

# The Iterated Auxiliary Particle Filter and its Application to Diffusion Bridges

doi: [10.1080/01621459.2016.1222291](https://doi.org/10.1080/01621459.2016.1222291)

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

University of Warwick

[a.m.johansen@warwick.ac.uk](mailto:a.m.johansen@warwick.ac.uk)

[www2.warwick.ac.uk/go/amjohansen/talks/](http://www2.warwick.ac.uk/go/amjohansen/talks/)

Kings College London, February 8th, 2018

Apologies, this is a little old. . .

but she isn't.

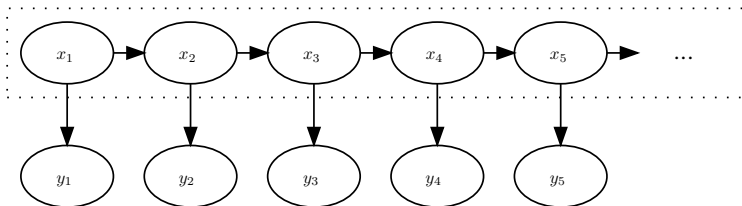


# Outline

- ▶ Background: SMC and Particle-MCMC
- ▶ Iterative Lookahead Methods
  - ▶ Motivation
  - ▶ Methodology
  - ▶ Toy Example: linear Gaussian
  - ▶ Application: stochastic volatility
  - ▶ Application: 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  $X_1^1, \dots, X_1^N \sim \mu_\theta$ .

For  $n > 1$ :

- ▶ Sample

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

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

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

# 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  $\theta$  [8, Section 4.2.1].
- ▶ Is valid for *very* general SMC algorithms.

## Twisting the HMM (a complement to [11])

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

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

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

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

and the sequence of non-negative functions  $(\tilde{\psi}_T \equiv 1)$ :

$$g_1^\psi(x_1) := g(x_1, y_1) \frac{\tilde{\psi}_1(x_1)}{\psi_1(x_1)} \tilde{\psi}_0, \quad g_t^\psi(x_t) := g(x_t, y_t) \frac{\tilde{\psi}_t(x_t)}{\psi_t(x_t)}.$$

## Proposition

For any sequence of bounded, continuous and positive functions  $\psi$ , let

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

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

The optimal choice is:

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

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



# Towards Iterative Auxiliary Particle Filters [6]

## $\psi$ -Auxiliary Particle Filter

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

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

## Necessary features of $\psi$

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

# A Recursive Approximation

## Proposition

The sequence  $\psi^*$  satisfies  $\psi_T^*(x_T) = g(x_T, y_T)$ ,  $x_T \in X$  and

$$\psi_t^*(x_t) = g(x_t, y_t) f(x_t, \psi_{t+1}^*), \quad x_t \in X, \quad t \in \{1, \dots, T-1\}.$$

---

## Algorithm 1 Recursive function approximations

---

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

1. Set  $\psi_t^i \leftarrow g(\xi_t^i, y_t) f(\xi_t^i, \psi_{t+1}^i)$  for  $i \in \{1, \dots, N\}$ .
  2. Choose  $\psi_t$  as a member of  $\Psi$  on the basis of  $\xi_t^{1:N}$  and  $\psi_t^{1:N}$ .
-

# Iterated Auxiliary Particle Filters

---

**Algorithm 2** An iterated auxiliary particle filter with parameters  $(N_0, k, \tau)$

---

1. Initialize: set  $\psi_t^0 = \mathbf{1}$ .  $l \leftarrow 0$ .
  2. Repeat:
    - 2.1 Run a  $\psi^l$ -APF with  $N_l$  particles; set  $\hat{Z}_l \leftarrow Z_{\psi^l}^{N_l}$ .
    - 2.2 If  $l > k$  and  $\text{sd}(\hat{Z}_{l-k:l}) / \text{mean}(\hat{Z}_{l-k:l}) < \tau$ , go to 3.
    - 2.3 Compute  $\psi^{l+1}$  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  $\psi^l$ -APF. Return  $\hat{Z} := Z_{\psi}^{N_l}$ .
-

# An Elementary Implementation

## Function Approximation

- ▶ Numerically obtain (a regularised solution to):

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

- ▶ 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
- ▶  $\tau = 0.5$

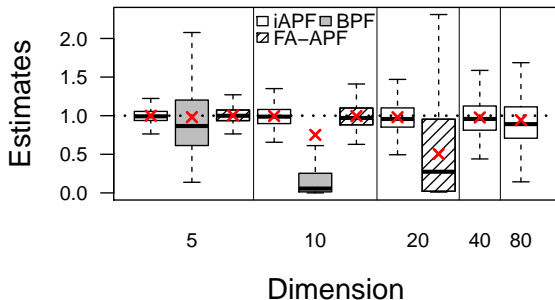
## Resampling

- ▶ Multinomial when  $\text{ESS} < N/2$ .

# A Linear Gaussian Model: Behaviour with Dimension

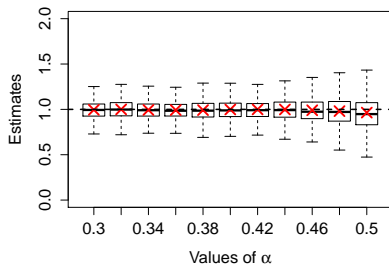
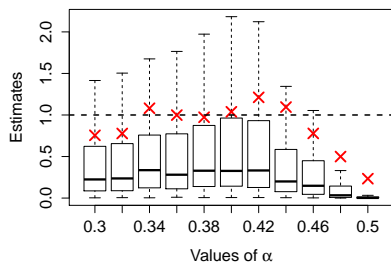
$$\begin{aligned} \mu &= \mathcal{N}(\cdot; \mathbf{0}, I_d) & f(x, \cdot) &= \mathcal{N}(\cdot; Ax, I_d) \\ \text{and } g(x, \cdot) &= \mathcal{N}(\cdot; x, I_d) & \text{where } A_{ij} &= 0.42^{|i-j|+1}, \end{aligned}$$

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



# Linear Gaussian Model: Sensitivity to Parameters

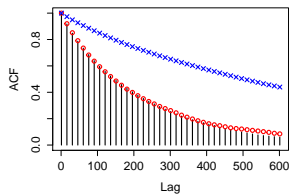
Fixing  $d = 10$ : Bootstrap ( $N = 50,000$ ) / iAPF ( $N_0 = 1,000$ )



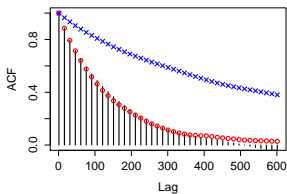
Box plots of  $\hat{\Sigma}$  for different values of the parameter  $\alpha$  using 1000 replicates.

# Linear Gaussian Model: PMMH Empirical Autocorrelations

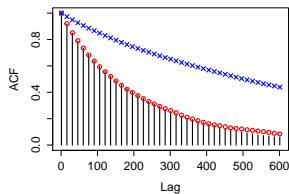
$A_{11}$



$A_{41}$



$A_{55}$



In this case:

$$d = 5$$

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

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

$$\text{and } g(x, \cdot) = \mathcal{N}(\cdot; x, 0.25I_d)$$

$$A = \begin{pmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 & 0 \\ 0.1 & 0.2 & 0.6 & 0 & 0 \\ 0.4 & 0.1 & 0.1 & 0.3 & 0 \\ 0.1 & 0.2 & 0.5 & 0.2 & 0 \end{pmatrix},$$

(unknown lower triangular matrix)

# Stochastic Volatility

- ▶ A simple stochastic volatility model is defined by:

$$\mu(\cdot) = \mathcal{N}(\cdot; 0, \sigma^2 / (1 - a)^2)$$

$$f(x, \cdot) = \mathcal{N}(\cdot; ax, \sigma^2)$$

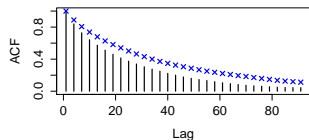
$$\text{and } g(x, \cdot) = \mathcal{N}(\cdot; 0, \beta^2 \exp(x)),$$

where  $a \in (0, 1)$ ,  $\beta > 0$  and  $\sigma^2 > 0$  are unknown.

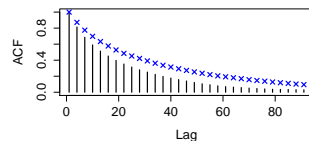
- ▶ Considered  $T = 945$  observations  $y_{1:T}$  corresponding to the mean-corrected daily returns for the GBP/USD exchange rate from 1/10/81 to 28/6/85.



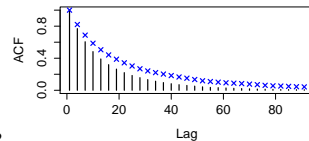
# Estimated PMCMC Autocorrelation



$a$

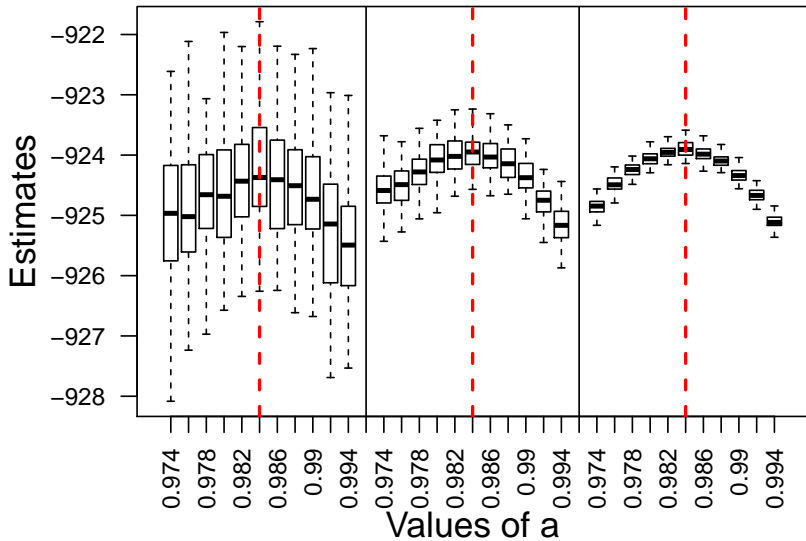


$\sigma$

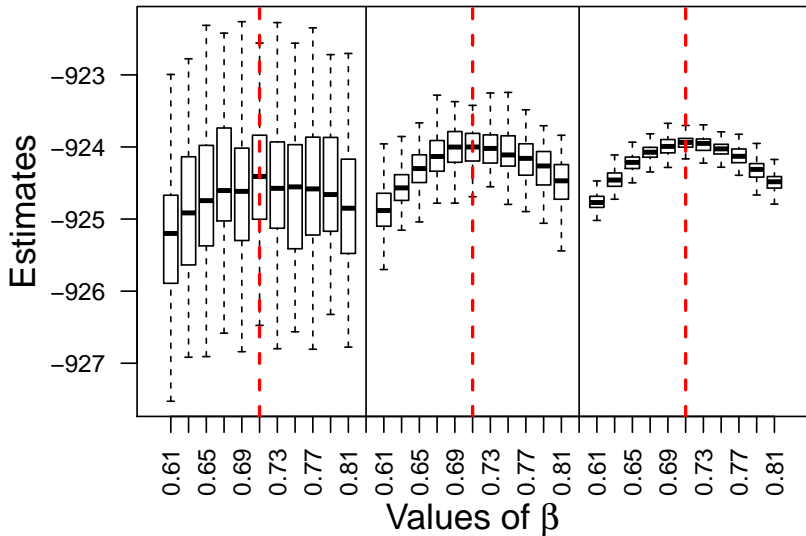


$\beta$

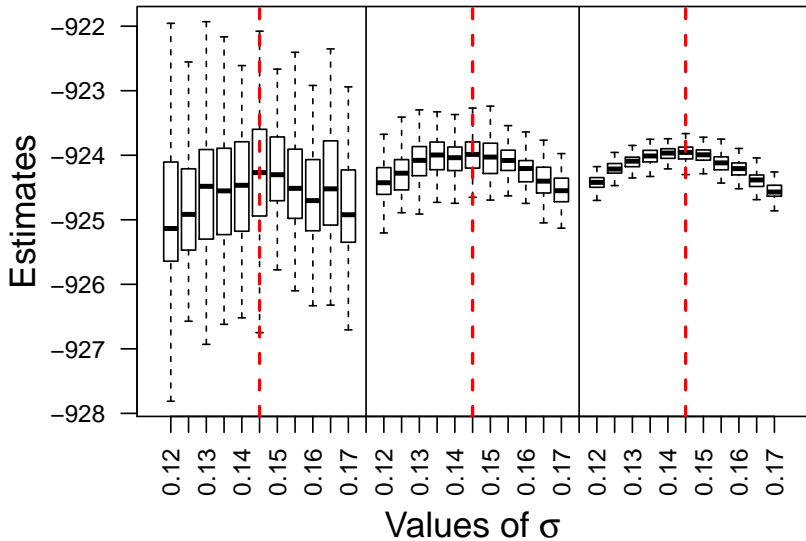
Bootstrap  $N = 1,000$ .  
iAPF  $N_0 = 100$ .  
Comparable cost.  
150,000 PMCMC iterations.



*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 motivated by 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  $g(x, \cdot) = \mathcal{N}(\cdot; 0, \exp(\text{diag}(x)))$ .

- ▶ We set  $\Phi = \text{diag}(\phi)$ , 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  $(m, \phi, U)$ , and  $T = \{102, 90\}$ .
- ▶ We conducted parameter estimation using particle MCMC.

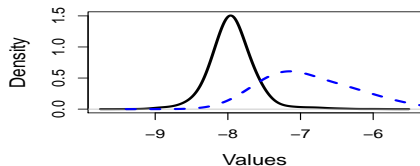
## Stochastic volatility: P-MCMC

- ▶ 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

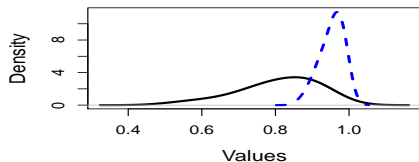
	$m_{\mathcal{L}}$	$\phi_{\mathcal{L}}$	$U_{\mathcal{L}}$	$U_{\mathcal{L},\epsilon}$
pre-crisis	408	112	218	116
post-crisis	175	129	197	120

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

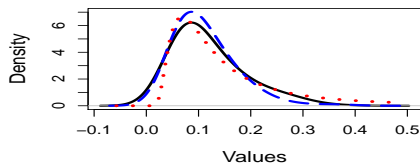
# Stochastic volatility: P-MCMC



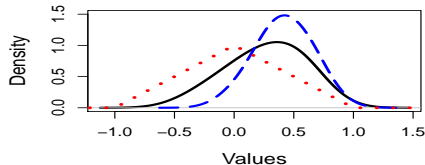
(a)  $m_E$



(b)  $\phi_{\text{pounds}}$



(c)  $U_E$



(d)  $U_{E,\epsilon}$

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

## A Natural Application — Guarniero, 2017 [5]

- ▶ We consider

$$d\bar{X}_s = a(\bar{X}_s) ds + b(\bar{X}_s) dW_s, \quad 0 \leq s \leq 1$$

with standard Brownian motion  $W$  and the condition  $\bar{X}_0 = \bar{x}_0$ .

- ▶ We are interested in (approximately)
  1. Simulating diffusion bridges, conditioning on the event  $\{\bar{X}_0 = \bar{x}_0, \bar{X}_1 = \bar{x}_1\}$ .
  2. Evaluation of transition densities, e.g.  $p(\bar{x}_0, \bar{x}_1)$ .
- ▶ We employ an Euler–Maruyama approximation defined by  $X_1 = \bar{x}_0$  and

$$X_t \sim \mathcal{N}(X_{t-1} + ha(X_{t-1}), hb(X_{t-1})b^T(X_{t-1})),$$

for  $t \in \{2, \dots, T\}$ , 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

$$X_t \sim \mathcal{N}(X_{t-1} + ha(X_{t-1}), hb(X_{t-1})b^T(X_{t-1})),$$

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

- ▶ If we want

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

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

$$g_T(\cdot) = \mathcal{N}(\bar{x}_1; x_T + ha(x_T)h, hb(x_T)b^T(X_T)).$$

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

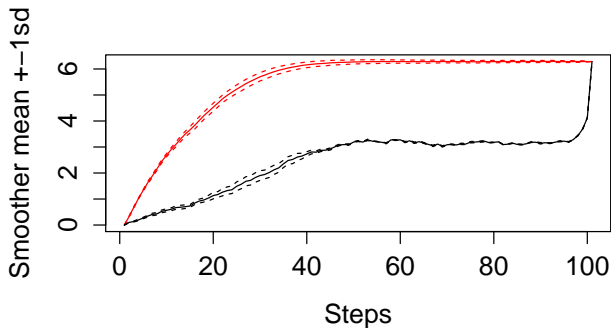
Cf. [9].

## Example

- ▶ We take

$$d\bar{X}_s = 50s \cdot \sin(\bar{X}_s) ds + 2dW_s,$$

$\bar{x}_0 = 0$  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  $2\pi$  from above than below.

# Multivariate Heston Model

A slight generalisation of Heston's SV model:

$$\begin{aligned} dv_t &= \theta_1(v_t - \theta_2) + \theta_3\sqrt{v_t}W_t \\ dS_t^{(i)} &= \mu_i S_t^{(i)} dt + S_t^{(i)}\sqrt{v_t}W_t^{(i)}, \quad i \in \{1, \dots, d\} \end{aligned}$$

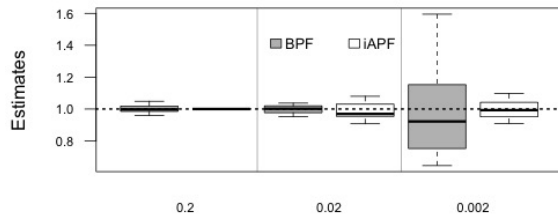
Compare BPF( $N = 10,000$ ) and iAPF( $N_0 = 100$ ) with parameters:

$$v_t = 0.5 \quad \theta_1 = 1 \quad \theta_2 = 0.5 \quad \theta_3 = 0.02$$

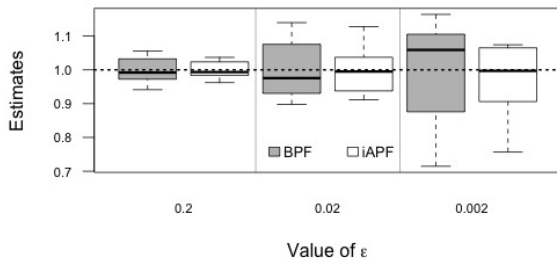
and discretization timescale  $10^{-5}$ .

We consider  $\text{corr}\left(W_t^{(i)}, W_t^{(j)}\right) = \mathbb{I}[|i - j| = 1] \rho$  for  $i \neq j$ .

$$d = 2$$

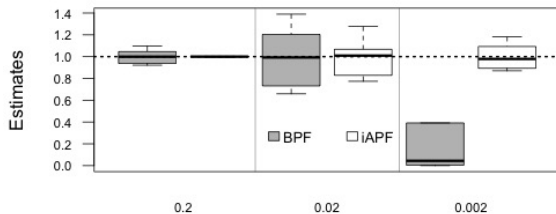


$$\rho = 0:$$

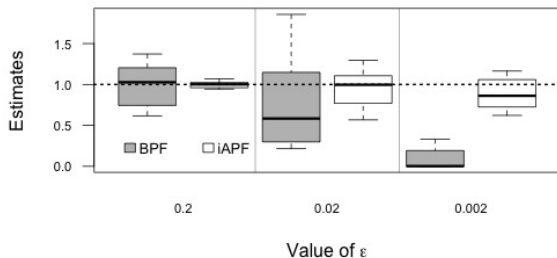


$$\rho = 0.25:$$

$d = 5$

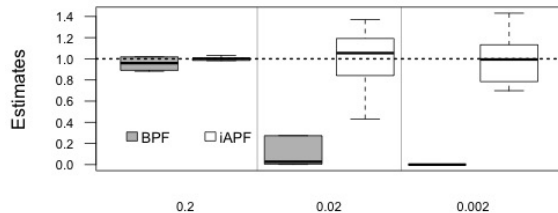


$\rho = 0$ :

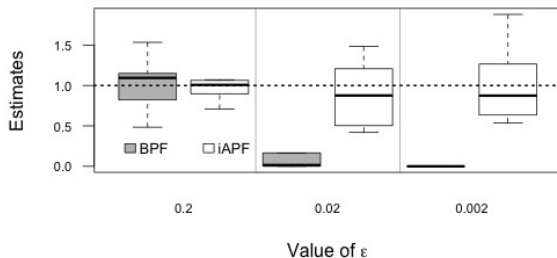


$\rho = 0.25$  :

$$d = 10$$



$$\rho = 0:$$



$$\rho = 0.25:$$

# Conclusions

- ▶ When using PMCMC we should exploit its flexibility.
- ▶ The iAPF can improve performance substantially in some settings.
- ▶ Closely-related ideas have also been proposed, e.g. [10, 7]
- ▶ The perspective here is one of *function approximation*.
- ▶ Extending the extent of its applicability / characterising it theoretically might be interesting.
- ▶ In principle any *function approximation* scheme can be employed: provided that  $f_t^\psi$  can be sampled from, and  $g_t^\psi$  evaluated pointwise.
- ▶ Gaussian transitions are quite common — e.g. discretized diffusions.

## 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. *The Iterated Auxiliary Particle Filter and Applications to State Space Models and Diffusion Processes*. Ph.d. thesis, University of Warwick, 2017.
- [6] P. Guarniero, A. M. Johansen, and A. Lee. The iterated auxiliary particle filter. *Journal of the American Statistical Association*, 112(520):1636–1647, 2017.
- [7] J. Heng, A. N. Bishop, G. Deligiannidis, and A. Doucet. Controlled sequential monte carlo. ArXiv Mathematics e-prints 1708.08396, ArXiv, 2017.
- [8] 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.
- [9] H.-C. Ruiz and H. J. Kappen. Particle smoothing for hidden diffusion processes: Adaptive path integral smoother. *IEEE Transactions on Signal Processing*, 65(12):3191–3203, 2017.
- [10] S. Thijssen and H. J. Kappen. Path integral control and state-dependent feedback. *Physical Review E*, 91(3), 2015.
- [11] N. Whiteley and A. Lee. Twisted particle filters. *Annals of Statistics*, 42(1):115–141, 2014.