Multilevel particle filter

Kody Law
& A. Jasra (NUS), K. Kamatani (Osaka), Y. Zhou (NUS)

OAK RIDGE National Laboratory

UQ annual workshop, KAUST, Thuwal, KSA

January 6, 2016

Outline

Orientation

Aim: Approximately sample from sequence of probability distributions $\eta_{\infty,m}$, which need to be approximated by some $\eta_{L,m}$, for $m = 1, 2, \ldots,$ each given by a Bayesian inversion.

Solution: The multilevel Monte Carlo (MLMC) framework is extended to the multilevel particle filter (MLPF).

- MLMC methods *reduce cost to* error$= \mathcal{O}(\varepsilon)$, can be used in the case that $\eta_{L,m}$ can be sampled from directly [G08].
- Here it is assumed that $\eta_{\infty,m}$ and $\eta_{L,m}$ cannot be sampled from directly.
- Particle filters are Monte Carlo based algorithms which provide consistent approximations of such distributions [AMGC02].

Outline

Outline

Single level Monte Carlo

Aim: Approximate $\eta(g) := \mathbb{E}_\eta(g)$ for $g : E \to \mathbb{R}$.

Monte Carlo approach

- Discretize the problem $\Rightarrow$ *approximate* distribution η_L.
- Sample $U_L^{(i)} \sim \eta_L$ i.i.d., and approximate

$$\eta_L(g) := \mathbb{E}_{\eta_L}(g) \approx \widehat{Y}_L^{N_L} := \frac{1}{N_L}\sum_{i=1}^{N_L} g(U_L^{(i)}).$$

- Mean square error (MSE) $\mathbb{E}\{\widehat{Y}_L - \mathbb{E}_{\eta_\infty}[g(U)]\}^2$ splits into

$$\underbrace{\mathbb{E}\{\widehat{Y}_L - \mathbb{E}_{\eta_L}[g(U)]\}^2}_{\text{variance}=\mathcal{O}(N_L^{-1})} + \{\underbrace{\mathbb{E}_{\eta_L}[g(U)] - \mathbb{E}_{\eta_\infty}[g(U)]}_{\text{bias}}\}^2$$

- Cost to achieve MSE$= \mathcal{O}(\varepsilon^2)$ is Cost$(U_L^{(i)}) \times \varepsilon^{-2}$.

Multilevel Monte Carlo I

Introduce a hierarchy of discretization levels $\{\eta_\ell\}_{\ell=1}^{L}$ and define $Y_\ell = \{\mathbb{E}_{\eta_\ell}[g(U)] - \mathbb{E}_{\eta_{\ell-1}}[g(U)]\}$, with $\eta_{-1} := 0$.
Observe the telescopic sum

$$\mathbb{E}_{\eta_L}[g(U)] = \sum_{\ell=0}^{L} Y_\ell .$$

Each term can be unbiasedly approximated by

$$Y_\ell^{N_\ell} = \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \{g(U_\ell^{(i)}) - g(U_{\ell-1}^{(i)})\}$$

where $g(U_{-1}^{(i)}) := 0$.

Multilevel Monte Carlo II

Multilevel Monte Carlo approach:

- Sample i.i.d. $(U_\ell, U_{\ell-1})^{(i)} \sim \bar{\eta}^\ell$, such that $\int \bar{\eta}^\ell du_{\ell-1,\ell} = \eta_{\ell,\ell-1}$, and approximate

$$\eta_L(g) \approx \widehat{Y}_{L,\mathrm{Multi}} := \sum_{\ell=0}^{L} Y_\ell^{N_\ell} .$$

- Mean square error (MSE) given by

$$\mathbb{E}\{\widehat{Y}_{L,\mathrm{Multi}} - \mathbb{E}_{\eta_\infty}[g(U)]\}^2 = \underbrace{\mathbb{E}\{\widehat{Y}_{L,\mathrm{Multi}} - \mathbb{E}_{\eta_L}[g(U)]\}^2}_{\mathrm{variance}=\sum_{\ell=0}^{L} V_\ell/N_\ell} + \underbrace{\{\mathbb{E}_{\eta_L}[g(U)] - \mathbb{E}_{\eta_\infty}[g(U)]\}^2}_{\mathrm{bias}} .$$

- Fix bias by choosing L. Minimize cost $C = \sum_{\ell=0}^{L} C_\ell N_\ell$ as a function of $\{N_\ell\}_{\ell=0}^{L}$ for fixed variance $\Rightarrow N_\ell \propto \sqrt{V_\ell / C_\ell}$.

Multilevel vs. Single level

Assume $h_\ell = M^{-\ell}$ and there are α, and $\beta > \zeta$ such that

(i) weak error $|\mathbb{E}[g(U_\ell) - g(U)]| = \mathcal{O}(h_\ell^\alpha)$.

(ii) strong error $\mathbb{E}|g(U_\ell) - g(U)|^2 = \mathcal{O}(h_\ell^\beta) \Rightarrow V_\ell = \mathcal{O}(h_\ell^\beta)$,

(iii) computational cost for a realisation of $g(U_\ell) - g(U_{\ell-1})$, $C_\ell = \mathcal{O}(h_\ell^{-\zeta})$.

In both cases, require $h_L^\alpha = \mathcal{O}(\varepsilon) \Rightarrow L \propto |\log \varepsilon|$.

- Single level cost $C = \mathcal{O}(\varepsilon^{-\zeta/\alpha - 2})$: cost per sample is $C_L \propto \varepsilon^{-\zeta/\alpha}$, and fixed $V \propto \varepsilon^2 \Rightarrow N_L \propto \varepsilon^{-2}$.
- Multilevel cost $C_{\rm ML} = \mathcal{O}(\varepsilon^{-2})$: $N_\ell \propto \varepsilon^{-2} h_\ell^{(\beta+\zeta)/2}$, so $V \propto \varepsilon^2 K_L$ and $C \propto \varepsilon^{-2} K_L$ for $K_L = \sum_{\ell=0}^L h_\ell^{(\beta-\zeta)/2} = \mathcal{O}(1)$ [G08] – cost of simulating a scalar random variable.
- Example: Milstein solution of SDE

$$C = \mathcal{O}(\varepsilon^{-3}) \quad \text{vs.} \quad C_{\rm ML} = \mathcal{O}(\varepsilon^{-2}).$$

Outline

Filtering Problem

Framework:

$$X_{n+1} \sim Q(X_n, \cdot)\,,$$

$$Y_n|X_n \text{ has density } G(y_n, x_n).$$

Objective: Approximate $\mathbb{E}(\varphi(X_n)|y_1, \ldots, y_n)$, where $y_k \in \mathbb{R}^m$ is a realization of Y_k and $\varphi : \mathbb{R}^d \to \mathbb{R}$.

The joint probability density of state and observations is

$$\prod_{i=1}^{n} G(y_i, x_i) Q(x_{(i-1)}, x_i).$$

We further assume we can only approximate $X^\ell_{n+1} \sim Q^\ell(X^\ell_n, \cdot)$, at a succession of resolutions $0 < \ell \to \infty$, where $Q^\infty := Q$.

Outline

Particle filter

$Q^L(x, \cdot)$: kernel associated to level L with initial condition x. This will be Euler-Maruyama discretization of an SDE.

Generate $\widehat{\eta}^{N_L}_{L,m} = \frac{1}{N_L} \sum_{i=1}^{N_L} \delta_{\widehat{U}^{L,i}_m} \approx \widehat{\eta}_{L,m}$ using

Particle filter algorithm:
For $i = 1, \ldots, N_L$, draw $\widehat{U}^{L,i}_0 \sim \mu_0$.
Initialize $m = 1$. **Do**

(i) **For** $i = 1, \ldots, N_L$, draw $U^{L,i}_m \sim Q^L(\widehat{U}^{L,i}_{m-1}, \cdot)$;

(ii) **For** $k = 1, \ldots, N_L$, draw $I^{L,k}_m$ according to multinomial distribution $\{w^i_m\}_{i=1}^{N_L}$, where $w^i_m := G_m(U^{L,i}) / \sum_{j=1}^{N_L} G_m(U^{L,j})$.

(iii) $\widehat{U}^{L,k}_m \leftarrow U^{L,I^{L,k}_m}_m$.

$m \leftarrow m + 1$

Outline

Multilevel particle filter

$M^\ell([x,y],\cdot)$: coupled kernel with marginals $Q^\ell(x,\cdot)$ and $Q^{\ell-1}(y,\cdot)$.

Generate $\widehat{\eta}^{ML}_{L,m} = \sum_{\ell=0}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} (\delta_{\widehat{U}^{\ell,i}_{m,1}} - \delta_{\widehat{U}^{\ell,i}_{m,2}}) \approx \widehat{\eta}_{L,m}$, with $\delta_{\widehat{U}^{0,i}_{m,2}} := 0$, using

Multilevel particle filter algorithm:

For $\ell = 0,1,\dots,L$ and $i = 1,\dots,N_\ell$, draw $\widehat{U}^{\ell,i}_{0,1} \sim \mu_0$, and let $\widehat{U}^{\ell,i}_{0,2} = \widehat{U}^{\ell,i}_{0,1}$.

Initialize $m = 1$. **Do**

(i) **For** $\ell = 0,1,\dots,L$ and $i = 1,\dots,N_\ell$, draw $(U^{\ell,i}_{m,1}, U^{\ell,i}_{m,2}) \sim M^\ell((\widehat{U}^{\ell,i}_{m-1,1}, \widehat{U}^{\ell,i}_{m-1,2}), \ \cdot\ )$;

(ii) **For** $\ell = 0,1,\dots,L$ and $k = 1,\dots,N_\ell$, draw $(I^{\ell,k}_{m,1}, I^{\ell,k}_{m,2})$ according to the coupled resampling procedure;

(iii) $(\widehat{U}^{\ell,k}_{m,1}, \widehat{U}^{\ell,k}_{m,2}) \leftarrow (U^{\ell,I^{\ell,k}_1}_{m,1}, U^{\ell,I^{\ell,k}_2}_{m,2})$ for $k = 1,\dots,N_\ell$.

$m \leftarrow m+1$

Coupled resampling I

Given $\{\{U_{m,1}^{\ell,i}, U_{m,2}^{\ell,i}\}_{i=1}^{N_\ell}\}_{\ell=0}^{L}$,

For $\ell = 0, 1, \ldots, L$ define

$$w_{m,1}^{\ell,i} = \frac{G_m(U_{m,1}^{\ell,i})}{\sum_{j=1}^{N_\ell} G_m(U_{m,1}^{\ell,i})} \qquad \text{and} \qquad w_{m,2}^{\ell,i} = \frac{G_m(U_{m,2}^{\ell,i})}{\sum_{j=1}^{N_\ell} G_m(U_{m,2}^{\ell,i})}.$$

Coupled resampling II

Coupled resampling procedure:

a. with probability $\alpha_m^\ell = \sum_{i=1}^{N_\ell} w_{m,1}^{\ell,i} \wedge w_{m,2}^{\ell,i}$, draw $I_{m,1}^{\ell,k}$ according to

$$\mathbb{P}(I_{m,1}^\ell = i) = \frac{w_{m,1}^{\ell,i} \wedge w_{m,2}^{\ell,i}}{\sum_{j=1}^{N_\ell} w_{m,1}^{\ell,j} \wedge w_{m,2}^{\ell,j}}, \qquad i = 1, \ldots, N_\ell.$$

and let $I_{m,2}^{\ell,k} = I_{m,1}^{\ell,k}$.

b. with probability $1 - \alpha_m^\ell$, draw $(I_{m,1}^{\ell,k}, I_{m,2}^{\ell,k})$ *independently* according to the probabilities

$$\mathbb{P}(I_{m,1}^\ell = i) = [w_{m,1}^{\ell,i} - w_{m,1}^{\ell,i} \wedge w_{m,2}^{\ell,i}]/(\sum_{j=1}^{N_\ell} w_{m,1}^{\ell,j} - w_{m,1}^{\ell,j} \wedge w_{m,2}^{\ell,j});$$

$$\mathbb{P}(I_{m,2}^\ell = i) = [w_{m,2}^{\ell,i} - w_{m,1}^{\ell,i} \wedge w_{m,2}^{\ell,i}]/(\sum_{j=1}^{N_\ell} w_{m,2}^{\ell,j} - w_{m,1}^{\ell,j} \wedge w_{m,2}^{\ell,j}),$$

for $i = 1, \ldots, N_\ell$.

Assuming 1-step strong error order β, weak error order α, and cost ζ (for Euler-Maruyama $\alpha = \beta = \zeta = 1$), the following theorem holds:

Theorem (JKLZ15)

Under suitable regularity assumptions on M^ℓ and G, for any $\varphi \in \mathcal{B}_b(\mathbb{R}^d) \cap Lip(\mathbb{R}^d)$

$$\mathbb{E}[\{\widehat{\eta}_m^{ML}(\varphi) - \widehat{\eta}_m^{L}(\varphi)\}^2] \lesssim \sum_{\ell=1}^{L} \frac{h_\ell^{\beta/2}}{N_\ell}$$

In particular, for $\beta/2 > \zeta$, L and $\{N_\ell\}_{\ell=0}^{L}$ can be chosen such that MSE$= \mathcal{O}(\varepsilon^2)$ for computational cost$= \mathcal{O}(\varepsilon^{-2})$.

Note that the coupled resampling effectively reduces rate $\beta \to \beta/2$.

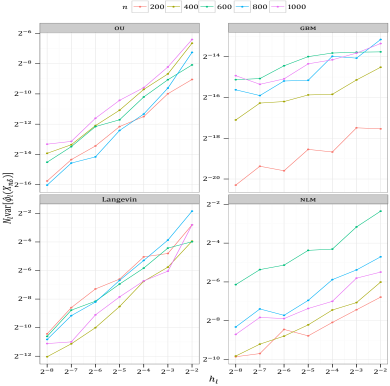
Numerical examples

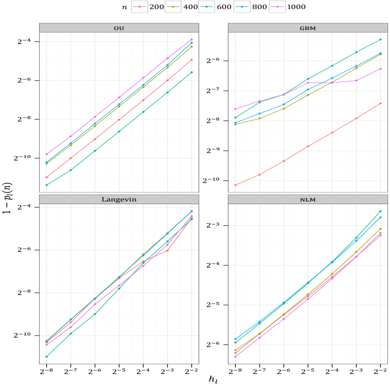
- $dX_t = a(X_t)dt + b(X_t)dW_t, \qquad X_0 = x_0,$
  with $X_t \in \mathbb{R}^d$, $t \geq 0$ and $\{W_t\}_{t\in[0,T]}$ a Brownian motion of appropriate dimension.
- Partial observations $\{y_1, \ldots, y_n\}$ available and $Y_k|X_k$ has a density function $G(y_k, x_k) =: G_k(x_k)$.
- Euler Marayuma discretization with $h_\ell = 2^{-\ell}$. For constant diffusion $\beta = 2$, non-constant $\beta = 1$.

Example	$a(x)$	$b(x)$	$G(y;x)$	$\varphi(x)$
OU	$\theta(\mu - x)$	σ	$\mathcal{N}(x, \tau^2)$	x
GBM	μx	σx	$\mathcal{N}(\log x, \tau^2)$	x
Langevin	$\frac{1}{2}\nabla \log \pi(x)$	σ	$\mathcal{N}(0, \tau^2 e^x)$	$\tau^2 e^x$
NLM	$\theta(\mu - x)$	$\frac{\sigma}{\sqrt{1+x^2}}$	$\mathcal{L}(x, s)$	x

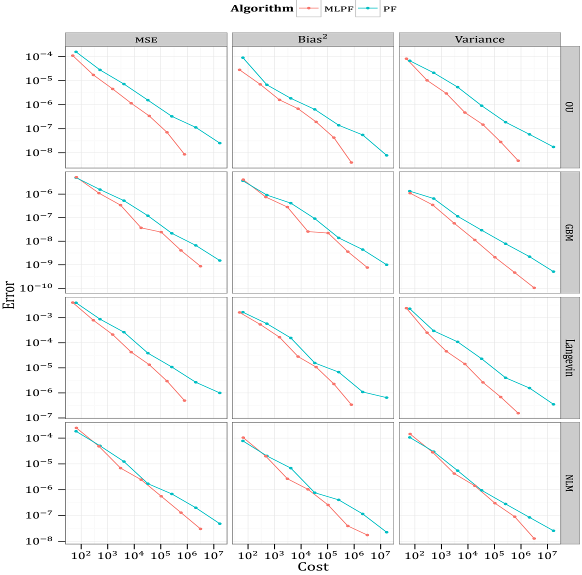
OAK RIDGE National Laboratory

Numerical examples: rates





Numerical examples: cost



Outline

Summary

- Multilevel particle filter (MLPF) can perform asymptotically as well as MLMC.
- Cost-to-ε can be asymptotically the same as for a scalar random variable!
- MLPF strong error is effectively reduced by coupled resampling $\beta \to \beta/2$.
- Examples have $\beta/2 \leq \zeta$. For $\beta/2 > \zeta$ optimal results can be obtained again.
- If $\zeta > 2\alpha$ then the optimal cost is $\varepsilon^{-\zeta/\alpha}$, the cost of a single simulation at the finest level.
- Many improvements and more complicated models are forthcoming.

References

- **[JKLZ15]**: Jasra, Kamatani, Law, Zhou. "Multilevel particle filter." http://arxiv.org/abs/1510.04977 (2015).
- **[BJLTZ15]**: Beskos, Jasra, Law, Tempone, Zhou. "Multilevel Sequential Monte Carlo samplers." http://arxiv.org/abs/1503.07259 (2015).
- **[HLT15]**: Hoel, Law, Tempone. "Multilevel ensemble Kalman filter." arxiv:1502.06069 (2015).
- **[G08]**: Giles. "Multilevel Monte Carlo." Op. Res., 56, 607-617 (2008).
- **[AMGC]** Arulampalam, Maskell, Gordon, Clapp. "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking." IEEE Tran Sig Proc 50:2, 174 (2002).
- **[D04]**: Del Moral. "Feynman-Kac Formulae." Springer: New York (2004).
- **[GCR]**: Cotter, Gregory, Reich. "Multilevel ensemble transform particle filter." arxiv:1509.00325 (2015).

Thank you!

New book!

Data Assimilation: A Mathematical Introduction. Kody Law, Andrew Stuart, Kostas Zygalakis. Springer (Oct. 2015).