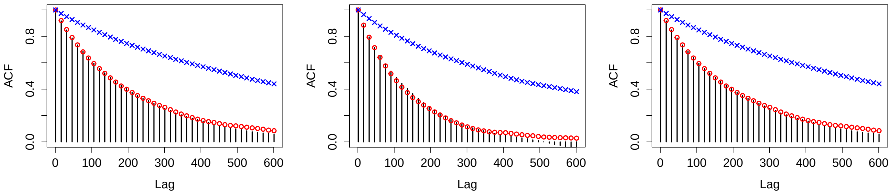
Fixing $d = 10$: Bootstrap ($N = 50,000$) / iAPF ($N_0 = 1,000$)



Box plots of $\frac{\hat{Z}}{Z}$ for different values of the parameter α using 1000 replicates.

Linear Gaussian Model: PMMH Empirical Autocorrelations

A_{11} A_{41} A_{55}



In this case:

$$
\begin{aligned}
d &= 5\\
\mu &= \mathcal{N}(\cdot; \mathbf{0}, I_d)\\
f(x,\cdot) &= \mathcal{N}(\cdot; Ax, I_d)\\
\text{and } g(x,\cdot) &= \mathcal{N}(\cdot; x, 0.25 I_d)
\end{aligned}
\qquad
A = \begin{pmatrix}
0.9 & 0 & 0 & 0 & 0\\
0.3 & 0.7 & 0 & 0 & 0\\
0.1 & 0.2 & 0.6 & 0 & 0\\
0.4 & 0.1 & 0.1 & 0.3 & 0\\
0.1 & 0.2 & 0.5 & 0.2 & 0
\end{pmatrix},
$$

(unknown lower triangular matrix)

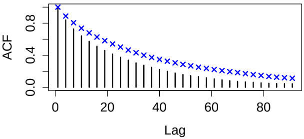
Stochastic Volatility

- A simple stochastic volatility model is defined by:

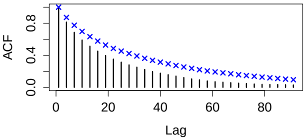
$$\begin{aligned} \mu(\cdot) =& \mathcal{N}(\cdot; 0, \sigma^2/(1-a)^2) \\ f(x, \cdot) =& \mathcal{N}(\cdot; ax, \sigma^2) \\ \text{and } g(x, \cdot) =& \mathcal{N}(\cdot; 0, \beta^2 \exp(x)), \end{aligned}$$

  where $a \in (0, 1)$, $\beta > 0$ and $\sigma^2 > 0$ are unknown.
- Considered $T = 945$ observations $y_{1:T}$ corresponding to the mean-corrected daily returns for the GBP/USD exchange rate from 1/10/81 to 28/6/85.

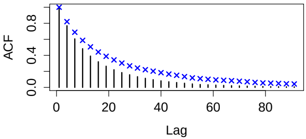
Estimated PMCMC Autocorrelation



a

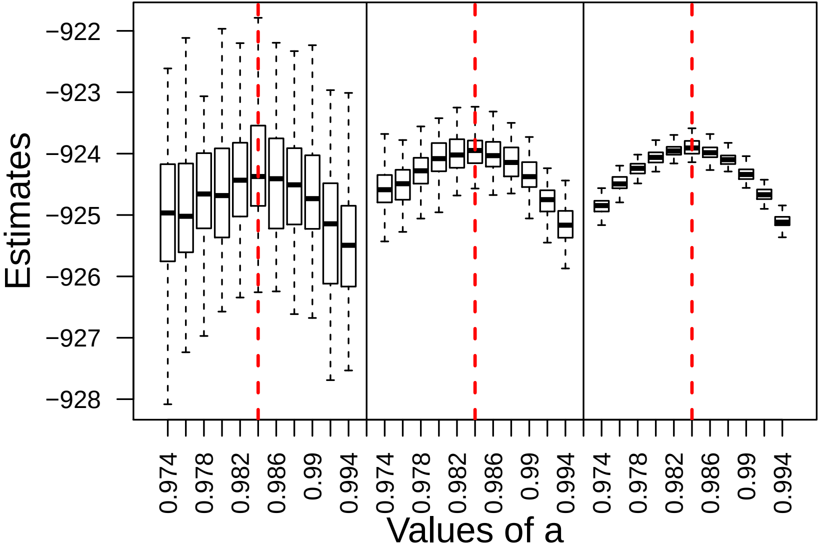


σ

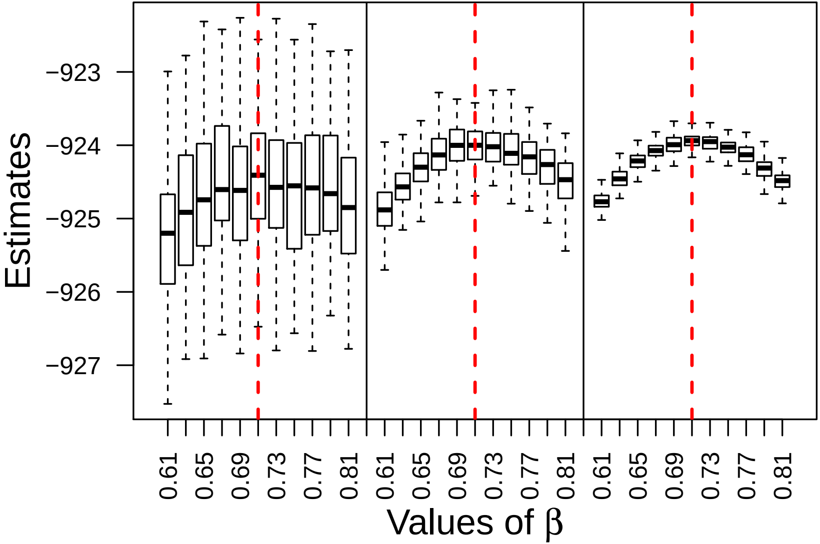


β

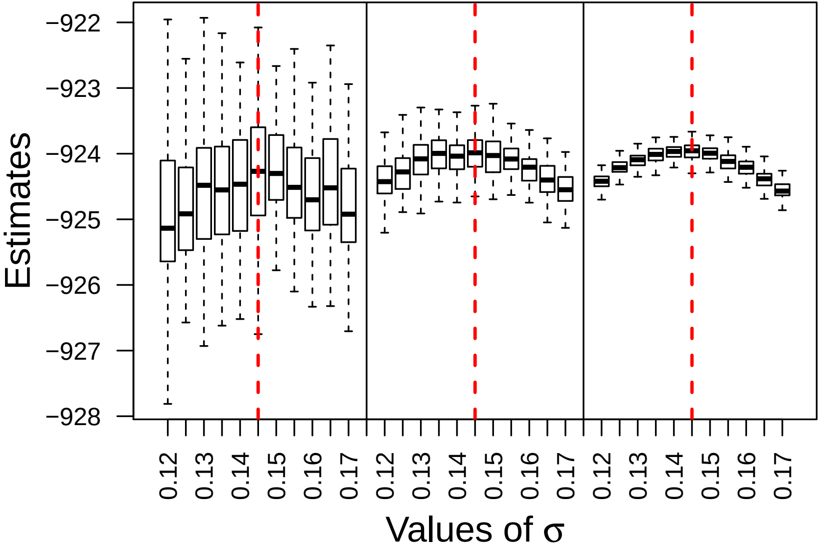
Boostrap $N = 1,000$.
iAPF $N_0 = 100$.
Comparable cost.
150,000 PMCMC iterations.



$Bootstrap : N = 1,000 \ / \ N = 10,000 \ / \ iAPF, N_0 = 100$



$Bootstrap: N = 1,000 \;/\; N = 10,000 \;/\; iAPF, N_0 = 100$



$Bootstrap : N = 1,000 \ / \ N = 10,000 \ / \ iAPF, N_0 = 100$

A More Challenging Stochastic volatility example

- The model is a multivariate stochastic volatility model from Chib et al. [3], with

$$\mu(\cdot) = \mathcal{N}(\cdot; m, U), \quad f(x, \cdot) = \mathcal{N}(\cdot; m + \Phi\,(x - m)\,, U),$$

  and $g(x, \cdot) = \mathcal{N}\left(\cdot; 0, \exp\left(\mathrm{diag}\,(x)\right)\right)$.
- We set $\Phi = \mathrm{diag}(\phi)$, and U is band-diagonal.
- The dataset is 20 international currencies, in the periods 3/2000–8/2008 (pre-crisis) and 9/2008–2/2016 (post-crisis).
- There are 79 parameters in (m, ϕ, U), and $T = \{102, 90\}$.

- We conducted parameter estimation using particle MCMC.

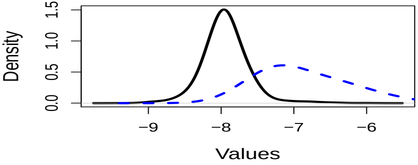
Stochastic volatility: P-MCMC

- The bootstrap particle filter systematically fails to provide reasonable marginal likelihood estimates in a feasible computational time.
- iAPF autocorrelation times sample size adjusted for autocorrelation

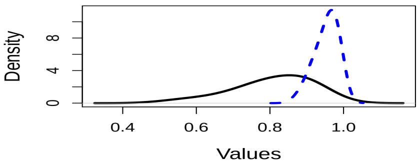
	$m_{£}$	$\phi_{£}$	$U_{£}$	$U_{£,€}$
pre-crisis	408	112	218	116
post-crisis	175	129	197	120

- Average number of particles at final iteration was about 1000.

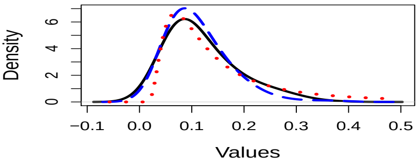
Stochastic volatility: P-MCMC



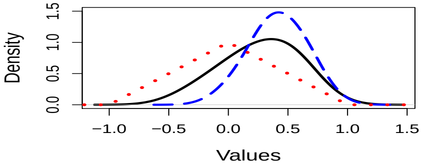
(a) $m_{£}$



(b) ϕ_{pounds}



(c) $U_{£}$



(d) $U_{£,€}$

Figure: Multivariate SV model: density estimates. Pre-crisis chain (solid), post-crisis chain (dashed) and prior density (dotted).

Ongoing work

- We consider

$$d\overline{X}_s = a\left(\overline{X}_s\right) ds + b\left(\overline{X}_s\right) dW_s, \qquad 0 \leq s \leq 1$$

  with standard Brownian motion W and the condition $\overline{X}_0 = \overline{x}_0$.
- We are interested in (approximately)

1. Simulating diffusion bridges, conditioning on the event $\left\{\overline{X}_0 = \overline{x}_0, \overline{X}_1 = \overline{x}_1\right\}$.
2. Evaluation of transition densities, e.g. $p\left(\bar{x}_0, \overline{x}_1\right)$.

- We employ an Euler–Maruyama approximation defined by $X_1 = \overline{x}_0$ and

$$X_t \sim \mathcal{N}\left(X_{t-1} + a\left(X_{t-1}\right) h, b^2\left(X_{t-1}\right) h\right),$$

  for $t \in \{2, \ldots, T\}$, with $T = 1/h$ so $X_T \approx \bar{X}_{1-h}$.

Model for a particle filter

- Euler–Maruyama approximation: $X_1 = \bar{x}_0$ and

$$X_t \sim \mathcal{N}\left(X_{t-1} + a\left(X_{t-1}\right)h, b^2\left(X_{t-1}\right)h\right),$$

  for $t \in \{2, \ldots, T\}$, and $T = 1/h$ so $X_T \approx \bar{X}_{1-h}$.

- If we want

$$p(\bar{x}_0, \bar{x}_1) \approx Z = \int_{\mathsf{X}^T} \mu_1\left(x_1\right) g_1\left(x_1\right) \prod_{t=2}^{T} f\left(x_{t-1}, x_t\right) g_t\left(x_t\right) dx_{1:T}$$

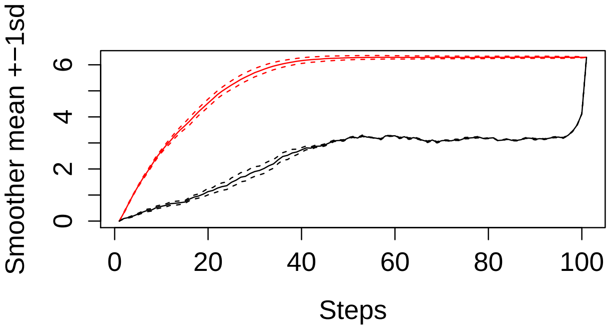
  we take $g_1 \equiv \cdots \equiv g_{T-1} \equiv 1$ and

$$g_T(x_T) = \mathcal{N}\left(\bar{x}_1; x_T + a\left(x_T\right)h, b^2\left(x_T\right)h\right).$$

- All the information comes at the end, if we run a standard particle filter.

Example

- We take
$$d\overline{X}_s = 50s \cdot \sin\left(\overline{X}_s\right) ds + 2dW_s,$$
$\bar{x}_0 = 0$ and $\bar{x}_1 = 2\pi$. [iAPF (red), BPF (black), $h = 1/100$]



- sin is negative on $(\pi, 2\pi) \Rightarrow$ more likely for the diffusion to approach 2π from above than below.

Conclusions

- To fully realise the potential of PMCMC we should exploit its flexibility.
- Even very simple variants on the standard particle filter can significantly improve performance.
- The iAPF can improve performance substantially in some settings.
- Extending the extent of its applicability / characterising it theoretically is ongoing work.
- In principle any *function approximation* scheme can be employed: provided that f_t^{ψ} can be sampled from, and g_t^{ψ} evaluated pointwise.
- Other [standard and less standard] ideas including blocking and tempering can also be readily employed (cf. [6]).

References

[1] C. Andrieu and G. O. Roberts. The pseudo-marginal approach for efficient Monte Carlo computations. *Annals of Statistics*, 37(2):697–725, 2009.

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

[3] S. Chib, Y. Omori, and M. Asai. Multivariate stochastic volatility. In T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. V. Mikosch, editors, *Handbook of Financial Time Series*, pages 365–400. Springer, 2009.

[4] 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.

[5] P. Guarniero, A. M. Johansen, and A. Lee. The iterated auxiliary particle filter. *Journal of the American Statistical Association*, 2016. In press.

[6] A. M. Johansen. On blocks, tempering and particle MCMC for systems identification. In Y. Zhao, editor, *Proceedings of 17th IFAC Symposium on System Identification*, pages 969–974, Beijing, China., 2015. IFAC. Invited submission.

[7] J. Owen, D. J. Wilkinson, and C. S. Gillespie. Likelihood free inference for markov processes: a comparison. *Statistical Applications in Genetics and Molecular Biology*, 2015. In press.

[8] N. Whiteley and A. Lee. Twisted particle filters. *Annals of Statistics*, 42(1): 115–141, 2014.