Bayesian parameter estimation using Multilevel and multi-index Monte Carlo

Kody Law

joint with A. Jasra (NUS), K. Kamatani (Osaka), Y. Xu (NUS*), & Y. Zhou (Cubist)

MANCHESTER 1824
The University of Manchester

Monash Workshop on Numerical Differential Equations and Applications
Monash University, AU

February 12, 2020

Outline

Orientation

Aim: Approximate posterior expectations of the state path and static parameters associated to an S(P)DE which must be finitely approximated.

Solution: Apply an approximate coupling strategy so that multi-index Monte Carlo (MIMC) methods can be used within a particle MCMC [B02, AR08, ADH10] and SMC^2 [CJP13].

- MLMC ($d = 1$) [H00, G08] and MIMC ($d > 1$) [HNT15] methods *reduce cost to* mean-squared error$= \mathcal{O}(\varepsilon^2)$;
- Recently this methodology has been applied to *inference*, mostly in cases where target can be evaluated up to a normalizing constant [HSS13, DKST15, HTL16, BJLTZ17].
- Here we can only simulate a non-negative unbiased estimator (utanc); using PMCMC we are able to sample consistently from an approximate coupling of successive targets [JKLZ18.i, JKLZ18.ii], and this is extended to the sequential context via SMC^2 [JLX19].

Outline

Example: expectation for SDE [G08]

Estimation of expectation of solution of intractable stochastic differential equation (SDE).

$$dX = f(X)dt + \sigma(X)dW, \quad X_0 = x_0 \ .$$

Aim: estimate $\mathbb{E}(g(X_T))$.
We need to

(1) Approximate, e.g. by Euler-Maruyama method with resolution h:

$$X_{n+1} = X_n + hf(X_n) + \sqrt{h}\sigma(X_n)\xi_n, \quad \xi_n \sim N(0,1).$$

(2) Sample $\{X_{N_T}^{(i)}\}_{i=1}^N$, $N_T = T/h$.

Multilevel Monte Carlo (MLMC)

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

- Single level estimator: $\frac{1}{N}\sum_{i=1}^{N} g(U_L^{(i)})$, $U_L^{(i)} \sim \eta_L$ i.i.d.
  Cost to achieve MSE$= \mathcal{O}(\varepsilon^2)$ is $C =$Cost$(U_L^{(i)}) \times \varepsilon^{-2}$.

- Multilevel estimator*: $\sum_{l=0}^{L} \frac{1}{N_l} \sum_{i=1}^{N_l} \{g(U_l^{(i)}) - g(U_{l-1}^{(i)})\}$, $(U_l, U_{l-1})^{(i)} \sim \bar{\eta}^l$ i.i.d. such that $\int \bar{\eta}^l du_{l-1,l} = \eta_{l,l-1}$ for $l = 0, \ldots, L$. (*$g(U_{-1}^{(i)}) := 0$)
  Cost is $C_{\rm ML} = \sum_{l=0}^{L} C_l N_l$, where C_l is the cost to obtain a sample from $\bar{\eta}^l$.
- Fix bias by choosing L. Minimize cost $C_{\rm ML}(\{N_l\}_{l=0}^{L})$ for fixed variance$= \sum_{l=0}^{L} V_l/N_l$, $\Rightarrow N_l \propto \sqrt{V_l/C_l}$.

- Example: Milstein solution of SDE for MSE$= \mathcal{O}(\varepsilon^2)$

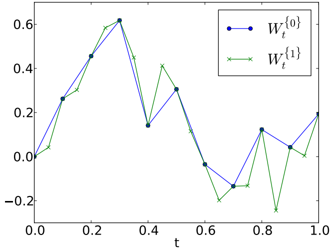
$$C = \mathcal{O}(\varepsilon^{-3}) \quad \text{vs.} \quad C_{\rm ML} = \mathcal{O}(\varepsilon^{-2}).$$

Illustration of pairwise coupling

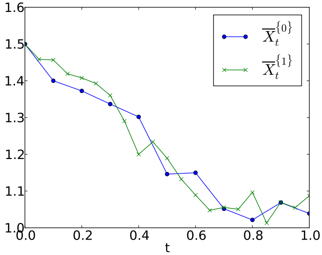
Pairwise coupling of trajectories of an SDE:

$$X^1_{n+1} = X^1_n + hf(X^1_n) + \sqrt{h}\sigma(X^1_n)\xi_n, \quad \xi_n \sim N(0,1), \quad n = 0, \dots, N_1$$

$$X^0_{n+1} = X^0_n + (2h)f(X^0_n) + \sqrt{h}\sigma(X^0_n)(\xi_{2n} + \xi_{2n+1}), \; n = 0, \dots, N_1/2.$$



(a) Wiener process $W^1_n = \sqrt{h}\sum_{i=0}^n \xi_n$, $W^0_n = W^1_{2n}$.



(b) Stochastic process driven by Wiener process.

Outline

Bayesian inference is about approximating integrals

Suppose we know how to evaluate $\gamma(x)$ for $x \in \mathsf{X}$.

Let

$$\eta(dx) = \frac{\gamma(x)dx}{\int_{\mathsf{X}} \gamma(x)dx} ,$$

and $\varphi : \mathsf{X} \to \mathbb{R}$, and suppose we want to estimate

$$\eta(\varphi) := \int_{\mathsf{X}} \varphi(x)\eta(dx) .$$

X may be quite high dimension, e.g. $\mathbb{R}^d$ with $d = 100$ easily, or even 1000, 10000, etc...

Monte Carlo

If we could obtain i.i.d. samples $x^i \sim \eta$, then we could use

$$\eta(\varphi) \approx \frac{1}{N} \sum_{i=1}^{N} \varphi(x^i) \, .$$

Convergence rate (of MSE) is $\mathcal{O}(1/N)$, independently of d.

Unfortunately we cannot get i.i.d. samples.

Importance sampling and ratio estimators

Suppose we can get i.i.d. samples $x^i \sim \nu$ where $0 < G(x) := \frac{\gamma(x)}{\nu(x)} < C$.

Then we can use the self-normalized importance sampling estimator

$$\eta(\varphi) \approx \frac{\sum_{i=1}^N G(x^i)\varphi(x^i)}{\sum_{i=1}^N G(x^i)} .$$

The rate will still be $\mathcal{O}(1/N)$, but typically with a constant $\mathcal{O}(e^d)$, depending on $\mathbb{E}(G(x) - \mathbb{E}G(x))^2$.

We may as well use quadrature.

Markov chain Monte Carlo

Suppose we can construct a Markov chain K, that is an operator with the property $K : \mathcal{B}(\mathsf{X}) \rightarrow \mathcal{B}(\mathsf{X})$ and $K^* : \mathcal{P}(\mathsf{X}) \rightarrow \mathcal{P}(\mathsf{X})$, where $\mathcal{B}(\mathsf{X})$ are bounded measurable functions and $\mathcal{P}(\mathsf{X})$ are probability measures, and such that

$$(\eta K)(dx) = \int_{\mathsf{X}} \eta(dx')K(x', dx) = \eta(dx)\,,$$

and for all $A \subset \mathsf{X}$, $x, x' \in \mathsf{X}$,

$$\int_A K(x, dz) \leq \int_A K(x', dz)\,.$$

Markov chain Monte Carlo

Then we can run the Markov chain to collect samples, $x^0 \in \mathsf{X}$ and $x^i \sim K(x^{i-1}, \cdot) = K^i(x^0, \cdot)$ and use these for Monte Carlo

$$\eta(\varphi) \approx \frac{1}{N} \sum_{i=N_b+1}^{N_b+N} \varphi(x^i) \, .$$

Again Monte Carlo provides rate $\mathcal{O}(1/N)$, but now under quite general conditions one may achieve polynomial constant $\mathcal{O}(d)$.

Example: Metropolis-Hastings

Let Q denote a Markov kernel on X.

1. Let $x^0 \in$ X.
2. Sample $x^* \sim Q(x^i, \cdot)$.
3. Set $x^{i+1} = x^*$ with probability:

$$\min\left\{1, \frac{\gamma(x^*)Q(x^*, x^i)}{\gamma(x^i)Q(x^i, x^*)}\right\} ,$$

   otherwise $x^{i+1} = x^i$.
4. Set $i = i + 1$ and return to the start of (2).

Outline

Parameter inference

Estimate the posterior expectation of a function φ of the joint path $X_{1:T}$ and *parameters* θ, of an intractable S(P)DE

$$dX = f_\theta(X)dt + \sigma_\theta(X)dW, \quad X_0 \sim \mu_\theta \,,$$

given noisy partial observations

$$Y_n \sim g_\theta(X_n, \cdot)\,, \quad n = 1, \ldots, T\,.$$

Aim: estimate $\mathbb{E}[\varphi(\theta, X_{0:T})|y_{1:T}]$, where $y_{1:T} := \{y_1, \ldots, y_T\}$.

The hidden process $\{X_n\}$ is a Markov chain.

Discretize with resolution h and denote the transition kernel $F_{\theta,h}\left(x_{p-1}, \mathrm{d}x_p\right)$ – this can be *simulated from*, but its density cannot be *evaluated*.

Return to ML (SDE, for simplicity)

The joint measure (suppressing fixed y_p in notation) is

$$\Pi_h(\mathrm{d}\theta, \mathrm{d}x_{0:n}) \propto \Pi(\mathrm{d}\theta)\mu_\theta(\mathrm{d}x_0) \prod_{p=1}^{n} g_\theta(x_p, y_p) F_{\theta,h}(x_{p-1}, \mathrm{d}x_p) ,$$

For $+\infty > h_0 > \cdots > h_L > 0$, we would like to compute

$$\mathbb{E}_{\Pi_{h_L}}[\varphi(\theta, X_{0:n})] = \sum_{l=0}^{L} \left\{ \mathbb{E}_{\Pi_{h_l}}[\varphi(\theta, X_{0:n})] - \mathbb{E}_{\Pi_{h_{l-1}}}[\varphi(\theta, X_{0:n})] \right\}$$

where $\mathbb{E}_{\Pi_{h_{-1}}}[\cdot] := 0$.

Outline

Approximate coupling

Consider a single pair $\mathbb{E}_{\Pi_h}[\varphi(\theta, X_{0:n})] - \mathbb{E}_{\Pi_{h'}}[\varphi(\theta, X_{0:n})]$, $h < h'$.

Let $z = (x, x')$ and let $Q_{\theta,h,h'}(z, \mathrm{d}\bar{z})$ be a coupling of $(F_{\theta,h}(x, \mathrm{d}\bar{x}), F_{\theta,h'}(x', \mathrm{d}\bar{x}'))$.

Let $G_{p,\theta}(z) = \max\{g_\theta(x, y_p), g_\theta(x', y_p)\}$.

We will sample from the joint coarse/fine filter

$$\Pi_{h,h'}(\mathrm{d}\theta, \mathrm{d}z_{0:n}) \propto \Pi(\mathrm{d}\theta)\nu_\theta(\mathrm{d}z_0) \prod_{p=1}^{n} G_{p,\theta}(z_p) Q_{\theta,h,h'}(z_{p-1}, \mathrm{d}z_p),$$

where ν_θ is the initial coupling

$$\nu_\theta(\mathrm{d}(x, x')) = \mu_\theta(\mathrm{d}x)\delta_x(\mathrm{d}x').$$

Change of measure

We have

$$\mathbb{E}_{\Pi_h}[\varphi(\theta, X_{0:n})] - \mathbb{E}_{\Pi_{h'}}[\varphi(\theta, X'_{0:n})] =$$

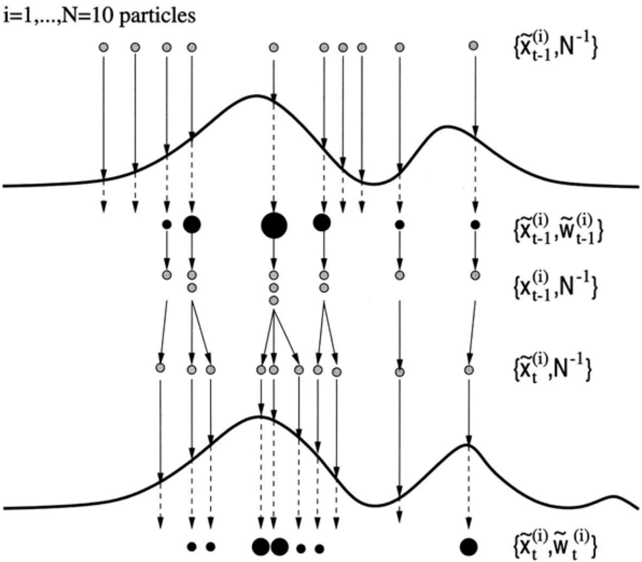
$$\frac{\mathbb{E}_{\Pi_{h,h'}}[\varphi(\theta, X_{0:n})H_{1,\theta}(\theta, Z_{0:n})]}{\mathbb{E}_{\Pi_{h,h'}}[H_{1,\theta}(\theta, Z_{0:n})]} - \frac{\mathbb{E}_{\Pi_{h,h'}}[\varphi(\theta, X'_{0:n})H_{2,\theta}(\theta, Z_{0:n})]}{\mathbb{E}_{\Pi_{h,h'}}[H_{2,\theta}(\theta, Z_{0:n})]}$$

where

$$\begin{aligned} H_{1,\theta}(\theta, z_{0:n}) &= \prod_{p=1}^{n} \frac{g_\theta(x_p, y_p)}{G_{p,\theta}(z_p)} \\ H_{2,\theta}(\theta, z_{0:n}) &= \prod_{p=1}^{n} \frac{g_\theta(x'_p, y_p)}{G_{p,\theta}(z_p)} . \end{aligned}$$

Outline

Sequential Importance Resampling [DDG01]



Particle filter, for fixed θ

Let $M \geq 1$ and θ be fixed, and introduce $a_{0:n-1}^{1:M} \in \{1, \ldots, M\}^{Mn}$. The bootstrap particle filter [Del04] approximates

$$\Pi_{h,h'}(\mathrm{d}z_{0:n}|\theta) \propto \nu_\theta(\mathrm{d}z_0) \prod_{p=1}^{n} G_{p,\theta}(z_p) Q_{\theta,h,h'}(z_{p-1}, \mathrm{d}z_p)$$

by sampling from

$$P(a_{0:n-1}^{1:M}, \mathrm{d}z_{0:n}^{1:M}|\theta) = \left(\prod_{i=1}^{M} \nu_\theta(\mathrm{d}z_0^i)\right) \prod_{p=1}^{n} \prod_{i=1}^{M} \left(\frac{G_{p-1,\theta}(z_{p-1}^{a_{p-1}^i})}{\sum_{j=1}^{M} G_{p-1,\theta}(z_{p-1}^j)} Q_{\theta,h,h'}(z_{p-1}^{a_{p-1}^i}, \mathrm{d}z_p^i) \right),$$

where $G_{0,\theta} := 1$,

i.e. $z_{p-1}^{a_{p-1}^i}$ is resampled with the probability $\frac{G_{p-1,\theta}(z_{p-1}^{a_{p-1}^i})}{\sum_{j=1}^{M} G_{p-1,\theta}(z_{p-1}^j)}$.

Particle marginal MH (PMMH) [ADH10]

Draw J with probability proportional to $G_{n,\theta}(z_n^j)$ for $j = 1, \ldots, M$. Let $\widehat{z}_n = z_n^J$, and trace its ancestral lineage

$$\widehat{z}_{n-1} = z_{n-1}^{a_{n-1}^J}, \quad \widehat{z}_{n-2} = z_{n-2}^{a_{n-2}^{a_{n-1}^J}}, \quad \text{and so on} .$$

Define $p^M_{h,h'}(y_{0:n}|\theta) = \prod_{p=1}^n \left(\frac{1}{M} \sum_{j=1}^M G_{p,\theta}(z_p^j) \right)$. Denote $U = (a^{1:M}_{0:n-1}, z^{1:M}_{0:n}, \theta)$.

Run a MH chain on *the extended state space* $\{U^i\}$ targeting $\propto P(a^{1:M}_{0:n-1}, z^{1:M}_{0:n}|\theta) p^M_{h,h'}(y_{0:n}|\theta)\Pi(d\theta)$. Draw $J \sim P(J|U^i)$ and construct $\widehat{z}^i_{0:n}$ as above. The target has the property [ADH10]

$$P(U,J) = P((U,J)\backslash(\widehat{z}_{0:n},\theta)|\widehat{z}_{0:n},\theta)\Pi_{h,h'}(\widehat{z}_{0:n},\theta) .$$

Particle marginal MH (PMMH) [ADH10]

1. Sample $\theta^0 \sim \pi(\mathrm{d}\theta)$ and $(a^{1:M}_{0:n-1}, z^{1:M}_{0:n})$ from particle filter $P(a^{1:M}_{0:n-1}, \mathrm{d}z^{1:M}_{0:n}|\theta^0)$, and store $p^M_{h,h'}(y_{0:n}|\theta^0)$.
2. Select a path $\widehat{z}^0_{0:n}$: draw z^j_n with probability proportional to $G_{n,\theta^0}(z^j_n)$, let $\widehat{z}^0_n = z^j_n$, and trace back its ancestral lineage
$$\widehat{z}^0_{n-1} = z^{a^j_{n-1}}_{n-1}, \quad \widehat{z}^0_{n-2} = z^{a^{a^j_{n-1}}_{n-2}}_{n-2}, \quad \text{and so on}; \quad \text{Set } i = 1\,.$$
3. Sample $\theta^*|\theta^{i-1}$ according to $R(\mathrm{d}\theta^*|\theta^{i-1}) = r(\theta^*|\theta^{i-1})\mathrm{d}\theta^*$, then sample from particle filter $P(a^{1:M}_{0:n-1}, \mathrm{d}z^{1:M}_{0:n}|\theta^*)$. Select one path $\widehat{z}^*_{0:n}$ as above.
4. Set $\theta^i = \theta^*$, $\widehat{z}^i_{0:n} = \widehat{z}^*_{0:n}$ with probability:
$$\min\left\{1, \frac{p^M_{h,h'}(y_{0:n}|\theta^*)}{p^M_{h,h'}(y_{0:n}|\theta^{i-1})} \frac{\pi(\theta^*)r(\theta^{i-1}|\theta^*)}{\pi(\theta^{i-1})r(\theta^*|\theta^{i-1})}\right\}$$
otherwise $\theta^i = \theta^{i-1}$, $\widehat{z}^i_{0:n} = \widehat{z}^{i-1}_{0:n}$.
5. Set $i = i + 1$ and return to the start of 3.

Outline

PMMH increment estimator

$$\frac{\frac{1}{N}\sum_{i=1}^{N}\varphi(\theta^i,\widehat{x}^i_{0:n})H_{1,\theta}(\theta^i,\widehat{z}^i_{0:n})}{\frac{1}{N}\sum_{i=1}^{N}H_{1,\theta}(\theta^i,\widehat{z}^i_{0:n})} - \frac{\frac{1}{N}\sum_{i=1}^{N}\varphi(\theta^i,\widehat{x}'^i_{0:n})H_{2,\theta}(\theta^i,\widehat{z}^i_{0:n})}{\frac{1}{N}\sum_{i=1}^{N}H_{2,\theta}(\theta^i,\widehat{z}^i_{0:n})}.$$

$$\xrightarrow[N\to\infty]{}$$

$$\frac{\mathbb{E}_{\Pi_{h,h'}}[\varphi(\theta,X_{0:n})H_{1,\theta}(\theta,Z_{0:n})]}{\mathbb{E}_{\Pi_{h,h'}}[H_{1,\theta}(\theta,Z_{0:n})]} - \frac{\mathbb{E}_{\Pi_{h,h'}}[\varphi(\theta,X'_{0:n})H_{2,\theta}(\theta,Z_{0:n})]}{\mathbb{E}_{\Pi_{h,h'}}[H_{2,\theta}(\theta,Z_{0:n})]}$$

$$=$$

$$\mathbb{E}_{\Pi_h}[\varphi(\theta,X_{0:n})] - \mathbb{E}_{\Pi_{h'}}[\varphi(\theta,X'_{0:n})]$$

Multilevel estimator

$$E_l^{N_l}(\varphi) = \frac{\frac{1}{N_l}\sum_{i=1}^{N_l}\varphi(\theta^i, , \widehat{x}_{0:n}^i)H_{1,\theta}(\theta^i, \widehat{z}_{0:n}^i)}{\frac{1}{N_l}\sum_{i=1}^{N_l}H_{1,\theta}(\theta^i, \widehat{z}_{0:n}^i)} - \frac{\frac{1}{N_l}\sum_{i=1}^{N_l}\varphi(\theta^i, \widehat{x}_{0:n}^{\prime i})H_{2,\theta}(\theta^i, \widehat{z}_{0:n}^i)}{\frac{1}{N_l}\sum_{i=1}^{N_l}H_{2,\theta}(\theta^i, \widehat{z}_{0:n}^i)}$$

is a consistent estimator of ($E_0 := \mathbb{E}_{\Pi_{h_0}}[\varphi(\theta, X_{0:n})]$)

$$E_l(\varphi) := \mathbb{E}_{\Pi_{h_l}}[\varphi(\theta, X_{0:n})] - \mathbb{E}_{\Pi_{h_{l-1}}}[\varphi(\theta, X_{0:n})] .$$

$\Rightarrow \sum_{l=0}^{L} E_l^{N_l}(\varphi)$ is a consistent estimator of $\mathbb{E}_{\Pi_{h_L}}[\varphi(\theta, X_{0:n})]$.

Multilevel estimator error analysis

Consider

$$\sum_{l=0}^{L} \bar{E}_l^{N_l}(\varphi) , \qquad \bar{E}_l^{N_l}(\varphi) = E_l^{N_l}(\varphi) - E_l(\varphi) .$$

One must bound

$$\mathbb{E}[(\sum_{l=0}^{L} \bar{E}_l^{N_l}(\varphi))^2] = \sum_{l=0}^{L} \left(\mathbb{E}[\bar{E}_l^{N_l}(\varphi)^2] + \mathbb{E}[\bar{E}_l^{N_l}(\varphi)] \sum_{q \neq l = 0}^{L} \mathbb{E}[\bar{E}_q^{N_q}(\varphi)] \right) .$$

Assumptions

(A1) $\forall\, y \in \mathsf{T}$, $\exists\, C > 0$ such that $\forall\, x \in \mathsf{S}$, $\theta \in \Theta$,

$$C \leq g_\theta(x, y) \leq C^{-1}.$$

And $\forall\, y \in \mathsf{T}$, $g_\theta(x, y)$ is globally Lipschitz on $\mathsf{S} \times \Theta$.

(A2) $\forall\, 0 \leq k \leq n$, $\exists\, \beta > 0$ such that $\forall$ $\varphi \in \mathcal{B}_b(\Theta \times \mathsf{S}^{k+1}) \cap \mathrm{Lip}(\Theta \times \mathsf{S}^{k+1})$ $\exists\, C > 0$

$$\int_{\Theta \times \mathsf{S}^{2k+2}} |\varphi(\theta, x_{0:k}) - \varphi(\theta, x'_{0:k})|^2 \Pi(\mathrm{d}\theta)\nu_\theta(\mathrm{d}z_0) \prod_{p=1}^{k} Q_{\theta,h,h'}(z_{p-1}, \mathrm{d}z_p) \leq C(h')^\beta.$$

(A3) $\forall\, n > 0$, $\exists\, \xi \in (0, 1)$ and $\nu \in \mathcal{P}(\mathsf{W})$ s.t. $\forall\, w \in \mathsf{W}$, $\varphi \in \mathcal{B}_b(\mathsf{W}) \cap \mathrm{Lip}(\mathsf{W})$, h, h', the PMMH kernel K satisfies:

$$\int_{\mathsf{W}} \varphi(w')K(w, dw') \geq \xi \int_{\mathsf{W}} \varphi(w')\nu(dw').$$

K is η-reversible, where η is the joint on the extended space.

Main result

Theorem (JKLZ18)

Assume (A1-3). Then $\forall\ n > 0$, $\exists\ \beta > 0$ *such that* $\forall$ $\varphi \in \mathcal{B}_b(\Theta \times \mathsf{S}^{n+1}) \cap \mathrm{Lip}(\Theta \times \mathsf{S}^{n+1})$ $\exists\ C > 0$ *such that*

$$\mathbb{E}[\bar{E}_l^{N_l}(\varphi)^2] \leq \frac{Ch_l^{\beta}}{N_l}, \qquad \mathbb{E}[\bar{E}_l^{N_l}(\varphi)] \leq \frac{Ch_l^{\beta/2}}{N_l},$$

and β *is from (A2).*

(A4) $\exists\, \gamma, \alpha, C > 0$, such that the cost to simulate $E_l^{N_l}$ is controlled by $\mathrm{C}(E_l^{N_l}) \leq C N_l h_l^{-\gamma}$, and the bias is controlled by

$$|\mathbb{E}_{\Pi_{h_L}}(\varphi(\theta, X_{0:n})) - \mathbb{E}_{\Pi_0}(\varphi(\theta, X_{0:n}))| \leq C h_L^{\alpha} \ .$$

Corollary

Assume (A1-4). $\forall\, n > 0$ *and* $\varphi \in \mathcal{B}_b(\Theta \times \mathsf{S}^{n+1}) \cap \mathrm{Lip}(\Theta \times \mathsf{S}^{n+1})$ $\exists$ $C > 0$ *such that* $\forall\, \epsilon > 0$ *one can choose* $(L, \{N_l\}_{l=0}^{L})$ *so*

$$\mathbb{E}\left[|\sum_{l=0}^{L} E_l^{N_l}(\varphi) - \mathbb{E}_{\Pi_0}(\varphi(\theta, X_{0:n}))|^2 \right] \leq C\epsilon^2 \ ,$$

with a total cost (per time step)

$$\mathrm{COST} \leq C \begin{cases} \epsilon^{-2}, & \text{if} \quad \beta > \gamma, \\ \epsilon^{-2}|\log(\epsilon)|^2, & \text{if} \quad \beta = \gamma, \\ \epsilon^{-\left(2+\frac{\gamma-\beta}{\alpha}\right)}, & \text{if} \quad \beta < \gamma. \end{cases}$$

Outline

SMC samplers

- Let $U_n = [u_0, \ldots, u_n] \in \mathsf{U}_n$ for $n = 0, 1, 2, \ldots$. Consider target distributions $\eta_n \propto \kappa_n$ on U_n.
- Interlace sequential importance resampling (selection) along the sequence, and mutation by MCMC kernels.
- This enables sampling sequentially with complexity $\mathcal{O}(n^2)$ per sample [1].
  - Initialize i.i.d. $U_0^i \sim \eta_0, i = 1, \ldots, N$, and $u_1^i \sim q(U_0^i, \cdot)$. For $n = 1, \ldots$
  - Resample $\{\widehat{U}_n^i\}_{i=1}^N$ according to the weights $\{w_n^i\}_{i=1}^N$, $w_n^i = G_n^i / \sum_{j=1}^N G_n^j$, $G_0 = 1$, and $G_n^i = \kappa_n(U_n^i)/[\kappa_{n-1}(U_{n-1}^i) q_n(U_{n-1}^i, u_n^i)]$.
  - Draw $U_{n+1}^i \sim M_{n+1}(\widehat{U}_n^i, \ \cdot\ )$, where

$$M_{n+1}(U_n, dU'_{n+1}) = K_n(U_n, dU'_n) \otimes q_{n+1}(U'_n, du'_{n+1})$$

is an MCMC kernel such that $\eta_n K_n = \eta_n$.

[1] In principle, under suitable assumptions [BCJ14]

Sequential Monte Carlo^2 (SMC^2)

- Define $U_n = (\theta, a_{0:n-1}^{1:M}, z_{0:n}^{1:M})$, and recall the PMMH target $P_n(U_n, J)$, which has the joint posterior as its marginal on $(\theta, \hat{z}_{0:n})$, where $\hat{z}_{0:n}$ is the ancestral path of the particle with index $J \sim \{G_{n,\theta}(z_n^j)/\sum_{i=1}^M G_{n,\theta}(z_n^i)\}$ at time n.
- Consider the PMMH marginal $P_n(U_n)$, which has as its marginal the posterior on θ.
- **SMC^2**: Run an SMC sampler with N samples $\{U_n^k\}_{k=1}^N$ targeting $P_n(U_n)$, using PMMH marginal kernels (without sampling J) as the mutations.
- When we wish to estimate expectations $\mathbb{E}_{\Pi_{h,h'}}(f(\theta, Z_{0:n}))$, we extend to $P_n(U_n, J) = P_n(J|U_n)P_n(U_n)$:
- For each $k = 1, \ldots, N$:
  - Sample $J^k \sim \{G_{n,\theta^k}(z_n^{j,k})/\sum_{i=1}^M G_{n,\theta^k}(z_n^{i,k})\}$.
  - Construct the ancestral lineage $\hat{z}_{0:n}^k$ (as before).
  - Estimate $\frac{1}{N}\sum_{k=1}^N f(\theta^k, \hat{z}_{0:n}^k)$.

Outline

SMLMC^2

- For each $l = 0, \ldots, L$:
- Run SMC^2 on $P_{l,n}(U_{l,n})$, then at a given time n (additional l index suppressed):
- For each $k = 1, \ldots, N_l$:
  - Sample $J^k \sim \{G_{n,\theta^k}(z_n^{j,k}) / \sum_{i=1}^M G_{n,\theta^k}(z_n^{i,k})\}$.
  - Construct the ancestral lineage $\widehat{z}_{0:n}^k$ (as before).
  - Estimate

$$E_l^{N_l}(\varphi) = \frac{\frac{1}{N_l}\sum_{i=1}^{N_l}\varphi(\theta^i, , \widehat{x}_{0:n}^i)H_{1,\theta}(\theta^i, \widehat{z}_{0:n}^i)}{\frac{1}{N_l}\sum_{i=1}^{N_l}H_{1,\theta}(\theta^i, \widehat{z}_{0:n}^i)} - \frac{\frac{1}{N_l}\sum_{i=1}^{N_l}\varphi(\theta^i, \widehat{x}_{0:n}'^i)H_{2,\theta}(\theta^i, \widehat{z}_{0:n}^i)}{\frac{1}{N_l}\sum_{i=1}^{N_l}H_{2,\theta}(\theta^i, \widehat{z}_{0:n}^i)}$$

- Construct the resulting MLMC estimator $\sum_{l=0}^L E_l^{N_l}(\varphi)$.
- We obtained a theorem (for MIMC) under very strong assumptions.
- Numerical results indicate the method works under weaker assumptions.

Outline

Ornstein-Uhlenbeck process

$$\begin{aligned} \mathrm{d}X_t &= \theta(\mu - X_t)\mathrm{d}t + \sigma \mathrm{d}W_t\,, \qquad X_0 = x_0\,, \\ Y_k | X_{\delta k} &\sim \mathcal{N}(X_{\delta k}, \tau^2)\,, \\ \theta &\sim \mathcal{G}(1,1)\,, \\ \sigma &\sim \mathcal{G}(1, 0.5)\,. \end{aligned}$$

- $\mathcal{N}(m, \tau^2)$ denotes the Normal with mean m and variance τ^2.
- $\mathcal{G}(a, b)$ denotes the Gamma with shape a and scale b.
- $x_0 = 0$, $\mu = 0$, $\delta = 0.5$, and $\tau^2 = 0.2$.
- 100 observations simulated with $\theta = 1$ and $\sigma = 0.5$.

Langevin SDE

$$
\begin{aligned}
\mathrm{d}X_t &= \frac{1}{2}\nabla \log \pi(X_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \qquad X_0 = x_0 \\
Y_k|X_k &\sim \mathcal{N}(0, \tau^2 \exp X_k)\,, \\
\theta &\sim \mathcal{G}(1,1)\,, \\
\sigma &\sim \mathcal{G}(1,0.5)\,.
\end{aligned}
$$

- $\pi(x)$ denotes the probability density function of a Student's t-distribution with θ degrees of freedom.
- $x_0 = 0$.
- 1,000 observations simulated with $\theta = 10$, $\sigma = 1$, and $\tau^2 = 1$.

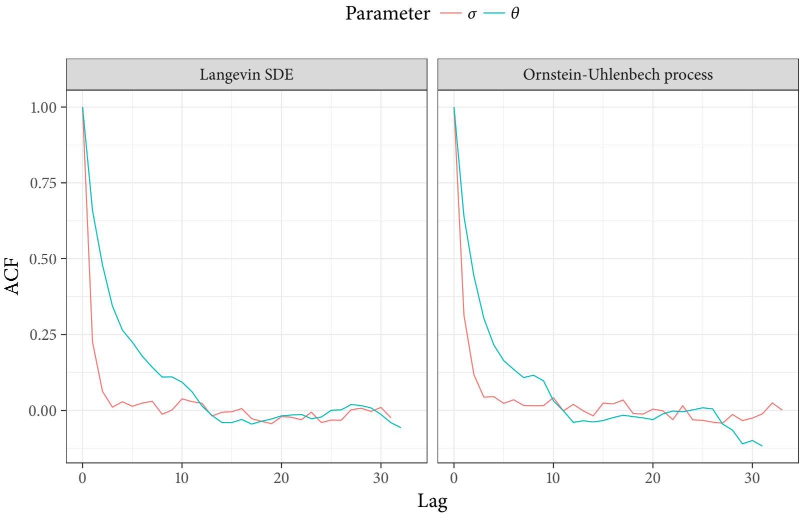


Figure: Autocorrelation of a typical PMCMC chain.

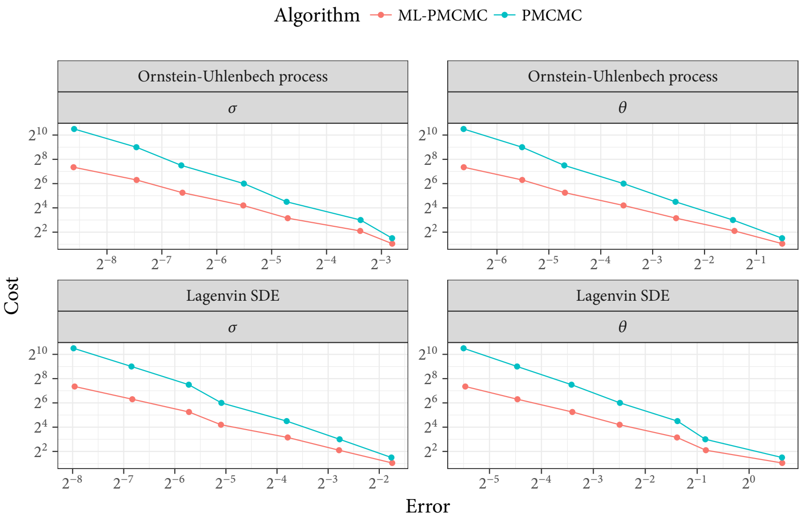


Figure: Cost vs. MSE for the 2 parameters for each of the 2 SDEs.

Model	Parameter	ML-PMCMC	PMCMC
OU	θ	-1.022	-1.463
	σ	-1.065	-1.522
Langevin	θ	-1.060	-1.508
	σ	-1.023	-1.481

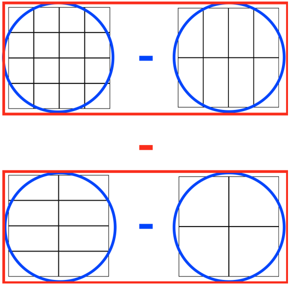
Table: Estimated rates of convergence of MSE with respect to cost for various parameters, fitted to the curves.

Outline

Multi-index Monte Carlo (MIMC) idea

- If spatio-temporal approximation dimension $d > 1$, then MIMC is preferable to MLMC [HNT15]. $\alpha \in \mathbb{N}^d$
- $\Delta_i \mathbb{E}_\alpha(\varphi(u)) = \mathbb{E}_\alpha(\varphi_\alpha(u)) - \mathbb{E}_{\alpha - e_i}(\varphi_{\alpha - e_i}(u))$, $\Delta = \Delta_d \cdots \Delta_1$,

$$\mathbb{E}(\varphi(u)) = \sum_{\alpha \in \mathbb{N}^d} \Delta \mathbb{E}_\alpha(\varphi_\alpha(u)) \approx \sum_{\alpha \in \mathcal{I}} \Delta \mathbb{E}_\alpha(\varphi_\alpha(u))$$



MIMC estimator

- Construct an empirical estimator of $\sum_{\alpha \in \mathcal{I}} \Delta \mathbb{E}_\alpha(\varphi_\alpha(u))$.
- **Key**: for each $\alpha \in \mathcal{I}$, obtain N_α i.i.d. samples from a suitably coupled target $(U_{\alpha(k_\alpha)}, U_{\alpha(k_\alpha - 1)}, \ldots, U_{\alpha(1)})^{(i)} \sim \bar{\eta}^\alpha$ such that $\int \bar{\eta}^\alpha du_{\alpha(1)} \cdots du_{\alpha(l-1)} du_{\alpha(l+1)} \cdots du_{\alpha(k_\alpha)} = \eta_{\alpha(l)}$ for $l = 0, \ldots, L$, where $k_\alpha \leq 2^d$.
- Similar to MLMC, one finds optimal $N_\alpha \propto \sqrt{V_\alpha / C_\alpha}$.
- **Optimal index set**[2] $\mathcal{I}$ consists of superlevel sets of $P_\alpha = B_\alpha / \sqrt{V_\alpha C_\alpha}$, where B_α is bias associated to increment α.
- Canonical complexity $\mathcal{O}(\varepsilon^{-2})$ can be obtained independent of dimension d for optimal index sets.
- For tensor product index sets an additional (dimension-dependent) constraint on the rates is required or else the cost is dominated by the single highest resolution sample.

[2] In the sense that any set with smaller work has larger bias.

Outline

Summary

- New approximate coupling strategy can be used to apply MLMC to PMCMC for static parameter estimation [JKLZ18.i].
- Same strategy can be employed for multi-index MCMC [JKLZ18.ii].
- Recently extended to MISMC^2 [JLX19.s].
- Exciting prospects of looking at
  - MIMC versions with appropriate index sets;
  - Other versions of MIMC $\cap$ SMC;
  - Links/comparisons with other methods, in cases where they are applicable
  - other practical enhancements.
- PhD and postdocs wanted – please inquire !
- **AIMS Foundations of Data Science now accepting papers !**

References

- **[G08]**: Giles. "Multilevel Monte Carlo path simulation." Op. Res., 56, 607-617 (2008).
- **[H00]** Heinrich. "Multilevel Monte Carlo methods." LSSC proceedings (2001).
- **[CGST11]** Cliffe, Giles, Scheichl, & Teckentrup. "MLMC and applications to elliptic PDEs." Computing and Visualization in Science, 14(1), 3 (2011).
- **[D04]**: Del Moral. "Feynman-Kac Formulae." Springer: New York (2004).
- **[B02]** Beaumont. "Estimation of population growth..." Genetics 164(3) 1139-1160 (2003).
- **[AR08]** Andrieu & Roberts. "The pseudo-marginal approach." Annals of Stat. 37(2) 697-725 (2009).
- **[ADH10]** Andrieu, Doucet, and Holenstein. "Particle MCMC methods." JRSSB 72(3) 269-342 (2010).

References

- **[JKLZ18.i]**: Jasra, Kamatani, Law, Zhou. "MLMC for static Bayesian parameter estimation." SISC 40, A887-A902 (2018).
- **[JKLZ18.ii]**: Jasra, Kamatani, Law, Zhou. "A Multi-Index Markov Chain Monte Carlo Method." IJUQ 8(1), 61-73 (2018).
- **[JLX19.s]**: Jasra, Law, Xu. "Multi-index SMC^2." Submitted.
- **[BJLTZ15]**: Beskos, Jasra, Law, Tempone, Zhou. "Multilevel Sequential Monte Carlo samplers." SPA 127:5, 1417–1440 (2017).
- **[HSS13]**: Hoang, Schwab, Stuart. Inverse Prob., 29, 085010 (2013).
- **[DKST13]**: Dodwell, Ketelsen, Scheichl, Teckentrup. "A hierarchical MLMCMC algorithm." SIAM/ASA JUQ 3(1) 1075-1108 (2015).

References

- **[CHLNT17]** Chernov, Hoel, Law, Nobile, Tempone. "Multilevel ensemble Kalman filtering for spatio-temporal processes." arXiv:1608.08558 (2017).
- **[JLS17]** Jasra, Law, Suciu. "Advanced Multilevel Monte Carlo Methods." arXiv:1704.07272 (2017). To appear in ISR (2019!)
- **[DJLZ16]**: Del Moral, Jasra, Law, Zhou. "Multilevel Sequential Monte Carlo samplers for normalizing constants." ToMACS 27(3), 20 (2017).
- **[JKLZ15]**: Jasra, Kamatani, Law, Zhou. "Multilevel particle filter." SINUM 55(6), 3068–3096 (2017).
- **[DDP18]**: Deligiannidis, Doucet, Pitt. "The Correlated Pseudo-Marginal Method." JRSSB 80 (5), 839–870 (2015).
- **[HLT15]**: Hoel, Law, Tempone. "Multilevel ensemble Kalman filter." SINUM 54(3), 1813–1839 (2016).
- **[GCR]**: Cotter, Gregory, Reich. "Multilevel ensemble transform particle filter."

Thank you

Warmly welcoming submissions!

FoDS

Foundations of Data Science

Editors in Chief

www.aimsciences.org

ISSN 2158-2491 (print)
ISSN 2639-8001 (online)

FoDS invites submissions focusing on advances in mathematical, statistical, and computational methods for data science. Results should significantly advance current understanding of data science, by algorithm development, analysis, and/or computational implementation which demonstrates behavior and applicability of the algorithm. Fields covered by the journal include, but are not limited to Bayesian Statistics, High Performance Computing, Inverse Problems, Data Assimilation, Machine Learning, Optimization, Topological Data Analysis, Spatial Statistics, Nonparametric Statistics, Uncertainty Quantification, and Data Centric Engineering. Expository and review articles are welcome. Papers which focus on applications in science and engineering are also encouraged, however the method(s) used should be applicable outside of one specific application domain.

- AIMS is a member of COPE.
- Publishes 4 issues a year in March, June, September and December.
- Publishes online only.
- Archived in Portico and CLOCKSS.
- FoDS is a publication of the American Institute of Mathematical Sciences.

American Institute of Mathematical Sciences
P.O. Box 2604, Springfield, MO 65801, USA
General@aimSciences.org; Phone/Fax: (417) 351-3204