Multilevel and Multi-index Monte Carlo methods for the McKean-Vlasov equation

Abdul-Lateef Haji-Ali* Raúl Tempone†

*Mathematical Institute, University of Oxford, United Kingdom
†King Abdullah University of Science and Technology (KAUST), Saudi Arabia.

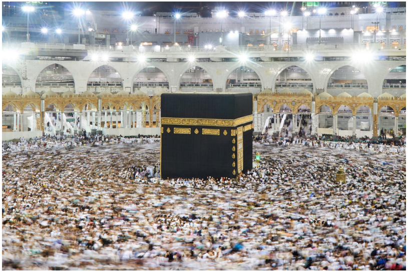
CEMRACS, Luminy, France

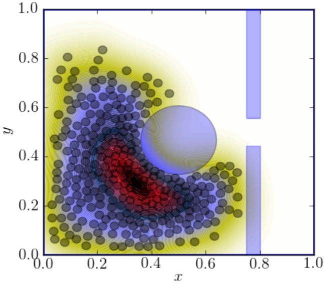
July 21, 2017

جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

Particle Systems in the Mean-field

- Particle systems are a collection of coupled, usually identical and simple, models that can be used to model complicated phenomena.
  - Molecular dynamics, Crowd simulation, Oscillators, ...
- Certain stochastic particles systems have a mean-field limit when the number of particles increase. Such limits can be useful to understand their complicated phenomena.





Main reference

This presentation is based on the manuscript

• "Multilevel and Multi-index Monte Carlo methods for the McKean-Vlasov equation" by A. L. Haji Ali and R. Tempone. *arXiv:1610.09934*, 2016. To appear in Statistics and Computing.

A McKean-Vlasov process is a stochastic process described by a SDE whose coefficients depend on the distribution of the solution itself. They relate to the Vlasov model for plasma evolution and were first studied by Henry McKean in 1966. For $0 < t \leq T$ the process $X(t)$ solves

$$\begin{aligned} dX(t) =& a(X(t), \mu(t))dW(t) + b(X(t), \mu(t))dt, \\ \mu(t) =& \mathcal{L}(X(t)) \end{aligned}$$

and $\mu(0)$ given. **Goal:** approximate $E[g(X(T))]$ for some given g.

Convergence to the mean-field

For particles $X_{p|P}$, $p = 1, \ldots, P$, (evolving in a system of size P) define "shadow" particles $X_{p|\infty}$, (evolving in a system of ∞ size)

$$X_{p|P}(t) = x_p^0 + \int_0^t \left(a(X_{p|P}(t)) + \frac{1}{P}\sum_{q=1}^{P} A(X_{p|P}(t), X_{q|P}(t)) \right) \mathrm{d}t + \sigma W_p(t)$$

$$X_{p|\infty}(t) = x_p^0 + \int_0^t \left(a(X_{p|\infty}(t)) + \int A(X_{p|\infty}(t), y)\mu_\infty(t)(\mathrm{d}y) \right) \mathrm{d}t + \sigma W_p(t),$$

with $\mu_\infty(t)$ the marginal distribution for $X_{p|\infty}(t)$ for any p.

Consistency: The initial values x_p^0, $p = 1, \ldots, P$, are i.i.d. from $\mu_\infty(0)$.

For $t > 0$, and all x, the pdf of the *marginal distribution* $\mu_\infty(t)$ of the infinite size system satisfies a nonlinear Fokker Planck equation

$$\partial_t \rho_\infty(t,x) + div\left(\rho_\infty(t,x)(a(x) + \rho_\infty(t,\cdot) * A(x,\cdot))\right) = \sum_i \frac{\sigma_i^2}{2} \partial_i^2 \rho_\infty(t,x)$$

with a given initial condition $\rho_\infty(0,\cdot)$ and suitable b.c.

Question: What about the rate of weak convergence?

$$\mathrm{E}\left[g(X_{p|P}(T)) - g(X_{p|\infty}(T))\right] \lesssim \ldots$$

Kuramoto oscillator model †

For $p = 1, 2, \ldots, P$ consider equally coupled oscillators with intrinsic natural frequencies ϑ_p that follow a system of Itô SDEs

$$\mathrm{d}X_{p|P}(t) = \left(\vartheta_p + \frac{1}{P} \sum_{q=1}^{P} \sin(X_{p|P}(t) - X_{q|P}(t)) \right) \mathrm{d}t + \sigma \mathrm{d}W_{p|P}(t)$$
$$X_{p|P}(0) = x^0_{p|P}$$

where we are interested in

$$\text{Total order} = \left(\frac{1}{P} \sum_{p=1}^{P} \cos\left(X_{p|P}(T) \right) \right)^2 + \left(\frac{1}{P} \sum_{p=1}^{P} \sin\left(X_{p|P}(T) \right) \right)^2;$$

a real number between zero and one that measures the level of synchronization of the coupled oscillators.

† Y. Kuramoto, Chemical Oscillations, Waves, and Turbulence, Springer, Berlin, 1984.

Kuramoto oscillator model †

For $p = 1, 2, \ldots, P$ consider equally coupled oscillators with intrinsic natural frequencies ϑ_p that follow a system of Itô SDEs

$$\mathrm{d}X_{p|P}(t) = \left(\vartheta_p + \frac{1}{P}\sum_{q=1}^{P} \sin(X_{p|P}(t) - X_{q|P}(t))\right)\mathrm{d}t + \sigma \mathrm{d}W_{p|P}(t)$$
$$X_{p|P}(0) = x^0_{p|P}$$

where we are interested in: $$\phi_P = \frac{1}{P}\sum_{p=1}^{P} \cos\left(X_{p|P}(T)\right),$$

Mean-field limit: $\phi_P \to \phi_\infty = \mathrm{E}\left[\cos(X_{p|\infty(T)})\right]$ as $P \uparrow \infty$

$$\mathrm{d}X_{p|\infty} = \left(\vartheta_p + \int_{\mathbb{R}} \sin(X_{p|\infty}(t) - y)\mu_\infty(t, \mathrm{d}y)\right)\mathrm{d}t + \sigma \mathrm{d}W_{p|P}(t)$$

† Y. Kuramoto, Chemical Oscillations, Waves, and Turbulence, Springer, Berlin, 1984.

Kuramoto oscillator model [†], Euler-Maruyama

For $p = 1, 2, \ldots, P$ consider equally coupled oscillators with intrinsic natural frequencies ϑ_p that follow a system of Itô SDEs

$$X_{p|P}^{n|N} - X_{p|P}^{n-1|N} = \left(\vartheta_p + \frac{1}{P}\sum_{q=1}^{P} \sin(X_{p|P}^{n|N} - X_{q|P}^{n|N})\right)\frac{T}{N} + \sigma \Delta W_{p|P}^{n|N}$$

$$X_{p|P}^{0|N} = x_{p|P}^{0}$$

where we are interested in: $$\phi_P^N = \frac{1}{P}\sum_{p=1}^{P} \cos\left(X_{p|P}^{N|N}\right),$$

Mean-field limit: $\phi_P \to \phi_\infty = \mathrm{E}\left[\cos(X_{p|\infty(T)})\right]$ as $P \uparrow \infty$

$$\mathrm{d}X_{p|\infty} = \left(\vartheta_p + \int_{\mathbb{R}} \sin(X_{p|\infty}(t) - y)\mu_\infty(t, \mathrm{d}y)\right)\mathrm{d}t + \sigma \mathrm{d}W_{p|P}(t)$$

[†] Y. Kuramoto, Chemical Oscillations, Waves, and Turbulence, Springer, Berlin, 1984.

Objective

Our objective is to build an estimator $\mathcal{A} \approx \phi_\infty$ with minimal work where

$$P(|\mathcal{A} - \phi_\infty| \leq \mathrm{TOL}) \;\geq 1 - \epsilon$$

for a given accuracy TOL and a given confidence level determined by $0 < \epsilon \ll 1$.

Objective

Our objective is to build an estimator $\mathcal{A} \approx \phi_\infty$ with minimal work where

$$P(|\mathcal{A} - \phi_\infty| \leq \mathrm{TOL}) \geq 1 - \epsilon$$

for a given accuracy TOL and a given confidence level determined by $0 < \epsilon \ll 1$. We instead impose the following, more restrictive, two constraints:

Bias constraint: $|\mathrm{E}[\mathcal{A}] - \phi_\infty| \leq \mathrm{TOL}/3,$

Statistical constraint: $P\left(|\mathcal{A} - \mathrm{E}[\mathcal{A}]| \geq 2\mathrm{TOL}/3\right)$

Objective

Our objective is to build an estimator $\mathcal{A} \approx \phi_\infty$ with minimal work where

$$P(|\mathcal{A} - \phi_\infty| \leq \mathrm{TOL}) \;\geq 1 - \epsilon$$

for a given accuracy TOL and a given confidence level determined by $0 < \epsilon \ll 1.$ We instead impose the following, more restrictive, two constraints:

$$\textbf{Bias constraint:} \qquad |\mathrm{E}[\mathcal{A}] - \phi_\infty| \leq \mathrm{TOL}/3,$$

$$\textbf{Variance constraint:} \qquad \mathrm{Var}[\mathcal{A}] \leq (2\mathrm{TOL}/3C_\epsilon)^2 \,.$$

assuming (at least asymptotic [a]) normality of the estimator, $\mathcal{A}$. Here, $0 < C_\epsilon$ is such that $\Phi(C_\epsilon) = 1 - \frac{\epsilon}{2}$, where Φ is the c.d.f. of a standard normal random variable.

[a] N. Collier, A.-L. Haji-Ali, E. von Schwerin, F. Nobile, and R. Tempone. "A continuation multilevel Monte Carlo algorithm". BIT Numerical Mathematics, 55(2):399-432, (2015).

Monte Carlo

The simplest (and most popular) estimator is the Monte Carlo estimator

$$\mathcal{A}_{\mathrm{MC}} = \frac{1}{M} \sum_{m=1}^{M} \phi_P^N(\boldsymbol{\omega}_P^m).$$

For a given P, N and M that we can choose to satisfy the error constraints and minimize the work. Here $\boldsymbol{\omega}_P^m = \left(\omega_p^m\right)_{p=1}^{P}$ and for each particle, ω_p^m denotes the independent, identically distributed (i.i.d.) samples of the set of underlying random variables that are used in calculating $X_{p|P}^{N|N}$, $1 \le p \le P$.

Monte Carlo work complexity

In our 1D example, we can check (at least numerically) that

Minimize total work: $\mathsf{Work}(\mathcal{A}_{\mathsf{MC}})$,

such that: $\mathsf{Bias}(\mathcal{A}_{\mathsf{MC}}) = \left|\phi_\infty - \mathrm{E}\left[\phi_P^N\right]\right| \leq \frac{\mathrm{TOL}}{3}$

and: $\mathrm{Var}[\mathcal{A}_{\mathsf{MC}}] = \frac{\mathrm{Var}\left[\phi_P^N\right]}{M} \leq \left(\frac{2\mathrm{TOL}}{3C_\epsilon}\right)^2$

Monte Carlo work complexity

In our 1D example, we can check (at least numerically) that

Minimize total work: $\mathrm{Work}(\mathcal{A}_{\mathsf{MC}}) = \mathcal{O}\left(MNP^2\right)$

such that: $\mathrm{Bias}(\mathcal{A}_{\mathsf{MC}}) = \mathcal{O}\left(N^{-1}\right) + \mathcal{O}\left(P^{-1}\right) \leq \frac{\mathrm{TOL}}{3}$

and: $\mathrm{Var}[\mathcal{A}_{\mathsf{MC}}] = \frac{\mathcal{O}\left(P^{-1}\right)}{M} \leq \left(\frac{2\mathrm{TOL}}{3C_\epsilon}\right)^2$

Monte Carlo work complexity

In our 1D example, we can check (at least numerically) that

$$\textbf{Minimize total work:} \quad \mathrm{Work}(\mathcal{A}_{\mathrm{MC}}) = \mathcal{O}\left(MNP^2\right)$$

$$\textbf{such that:} \quad \mathrm{Bias}(\mathcal{A}_{\mathrm{MC}}) = \mathcal{O}\left(N^{-1}\right) + \mathcal{O}\left(P^{-1}\right) \leq \frac{\mathrm{TOL}}{3}$$

$$\textbf{and:} \quad \mathrm{Var}[\mathcal{A}_{\mathrm{MC}}] = \frac{\mathcal{O}\left(P^{-1}\right)}{M} \leq \left(\frac{2\mathrm{TOL}}{3C_\epsilon}\right)^2$$

In this case, we choose

$$P = \mathcal{O}\left(\mathrm{TOL}^{-1}\right), \quad N = \mathcal{O}\left(\mathrm{TOL}^{-1}\right), \quad M = \mathcal{O}\left(\mathrm{TOL}^{-1}\right)$$

and the total cost of a naive Monte Carlo is $\mathcal{O}\left(\mathrm{TOL}^{-4}\right)$.

Observe: The cost of a "single cloud" naive method with $M = 1$ is $\mathcal{O}\left(\mathrm{TOL}^{-5}\right)$

Following (Heinrich, 2001) and (Giles, 2008), For a given $L \in \mathbb{N}$, define two hierarchies $\{N_\ell\}_{\ell=1}^L$ and $\{P_\ell\}_{\ell=1}^L$ satisfying $P_{\ell-1} \leq P_\ell$ and $N_{\ell-1} \leq N_\ell$ for all ℓ.
Recall the telescopic decomposition

$$\phi_\infty \approx \mathrm{E}\left[\phi_{P_L}^{N_L}\right] = \mathrm{E}\left[\phi_{P_0}^{N_0}\right] + \sum_{\ell=1}^{L} \mathrm{E}\left[\phi_{P_\ell}^{N_\ell} - \varphi_{P_{\ell-1}}^{N_{\ell-1}}\right]$$

KAUST

Following (Heinrich, 2001) and (Giles, 2008), For a given $L \in \mathbb{N}$, define two hierarchies $\{N_\ell\}_{\ell=1}^{L}$ and $\{P_\ell\}_{\ell=1}^{L}$ satisfying $P_{\ell-1} \le P_\ell$ and $N_{\ell-1} \le N_\ell$ for all ℓ.
Recall the telescopic decomposition

$$\phi_\infty \approx \mathrm{E}\left[\phi_{P_L}^{N_L}\right] = \mathrm{E}\left[\phi_{P_0}^{N_0}\right] + \sum_{\ell=1}^{L} \mathrm{E}\left[\phi_{P_\ell}^{N_\ell} - \varphi_{P_{\ell-1}}^{N_{\ell-1}}\right] = \sum_{\ell=0}^{L} \mathrm{E}[\Delta_\ell \phi].$$

$$\text{where} \quad \Delta_\ell \phi = \begin{cases} \phi_{P_0}^{N_0} & \text{if } \ell = 0, \\ \phi_{P_\ell}^{N_\ell} - \varphi_{P_{\ell-1}}^{N_{\ell-1}} & \text{if } \ell > 0. \end{cases}$$

Here, we assume that the auxiliary estimator φ satisfies

$$\mathrm{E}\left[\varphi_{P_{\ell-1}}^{N_{\ell-1}}\right] = \mathrm{E}\left[\phi_{P_{\ell-1}}^{N_{\ell-1}}\right]$$

Then, using Monte Carlo to approximate each level independently, the MLMC estimator can be written as

$$\mathcal{A}_{\mathrm{MLMC}} = \sum_{\ell=0}^{L} \frac{1}{M_\ell} \sum_{m=1}^{M_\ell} \Delta_\ell \phi(\omega_{P_\ell}^{\ell,m}).$$

where M_ℓ is optimally chosen. **High correlation is crucial** (between the pairs $(N_\ell, N_{\ell-1})$? $(P_\ell, P_{\ell-1})$?) to ensure that

$$\mathrm{Var}[\Delta_\ell \phi]$$

goes to zero sufficiently fast.

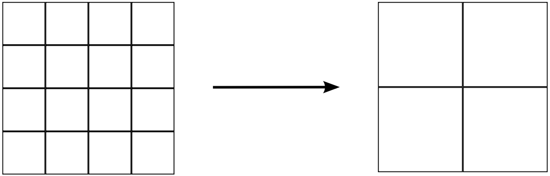
Variance reduction

Recall:

$$\mathrm{Var}[\mathcal{A}_{MC}] = \frac{1}{M_L}\mathrm{Var}[S_L].$$

$$\mathrm{Var}[\mathcal{A}_{\mathrm{MLMC}}] = \frac{1}{M_0}\mathrm{Var}[S_0] + \sum_{\ell=1}^{L}\frac{1}{M_\ell}\mathrm{Var}[\Delta_\ell S].$$

Main point: MLMC reduces the variance of the deepest level using samples on coarser (less expensive) levels!



Recall: MLMC optimal work complexity [‡] [§]

$$
\begin{aligned}
\textbf{Bias:} \quad & |\mathrm{E}[\Delta_\ell S]| = \mathcal{O}\left(\exp(-w\ell)\right), \\
\textbf{Variance:} \quad & \mathrm{Var}[\Delta_\ell S] = \mathcal{O}\left(\exp(-s\ell)\right), \\
\textbf{Work:} \quad & \mathrm{Work}[\Delta_\ell S] = \mathcal{O}\left(\exp(\gamma\ell)\right).
\end{aligned}
$$

[‡] Cliffe, K.A. and Giles, M.B. and Scheichl, R. and Teckentrup, A. Computing and Visualization in Science, "Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients" (2011).

[§] Giles, Acta Numerica 2015.

Recall: MLMC optimal work complexity [‡] [§]

$$
\begin{aligned}
\textbf{Bias:} \quad & |\mathrm{E}[\Delta_\ell S]| = \mathcal{O}\left(\exp(-w\ell)\right), \\
\textbf{Variance:} \quad & \mathrm{Var}[\Delta_\ell S] = \mathcal{O}\left(\exp(-s\ell)\right), \\
\textbf{Work:} \quad & \mathrm{Work}[\Delta_\ell S] = \mathcal{O}\left(\exp(\gamma\ell)\right).
\end{aligned}
$$

The optimal work of MLMC is

$$
\begin{cases}
\mathcal{O}\left(\mathrm{TOL}^{-2}\right) & s > \gamma \\
\mathcal{O}\left(\mathrm{TOL}^{-2}\right)\left(\log\left(\mathrm{TOL}^{-1}\right)\right)^2 & s = \gamma \\
\mathcal{O}\left(\mathrm{TOL}^{-2-\frac{\gamma - s}{w}}\right) & s < \gamma
\end{cases}
$$

Recall the total cost of Monte Carlo is

$$
\mathcal{O}\left(\mathrm{TOL}^{-2-\frac{\gamma}{w}}\right)
$$

[‡] Cliffe, K.A. and Giles, M.B. and Scheichl, R. and Teckentrup, A. Computing and Visualization in Science, "Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients" (2011).

[§] Giles, Acta Numerica 2015.

- In the following, we look at different settings in which either P_ℓ or N_ℓ depends on ℓ while the other parameter is constant for all ℓ.
- We begin by recalling the optimal convergence rates of MLMC when applied to a generic real valued random variable, Y, for the case when there are two discretization parameters:
  - ℓ, that is a function of the level,
  - $\mathcal{L}$, that is fixed for all levels.

Corollary (Optimal MLMC complexity)

Let $Y_{\mathcal{L},\ell}$ be an approximation of Y for every $(\mathcal{L},\ell) \in \mathbb{N}^2$. Consider the MLMC estimator

$$\mathcal{A}_{\mathsf{MLMC}}(L,\mathcal{L}) = \sum_{\ell=0}^{L} \frac{1}{M_\ell} \sum_{m=1}^{M_\ell} (Y_{\mathcal{L},\ell} - Y_{\mathcal{L},\ell-1})$$

with $Y_{\mathcal{L},-1} = 0$ and assume the following

1. $\left|\mathrm{E}\left[Y - Y_{\mathcal{L},\ell}\right]\right| \lesssim \exp(-\widetilde{w}\mathcal{L}) + \exp(-w\ell)$
2. $\mathrm{Var}\left[Y_{\mathcal{L},\ell} - Y_{\mathcal{L},\ell-1}\right] \lesssim \exp(-\widetilde{c}\mathcal{L})\exp(-s\ell)$
3. $W Y_{\mathcal{L},\ell} - Y_{\mathcal{L},\ell-1} \lesssim \exp(\widetilde{\gamma}\mathcal{L})\exp(\gamma\ell)$.

The optimal work of MLMC in this setting is

$$W(\mathcal{A}_{\mathsf{MLMC}}) \lesssim \begin{cases} \mathrm{TOL}^{-(2-\widetilde{c})-\frac{\widetilde{\gamma}}{\widetilde{w}}} & \text{if } s > \gamma, \\ \mathrm{TOL}^{-(2-\widetilde{c})-\frac{\widetilde{\gamma}}{\widetilde{w}}} \log\left(\mathrm{TOL}^{-1}\right)^2 & \text{if } s = \gamma, \\ \mathrm{TOL}^{-(2-\widetilde{c})-\frac{\widetilde{\gamma}}{\widetilde{w}}-\frac{\gamma-s}{w}} & \text{if } s < \gamma. \end{cases}$$

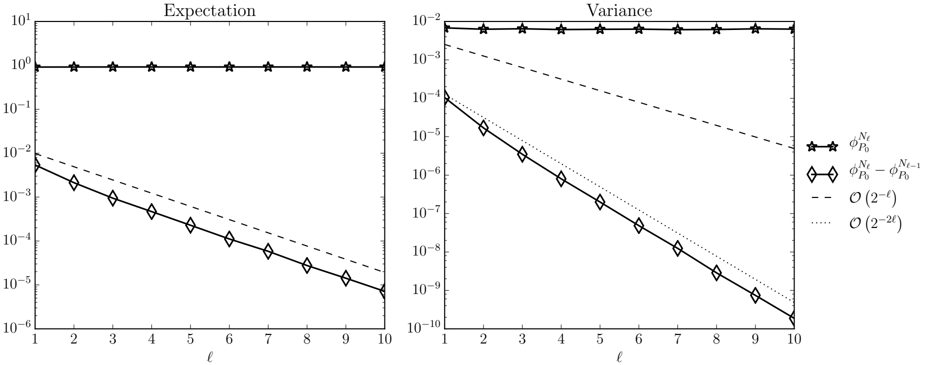
MLMC in number of time-steps, N

- $P_\ell = P_L$ and $N_\ell = 2^\ell$.
- Build correlated samples by using the same Brownian paths discretized with different meshes 2^ℓ and $2^{\ell-1}$ (Recall that we are using Euler-Maruyama discretization).

$$\varphi_{P_L}^{N_{\ell-1}}(\boldsymbol{\omega}_{P_L}^{\ell,m}) = \phi_{P_L}^{N_{\ell-1}}(\boldsymbol{\omega}_{P_L}^{\ell,m})$$

MLMC in number of time-steps, N

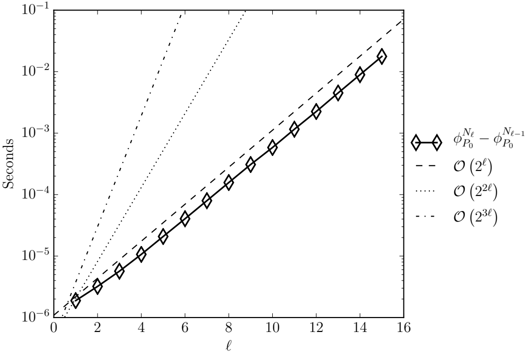
- $P_\ell = P_L$ and $N_\ell = 2^\ell$.



Bias, $w = 1, s = 2$

MLMC in number of time-steps, N

- $P_\ell = P_L$ and $N_\ell = 2^\ell$.



Work, $\gamma = 1$

MLMC in number of time-steps, N

- $P_\ell = P_L$ and $N_\ell = 2^\ell$.

- **Summary:** $w = 1, s = 2, \gamma = 1$
- Fixing P_L, the optimal work of biased MLMC is $\mathcal{O}\left(\mathrm{TOL}^{-2}\right)$.

MLMC in number of time-steps, N

- $P_\ell = P_L$ and $N_\ell = 2^\ell$.

- **Summary:** $w = 1, s = 2, \gamma = 1$
- Fixing P_L, the optimal work of biased MLMC is $\mathcal{O}\left(\mathrm{TOL}^{-2}\right)$.
- To control bias $\mathcal{O}\left(P_L^{-1}\right)$, choose $P_L = \mathcal{O}\left(\mathrm{TOL}^{-1}\right)$
  - Cost per sample: $\mathcal{O}\left(P_L^2 N_\ell\right)$
  - Variance: $\mathcal{O}\left(P_L^{-1} N_\ell^{-1}\right)$
  - **Summary:** $\tilde{w} = 1, \tilde{c} = 1, \tilde{\gamma} = 2$

  then the total cost becomes

$$\mathcal{O}\left(\mathrm{TOL}^{-3}\right) = \mathcal{O}\left(\mathrm{TOL}^{-(2-\tilde{c})-\frac{\tilde{\gamma}}{\tilde{w}}}\right).$$

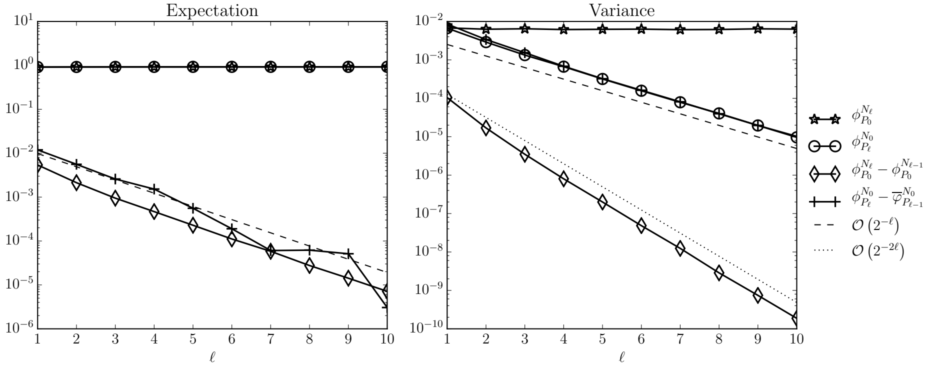
MLMC in number of particles, P

- $P_\ell = 2^\ell$ and $N_\ell = N_L$.
- Build correlated samples by sampling 2^ℓ and sub-sampling $2^{\ell-1}$ particles out of them (e.g. the first $2^{\ell-1}$). Use the same initial conditions, Brownian paths or any other random variables associated to a particle.

$$\varphi^{N_L}_{P_{\ell-1}}(\boldsymbol{\omega}^{\ell,m}_{P_\ell}) = \phi^{N_L}_{P_{\ell-1}}\left(\left(\omega^{\ell,m}_p\right)_{p=1}^{P_{\ell-1}}\right)$$

MLMC in number of particles, P

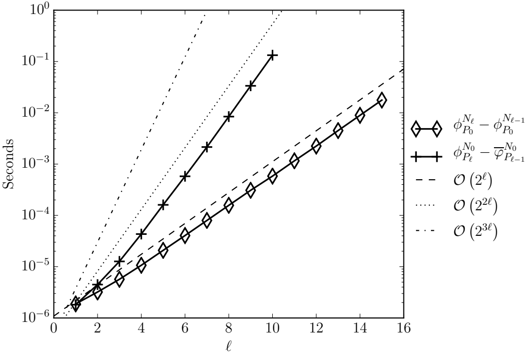
- $P_\ell = 2^\ell$ and $N_\ell = N_L$.



$$s = w = 1$$

MLMC in number of particles, P

- $P_\ell = 2^\ell$ and $N_\ell = N_L$.



Work, $\gamma = 2$

MLMC in number of particles, P

- $P_\ell = 2^\ell$ and $N_\ell = N_L$.
- **Summary:** $w = 1, s = 1, \gamma = 2$
- Fixing N_L, the optimal work of biased MLMC is $\mathcal{O}\left(\mathrm{TOL}^{-3}\right)$.

KAUST

MLMC in number of particles, P

- $P_\ell = 2^\ell$ and $N_\ell = N_L$.

- **Summary:** $w = 1, s = 1, \gamma = 2$
- Fixing N_L, the optimal work of biased MLMC is $\mathcal{O}\left(\mathrm{TOL}^{-3}\right)$.
- To control bias $\mathcal{O}\left(N_L^{-1}\right)$, choose $N_L = \mathcal{O}\left(\mathrm{TOL}^{-1}\right)$
  - Cost per sample: $\mathcal{O}\left(N_L P_\ell\right)$
  - Variance: $\mathcal{O}\left(P_\ell^{-1}\right)$
  - **Summary:** $\tilde{w} = 1, \tilde{c} = 0, \tilde{\gamma} = 1$

  then the total cost becomes

$$\mathcal{O}\left(\mathrm{TOL}^{-4}\right) = \mathcal{O}\left(\mathrm{TOL}^{-(2+\frac{\gamma-s}{w})-\frac{\tilde{\gamma}}{\tilde{w}}}\right).$$

- No advantage of MLMC over MC? Need a more correlated estimator!

Summary

Method	Work complexity
Monte Carlo	$\mathcal{O}\left(\mathrm{TOL}^{-4}\right)$
MLMC in N	$\mathcal{O}\left(\mathrm{TOL}^{-3}\right)$
MLMC in P	$\mathcal{O}\left(\mathrm{TOL}^{-4}\right)$

Reducing the variance w.r.t P; partitioning sampler

- The crucial element is how fast $\mathrm{Var}[\Delta\phi_\ell]$ goes to zero compared to how much it costs to compute $\Delta\phi_\ell$.
- Better choice of $\varphi(\cdot)$ to reduce $\mathrm{Var}[\Delta_\ell\phi]$.

Reducing the variance w.r.t P; partitioning sampler

- The crucial element is how fast $\mathrm{Var}[\Delta\phi_\ell]$ goes to zero compared to how much it costs to compute $\Delta\phi_\ell$.
- Better choice of $\varphi(\cdot)$ to reduce $\mathrm{Var}[\Delta_\ell\phi]$.
- If $P_{\ell-1} = P_\ell/2$, then given a group of P_ℓ particles, we can subsample *two identically-distributed, independent groups* that have $P_{\ell-1}$ particles and average the QoI

$$\varphi^{N_L}_{P_{\ell-1}}(\boldsymbol{\omega}^{\ell,m}_{P_\ell}) = \frac{1}{2}\left(\phi^{N_L}_{P_{\ell-1}}\left(\left(\omega^{\ell,m}_p\right)^{P_{\ell-1}}_{p=1}\right) + \phi^{N_L}_{P_{\ell-1}}\left(\left(\omega^{\ell,m}_p\right)^{P_\ell}_{p=1+P_{\ell-1}}\right)\right)$$

MLMC in number of particles P, with partitioning samplers

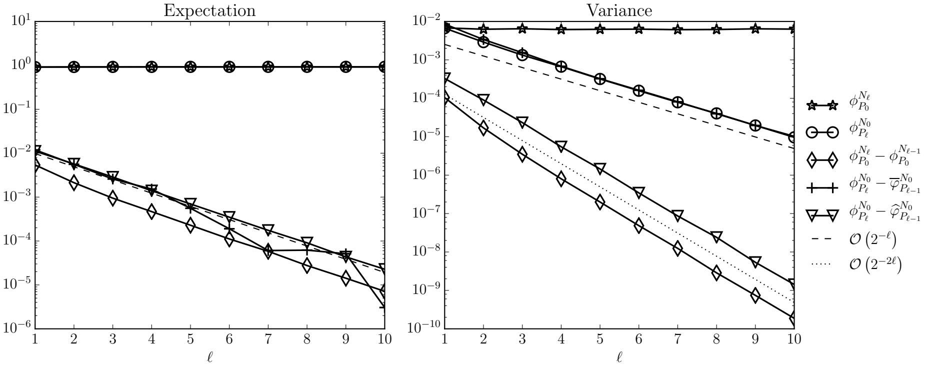
- $P_\ell = 2^\ell$ and $N_\ell = N_L$.

- Build correlated samples by sampling 2^ℓ and sub-sampling two identically-distributed, independent groups of $2^{\ell-1}$ particles out of them. Use the same initial conditions, Brownian paths or any other random variables associated to a particle.

$$\varphi^{N_L}_{P_{\ell-1}}(\boldsymbol{\omega}^{\ell,m}_{P_\ell}) = \frac{1}{2}\left(\phi^{N_L}_{P_{\ell-1}}\left(\left(\omega^{\ell,m}_p\right)^{P_{\ell-1}}_{p=1}\right) + \phi^{N_L}_{P_{\ell-1}}\left(\left(\omega^{\ell,m}_p\right)^{P_\ell}_{p=1+P_{\ell-1}}\right)\right)$$

MLMC in number of particles P, with partitioning samplers

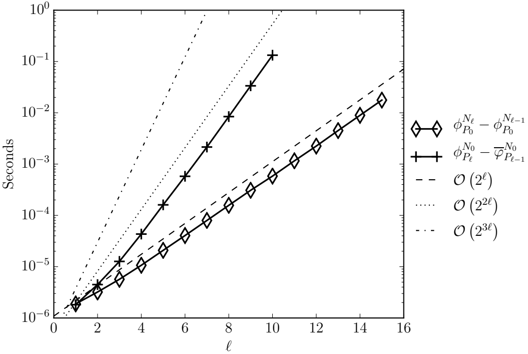
- $P_\ell = 2^\ell$ and $N_\ell = N_L$.



Bias, $w = 1, s = 2$

MLMC in number of particles P, with partitioning samplers

- $P_\ell = 2^\ell$ and $N_\ell = N_L$.



Work, $\gamma = 2$

MLMC in number of particles P, with partitioning samplers

- $P_\ell = 2^\ell$ and $N_\ell = N_L$.

- **Summary:** $w = 1, s = 2, \gamma = 2$
- Fixing N_L, the optimal work of biased MLMC is $\mathcal{O}\left(\mathrm{TOL}^{-3}\right)$.

MLMC in number of particles P, with partitioning samplers

- $P_\ell = 2^\ell$ and $N_\ell = N_L$.

- **Summary:** $w = 1, s = 2, \gamma = 2$
- Fixing N_L, the optimal work of biased MLMC is $\mathcal{O}\left(\mathrm{TOL}^{-3}\right)$.
- To control bias $\mathcal{O}\left(N_L^{-1}\right)$, choose $N_L = \mathcal{O}\left(\mathrm{TOL}^{-1}\right)$
    - Cost per sample: $\mathcal{O}\left(N_L P_\ell\right)$
    - Variance: $\mathcal{O}\left(P_\ell^{-1}\right)$
    - **Summary:** $\tilde{w} = 1, \tilde{c} = 0, \tilde{\gamma} = 1$

  then the total cost becomes

$$\mathcal{O}\left(\mathrm{TOL}^{-3}\left(\log \mathrm{TOL}\right)^2\right) = \mathcal{O}\left(\mathrm{TOL}^{-2-\frac{\tilde{\gamma}}{\tilde{w}}} \log\left(\mathrm{TOL}^{-1}\right)^2\right).$$

- Can we do better?

Summary

Method	Work complexity
Monte Carlo	$\mathcal{O}\left(\mathrm{TOL}^{-4}\right)$
MLMC in N	$\mathcal{O}\left(\mathrm{TOL}^{-3}\right)$
MLMC in P	$\mathcal{O}\left(\mathrm{TOL}^{-4}\right)$
MLMC in P, partitioning	$\mathcal{O}\left(\mathrm{TOL}^{-3}\log(\mathrm{TOL}^{-1})^2\right)$

MLMC in P and N, with partitioning samplers

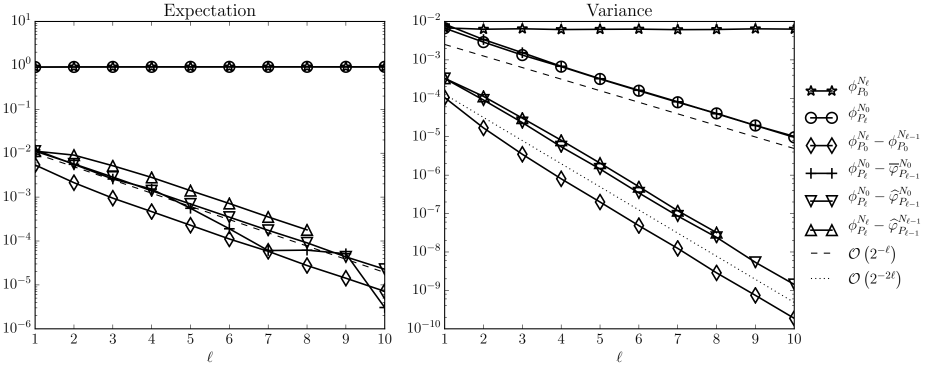
- $P_\ell = 2^\ell$ and $N_\ell = 2^\ell$.

- Build correlated samples by
  - Sampling 2^ℓ and sub-sampling two identically-distributed, independent groups of $2^{\ell-1}$ particles out of them. Use the same initial conditions, Brownian paths or any other random variables associated to a particle
  - At the same time, by using the same Brownian paths discretized with different meshes 2^ℓ and $2^{\ell-1}$.

$$\varphi_{P_{\ell-1}}^{N_{\ell-1}}(\boldsymbol{\omega}_{P_\ell}^{\ell,m}) = \frac{1}{2}\left(\phi_{P_{\ell-1}}^{N_{\ell-1}}\left(\left(\omega_p^{\ell,m}\right)_{p=1}^{P_{\ell-1}}\right) + \phi_{P_{\ell-1}}^{N_{\ell-1}}\left(\left(\omega_p^{\ell,m}\right)_{p=1+P_{\ell-1}}^{P_\ell}\right)\right)$$

MLMC in P and N, with partitioning samplers

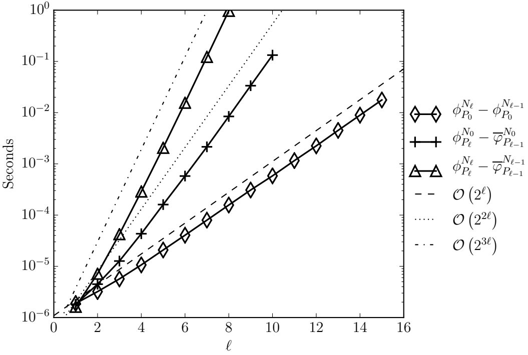
- $P_\ell = 2^\ell$ and $N_\ell = 2^\ell$.



$$w = 1, s = 2$$

MLMC in P and N, with partitioning samplers

- $P_\ell = 2^\ell$ and $N_\ell = 2^\ell$.



Work, $\gamma = 3$

MLMC in P and N, with partitioning samplers

- $P_\ell = 2^\ell$ and $N_\ell = 2^\ell$.

- **Summary:** $w = 1, s = 2, \gamma = 3$
- The optimal work of asymptotically unbiased MLMC is

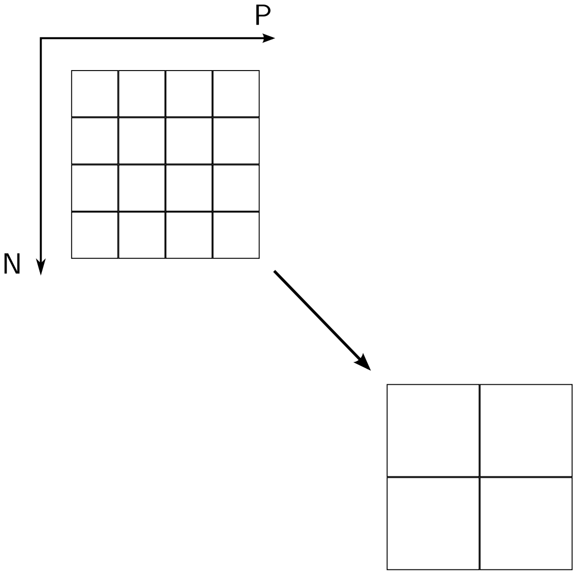
$$\mathcal{O}\left(\mathrm{TOL}^{-3}\right) = \mathcal{O}\left(\mathrm{TOL}^{-(2+\frac{\gamma-s}{w})}\right).$$

- The number of particles is not helping. Can we do better?

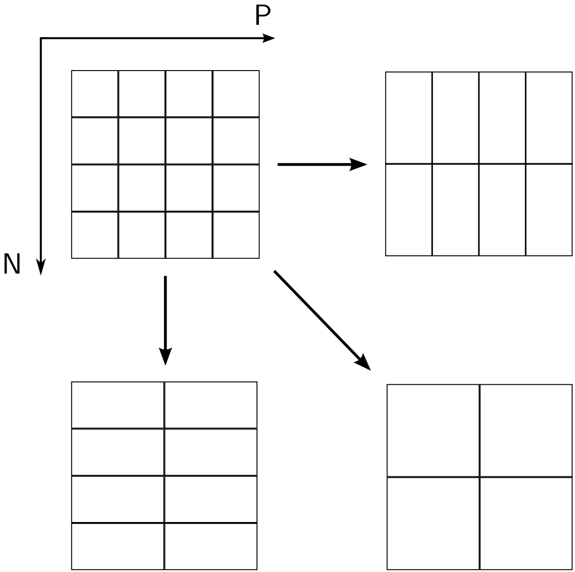
Summary

Method	Work complexity
Monte Carlo	$\mathcal{O}\left(\mathrm{TOL}^{-4}\right)$
MLMC in N	$\mathcal{O}\left(\mathrm{TOL}^{-3}\right)$
MLMC in P	$\mathcal{O}\left(\mathrm{TOL}^{-4}\right)$
MLMC in P, partitioning	$\mathcal{O}\left(\mathrm{TOL}^{-3} \log(\mathrm{TOL}^{-1})^2\right)$
MLMC in P and N	$\mathcal{O}\left(\mathrm{TOL}^{-4}\right)$
MLMC in P and N, partitioning	$\mathcal{O}\left(\mathrm{TOL}^{-3}\right)$

Variance reduction: MLMC



Variance reduction: Further potential



MIMC Estimator[¶]

For $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) \in \mathbb{N}^2$, let $P_{\alpha_1} = 2^{\alpha_1}$ and $N_{\alpha_2} = 2^{\alpha_2}$. Define the first order mixed difference

$$\begin{aligned} \boldsymbol{\Delta}_{\boldsymbol{\alpha}} \phi &= \Delta_{1,\boldsymbol{\alpha}}(\Delta_{2,\boldsymbol{\alpha}} \phi) \\ &= (\phi_{P_{\alpha_1}}^{N_{\alpha_2}} - \phi_{P_{\alpha_1}}^{N_{\alpha_2 - 1}}) - (\varphi_{P_{\alpha_1 - 1}}^{N_{\alpha_2}} - \varphi_{P_{\alpha_1 - 1}}^{N_{\alpha_2 - 1}}). \end{aligned}$$

[¶] A.-L. Haji-Ali, F. Nobile, and R. Tempone. "Multi-Index Monte Carlo: When Sparsity Meets Sampling". Numerische Mathematik, 1-40, (2015)

MIMC Estimator[¶]

For $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) \in \mathbb{N}^2$, let $P_{\alpha_1} = 2^{\alpha_1}$ and $N_{\alpha_2} = 2^{\alpha_2}$. Define the first order mixed difference

$$\begin{aligned} \boldsymbol{\Delta}_{\boldsymbol{\alpha}} \phi &= \Delta_{1,\boldsymbol{\alpha}}(\Delta_{2,\boldsymbol{\alpha}} \phi) \\ &= (\phi_{P_{\alpha_1}}^{N_{\alpha_2}} - \phi_{P_{\alpha_1}}^{N_{\alpha_2 - 1}}) - (\varphi_{P_{\alpha_1 - 1}}^{N_{\alpha_2}} - \varphi_{P_{\alpha_1 - 1}}^{N_{\alpha_2 - 1}}). \end{aligned}$$

Then the MIMC estimator can be written as

$$\mathcal{A}_{\text{MIMC}} = \sum_{\boldsymbol{\alpha} \in \mathcal{I}} \frac{1}{M_{\boldsymbol{\alpha}}} \sum_{m=1}^{M_\alpha} \boldsymbol{\Delta}_{\boldsymbol{\alpha}} \phi(\omega_{\boldsymbol{\alpha},m})$$

for some *properly chosen* index set $\mathcal{I} \subset \mathbb{N}^2$ and optimal number of samples $M_{\boldsymbol{\alpha}}$ for every $\boldsymbol{\alpha} \in \mathcal{I}$.

[¶] A.-L. Haji-Ali, F. Nobile, and R. Tempone. "Multi-Index Monte Carlo: When Sparsity Meets Sampling". Numerische Mathematik, 1-40, (2015)

MIMC optimal work complexity

$$\begin{aligned} \textbf{Bias:} \quad & |\mathrm{E}[\Delta_{\boldsymbol{\alpha}} \phi]| = \mathcal{O}\left(\exp(-w_1 \alpha_1 - w_2 \alpha_2)\right), \\ \textbf{Variance:} \quad & \mathrm{Var}[\Delta_{\boldsymbol{\alpha}} \phi] = \mathcal{O}\left(\exp(-s_1 \alpha_1 - s_2 \alpha_2)\right), \\ \textbf{Work:} \quad & \mathrm{Work}[\Delta_{\boldsymbol{\alpha}} \phi] = \mathcal{O}\left(\exp(\gamma_1 \alpha_1 + \gamma_2 \alpha_2)\right). \end{aligned}$$

MIMC optimal work complexity

$$\begin{aligned} \textbf{Bias:} \quad |\mathrm{E}[\Delta_{\boldsymbol{\alpha}}\phi]| &= \mathcal{O}\left(\exp(-w_1\alpha_1 - w_2\alpha_2)\right), \\ \textbf{Variance:} \quad \mathrm{Var}[\Delta_{\boldsymbol{\alpha}}\phi] &= \mathcal{O}\left(\exp(-s_1\alpha_1 - s_2\alpha_2)\right), \\ \textbf{Work:} \quad \mathrm{Work}[\Delta_{\boldsymbol{\alpha}}\phi] &= \mathcal{O}\left(\exp(\gamma_1\alpha_1 + \gamma_2\alpha_2)\right). \end{aligned}$$

The optimal set $\mathcal{I}$

$$\mathcal{I}(L) = \Big\{(\alpha_1, \alpha_2) \in \mathbb{N}^2 \ : \ (2w_1 + \gamma_1 - s_1)\,\alpha_1 + (2w_2 + \gamma_2 - s_2)\alpha_2 \leq L\Big\}$$

MIMC optimal work complexity

$$\begin{aligned} \textbf{Bias:} \quad & |\mathrm{E}[\Delta_{\boldsymbol{\alpha}}\phi]| = \mathcal{O}\left(\exp(-w_1\alpha_1 - w_2\alpha_2)\right), \\ \textbf{Variance:} \quad & \mathrm{Var}[\Delta_{\boldsymbol{\alpha}}\phi] = \mathcal{O}\left(\exp(-s_1\alpha_1 - s_2\alpha_2)\right), \\ \textbf{Work:} \quad & \mathrm{Work}[\Delta_{\boldsymbol{\alpha}}\phi] = \mathcal{O}\left(\exp(\gamma_1\alpha_1 + \gamma_2\alpha_2)\right). \end{aligned}$$

The optimal set $\mathcal{I}$

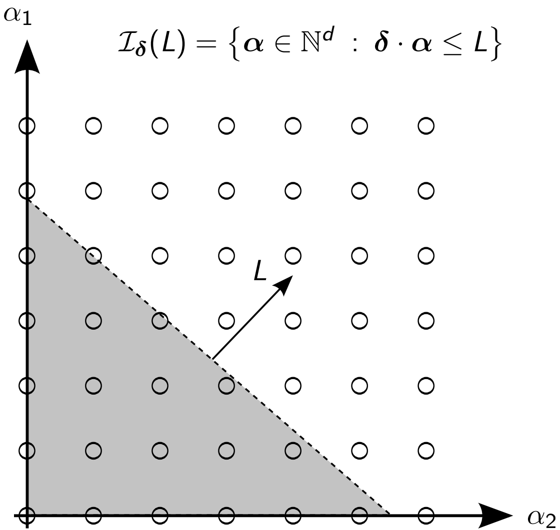
$$\mathcal{I}(L) = \Big\{(\alpha_1, \alpha_2) \in \mathbb{N}^2 \;: \\ (2w_1 + \gamma_1 - s_1)\,\alpha_1 + (2w_2 + \gamma_2 - s_2)\alpha_2 \le L\Big\}$$

Letting $\zeta = \max\left(\frac{\gamma_1 - s_1}{2w_1}, \frac{\gamma_2 - s_2}{2w_2}\right)$, then the optimal work of MIMC is

$$\mathcal{O}\left(\mathrm{TOL}^{-2(1+\max(0,\zeta))} \log\left(\mathrm{TOL}^{-1}\right)^{\mathfrak{p}}\right)$$

for $\mathfrak{p} \ge 0$.

Optimal choice of $\mathcal{I}$: Total degree set

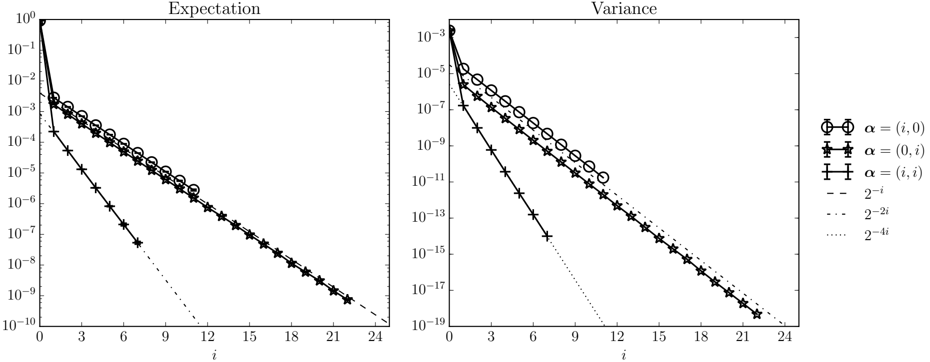


MIMC, with partitioning samplers

- Let $P_{\alpha_1} = 2^{\alpha_1}$ and $N_{\alpha_2} = 2^{\alpha_2}$.
- Build correlated samples by
  - Sampling 2^{α_1} and sub-sampling two identically-distributed, independent groups of $2^{\alpha_1 - 1}$ particles out of them.
  - At the same time, by using the same Brownian paths discretized with different meshes 2^{α_2} and $2^{\alpha_2 - 1}$.
  - Use MIMC levels: Mixed differences!

MIMC, with partitioning samplers

- Let $P_{\alpha_1} = 2^{\alpha_1}$ and $N_{\alpha_2} = 2^{\alpha_2}$.

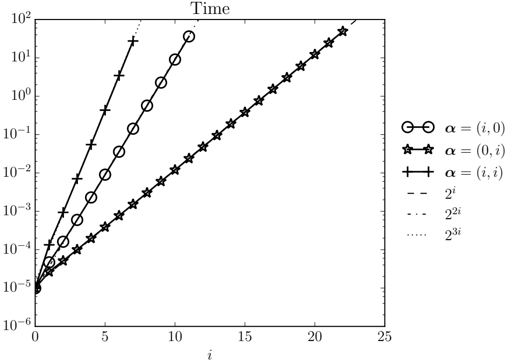


$w_1, w_2 = 1, s_1 = s_2 = 2$

Notice higher rates for mixed difference.

MIMC, with partitioning samplers

- Let $P_{\alpha_1} = 2^{\alpha_1}$ and $N_{\alpha_2} = 2^{\alpha_2}$.



$$\gamma_1 = 2, \gamma_2 = 1$$

MIMC, with partitioning samplers

- Let $P_{\alpha_1} = 2^{\alpha_1}$ and $N_{\alpha_2} = 2^{\alpha_2}$.
- **Summary:**

$$\left.\begin{array}{c} w_1 = w_2 = 1 \\ s_1 = s_2 = 2 \\ \gamma_1 = 2\gamma_2 = 2 \end{array}\right\} \Longrightarrow \zeta = \max\left(\frac{\gamma_1 - s_1}{2w_1}, \frac{\gamma_2 - s_2}{2w_2}\right) = 0$$

- The optimal set

$$\mathcal{I}(L) = \left\{(\alpha_1, \alpha_2) \in \mathbb{N}^2 \ : \ 2\alpha_1 + 3\alpha_2 \leq L\right\}$$

- The optimal work of the asymptotically unbiased MIMC is

$$\mathcal{O}\left(\mathrm{TOL}^{-2} \log\left(\mathrm{TOL}^{-1}\right)^2\right)$$

Summary

Method	Work complexity
Monte Carlo	$\mathcal{O}\left(\mathrm{TOL}^{-4}\right)$
MLMC in N	$\mathcal{O}\left(\mathrm{TOL}^{-3}\right)$
MLMC in P	$\mathcal{O}\left(\mathrm{TOL}^{-4}\right)$
MLMC in P, partitioning	$\mathcal{O}\left(\mathrm{TOL}^{-3}\log(\mathrm{TOL}^{-1})^2\right)$
MLMC in P and N	$\mathcal{O}\left(\mathrm{TOL}^{-4}\right)$
MLMC in P and N, partitioning	$\mathcal{O}\left(\mathrm{TOL}^{-3}\right)$
MIMC	$\mathcal{O}\left(\mathrm{TOL}^{-2}\log\left(\mathrm{TOL}^{-1}\right)^2\right)$

Numerical Example: MIMC vs. MLMC

$$X_{p|P}^{n|N} - X_{p|P}^{n-1|N} = \left(\vartheta_p + \frac{0.4}{P} \sum_{q=1}^{P} \sin(X_{p|P}^{n|N} - X_{q|P}^{n|N}) \right) \frac{T}{N} + 0.4 \Delta W_{p|P}^{n|N}$$

$$X_{p|P}^{0|N} \sim \mathcal{N}(0, 0.2)$$

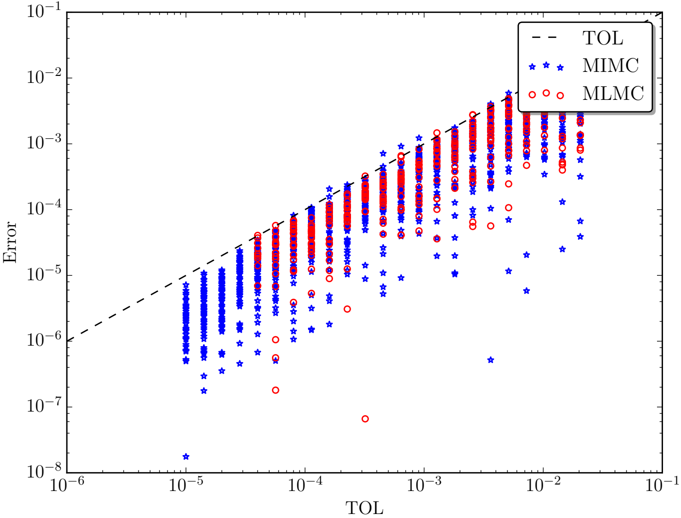
where $\vartheta_p \sim \mathcal{U}(-0.2, 0.2)$. The quantity of interest is

$$\phi_P^N = \frac{1}{P} \sum_{p=1}^{P} \cos\left(X_{p|P}^{N|N} \right).$$

for $T = 1$.

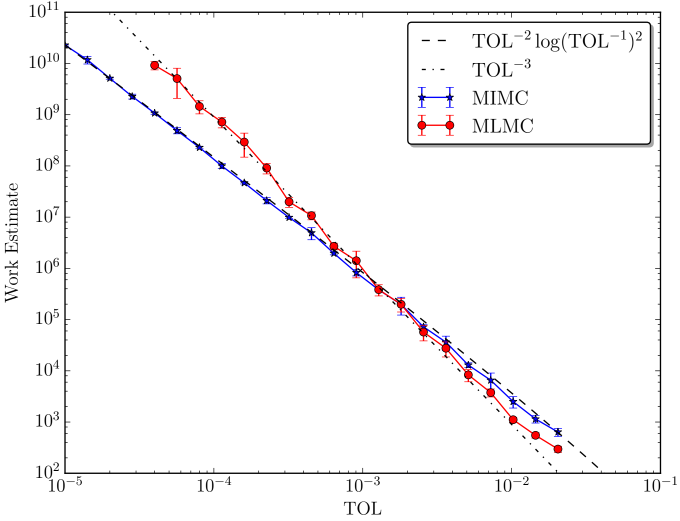
Numerical Example: MIMC vs. MLMC

for $T = 1$.



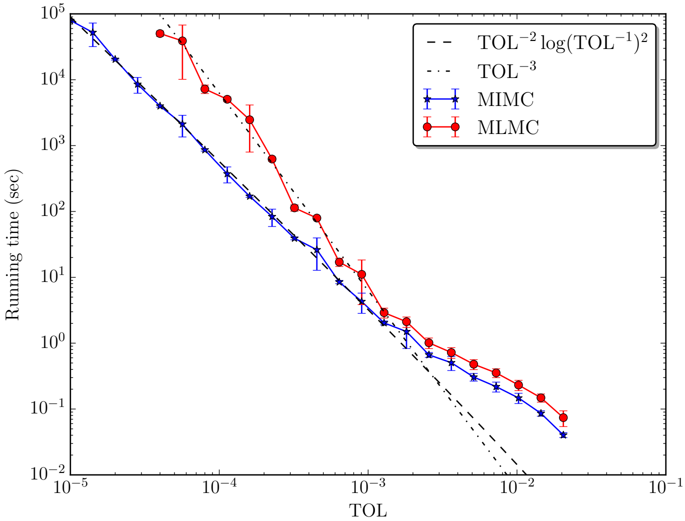
Numerical Example: MIMC vs. MLMC

for $T = 1$.



Numerical Example: MIMC vs. MLMC

for $T = 1$.



Conclusions

- MLMC applied to particle systems in the mean-field limit is not trivial since the variance of the quantity of interest depends on the number of particles, P.
- The partitioning estimator achieves much better L^2 convergence rate w.r.t. P.
- Considerable computational saving by using MIMC, from TOL^{-4} to $\mathrm{TOL}^{-2} \log^2(\mathrm{TOL})$.
- Other applications: higher dimension particle systems (e.g. crowd flow).
- MLMC and MIMC Theory: working on estimates on mixed differences.