Multilevel sequential Monte Carlo methods

Kody Law

Oak Ridge National Laboratory

SMC2017, Uppsala, Sweden

September 2, 2017

Outline

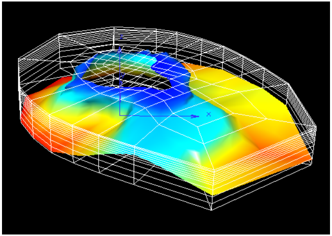
Outline

1. Introduction
2. Multilevel Monte Carlo sampling
3. Bayesian inference problem
4. Sequential Monte Carlo samplers
5. Multilevel Sequential Monte Carlo (MLSMC) samplers
6. Other MLMC algorithms for inference
   - Importance sampling (IS) strategy
   - Coupled algorithm (CA) strategy
   - Approximate coupling (AC) strategy
7. Summary

Inverse Problems

Data

Parameter



$$y = G(u) + e$$

forward model (PDE)

observation/model errors

$y \in \mathbb{R}^M$	Data y may be limited in number, noisy, and indirect.
$u \in E$	Parameter u often a function, discretized.
$G : E \to \mathbb{R}^M$	Needs to be **approximated**.

Oak Ridge National Laboratory

Bayesian inversion

Goal of Bayesian inference: given observed data y, find the posterior distribution

$$\mathbb{P}(du|y) = \frac{\mathbb{P}(y|u)\mathbb{P}_0(du)}{\int_E \mathbb{P}(y|u)\mathbb{P}_0(du)},$$

and estimate *quantities of interest*, such as, for $g : E \to \mathbb{R}$,

- expected value, $\int_E g(u)\mathbb{P}(du|y)$;
- variance, $\int_E g(u)^2\mathbb{P}(du|y) - (\int_E g(u)\mathbb{P}(du|y))^2$;
- probability of exceeding some value $\int_{\{u\in E; g(u)>R\}} \mathbb{P}(du|y)$.

There exists a connection with a classical inverse problems: identify most probable value, $\sup_{u\in E} \lim_{\delta\to 0} \frac{\mathbb{P}(B_\delta(u)|y)}{\mathbb{P}(B_\delta(v)|y)}$.

Orientation

Aim: Approximate expectations with respect to a probability distribution η_∞, which needs to be approximated by some η_L, and can only be evaluated up to a normalizing constant.
Solution: The multilevel Monte Carlo (MLMC) framework is extended to Sequential Monte Carlo (SMC) samplers, yielding the MLSMC sampler for Bayesian inference problems (and several other MC methods).

- MLMC methods *reduce cost to* error$= \mathcal{O}(\varepsilon)$, can be used in the case that η_L can be sampled from directly [G08].
- Here it is assumed that η_L cannot be sampled from directly, but can be evaluated up to a normalizing constant (e.g. Bayesian inference problems).
- SMC samplers are a general class of algorithms which are effective for sampling from such distributions [DDJ06,C02].

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

Single level Monte Carlo

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

Monte Carlo approach

- Discretize the space $\Rightarrow$ *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) $\mathbb{E}\{\widehat{Y}_L^{N_L} - \mathbb{E}_{\eta_\infty}[g(U)]\}^2$ splits into

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

- 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 $\{\eta_l\}_{l=1}^{L}$ and define $Y_l = \{\mathbb{E}_{\eta_l}[g(U)] - \mathbb{E}_{\eta_{l-1}}[g(U)]\}$, with $\eta_{-1} := 0$.
Observe the telescopic sum

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

Each term can be unbiasedly approximated by

$$Y_l^{N_l} = \frac{1}{N_l}\sum_{i=1}^{N_l}\{g(U_l^{(i)}) - g(U_{l-1}^{(i)})\}$$

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

Multilevel Monte Carlo II

Multilevel Monte Carlo approach:

- Sample i.i.d. $(U_l, U_{l-1})^{(i)} \sim \bar{\eta}^l$, such that $\int \bar{\eta}^l du_{l-1,l} = \eta_{l,l-1}$, and approximate
$$\eta_L(g) \approx \widehat{Y}_{L,\text{Multi}} := \sum_{l=0}^{L} Y_l^{N_l} .$$
- Mean square error (MSE) given by
$$\mathbb{E}\{\widehat{Y}_{L,\text{Multi}} - \mathbb{E}_{\eta_\infty}[g(U)]\}^2 =$$
$$\underbrace{\mathbb{E}\{\widehat{Y}_{L,\text{Multi}} - \mathbb{E}_{\eta_L}[g(U)]\}^2}_{\text{variance}=\sum_{l=0}^{L} V_l/N_l} + \underbrace{\{\mathbb{E}_{\eta_L}[g(U)] - \mathbb{E}_{\eta_\infty}[g(U)]\}^2}_{\text{bias}} .$$
- Fix bias by choosing L. Minimize cost $C = \sum_{l=0}^{L} C_l N_l$ as a function of $\{N_l\}_{l=0}^{L}$ for fixed variance $\Rightarrow N_l \propto \sqrt{V_l/C_l}$.

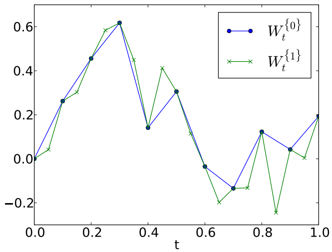
OAK RIDGE National Laboratory

Illustration of pairwise coupling

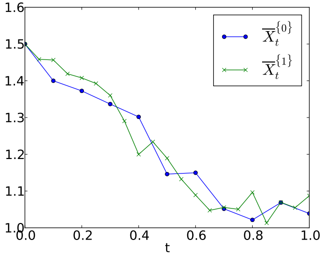
Pairwise coupling of trajectories of an SDE:

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

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



(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.

Multilevel vs. Single level

Assume $h_l = 2^{-l}$ and there are α, and $\beta > \zeta$ such that

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

(ii) strong error $\mathbb{E}|g(U_l) - g(U)|^2 = \mathcal{O}(h_l^{\beta}) \Rightarrow V_l = \mathcal{O}(h_l^{\beta})$,

(iii) computational cost for a realization of $g(U_l) - g(U_{l-1})$, $C_l \propto h_l^{-\zeta}$.

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 $C_{\mathrm{ML}} = \mathcal{O}(\varepsilon^{-2})$: $N_l \propto \varepsilon^{-2} K_L h_l^{(\beta+\zeta)/2}$, so $V \propto \varepsilon^2$ and $C \propto \varepsilon^{-2} K_L^2$ for $K_L = \sum_{l=0}^{L} h_l^{(\beta-\zeta)/2} = \mathcal{O}(1)$ [G08] – cost of simulating a scalar random variable.
- Example: Milstein solution of SDE

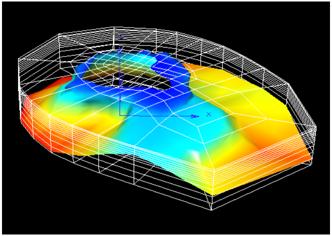
$$C = \mathcal{O}(\varepsilon^{-3}) \quad \text{vs.} \quad C_{\mathrm{ML}} = \mathcal{O}(\varepsilon^{-2}).$$

Outline

Inverse Problems

Data

Parameter



$$y = G(u) + e$$

forward model (PDE)

observation/model errors

$y \in \mathbb{R}^M$	Data y may be limited in number, noisy, and indirect.
$u \in E$	Parameter u often a function, discretized.
$G : E \to \mathbb{R}^M$	Needs to be **approximated**.

Example forward problem

Let $V := H^1(\Omega) \subset L^2(\Omega) \subset H^{-1}(\Omega) =: V^*$, $\Omega \subset \mathbb{R}^d$ with $\partial\Omega$ convex, and $f \in V^*$. Consider

$$\begin{aligned} -\nabla \cdot (\widehat{u}\nabla p) &= f, \quad \text{on } \Omega \\ p &= 0, \quad \text{on } \partial\Omega, \end{aligned}$$

where

$$\widehat{u}(x) = \bar{u}(x) + \sum_{k=1}^{K} u_k \sigma_k \Phi_k(x) \ .$$

Define $u = \{u_k\}_{k=1}^K \in E := \prod_{k=1}^K [-1, 1]$, with $u_k \sim U[-1, 1]$ i.i.d. This determines the prior distribution for u. Assume $\bar{u}, \Phi_k \in C^\infty$, and $\|\Phi_k\|_\infty = 1$ for all k, and require

$$\inf_x \widehat{u}(x) \geq \inf_x \bar{u}(x) - \sum_{k=1}^{K} \sigma_k \geq u_* > 0.$$

Bayesian inverse problem

Let $p(\cdot; u)$ denote the weak solution for parameter value u, and define

$$\mathcal{G}(p) = [g_1(p), \cdots, g_M(p)]^\top ,$$

where $g_m \in V^*$ for $m = 1, \ldots, M$.

DATA : $$y = \mathcal{G}(p(\cdot; u)) + e, \quad e \sim N(0, \Gamma), \quad e \perp u.$$

The *unnormalized* density of $u|y$ over $u \in E$ is given by

$$\kappa(u) = e^{-\Phi[\mathcal{G}(p(\cdot;u))]} ; \quad \Phi(\mathcal{G}) = \frac{1}{2} |\mathcal{G} - y|_\Gamma^2 .$$

TARGET : $$\eta(u) = \frac{\kappa(u)}{Z}, \quad Z = \int_E \kappa(u) du.$$

Outline

SMC sampler algorithm

Distributions η_l dictated by an accuracy parameter h_l (here FEM mesh diameter) $\infty > h_0 > h_1 \cdots > h_\infty = 0$. Approximate $\mathbb{E}_{\eta_L}[g(U)] = \eta_L(g) = \int_E g(u)\eta_L(u)du$.

Idea: interlace sequential importance resampling (selection) along the hierarchy, and mutation by MCMC kernels.

- Initialize i.i.d. $U_0^i \sim \eta_0, i = 1, \ldots, N$. For $l \in \{0, \ldots, L-1\}$:
- Resample $\{\widehat{U}_l^i\}_{i=1}^N$ according to the weights $\{w_l^i\}_{i=1}^N$, $w_l^i = G_l^i / \sum_{j=1}^N G_l^j$, $G_l^i = (\kappa_{l+1}/\kappa_l)(U_l^i)$.
- Draw $U_{l+1}^i \sim M_{l+1}(\widehat{U}_l^i, \cdot)$, where M_{l+1} is an MCMC kernel such that $\eta_{l+1} M_{l+1} = \eta_{l+1}$.

For $g : E \to \mathbb{R}$, $l \in \{0, \ldots, L\}$, we have the following estimators

$$\mathbb{E}_{\eta_l}[g(U)] \approx \eta_l^N(g) := \frac{1}{N}\sum_{i=1}^N g(U_l^i) \ .$$

Outline

MLSMC sampler

Notice

$$\mathbb{E}_{\eta_L}[g(U)] = \mathbb{E}_{\eta_0}[g(U)] + \sum_{l=1}^{L} \left\{ \mathbb{E}_{\eta_l}[g(U)] - \mathbb{E}_{\eta_{l-1}}[g(U)] \right\}$$

$$= \mathbb{E}_{\eta_0}[g(U)] + \sum_{l=1}^{L} \mathbb{E}_{\eta_{l-1}}\left[\left(\frac{\kappa_l(U) Z_{l-1}}{\kappa_{l-1}(U) Z_l} - 1 \right) g(U) \right]. \quad \dagger$$

Idea: Approximate $\dagger$ using SMC sample hierarchy.

Key: Subsample $(U_0^{1:N_0}, \ldots, U_{L-1}^{1:N_{L-1}})$ as in single level SMC, but with $+\infty > N_0 \geq N_1 \cdots \geq N_{L-1} \geq 1$ appropriately chosen.

MLSMC estimator

The MLSMC consistent estimator of $\eta_L(g)$ is given by

$$\widehat{Y} := \eta_0^{N_0}(g) + \sum_{l=1}^{L} \left\{ \frac{\eta_{l-1}^{N_{l-1}}(gG_{l-1})}{\eta_{l-1}^{N_{l-1}}(G_{l-1})} - \eta_{l-1}^{N_{l-1}}(g) \right\}.$$

i) the $L+1$ terms above are *not* unbiased estimates of $\mathbb{E}_{\eta_l}[g(U)] - \mathbb{E}_{\eta_{l-1}}[g(U)]$, so decompose MSE as:

$$\mathbb{E}\left[\{\widehat{Y} - \mathbb{E}_{\eta_\infty}[g(U)]\}^2\right] \leq$$
$$2\,\mathbb{E}\left[\{\widehat{Y} - \mathbb{E}_{\eta_L}[g(U)]\}^2\right] + 2\,\{\mathbb{E}_{\eta_L}[g(U)] - \mathbb{E}_{\eta_\infty}[g(U)]\}^2 \;.$$

ii) the same $L+1$ estimates are *not* independent, so a more complex error analysis will be required to characterize $\mathbb{E}[\{\widehat{Y} - \mathbb{E}_{\eta_L}[g(U)]\}^2]$.

Oak Ridge National Laboratory

Assumptions

(A1) There exist $0 < \underline{C} < \overline{C} < +\infty$ such that

$$\begin{aligned} \sup_{1\leq l\leq L} \sup_{u\in E} G_l(u) &\leq \overline{C}\,, \\ \inf_{1\leq l\leq L} \inf_{u\in E} G_l(u) &\geq \underline{C}\,. \end{aligned}$$

(A2) There exist a $\rho \in (0,1)$ such that for any $1 \leq p \leq L-1$, $(u,v) \in E^2$, $A \in \sigma(E)$,

$$\int_A M_p(u, du') \geq \rho \int_A M_p(v, du').$$

(A3) There is a $\beta > 0$ such that

$$V_l := \| \tfrac{Z_{l-1}}{Z_l} G_{l-1} - 1 \|_\infty^2 = \mathcal{O}(h_l^\beta)\,.$$

Main result

Theorem (BJLTZ16 – Stoch. Proc. App.)

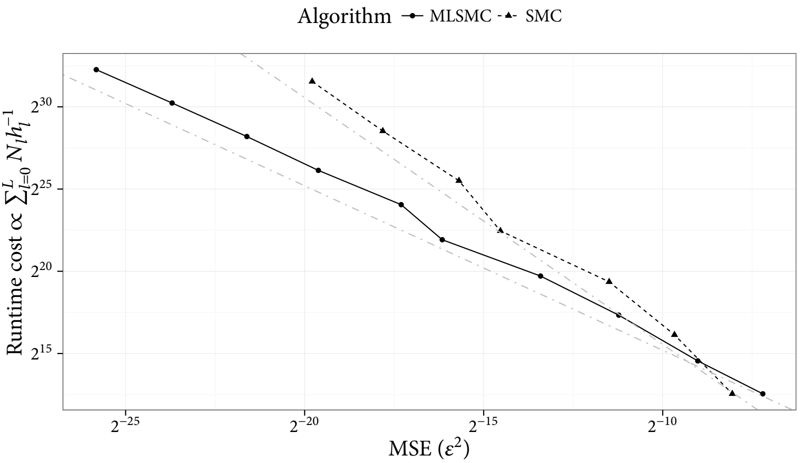
Assume (A1-3). For any $g : E \to \mathbb{R}$ bounded

$$\mathbb{E}[\{\widehat{Y} - \mathbb{E}_{\eta_L}[g(U)]\}^2] = \frac{V}{2}$$

$$\lesssim \frac{1}{N_0} + \sum_{l=1}^{L} \left(\frac{V_l}{N_l} + \left(\frac{V_l}{N_l}\right)^{1/2} \sum_{q=l+1}^{L} \frac{V_q^{1/2}}{N_q} \right) .$$

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

Runtime cost as a function of error

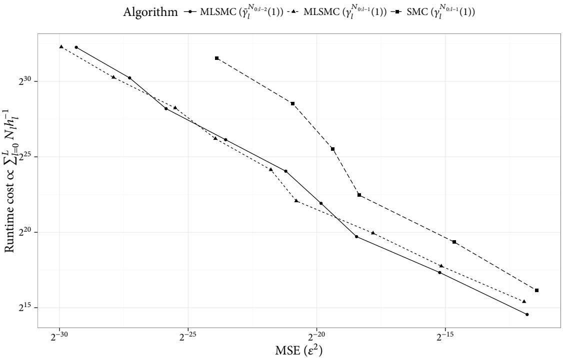


OAK RIDGE National Laboratory

Outline

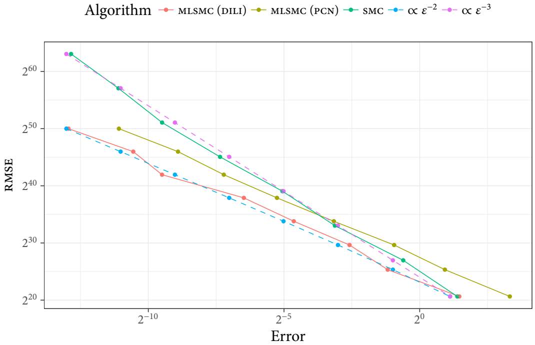
(IS) MLSMC sampler for normalizing constants

- In case $g = 1$, the original estimator does not make sense.
- Two *unbiased* estimators proposed which provide the optimal rate with a logarithmic penalty on the cost: MSE $\mathcal{O}(\varepsilon^2)$ for cost $\mathcal{O}(|\log \varepsilon|\varepsilon^{-2})$ [DJLZ16 – Trans. Mod. Comp. Sim.] Here one can also construct estimators of Rhee&Glynn type.



(IS) MLSMC samplers with DILI mutations

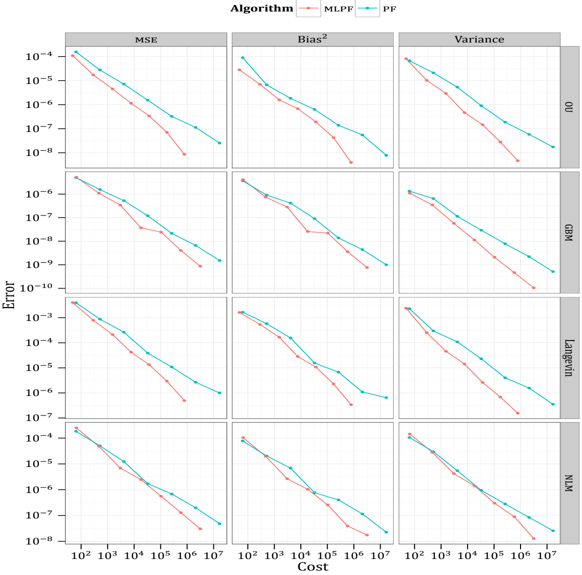
- Posterior over function-space, levels include refinement in parameter and model $\widehat{\eta}_l(u_{0:l})$.
- Covariance-based LIS (cLIS) introduced and incorporated in DILI proposals [CLM16, BJLMZ17] – substantial reduction in cost.



(CA) ML particle filter (MLPF) for SDE

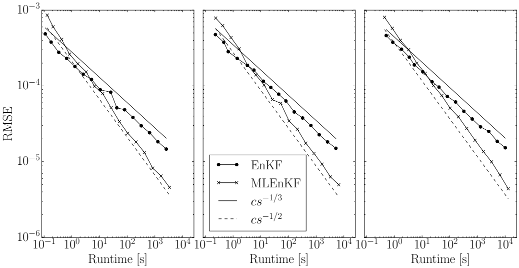
- Filtering involves a sequence of Bayesian inversions, separated by propagation in time (in this case through an SDE).
- Let $\widehat{\eta}_{0,m}, \ldots, \widehat{\eta}_{L,m}, \ldots, \widehat{\eta}_{\infty,m}$ denote the time m filtering distribution at a hierarchy of levels (time discretization).
- Coupled traditional SMC algorithms (particle filters) can be used for each level.
- Mutation M^{ℓ} is now coupled propagation of a pair of initial conditions through an SDE discretized at two successive mesh-refinements, for $\ell = 0, \ldots, L$.
- Selection is performed by novel pairwise coupled resampling which preserves marginals.
- MLMC results carry over with somewhat weaker rate $\beta \to \beta/2$ [JKLZ15].

(CA) MLPF numerical experiments



(CA) ML ensemble Kalman filter (MLEnKF) for SDE

- EnKF uses sample covariance from an ensemble of particles to approximate a linear Gaussian Bayesian update, given by an affine transformation of particles.
- Multilevel approximation of the covariance improves cost for MSE $\mathcal{O}(\varepsilon^2)$ [HLT16].
  - Best theoretical bound (at step n): $\mathcal{O}(|\log \varepsilon|^{2n}\varepsilon^{-2})$.
  - Numerically (uniformly in n): $\mathcal{O}(\varepsilon^{-2})$.

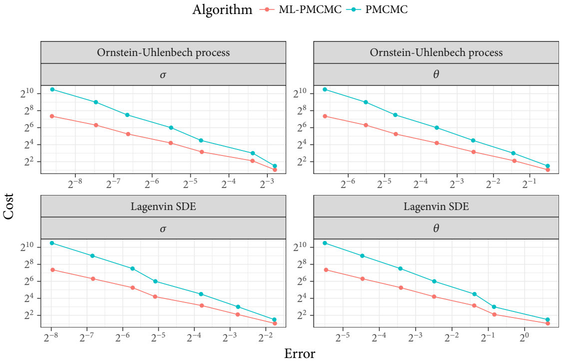


(AC) Parameter estimation for SDE with pMCMC

- Aim: estimate $\mathbb{E}[\varphi(\theta)|y]$, where y is a finite set of partial observations of the SDE X_t^θ on $[0, T]$, parameterized by θ.
- particle MCMC: Iterate
  - propose $\theta' \sim q(\theta, \theta')$,
  - simulate $\{X^{\theta',i}\}_{i=1}^M \approx \pi(X|\theta', y)$ with particle filter,
  - compute non-negative and unbiased estimator $p^M(y|\theta') = \prod_{p=1}^n (\frac{1}{M}\sum_{i=1}^M g_p(X_p^{\theta',i}))$,
  - accept/reject according to

$$1 \wedge \frac{p^M(y|\theta')\pi(\theta')q(\theta',\theta)}{p^M(y|\theta)\pi(\theta)q(\theta,\theta')}.$$

- MLMC version [JKLZ16]:
  - Construct **approximate coupling** $\tilde{\pi}_{l,l-1}(\theta, X^l, X^{l-1})$: usual coupled forward kernel, and coupled selection function $G_{p,\theta}(X^l, X^{l-1}) = \max\{g_{p,\theta}(X^l), g_{p,\theta}(X^{l-1})\}$.
  - Let $H_l(\theta, X^l, X^{l-1}) = \prod_{p=1}^{n} g_{p,\theta}(X^l)/G_{p,\theta}(X^l, X^{l-1})$. Then $\mathbb{E}_{\pi_l}[\varphi(\theta)] - \mathbb{E}_{\pi_{l-1}}[\varphi(\theta)] =$
  $$\frac{\mathbb{E}_{\tilde{\pi}_{l,l-1}}[\varphi(\theta)H_l(\theta,X^l,X^{l-1})]}{\mathbb{E}_{\tilde{\pi}_{l,l-1}}[H_l(\theta,X^l,X^{l-1})]} - \frac{\mathbb{E}_{\tilde{\pi}_{l,l-1}}[\varphi(\theta^{l-1})H_{l-1}(\theta,X^l,X^{l-1})]}{\mathbb{E}_{\tilde{\pi}_{l,l-1}}[H_{l-1}(\theta,X^l,X^{l-1})]}.$$
- Optimal results hold with **same** rate as forward.



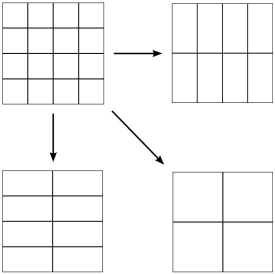
(AC) Multi-index Markov chain Monte Carlo (MIMCMC)

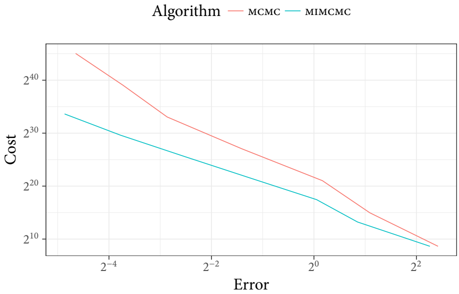
- If spatio-temporal approximation dimension $d > 1$, then MIMC is preferable to MLMC [HNT15]. $\alpha \in \mathbb{N}^d$
- $\Delta_i \mathbb{E}_\alpha(\varphi(u)) = \mathbb{E}_\alpha(\varphi(u)) - \mathbb{E}_{\alpha - e_i}(\varphi(u))$, $\Delta = \Delta_d \cdots \Delta_1$,

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

- **Approximate coupling** can be applied to the 2^d probability measures in each summand.

- Optimal results hold for appropriate regularity [JKLZ17]





Outline

Summary

- MLSMC sampler can perform as well as MLMC.
- For our example $\beta > \zeta$. If $\beta \leq \zeta$, cost is somewhat higher, analogous to standard MLMC.
- If $\zeta > 2\alpha$ then the optimal cost is $\varepsilon^{-\zeta/\alpha}$, the cost of a single simulation at the finest level.
- New **importance sampling**: MLSMC with DILI mutations.
- **Coupled algorithms**: MLPF strong error is effectively reduced by coupled resampling $\beta \to \beta/2$.
- **Coupled algorithms**: MLEnKF has a spurious n-dependent logarithmic penalty $|\log \varepsilon|^{2n}$ on cost.
- New **approximate couplings**: ML PMCMC for SDE parameter estimation preserves strong error β.
- New **approximate couplings**: MIMCMC can perform as well as MIMC; also for $d = 1$ new MLMCMC.
- We are keen to do more applications (on HPC) ! Looking for students/postdocs with similar interests.

References

- **[BJLTZ15]**: Beskos, Jasra, Law, Tempone, Zhou. "Multilevel Sequential Monte Carlo samplers." SPA 127:5, 1417–1440 (2017).
- **[DJLZ16]**: Del Moral, Jasra, Law, Zhou. "Multilevel Sequential Monte Carlo samplers for normalizing constants." To appear in TOMACS (2017).
- **[JKLZ15]**: Jasra, Kamatani, Law, Zhou. "Multilevel particle filter." arxiv:1510.04977 (2015).
- **[JLZ16]**: Jasra, Law, and Zhou. "Forward and Inverse Uncertainty Quantification using Multilevel Monte Carlo Algorithms for an Elliptic Nonlocal Equation." Int. J. Unc. Quant., 6(6), 501–514 (2016)
- **[HLT15]**: Hoel, Law, Tempone. "Multilevel ensemble Kalman filter." SIAM J. Numer. Anal., 54(3), 1813–1839 (2016).

References

- **[G08]**: Giles. "Multilevel Monte Carlo path simulation." Op. Res., 56, 607-617 (2008).
- **[H00]** Heinrich. "Multilevel Monte Carlo methods." LSSC proceedings (2001).
- **[HNT15]** Haji-Ali, Nobile, Tempone. "MIMC: sparsity meets sampling." Numerische Mathematik, 132, 767-806 (2016).
- **[DDJ06]**: Del Moral, Doucet, Jasra. "Sequential Monte Carlo samplers." J. R. Statist. Soc. B, 68, 411-436 (2006).
- **[C02]**: Chopin. "A sequential particle filter method for static models." Biometrika 89:3 539–552 (2002).
- **[D04]**: Del Moral. "Feynman-Kac Formulae." Springer: New York (2004).
- **[CLM16]**: Cui, Law, Marzouk. "DILI MCMC." J. Comp. Phys. 304, 109-137 (2016).

References

- **[CHLNT17]** Chernov, Hoel, Law, Nobile, Tempone. "Multilevel ensemble Kalman filtering for spatially extended models." arXiv:1608.08558 (2017).
- **[JKLZ17]** Jasra, Kamatani, Law, Zhou. "MLMC for static Bayesian parameter estimation." arXiv:1608.08558 (2017).
- **[JKLZ17]** Jasra, Kamatani, Law, Zhou. "A Multi-Index Markov Chain Monte Carlo Method." arXiv:1704.00117 (2017).
- **[JLS17]** Jasra, Law, Suciu. "Advanced Multilevel Monte Carlo Methods." arXiv:1704.07272 (2017).
- **[BJLMZ17]**: Beskos, Jasra, Law, Marzouk, Zhou. "MLSMC samplers with DILI mutations." arXiv:1703.04866 (2017).

Thank you