Multi-Level and Multi-index Monte Carlo (and Multi-index Stochastic Collocation)

Raúl Tempone

King Abdullah University of Science and Technology (KAUST), Saudi Arabia.
stochastic_numerics.kaust.edu.sa sri-uq.kaust.edu.sa
International Conference on Monte Carlo techniques
July 8th, 2016

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

Overview of the talk

Monte Carlo and extensions

Motivational Example: Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and $\mathcal{D}$ be a bounded convex polygonal domain in $\mathbb{R}^d$.
The solution $u : \mathcal{D} \times \Omega \to \mathbb{R}$ here solves almost surely (a.s.) the following equation:

$$\begin{aligned} -\nabla \cdot (a(\boldsymbol{x};\omega)\nabla u(\boldsymbol{x};\omega)) &= f(\boldsymbol{x};\omega) && \text{for } \boldsymbol{x} \in \mathcal{D}, \\ u(\boldsymbol{x};\omega) &= 0 && \text{for } \boldsymbol{x} \in \partial\mathcal{D}. \end{aligned}$$

Goal: to approximate $\mathrm{E}[S] \in \mathbb{R}$ where $S = \Psi(u)$ for some sufficiently "smooth" a, f and functional Ψ.

Monte Carlo and extensions

Motivational Example: Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and $\mathcal{D}$ be a bounded convex polygonal domain in $\mathbb{R}^d$.
The solution $u : \mathcal{D} \times \Omega \to \mathbb{R}$ here solves almost surely (a.s.) the following equation:

$$
\begin{aligned}
-\nabla \cdot (a(\boldsymbol{x};\omega)\nabla u(\boldsymbol{x};\omega)) &= f(\boldsymbol{x};\omega) && \text{for } \boldsymbol{x} \in \mathcal{D}, \\
u(\boldsymbol{x};\omega) &= 0 && \text{for } \boldsymbol{x} \in \partial\mathcal{D}.
\end{aligned}
$$

Goal: to approximate $\mathrm{E}[S] \in \mathbb{R}$ where $S = \Psi(u)$ for some sufficiently "smooth" a, f and functional Ψ.
Later, in our numerical example we use

$$
S = 100 \left(2\pi\sigma^2\right)^{\frac{-3}{2}} \int_{\mathcal{D}} \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}_0\|_2^2}{2\sigma^2}\right) u(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}.
$$

for $\boldsymbol{x}_0 \in \mathcal{D}$ and $\sigma > 0$.

Monte Carlo (Metropolis and Ulam, 1949)

Recall the Monte Carlo method and its error splitting:

$$\mathrm{E}[\Psi(u(\boldsymbol{y}))] - \frac{1}{M}\sum_{m=1}^{M}\Psi(u_h(\boldsymbol{y}(\omega_m))) = \mathcal{E}^{\Psi}_{bias}(h) + \mathcal{E}^{\Psi}_{stat}(M)$$

$$|\mathcal{E}^{\Psi}_{bias}(h)| = \underbrace{|E[\Psi(u(\boldsymbol{y})) - \Psi(u_h(\boldsymbol{y}))]|}_{\text{discretization error}} \leq C h^{w}$$

$$|\mathcal{E}^{\Psi}_{stat}(M)| = \underbrace{|E[\Psi(u_h(\boldsymbol{y}))] - \frac{1}{M}\sum_{m=1}^{M}\Psi(u_h(\boldsymbol{y}(\omega_m)))|}_{\text{statistical error}} \lesssim \quad c_0 \frac{\mathrm{std}[\Psi(u_h)]}{\sqrt{M}}$$

The last approximation is motivated by the Central Limit Theorem.

$$P\left(|\mathcal{E}^{\Psi}_{stat}(M)| \leq c_0 \frac{\mathrm{std}[\Psi(u_h)]}{\sqrt{M}}\right) \approx 1 - \epsilon$$

Assume: computational work for each $u(\boldsymbol{y}(\omega_m))$ is $\mathcal{O}(h^{-d\gamma})$.

$$\text{Total work}: \quad M h^{-d\gamma}$$

$$\text{Total error}: \quad |\mathcal{E}^{\Psi}_{bias}(h)| + |\mathcal{E}^{\Psi}_{stat}(M)| \leq C_1 h^w + \frac{C_2}{\sqrt{M}}$$

We want now to choose optimally h and M. Here we minimize the computational work subject to an accuracy constraint, i.e. we solve

$$\begin{cases} \min_{h,M} \;\; M\, h^{-d\gamma} \\ \text{s.t.} \;\; C_1 h^w + \frac{C_2}{\sqrt{M}} \leq \mathrm{TOL} \end{cases}$$

We can interpret the above as a tolerance splitting into statistical and space discretization tolerances, $\mathrm{TOL} = \mathrm{TOL}_S + \mathrm{TOL}_h$, such that

$$\mathrm{TOL}_h = \frac{\mathrm{TOL}}{(1 + 2w/(d\gamma))} \;\; \text{and} \;\; \mathrm{TOL}_S = \mathrm{TOL}\left(1 - \frac{1}{(1 + 2w/(d\gamma))}\right).$$

The resulting complexity (error versus computational work) is then

$$W \propto \mathrm{TOL}^{-(2+d\gamma/w)}$$

Numerical Approximation

We assume:

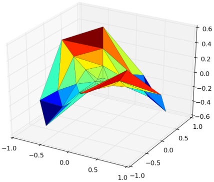
- $\mathcal{D} = \prod_{i=1}^{d} [0, D_i]$ for $D_i \subset \mathbb{R}_+$ be a hypercube domain in $\mathbb{R}^d$.
- we have an approximation of u (FEM, FD, FV, ...) based on discretization parameters h_i for $i = 1 \ldots d$. Here

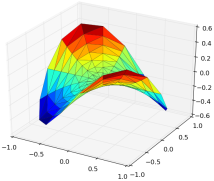
$$h_i = h_{i,0}\, \beta_i^{-\alpha_i},$$

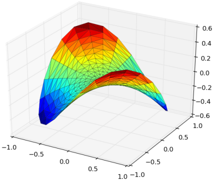
with $\beta_i > 1$ and the multi-index

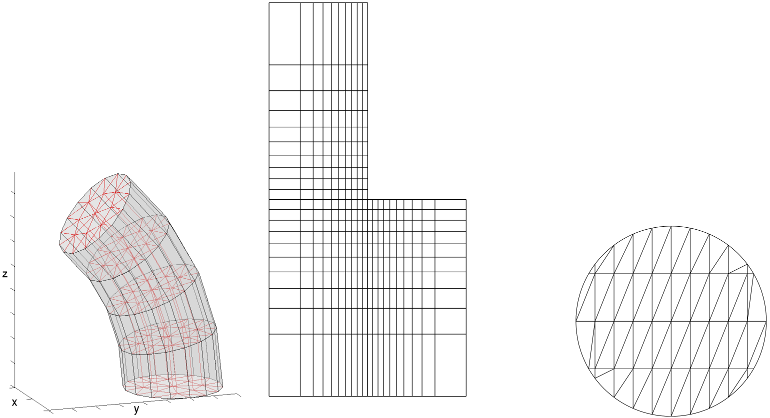
$$\boldsymbol{\alpha} = (\alpha_i)_{i=1}^{d} \in \mathbb{N}^d.$$

Notation: $S_{\boldsymbol{\alpha}}$ is the approximation of S calculated using a discretization defined by $\boldsymbol{\alpha}$.









Left: Tensor domain, cylinder.
Center: Non-tensor domain immersed in a tensor box.
Right: Non-tensor domain with a structured mesh.

Multilevel Monte Carlo (MLMC)

(Heinrich, 1998) and (Giles, 2008)

Take $\beta_i = \beta$ and for each $\ell = 1, 2, \ldots$ use discretizations with $\boldsymbol{\alpha} = (\ell, \ldots, \ell)$. Recall the standard MLMC difference operator

$$\widetilde{\Delta} S_\ell = \begin{cases} S_{\mathbf{0}} & \text{if } \ell = 0, \\ S_{\ell \cdot \mathbf{1}} - S_{(\ell-1)\cdot \mathbf{1}} & \text{if } \ell > 0. \end{cases}$$

Multilevel Monte Carlo (MLMC)

(Heinrich, 1998) and (Giles, 2008)

Take $\beta_i = \beta$ and for each $\ell = 1, 2, \ldots$ use discretizations with $\boldsymbol{\alpha} = (\ell, \ldots, \ell)$. Recall the standard MLMC difference operator

$$\widetilde{\Delta} S_\ell = \begin{cases} S_{\mathbf{0}} & \text{if } \ell = 0, \\ S_{\ell \cdot \mathbf{1}} - S_{(\ell-1)\cdot \mathbf{1}} & \text{if } \ell > 0. \end{cases}$$

Observe the telescopic identity

$$\mathrm{E}[S] \approx \mathrm{E}[S_{L \cdot \mathbf{1}}] = \sum_{\ell=0}^{L} \mathrm{E}\left[\widetilde{\Delta} S_\ell\right].$$

Multilevel Monte Carlo (MLMC)

(Heinrich, 1998) and (Giles, 2008)

Take $\beta_i = \beta$ and for each $\ell = 1, 2, \ldots$ use discretizations with $\boldsymbol{\alpha} = (\ell, \ldots, \ell)$. Recall the standard MLMC difference operator

$$\widetilde{\Delta} S_\ell = \begin{cases} S_{\mathbf{0}} & \text{if } \ell = 0, \\ S_{\ell \cdot \mathbf{1}} - S_{(\ell-1)\cdot \mathbf{1}} & \text{if } \ell > 0. \end{cases}$$

Observe the telescopic identity

$$\mathrm{E}[S] \approx \mathrm{E}[S_{L \cdot \mathbf{1}}] = \sum_{\ell=0}^{L} \mathrm{E}\left[\widetilde{\Delta} S_\ell\right].$$

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

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

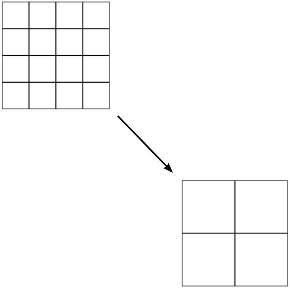
Variance reduction: MLMC

Recall: With Monte Carlo we have to satisfy

$$\mathrm{Var}[A_{MC}] = \frac{1}{M_L}\mathrm{Var}[S_L] \approx \frac{1}{M_L}\mathrm{Var}[S] \leq \mathrm{TOL}^2.$$

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

$$\mathrm{Var}[A_{\mathrm{MLMC}}] = \frac{1}{M_0}\mathrm{Var}[S_0] + \sum_{\ell=1}^{L} \frac{1}{M_\ell}\mathrm{Var}[\Delta S_\ell] \leq \mathrm{TOL}^2.$$



Observe: Level 0 in MLMC is usually determined by *both* stability and accuracy, i.e. $\mathrm{Var}[\Delta S_1] << \mathrm{Var}[S_0] \approx \mathrm{Var}[S] < \infty$.

Classical assumptions for MLMC

For every ℓ, we assume the following:

Assumption $\tilde{1}$ (Bias): $$|\mathrm{E}[S - S_\ell]| \lesssim \beta^{-w\ell},$$

Assumption $\tilde{2}$ (Variance): $$V_\ell = \mathrm{Var}\left[\widetilde{\Delta} S_\ell\right] \lesssim \beta^{-s\ell},$$

Assumption $\tilde{3}$ (Work): $$W_\ell = \mathrm{Work}(\widetilde{\Delta} S_\ell) \lesssim \beta^{d\gamma\ell},$$

for positive constants γ, w and $s \leq 2w$.

Example: Our smooth linear elliptic PDE example approximated with Multilinear piecewise cont. FEM: $2w = s = 4, 1 \leq \gamma \leq 3$.

Work of MLMC: $$\mathrm{Work}(\mathrm{MLMC}) = \sum_{\ell=0}^{L} M_\ell W_\ell$$

Choose the samples $(M_\ell)_{\ell=0}^{L}$ optimally so $\mathrm{Var}[\mathcal{A}_{\mathrm{MLMC}}] \lesssim \mathrm{TOL}^2$.

Optimal Work of MLMC: $$\mathrm{Work}(\mathrm{MLMC}) \lesssim \mathrm{TOL}^{-2} \left(\sum_{\ell=0}^{L} \sqrt{V_\ell W_\ell} \right)^2$$

MLMC Computational Complexity

Choose the number of levels $L(\mathrm{TOL})$ to bound the bias

$$|\mathrm{E}[S - S_L]| \lesssim \beta^{-Lw} \leq C\mathrm{TOL} \quad \Rightarrow \quad L \geq \frac{\log(\mathrm{TOL}^{-1}) - \log(C)}{w \log(\beta)},$$

Then the optimal work satisfies (Giles et al., 2008, 2011):

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

Recall: $\mathrm{Work(MC)} = \mathcal{O}\left(\mathrm{TOL}^{-(2+\frac{d\gamma}{w})}\right)$.

Questions related to MLMC

1. How to choose the mesh hierarchy $\boldsymbol{h}_\ell$? **[H-ASNT, 2015]**
2. How to efficiently and reliably estimate V_ℓ? How to find the correct number of levels, L? **[CH-ASNT, 2015]**
3. Can we do better? Especially for $d > 1$? **[H-ANT, 2015]**

[H-ASNT, 2015] A.-L. Haji-Ali, E. von Schwerin, F. Nobile, and R. T. "Optimization of mesh hierarchies in Multilevel Monte Carlo samplers". arXiv:1403.2480, Stochastic Partial Differential Equations: Analysis and Computations, 4(1):76–112, (2016).

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

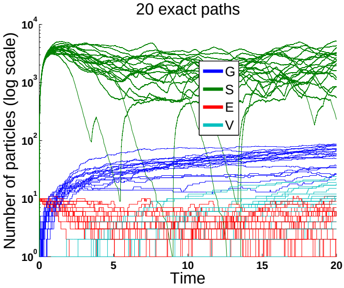
[H-ANT, 2015] A.-L. Haji-Ali, F. Nobile, and R. T. "Multi-Index Monte Carlo: When Sparsity Meets Sampling". Numerische Mathematik, 806:132–767, (2016).

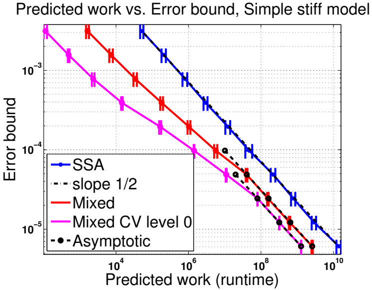
Time adaptivity for MLMC in Itô SDEs: Stopping with optimal asymptotic Accuracy and Efficiency

- Adaptive Multilevel Monte Carlo Simulation, by H. Hoel, E. von Schwerin, A. Szepessy and R. T., Numerical Analysis of Multiscale Computations, **82**, Lect. Notes Comput. Sci. Eng., (2011).
- Implementation and Analysis of an Adaptive Multi Level Monte Carlo Algorithm, by H. Hoel, E. von Schwerin, A. Szepessy and R. T., Monte Carlo Methods and Applications. **20**, (2014).
- Construction of a mean square error adaptive Euler-Maruyama method with applications in multilevel Monte Carlo, by H. Hoel, J. Häppöla, and R. T. To appear in MC and Q-MC Methods 2014, Springer Verlag, (2016).

Hybrid MLMC for Stochastic Reaction Networks

- A. Moraes, R. T., and P. Vilanova. Multilevel hybrid Chernoff tau-leap. BIT Numerical Mathematics, April 2015.
- A. Moraes, R. T., and P. Vilanova. A multilevel adaptive reaction-splitting simulation method for stochastic reaction networks. *arXiv:1406.1989*. To appear in SIAM Journal on Scientific Computing (SISC), 2016.
- C. Ben Hammouda, A. Moraes and R. T. Multilevel drift-implicit tau-leap, *arXiv:1512.00721*. To appear in the Journal of Numerical Algorithms, 2016.

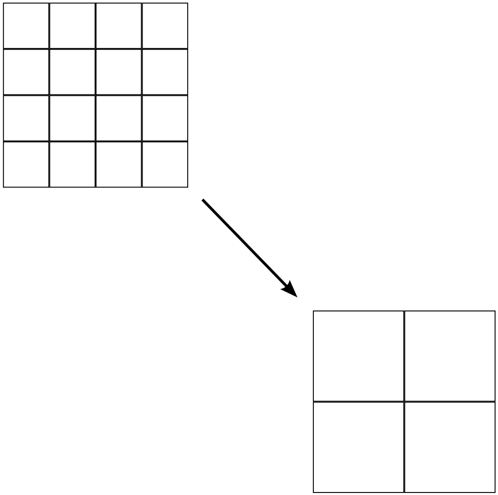




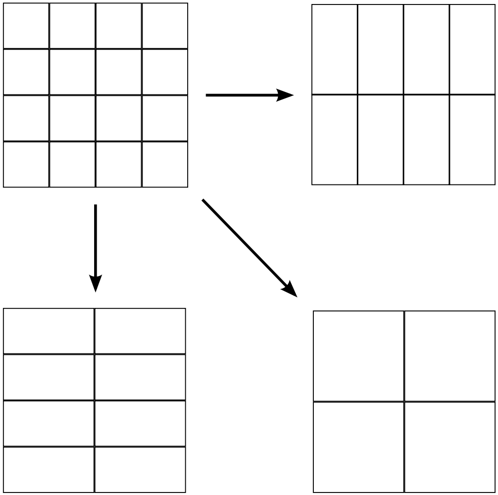
Kurtz representation: $X(t) = x_0 + \sum_{j=1}^{J} Y_j \left(\int_0^t a_j(X(s))ds \right) \nu_j$, Tau-Leap: $\bar{X}_{n+1} = \bar{X}_n + \sum_{j=1}^{J} \mathcal{P}_j(a_n \Delta t) \nu_j$

with independent unit-rate Poisson processes $\{Y_j(t)\}_{t \geq 0}$ and reaction channels $\{a_j, \nu_j\}$.

Variance reduction: MLMC



Variance reduction: Further potential



Multi-Index Monte Carlo (MIMC) (Haji Ali, Nobile, T. 2015)

For $i = 1, \ldots, d$, define the first order difference operators

$$\Delta_i S_{\boldsymbol{\alpha}} = \begin{cases} S_{\boldsymbol{\alpha}} & \text{if } \alpha_i = 0, \\ S_{\boldsymbol{\alpha}} - S_{\boldsymbol{\alpha} - \boldsymbol{e}_i} & \text{if } \alpha_i > 0, \end{cases}$$

and construct the first order mixed difference

$$\Delta S_{\boldsymbol{\alpha}} = \left(\otimes_{i=1}^{d} \Delta_i \right) S_{\boldsymbol{\alpha}}.$$

Multi-Index Monte Carlo (MIMC)

(Haji Ali, Nobile, T. 2015)

For $i = 1, \dots, d$, define the first order difference operators

$$\Delta_i S_{\boldsymbol{\alpha}} = \begin{cases} S_{\boldsymbol{\alpha}} & \text{if } \alpha_i = 0, \\ S_{\boldsymbol{\alpha}} - S_{\boldsymbol{\alpha}-\boldsymbol{e}_i} & \text{if } \alpha_i > 0, \end{cases}$$

and construct the first order mixed difference

$$\Delta S_{\boldsymbol{\alpha}} = \left(\otimes_{i=1}^{d} \Delta_i\right) S_{\boldsymbol{\alpha}}.$$

Then the MIMC estimator can be written as

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

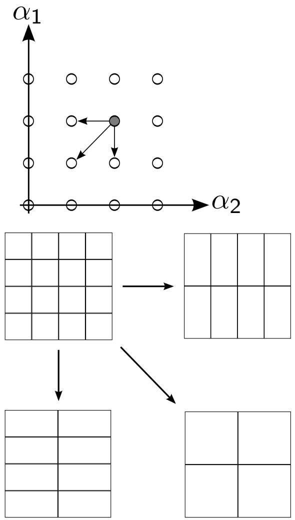
for some *properly chosen* index set $\mathcal{I} \subset \mathbb{N}^d$ and samples $(M_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathcal{I}}$.

Example: On mixed differences

Consider $d = 2$. In this case, letting $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$, we have

$$\begin{aligned}\Delta S_{(\alpha_1,\alpha_2)} &= \Delta_2(\Delta_1 S_{(\alpha_1,\alpha_2)}) \\ &= \Delta_2\left(S_{\alpha_1,\alpha_2} - S_{\alpha_1-1,\alpha_2}\right) \\ &= \left(S_{\alpha_1,\alpha_2} - S_{\alpha_1-1,\alpha_2}\right) \\ &\quad - \left(S_{\alpha_1,\alpha_2-1} - S_{\alpha_1-1,\alpha_2-1}\right).\end{aligned}$$

Notice that in general, $\Delta S_{\boldsymbol{\alpha}}$ requires 2^d evaluations of S at different discretization parameters, the largest work of which corresponds precisely to the index appearing in $\Delta S_{\boldsymbol{\alpha}}$, namely $\boldsymbol{\alpha}$.



Our objective is to build an estimator $\mathcal{A} = \mathcal{A}_{\text{MIMC}}$ where

$$P(|\mathcal{A} - \mathrm{E}[S]| \leq \mathrm{TOL}) \;\geq 1 - \epsilon \qquad (1)$$

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:

$$\textbf{Bias constraint:} \qquad |\mathrm{E}[\mathcal{A} - S]| \leq (1-\theta)\mathrm{TOL}, \qquad (2)$$

$$\textbf{Statistical constraint:} \qquad P\left(|\mathcal{A} - \mathrm{E}[\mathcal{A}]| \leq \theta\mathrm{TOL}\right) \geq 1 - \epsilon. \qquad (3)$$

For a given fixed $\theta \in (0,1)$. Moreover, motivated by the asymptotic normality of the estimator, $\mathcal{A}$, we approximate (3) by

$$\mathrm{Var}[\mathcal{A}] \leq \left(\frac{\theta\mathrm{TOL}}{C_\epsilon}\right)^2. \qquad (4)$$

Here, $0 < C_\epsilon$ is such that $\Phi(C_\epsilon) = 1 - \frac{\epsilon}{2}$, where Φ is the cumulative distribution function of a standard normal random var.

Given variance and work estimates we can already optimize for the optimal number of samples $M^*_{\boldsymbol{\alpha}} \in \mathbb{R}$ to satisfy the variance constraint (4)

$$M^*_{\boldsymbol{\alpha}} = \left(\frac{C_\epsilon}{\theta \mathrm{TOL}}\right)^2 \sqrt{\frac{V_{\boldsymbol{\alpha}}}{W_{\boldsymbol{\alpha}}}} \left(\sum_{\boldsymbol{\alpha} \in \mathcal{I}} \sqrt{V_{\boldsymbol{\alpha}} W_{\boldsymbol{\alpha}}}\right).$$

Taking $M^*_{\boldsymbol{\alpha}} \leq M_{\boldsymbol{\alpha}} \leq M^*_{\boldsymbol{\alpha}} + 1$ such that $M_{\boldsymbol{\alpha}} \in \mathbb{N}$ and substituting in the total work gives

$$\mathrm{Work}(\mathcal{I}) \leq \left(\frac{C_\epsilon}{\theta \mathrm{TOL}}\right)^2 \left(\sum_{\boldsymbol{\alpha} \in \mathcal{I}} \sqrt{V_{\boldsymbol{\alpha}} W_{\boldsymbol{\alpha}}}\right)^2 + \underbrace{\sum_{\boldsymbol{\alpha} \in \mathcal{I}} W_{\boldsymbol{\alpha}}}_{\text{Min. cost of } \mathcal{I}} .$$

Observe: The work now depends on $\mathcal{I}$ only.

MIMC general analysis framework

Question: How do we find optimal index set $\mathcal{I}$ for MIMC?

$$\min_{\mathcal{I}\subset\mathbb{N}^d} \mathit{Work}(\mathcal{I}) \quad \text{such that Bias} = \sum_{\boldsymbol{\alpha}\notin\mathcal{I}} E_{\boldsymbol{\alpha}} \leq (1-\theta)\mathrm{TOL},$$

Assumption: MIMC work is *not* dominated by the work to compute a single sample corresponding to each $\boldsymbol{\alpha}$.

Then, minimizing equivalently $\sqrt{\mathit{Work}(\mathcal{I})}$, the previous min problem can be recast into a knapsack problem with profits defined for each multi-index $\boldsymbol{\alpha}$.

The corresponding $\boldsymbol{\alpha}$ profit is

$$\mathcal{P}_{\boldsymbol{\alpha}} = \frac{\text{Bias contribution}}{\text{Work contribution}} = \frac{E_{\boldsymbol{\alpha}}}{\sqrt{V_{\boldsymbol{\alpha}} W_{\boldsymbol{\alpha}}}}$$

MIMC general analysis framework

Define the total error associated with an index-set $\mathcal{I}$ as

$$\mathfrak{E}(\mathcal{I}) = \sum_{\boldsymbol{\alpha} \notin \mathcal{I}} E_{\boldsymbol{\alpha}}$$

and the corresponding total work estimate as

$$\mathfrak{W}(\mathcal{I}) = \sum_{\boldsymbol{\alpha} \in \mathcal{I}} \sqrt{V_{\boldsymbol{\alpha}} W_{\boldsymbol{\alpha}}}.$$

Then we can show the following optimality result with respect to $\mathfrak{E}(\mathcal{I})$ and $\mathfrak{W}(\mathcal{I})$, namely:

Lemma (Optimal profit sets)

The index-set

$$\mathcal{I}(\nu) = \{\boldsymbol{\alpha} \in \mathbb{N}^d : \mathcal{P}_{\boldsymbol{\alpha}} \geq \nu\}$$

for $\mathcal{P}_{\boldsymbol{\alpha}} = \frac{E_{\boldsymbol{\alpha}}}{\sqrt{V_{\boldsymbol{\alpha}} W_{\boldsymbol{\alpha}}}}$ *is optimal in the sense that any other index-set,* $\tilde{\mathcal{I}}$, *with smaller work,* $\mathfrak{W}(\tilde{\mathcal{I}}) < \mathfrak{W}(\mathcal{I}(\nu))$, *leads to a larger error,* $\mathfrak{E}(\tilde{\mathcal{I}}) > \mathfrak{E}(\mathcal{I}(\nu))$.

MIMC general analysis framework

Once the shape of $\mathcal{I}$ is determined, we find $\mathcal{I}(\mathrm{TOL})$ by the bias

$$\mathfrak{E}(\mathcal{I}(\mathrm{TOL})) = \sum_{\boldsymbol{\alpha} \notin \mathcal{I}(\mathrm{TOL})} E_{\boldsymbol{\alpha}} \leq (1-\theta)\mathrm{TOL}$$

yielding the corresponding computational work

$$\left(\frac{C_\epsilon}{\theta \mathrm{TOL}}\right)^2 \left(\sum_{\boldsymbol{\alpha} \in \mathcal{I}(\mathrm{TOL})} \sqrt{V_{\boldsymbol{\alpha}} W_{\boldsymbol{\alpha}}}\right)^2 \lesssim \mathrm{TOL}^{-(2+p)}$$

Particular assumptions for MIMC For every $\boldsymbol{\alpha}$, assume

Assumption 1 (Bias) : $E_{\boldsymbol{\alpha}} = |\mathrm{E}[\Delta S_{\boldsymbol{\alpha}}]| \lesssim \prod_{i=1}^{d} \beta_i^{-\alpha_i w_i}$

Assumption 2 (Variance) : $V_{\boldsymbol{\alpha}} = \mathrm{Var}[\Delta S_{\boldsymbol{\alpha}}] \lesssim \prod_{i=1}^{d} \beta_i^{-\alpha_i s_i}$,

Assumption 3 (Work) : $W_{\boldsymbol{\alpha}} = \mathrm{Work}(\Delta S_{\boldsymbol{\alpha}}) \lesssim \prod_{i=1}^{d} \beta_i^{\alpha_i \gamma_i}$,

For positive constants $\gamma_i, w_i, s_i \leq 2w_i$ and for $i = 1 \ldots d$.

Particular optimal index-set for MIMC

In particular, under **Assumptions 1-3**, the optimal index-set can be written (by the profit-thresholding Lemma defining $\mathcal{I}$) as

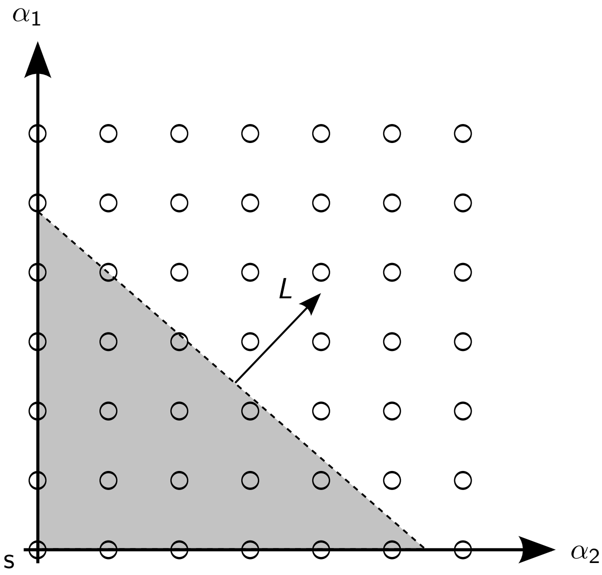
$$\mathcal{I}_{\boldsymbol{\delta}}(L) = \{\boldsymbol{\alpha} \in \mathbb{N}^d : \boldsymbol{\alpha} \cdot \boldsymbol{\delta} = \sum_{i=1}^{d} \boldsymbol{\alpha}_i \delta_i \leq L\}. \tag{5}$$

Here $L \in \mathbb{R}$,

$$\delta_i = \frac{\log(\beta_i)(w_i + \frac{\gamma_i - s_i}{2})}{C_{\boldsymbol{\delta}}}, \quad \text{for all } i \in \{1 \cdots d\},$$

$$\text{and} \qquad C_{\boldsymbol{\delta}} = \sum_{j=1}^{d} \log(\beta_j)(w_j + \frac{\gamma_j - s_j}{2}). \tag{6}$$

Observe that $0 < \delta_i \leq 1$, since $s_i \leq 2w_i$ and $\gamma_i > 0$. Moreover, $\sum_{i=1}^{d} \delta_i = 1$.


α_1
α_2
L
s

MIMC work estimate for particular assumptions

$$\eta = \min_{i\in\{1\cdots d\}} \frac{\log(\beta_i) w_i}{\delta_i}, \quad \zeta = \max_{i\in\{1\cdots d\}} \frac{\gamma_i - s_i}{2w_i}, \quad \mathfrak{z} = \#\{i \in \{1\cdots d\} : \frac{\gamma_i - s_i}{2w_i} = \zeta\}.$$

Theorem (Work estimate with optimal weights)

Let the total-degree index set $\mathcal{I}_\delta(L)$ be given by (5) and (6), taking

$$L = \frac{1}{\eta}\left(\log(\mathrm{TOL}^{-1}) + (\mathfrak{z}-1)\log\left(\frac{1}{\eta}\log(\mathrm{TOL}^{-1})\right) + C\right).$$

Under **Assumptions 1-3**, *the bias constraint in (2) is satisfied asymptotically and the total work, $W(\mathcal{I}_\delta)$, of the MIMC estimator, $\mathcal{A}$, subject to the variance constraint (4) satisfies:*

$$\limsup_{\mathrm{TOL}\downarrow 0} \frac{W(\mathcal{I}_\delta)}{\mathrm{TOL}^{-2-2\max(0,\zeta)}\left(\log\left(\mathrm{TOL}^{-1}\right)\right)^{\mathfrak{p}}} < \infty,$$

where $0 \le \mathfrak{p} \le 3d + 2(d-1)\zeta$ is known and depends on $d, \boldsymbol{\gamma}, \boldsymbol{w}, \boldsymbol{s}$ and $\boldsymbol{\beta}$.

Powers of the logarithmic term

$$\xi = \min_{i \in \{1 \cdots d\}} \frac{2w_i - s_i}{\gamma_i}, \qquad d_2 = \#\{i \in \{1 \cdots d\} : \gamma_i = s_i\},$$

$$\zeta = \max_{i \in \{1 \cdots d\}} \frac{\gamma_i - s_i}{2w_i}, \qquad \mathfrak{z} = \#\{i \in \{1 \cdots d\} : \frac{\gamma_i - s_i}{2w_i} = \zeta\}.$$

Cases for $\mathfrak{p}$:

A) if $\zeta \leq 0$ and $\zeta < \xi$, or $\zeta = \xi = 0$ then $\mathfrak{p} = 2d_2$.

B) if $\zeta > 0$ and $\xi > 0$ then $\mathfrak{p} = 2(\mathfrak{z} - 1)(\zeta + 1)$.

C-D) if $\zeta \geq 0$ and $\xi = 0$ then $\mathfrak{p} = d - 1 + 2(\mathfrak{z} - 1)(\zeta + 1)$.

KAUST

Fully Isotropic Case: Smooth noise case

Assume $w_i = w$, $s_i = 2w$, $\beta_i = \beta$ and $\gamma_i = \gamma$ for all $i \in \{1 \cdots d\}$ and $d \geq 3$. Then the optimal work is

$$\text{Work(MC)} = \mathcal{O}\left(\text{TOL}^{-2-\frac{d\gamma}{w}}\right).$$

$$\text{Work(MLMC)} = \begin{cases} \mathcal{O}\left(\text{TOL}^{-2}\right), & 2w > d\gamma, \\ \mathcal{O}\left(\text{TOL}^{-2}\left(\log\left(\text{TOL}^{-1}\right)\right)^2\right), & 2w = d\gamma, \\ \mathcal{O}\left(\text{TOL}^{-\frac{d\gamma}{w}}\right), & 2w < d\gamma. \end{cases}$$

$$\text{Work(MIMC)} = \begin{cases} \mathcal{O}\left(\text{TOL}^{-2}\right), & 2w > \gamma, \\ \mathcal{O}\left(\text{TOL}^{-2}\left(\log\left(\text{TOL}^{-1}\right)\right)^{3(d-1)}\right), & 2w = \gamma, \\ \mathcal{O}\left(\text{TOL}^{-\frac{\gamma}{w}}\left(\log\left(\text{TOL}^{-1}\right)\right)^{(d-1)(1+\gamma/w)}\right), & 2w < \gamma, \end{cases}$$

Up to a multiplicative logarithmic term, Work(MIMC) is the same as solving just a **one dimensional** *deterministic problem.*

Three dimensional PDE problem description

$$-\nabla \cdot (a(\boldsymbol{x};\omega)\nabla u(\boldsymbol{x};\omega)) = 1 \qquad \text{for } \boldsymbol{x} \in (0,1)^3,$$
$$u(\boldsymbol{x};\omega) = 0 \qquad \text{for } \boldsymbol{x} \in \partial(0,1)^3,$$

where $$a(\boldsymbol{x};\omega) = 1 + \exp\left(2Y_1\Phi_{121}(\boldsymbol{x}) + 2Y_2\Phi_{877}(\boldsymbol{x})\right).$$

Here, Y_1 and Y_2 are i.i.d. uniform random variables in the range $[-1,1]$. We also take

$$\Phi_{ijk}(\boldsymbol{x}) = \phi_i(x_1)\phi_j(x_2)\phi_k(x_3),$$

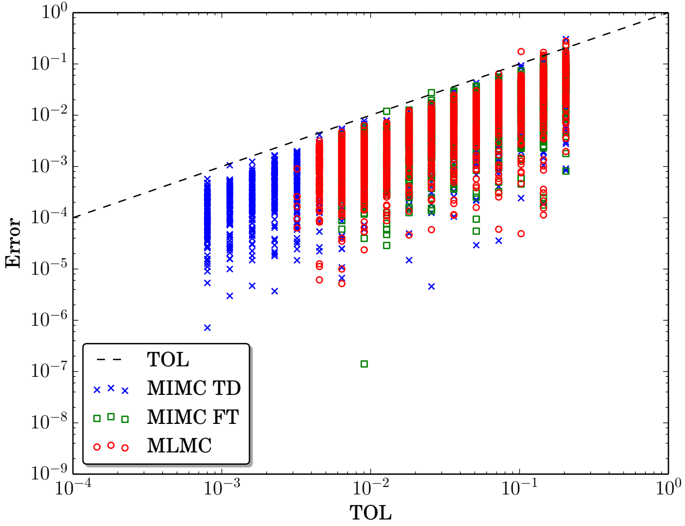
and $$\phi_i(x) = \begin{cases} \cos\left(\frac{i}{2}\pi x\right) & i \text{ is even,} \\ \sin\left(\frac{i+1}{2}\pi x\right) & i \text{ is odd,} \end{cases}$$

Finally, the quantity of interest, S, is

$$S = 100\left(2\pi\sigma^2\right)^{\frac{-3}{2}} \int_{\mathcal{D}} \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x}_0\|_2^2}{2\sigma^2}\right) u(\boldsymbol{x}) d\boldsymbol{x},$$

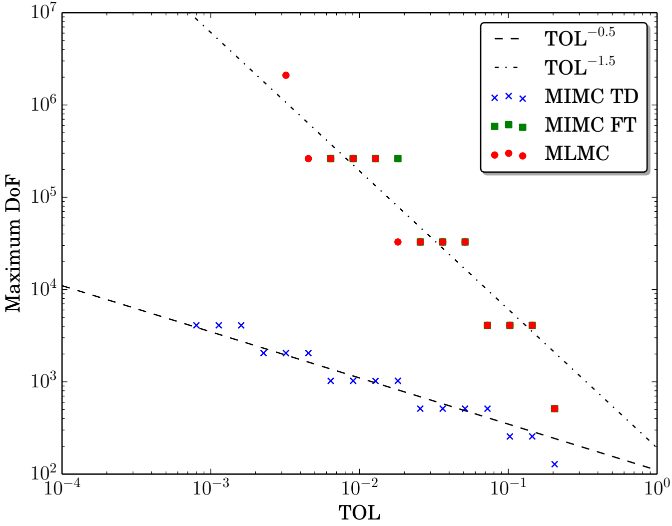
and the selected parameters are $\sigma = 0.04$ and $\boldsymbol{x}_0 = [0.5, 0.2, 0.6]$. We have $\gamma_i = 2$, $w_i = 2$, and $s_i = 4$.

Numerical test: Computational Errors



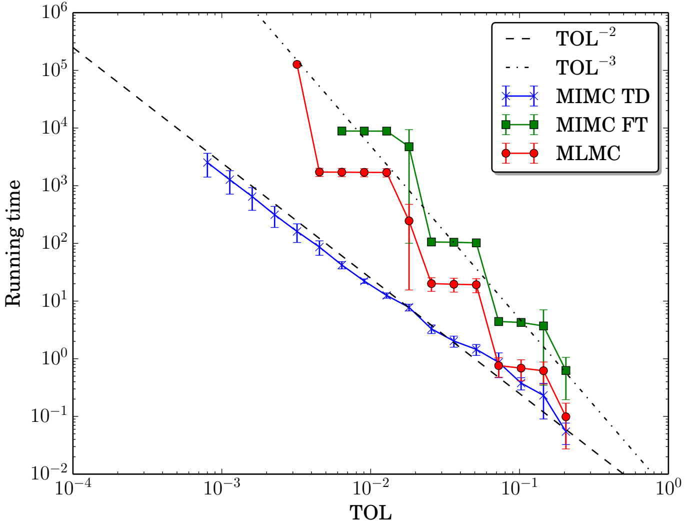
Several runs for different TOL values. Error is satisfied in probability but not over-killed.

Numerical test: Maximum degrees of freedom



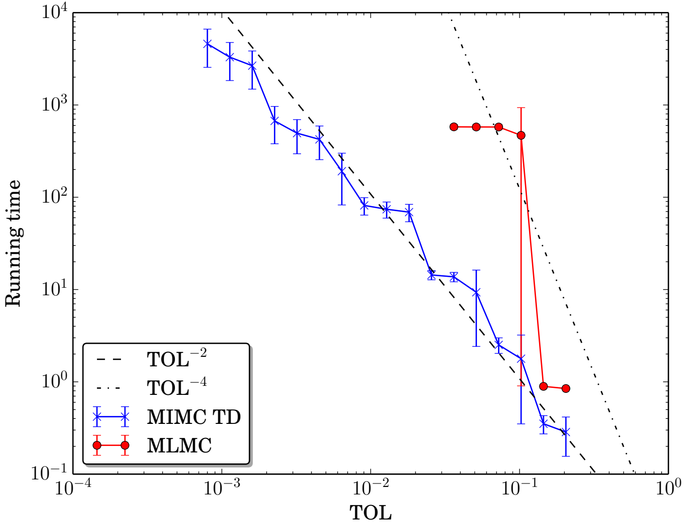
Maximum number of degrees of freedom of a sample PDE solve for different TOL values. This is an indication of required memory.

Numerical test: Running time, 3D problem



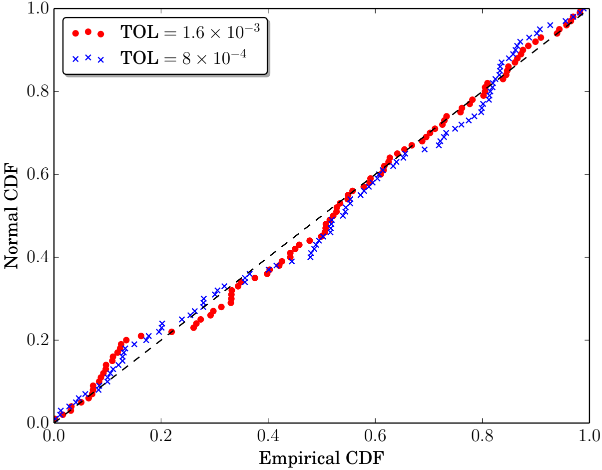
Recall that the work complexity of Monte Carlo is $\mathcal{O}\left(\mathrm{TOL}^{-5}\right)$

Numerical test: Running time, 4D problem



A similar PDE problem with $d=4$. The work complexity of Monte Carlo is $\mathcal{O}\left(\mathrm{TOL}^{-6}\right)$

Numerical test: QQ-plot



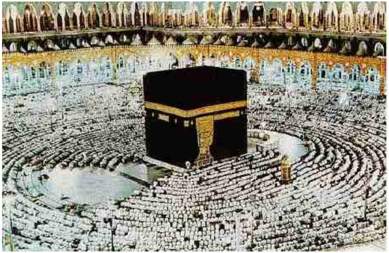
Numerical verification of asymptotic normality of the MIMC estimator. A corresponding statement and proof of the normality of an MIMC estimator can be found in (Haji-Ali et al. 2015).

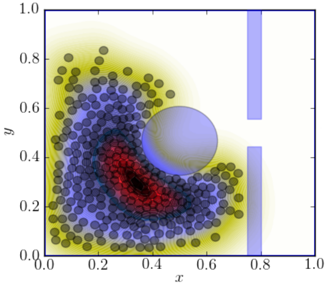
MIMC Conclusions and Extra Points

- MIMC is a generalization of MLMC and performs better, especially in higher dimensions, provided mixed regularity between discretization parameters.
- MIMC general analysis framework, identifying optimal index-set through profit thresholding. Each set of regularity assumptions yield its optimal index-set and related complexity.
- A MIMC direction does not have to be a spatial dimension. It can represent any form of discretization parameter!
  Example: 1-DIM Stochastic Particle Systems, MIMC brings complexity down from $\mathcal{O}(\mathrm{TOL}^{-4})$ to $\mathcal{O}(\mathrm{TOL}^{-2} \log\left(\mathrm{TOL}^{-1}\right)^2)$.
  "A study of Monte Carlo methods for weak approximations of stochastic particle systems in the mean-field", by A. L. Haji Ali and R. T. May 2016.
- Observe, connection to Ensemble Kalman Filter (EnKF): ML-MIMC can compute other statistics, for instance the *covariance*.

Stochastic 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 particles systems approach a mean-field limit as the number of particles increase. Such limits can be useful to understand their complicated phenomena.





Kuramoto oscillator model †

For $p = 1, 2, \ldots, P$

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

$$X_{p|P}(0) = x^0_{p|P}$$

where we are interested in

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

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

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

Kuramoto oscillator model [†]

For $p = 1, 2, \ldots, P$

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

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 $P \uparrow \infty$

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

Kuramoto oscillator model †, Euler-Maruyama

For $p = 1, 2, \dots, P$

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

$$X^{0|N}_{p|P} = x^0_{p|P}$$

where we are interested in

$$\phi^N_P = \frac{1}{P} \sum_{p=1}^{P} \cos\left(X^{N|N}_{p|P} \right),$$

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

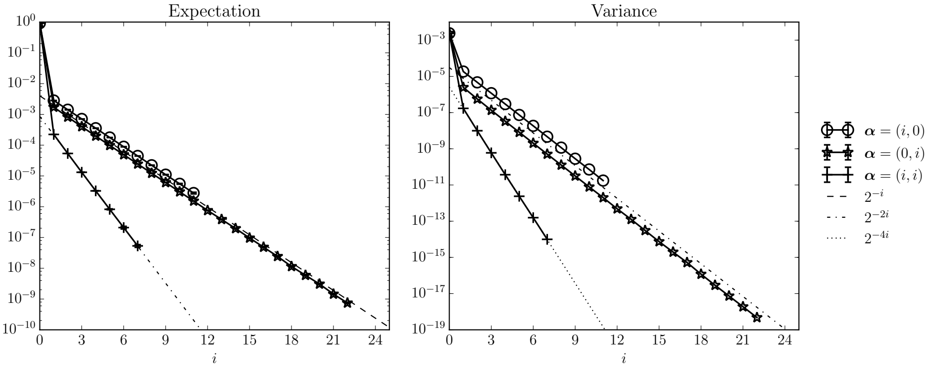
† Y. Kuramoto, Chemical Oscillations, Waves, and Turbulence, Springer, Berlin, 1984.

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 $N_{\alpha_2} = 2^{\alpha_2}$.

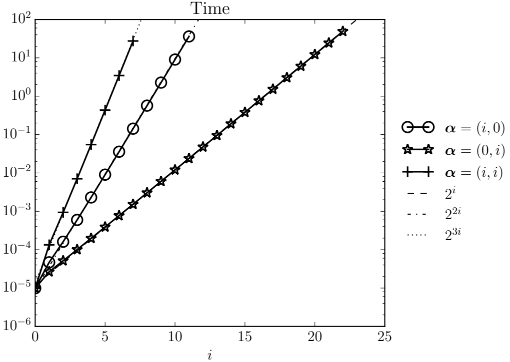


$w_1, w_2 = 1, s_1 = s_2 = 2$

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:**

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

- 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

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

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

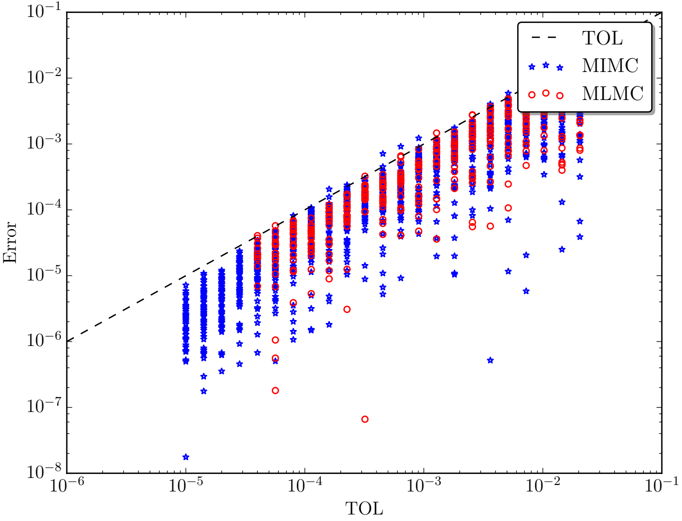
where $\vartheta_p \sim \mathcal{U}(-0.2, 0.2)$. The quantity of interested is

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

for $T = 1$.

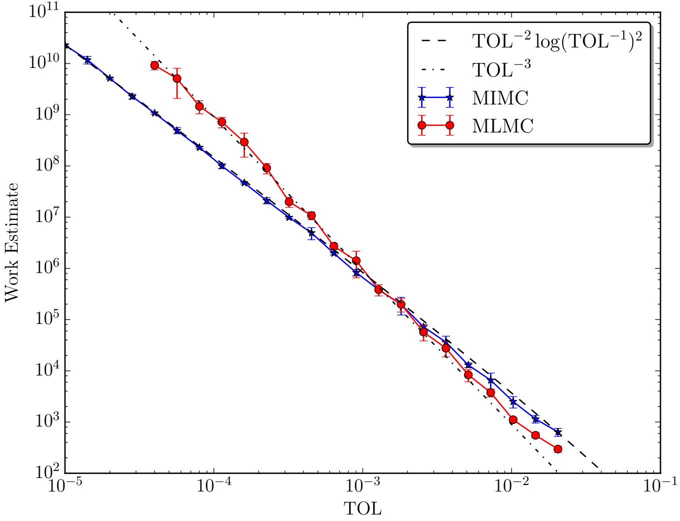
Numerical Example: MIMC vs. MLMC

for $T = 1$.



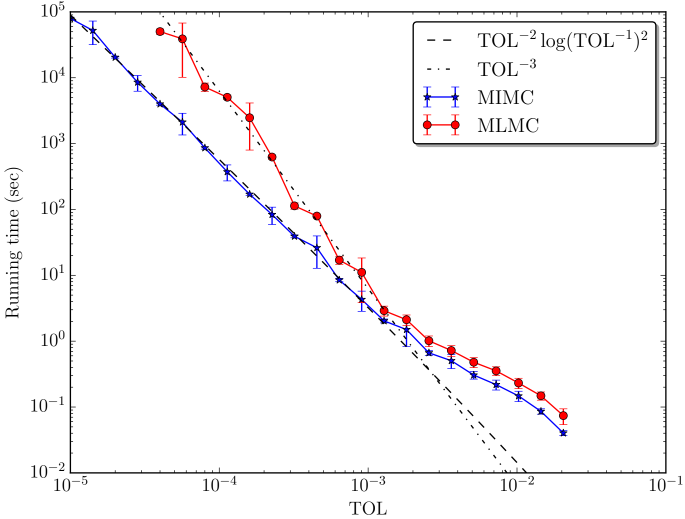
Numerical Example: MIMC vs. MLMC

for $T = 1$.



Numerical Example: MIMC vs. MLMC

for $T = 1$.



Multi Level computation of the Covariance, $Cov[S]$

Let now S be a vector valued output quantity of the solution of a stochastic differential equation, and S_ℓ its approximation based on a level ℓ discretization. Our goal is to approximate the covariance.

Monte Carlo: based on M iid samples, $\{S_{L,i}\}_{i=1}^M$, compute the sample mean and sample covariance

$$E[S_L; M] = \frac{1}{M} \sum_{m=1}^{M} S_{L,i},$$

$$Cov[S_L; M] = \frac{1}{M-1} \sum_{m=1}^{M} (S_{L,i} - E[S_L; M])(S_{L,i} - E[S_L; M])^T$$

Multilevel Monte Carlo: [Bierig-Chernov,2014]

$$Cov_{ML} = \sum_{\ell=0}^{L} \{Cov[S_\ell; M_\ell] - Cov[S_{\ell-1}; M_\ell]\}$$

Observe: both $Cov[S_\ell; M_\ell]$ and $Cov[S_{\ell-1}; M_\ell]$ use the *same noise.*

Filtering problem description

Consider the underlying and unobservable (stochastic) dynamics and observations,

$$\begin{aligned} u_{n+1} &= \Psi(u_n), \\ y_{n+1} &= H u_{n+1} + \gamma_{n+1}, \quad \gamma_{n+1} \sim N(0, \Gamma). \end{aligned}$$

Assume $u_0 \in L^p(\Omega)$ for any $p \geq 1$ and $H \in \mathbb{R}^{k \times d}$.
The observation noise is iid and independent of the noise driving the dynamics.

Objective: Let $Y_n := (y_1, y_2, \ldots, y_n)$ and let Y_n^{obs} be a sequence of *fixed* observations. Construct an efficient method for tracking $u_n | (Y_n = Y_n^{obs})$. That is, approximate

$$\mathrm{E}\left[\phi(u_n) | Y_n = Y_n^{obs}\right]$$

for a given observable $\phi : \mathbb{R}^d \to \mathbb{R}$.

Ensemble Kalman Filtering (Evensen 94)

Predict

1. Compute (numerical solutions of) M particle paths one step forward

$$\widehat{v}_{n+1,i} = \Psi(v_{n,i}, \omega_i) \quad \text{for } i = 1, 2, \ldots, M.$$

2. Compute their sample mean and covariance

$$\widehat{m}_{n+1}^{\mathrm{MC}} = E_M[\widehat{v}_{n+1}]$$
$$\widehat{C}_{n+1}^{\mathrm{MC}} = \mathrm{Cov}_M[\widehat{v}_{n+1}]$$

$$\text{where} \quad E_M[\widehat{v}_{n+1}] := \frac{1}{M} \sum_{i=1}^{M} \widehat{v}_{n+1,i}$$

$$\text{and} \quad \mathrm{Cov}_M[\widehat{v}_{n+1}] := E_M[\widehat{v}_{n+1}\widehat{v}_{n+1}^T] - E_M[\widehat{v}_{n+1}](E_M[\widehat{v}_{n+1}])^T.$$

Ensemble Kalman Filtering II

Update

1. Generate signal observations for the ensemble of particles

$$\tilde{y}_{n+1,i} = y^{obs}_{n+1} + \gamma_{n+1,i} \quad \text{for } i = 1, 2 \ldots, M,$$

   with i.i.d. $\gamma_{n+1,1} \sim N(0, \Gamma)$.

2. Use signal observations to update, for $i = 1, 2 \ldots, M$, particle paths

$$\begin{aligned} v_{n+1,i} &= (I - K^{\mathrm{MC}}_{n+1} H)\widehat{v}_{n+1,i} + K^{\mathrm{MC}}_{n+1} \tilde{y}_{n+1,i}, \\ \text{where} \quad K^{\mathrm{MC}}_{n+1} &= \widehat{C}^{\mathrm{MC}}_{n+1} H^T (H \widehat{C}^{\mathrm{MC}}_{n+1} H^T + \Gamma)^{-1}. \end{aligned}$$

Note: After the first update step, all particles are correlated due to K^{MC}_{n+1}.

Reducing computational cost of EnKF with MLEnKF

Idea: In Multilevel EnKF, we aim to produce similar computational gains wrt EnKF as Multilevel MC does wrt MC. The Multilevel approximation is done to the state covariance!

- H. Hoel, K. J. H. Law, R. T., "Multilevel Ensemble Kalman Filtering". Accepted for publication, SINUM (2016).
- A. Beskos, Ajay Jasra, K. Law, R. T. and Y. Zhou, Multilevel Sequential Monte Carlo Samplers. Submitted, 2015.

Multilevel EnKF (MLEnKF)

Prediction Step

- Compute an ensemble of particle paths on a hierarchy of accuracy levels

$$\widehat{v}^{\ell-1}_{n+1,i} = \Psi^{\ell-1}(v^{\ell-1}_{n,i}, \omega_{\ell,i}), \quad \widehat{v}^{\ell}_{n+1,i} = \Psi^{\ell}(v^{\ell}_{n,i}, \omega_{\ell,i}),$$

  for the levels $\ell = 0, 1, \dots, L$ and $i = 1, 2, \dots, M_\ell$.

- Multilevel approximation of mean and covariance matrices:

$$\widehat{m}^{\mathrm{ML}}_{n+1} = \sum_{\ell=0}^{L} E_{M_\ell}[\widehat{v}^{\ell}_{n+1} - \widehat{v}^{\ell-1}_{n+1}],$$

$$\widehat{C}^{\mathrm{ML}}_{n+1} = \sum_{\ell=0}^{L} \left\{ \mathrm{Cov}_{M_\ell}[\widehat{v}^{\ell}_{n+1}] - \mathrm{Cov}_{M_\ell}[\widehat{v}^{\ell-1}_{n+1}] \right\}$$

Notice the MLMC telescoping properties hold by construction.

Multi Level EnKF update step

Update Step

For $\ell = 0, 1, \ldots, L$ and $i = 1, 2, \ldots, M_\ell$,

$$\begin{aligned}
\tilde{y}^{\ell}_{n+1,i} &= y^{obs}_{n+1} + \gamma^{\ell}_{n+1,i}, \qquad \text{i.i.d. } \gamma^{\ell}_{n+1,i} \sim N(0, \Gamma) \\
v^{\ell-1}_{n+1,i} &= (I - K^{\mathrm{ML}}_{n+1} H)\widehat{v}^{\ell-1}_{n+1,i} + K^{\mathrm{ML}}_{n+1}\tilde{y}^{\ell}_{n+1,i}, \\
v^{\ell}_{n+1,i} &= (I - K^{\mathrm{ML}}_{n+1} H)\widehat{v}^{\ell}_{n+1,i} + K^{\mathrm{ML}}_{n+1}\tilde{y}^{\ell}_{n+1,i},
\end{aligned}$$

$$\text{where} \quad K^{\mathrm{ML}}_{n+1} = \widehat{C}^{\mathrm{ML}}_{n+1} H^{\mathsf{T}} (H \widehat{C}^{\mathrm{ML}}_{n+1} H^{\mathsf{T}} + \Gamma)^{-1}.$$

Beyond MIMC: Multi-Index Stochastic Collocation

- Can we do even better if additional smoothness is available?

[MISC1, 2015] A.-L. Haji-Ali, F. Nobile, L. Tamellini and R. T. "Multi-Index Stochastic Collocation for random PDEs". arXiv:1508.07467. Computers and Mathematics with Applications, Vol. 306, pp. 95–122, July 2016.

[MISC2, 2015] A.-L. Haji-Ali, F. Nobile, L. Tamellini and R. T. "Multi-Index Stochastic Collocation convergence rates for random PDEs with parametric regularity". arXiv:1511.05393v1. Submitted, November 2015.

Idea: Use sparse quadrature to carry the integration in MIMC!

MISC Assumptions

For some strictly positive constant Q_W, g_j, w_i, C_{work} and γ_i for $i = 1 \dots d$ and $j = 1 \dots n$, there holds

$$\left| \boldsymbol{\Delta}^n \left(\boldsymbol{\Delta}^d S_{\boldsymbol{\alpha},\boldsymbol{\tau}} \right) \right| \leq Q_W \left(\prod_{j=1}^{n} \exp(-g_j \tau_j) \right) \left(\prod_{i=1}^{d} \exp(-w_i \alpha_i) \right).$$

$$\text{Work} \left(\boldsymbol{\Delta}^n \left(\boldsymbol{\Delta}^d S_{\boldsymbol{\alpha},\boldsymbol{\tau}} \right) \right) \leq C_{\text{work}} \left(\prod_{j=1}^{n} \tau_j \right) \left(\prod_{i=1}^{d} \exp(\gamma_i \alpha_i) \right).$$

This a simplified presentation that can be easily generalized to nested points.

MISC work estimate

Theorem (Work estimate with optimal weights)

[MISC1, 2015] *Under (our usual) assumptions on the error and work convergence there exists an index-set $\mathcal{I}$ such that*

$$\lim_{\mathrm{TOL}\downarrow 0} \frac{|\mathcal{A}_{MISC}(\mathcal{I}) - \mathrm{E}[S]|}{\mathrm{TOL}} \leq 1$$

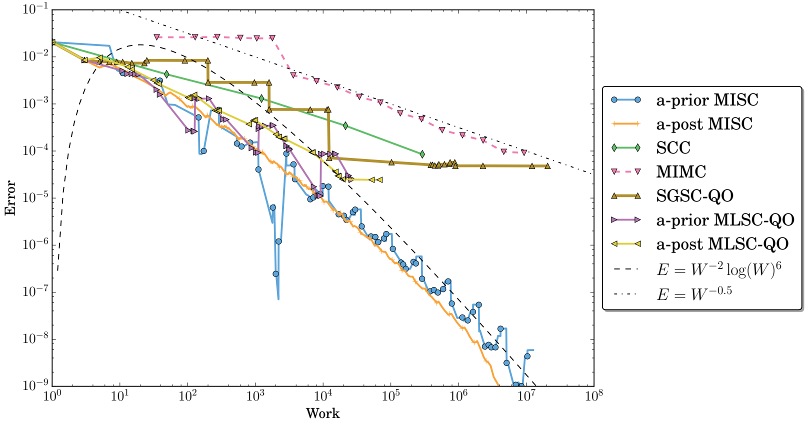
$$\text{and} \quad \lim_{\mathrm{TOL}\downarrow 0} \frac{\mathrm{Work}[\mathcal{A}_{MISC}(\mathcal{I})]}{\mathrm{TOL}^{-\zeta} \left(\log\left(\mathrm{TOL}^{-1}\right)\right)^{(\mathfrak{z}-1)(\zeta+1)}} = C(n,d) < \infty \tag{7}$$

where $\zeta = \max_{i=1}^{d} \frac{\gamma_i}{w_i}$ *and* $\mathfrak{z} = \#\{i = 1, \ldots d \ : \ \frac{w_i}{\gamma_i} = \zeta\}$.

Note that the rate is independent of the number of random variables n. Moreover, d appears only in the logarithmic power.

MISC numerical comparison **[MISC1, 2015]**

Comparison with MIMC and Quasi Optimal (QO) Single & Multilevel Level Sparse Grid Stochastic Collocation



MISC (parametric regularity, $N = \infty$) [MISC2, 2015]

We use MISC to compute on a hypercube domain $B \subset \mathbb{R}^d$

$$\begin{aligned} -\nabla \cdot (a(\boldsymbol{x},\boldsymbol{y})\,\nabla u(\boldsymbol{x},\boldsymbol{y})) &= f(\boldsymbol{x}) \quad \text{in} \quad B \\ u(\boldsymbol{x},\boldsymbol{y}) &= 0 \quad \text{on} \quad \partial B, \end{aligned}$$

where

$$a(\boldsymbol{x},\boldsymbol{y}) = e^{\kappa(\boldsymbol{x},\boldsymbol{y})}, \text{ with } \kappa(\boldsymbol{x},\boldsymbol{y}) = \sum_{j\in\mathbb{N}_+} \psi_j(\boldsymbol{x}) y_j.$$

Here, $\boldsymbol{y}$ are iid uniform and the regularity of a (and hence u) is determined through the decay of the norm of the derivatives of $\psi_j \in C^\infty(\mathcal{B})$. Given the sequences

$$b_{s,j} = \max_{\boldsymbol{s}\in\mathbb{N}^d:|\boldsymbol{s}|\le s} \|D^{\boldsymbol{s}}\psi_j\| L^\infty(B)\,, \qquad j \ge 1,$$

we assume that there exist $0 < p_0 \le p_s < \frac{1}{2}$ s.t. $\{b_{s,j}\}_{j\in\mathbb{N}_+} \in \ell^{p_s}$.

KAUST

Theorem (MISC convergence theorem)

[MISC2, 2015] *Under technical assumptions the profit-based MISC estimator built using Stochastic Collocation over Clenshaw-Curtis points and piecewise multilinear finite elements for solving the deterministic problems, we have, for* $\delta > 0$,

$$\left|\mathrm{E}[S] - \mathcal{A}_{MISC}[S]\right| \leq \tilde{C}_P(\delta)\,\mathrm{Work}[\mathcal{A}_{MISC}[S]]^{-r_{\mathrm{MISC}}+\delta}\,.$$

The rate r_{MISC} *is as follows:*

Case 1 *if* $\frac{\gamma}{r_{\mathrm{FEM}}+\gamma} \geq \frac{p_s}{1-p_s}$, *then* $r_{\mathrm{MISC}} = \frac{r_{\mathrm{FEM}}}{\gamma}$,

Case 2 *if* $\frac{\gamma}{r_{\mathrm{FEM}}+\gamma} \leq \frac{p_s}{1-p_s}$, *then*

$$r_{\mathrm{MISC}} = \left(\frac{1}{p_0} - 2\right)\left(\gamma\frac{p_s - p_0}{r_{\mathrm{FEM}}\,p_0\,p_s} + 1\right)^{-1}.$$

Ideas for proofs in **[MISC2, 2015]**

- Shift theorem: From regularity of a and f to regularity of $u \in H^{1+s}(B) \Rightarrow u \in \mathcal{H}^{1+q}_{mix}(B), \quad$ for $0 < q < s/d$.
- Extend holomorphically $u(\cdot, \boldsymbol{z}) \in H^{1+r}(B)$ on polyellipse $\boldsymbol{z} \in \Sigma_r$ (use p_r summability of $\boldsymbol{b}_r$) to get stochastic rates and estimates for Δ.
- Use weighted summability of knapsack profits to prove convergence rates.

Example: log uniform field with parametric regularity **[MISC2, 2015]**

Here, the regularity of $\kappa = \log(a)$ is determined through $\nu > 0$

$$\kappa(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\boldsymbol{k} \in \mathbb{N}^d} A_{\boldsymbol{k}} \sum_{\boldsymbol{\ell} \in \{0,1\}^d} y_{\boldsymbol{k},\boldsymbol{\ell}} \prod_{j=1}^{d} \left(\cos\left(\frac{\pi}{L} k_j x_j\right)\right)^{\ell_j} \left(\sin\left(\frac{\pi}{L} k_j x_j\right)\right)^{1-\ell_j},$$

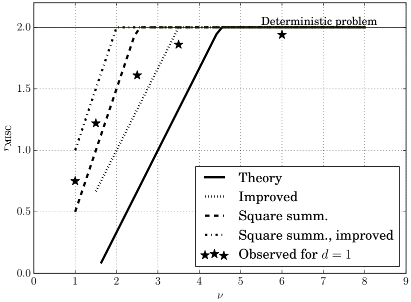
where the coefficients $A_{\boldsymbol{k}}$ are taken as

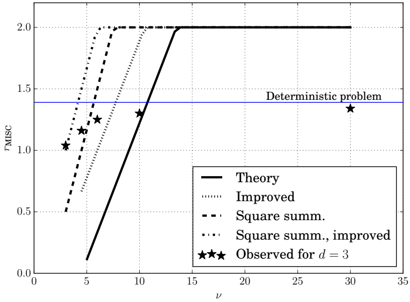
$$A_{\boldsymbol{k}} = \left(\sqrt{3}\right) 2^{\frac{|\boldsymbol{k}|_0}{2}} (1 + |\boldsymbol{k}|^2)^{-\frac{\nu + d/2}{2}}.$$

We have

$$p_0 > \left(\frac{\nu}{d} + \frac{1}{2}\right)^{-1} \quad \text{and} \quad p_s > \left(\frac{\nu - s}{d} + \frac{1}{2}\right)^{-1}.$$

Application of main theorem **[MISC2, 2015]**

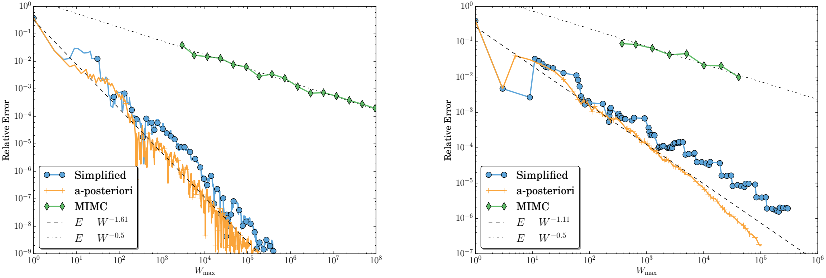




$$Error \propto Work^{-r_{MISC}(\nu,d)}$$

A similar analysis shows the corresponding ν-dependent convergence rates of MIMC but based on ℓ^2 summability of $\boldsymbol{b}_s$ and Fernique type of results.

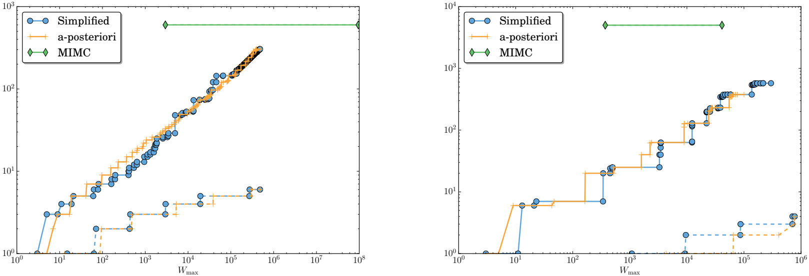
MISC numerical results **[MISC2, 2015]**



Left: $d = 1, \nu = 2.5$. Right: $d = 3, \nu = 4.5$.

$$Error \propto Work^{-r_{MISC}(\nu,d)}$$

MISC numerical results **[MISC2, 2015]**

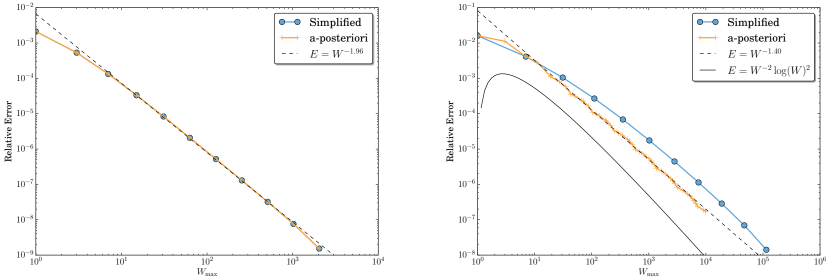


Left: $d = 1, \nu = 2.5$. Right: $d = 3, \nu = 4.5$.

$$Error \propto Work^{-r_{MISC}(\nu,d)}$$

Deterministic runs, numerical results **[MISC2, 2015]**

These plots shows the non-asymptotic effect of the logarithmic factor for $d > 1$ (as discussed in [Thm. 1]**[MISC1, 2015]**) on the linear convergence fit in log-log scale.



Left: $d = 1$. Right: $d = 3$.