Multi-Index Monte Carlo (MIMC) and Multi-Index Stochastic Collocation (MISC)

When sparsity meets sampling

A.–L. Haji–Ali* F. Nobile+ L. Tamellini+ **R. Tempone***

*King Abdullah University of Science and Technology (KAUST), Saudi Arabia.
+Ecole Polytechnique Fédérale de Lausanne, Switzerland.
sri-uq.kaust.edu.sa, stochastic_numerics.kaust.edu.sa

February 22, 2016, Data Assimilation and Inverse Problems EQUIP/AIB Meeting, Warwick, UK

KAUST

Motivational Example: Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and $\mathcal{D} = \prod_{i=1}^{d} [0, D_i]$ for $D_i \subset \mathbb{R}_+$ be a hypercube 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.

Motivational Example: Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and $\mathcal{D} = \prod_{i=1}^{d} [0, D_i]$ for $D_i \subset \mathbb{R}_+$ be a hypercube domain in $\mathbb{R}^d$.
The solution $u : \mathcal{D} \times \Omega \to \mathbb{R}$ here solves almost surely (a.s.) the following equation:

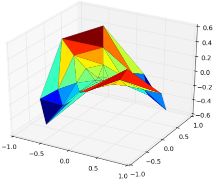
$$\begin{aligned} -\nabla \cdot (a(\boldsymbol{x};\omega)\nabla u(\boldsymbol{x};\omega)) &= f(\boldsymbol{x};\omega) \qquad \text{for } \boldsymbol{x} \in \mathcal{D}, \\ u(\boldsymbol{x};\omega) &= 0 \qquad \text{for } \boldsymbol{x} \in \partial\mathcal{D}. \end{aligned}$$

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

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

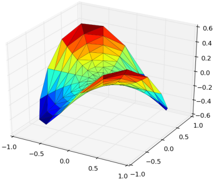
for $\boldsymbol{x}_0 \in \mathcal{D}$ and $\sigma > 0$.

Numerical Approximation



We assume we have an approximation of u (FEM, FD, FV, ...) based on discretization parameters h_i for $i = 1 \ldots d$. Here

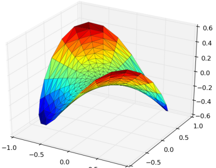
$$h_i = h_{i,0}\, \beta_i^{-\alpha_i},$$



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 $\boldsymbol{\alpha}$.



Monte Carlo complexity analysis

Recall the Monte Carlo method and its error splitting:
$\mathrm{E}[\Psi(u(\boldsymbol{y}))] - \frac{1}{M}\sum_{m=1}^{M} \Psi(u_h(\boldsymbol{y}(\omega_m))) = \mathcal{E}^{\Psi}(h) + \mathcal{E}_h^{\Psi}(M)$ with

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

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

The last approximation is motivated by the Central Limit Theorem.
Assume: computational work for each $u(\boldsymbol{y}(\omega_m))$ is $\mathcal{O}(h^{-d\gamma})$.

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

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

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

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

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

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

The resulting complexity (error versus computational work) is then

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

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

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

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

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

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

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

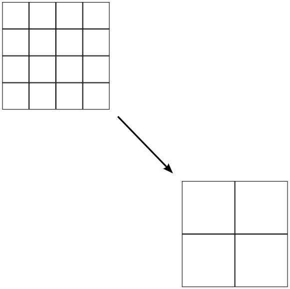
Variance reduction: MLMC

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!

$$\mathrm{Var}[A_{\mathrm{MLMC}}] = \frac{1}{M_0}\mathrm{Var}[S_{\mathbf{0}}] + \sum_{\ell=1}^{L} \frac{1}{M_\ell}\mathrm{Var}[\Delta S_\ell] \leq \mathrm{TOL}^2.$$



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

Classical assumptions for MLMC

For every ℓ, we assume the following:

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

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

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

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

$$\mathrm{Work}(\mathrm{MLMC}) = \sum_{\ell=0}^{L} M_\ell \,\mathrm{Work}(\widetilde{\Delta} S_\ell) \lesssim \sum_{\ell=0}^{L} M_\ell \,\beta^{d\gamma\ell},$$

Example: Our smooth linear elliptic PDE example approximated with Multilinear piecewise continuous FEM:

$$2w = s = 4 \text{ and } 1 \leq \gamma \leq 3.$$

MLMC Computational Complexity

We choose the number of levels to bound the bias

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

and choose the samples $(M_\ell)_{\ell=0}^{L}$ optimally to bound $\mathrm{Var}[\mathcal{A}_{\mathrm{MLMC}}] \lesssim \mathrm{TOL}^2$, then the optimal work satisfies (Giles et al., 2008, 2011):

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

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

Questions related to MLMC

- How to choose the mesh hierarchy $\boldsymbol{h}_\ell$? **[H-ASNT, 2015]**
- How to efficiently and reliably estimate V_ℓ? How to find the correct number of levels, L? **[CH-ASNT, 2015]**
- 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, Accepted (2015).

[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).

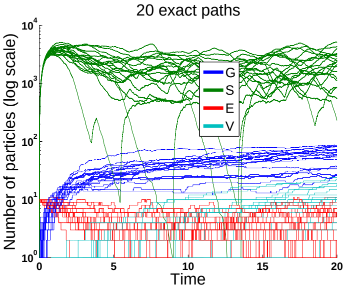
[H-ANT, 2015] A.-L. Haji-Ali, F. Nobile, and R. T. "Multi-Index Monte Carlo: When Sparsity Meets Sampling". arXiv:1405.3757, Numerische Mathematik, Accepted (2015).

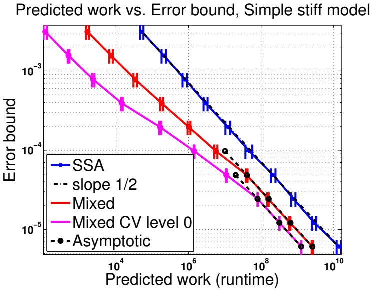
Time adaptivity for MLMC in Itô SDEs:

- 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. Häppölä, and R. T. To appear in MC and Q-MC Methods 2014, Springer Verlag, (2016).

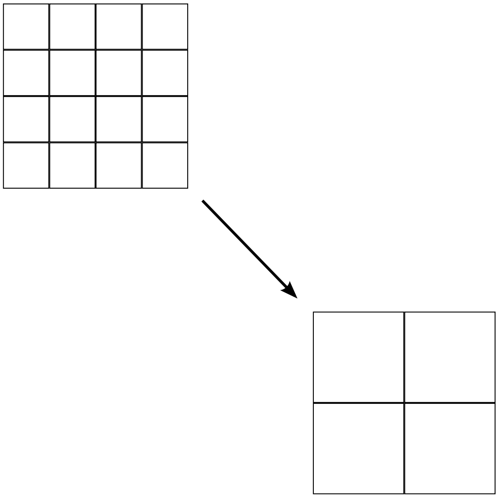
Hybrid MLMC for Stochastic Reaction Networks

- 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*. Submitted, (2014).
- Multilevel drift-implicit tau-leap, by C. Ben Hammouda, A. Moraes and R. T. *arXiv:1512.00721*. Submitted (2015).

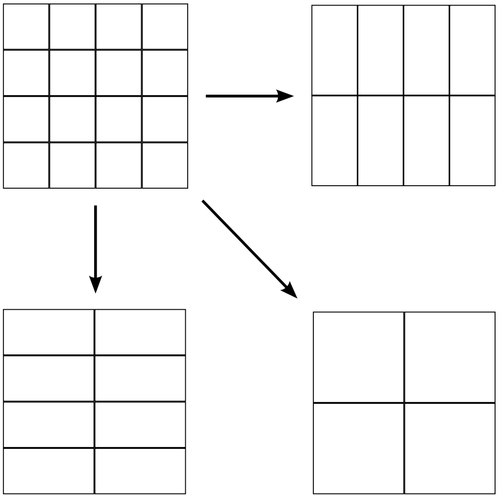




Variance reduction: MLMC



Variance reduction: Further potential



MIMC Estimator

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

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

MIMC Estimator

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

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

Then the MIMC estimator can be written as

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

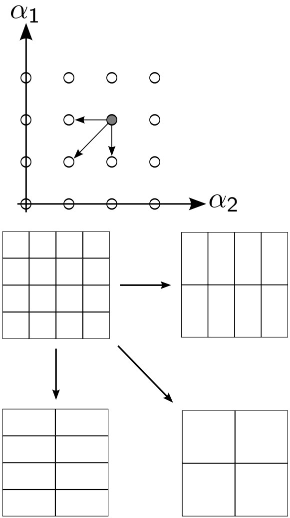
for some *properly chosen* index set $\mathcal{I} \subset \mathbb{N}^d$ and samples $(M_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathcal{I}}$.

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

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

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:

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

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

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

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

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.

Assumptions for MIMC

For every $\boldsymbol{\alpha}$, we assume the following

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 $i = 1 \ldots d$.

$$\mathrm{Work}(\mathrm{MIMC}) = \sum_{\boldsymbol{\alpha} \in \mathcal{I}} M_{\boldsymbol{\alpha}} W_{\boldsymbol{\alpha}} \lesssim \sum_{\boldsymbol{\alpha} \in \mathcal{I}} M_{\boldsymbol{\alpha}} \left(\prod_{i=1}^{d} \beta_i^{\alpha_i \gamma_i} \right).$$

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)

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

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

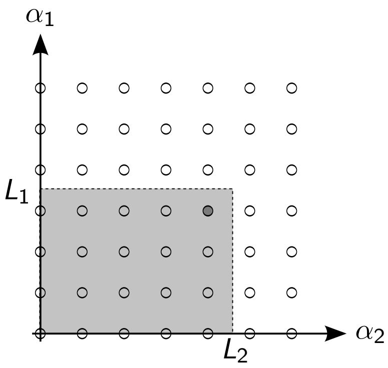
$$\mathrm{Work}(\mathcal{I}) \leq C_\epsilon^2 \theta^{-2} \mathrm{TOL}^{-2} \left(\sum_{\boldsymbol{\alpha} \in \mathcal{I}} \sqrt{V_{\boldsymbol{\alpha}} W_{\boldsymbol{\alpha}}} \right)^2 + \underbrace{\sum_{\boldsymbol{\alpha} \in \mathcal{I}} W_{\boldsymbol{\alpha}}}_{\text{Min. cost of } \mathcal{I}}.$$

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

An obvious choice of $\mathcal{I}$ is the Full Tensor index-set

$$\mathcal{I}(\boldsymbol{L}) = \{\boldsymbol{\alpha} \in \mathbb{N}^d : \alpha_i \leq L_i \text{ for } i \in \{1 \cdots d\}\}$$

for some $\boldsymbol{L} \in \mathbb{R}^d$.

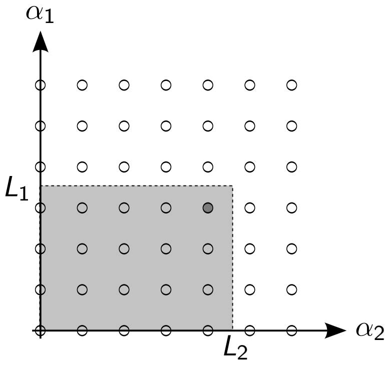


An obvious choice of $\mathcal{I}$ is the Full Tensor index-set

$$\mathcal{I}(\boldsymbol{L}) = \{\boldsymbol{\alpha} \in \mathbb{N}^d : \alpha_i \leq L_i \text{ for } i \in \{1 \cdots d\}\}$$

for some $\boldsymbol{L} \in \mathbb{R}^d$.

It turns out, unsurprisingly, that Full Tensor (FT) index-sets impose restrictive conditions on the weak rates w_i and yield sub-optimal complexity rates.



Question: How do we find optimal index set $\mathcal{I}$ for MIMC?
Then the MIMC work depends only on $\mathcal{I}$ and our goal is to solve

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

We assume that the work of MIMC is *not* dominated by the work to compute a single sample corresponding to each $\boldsymbol{\alpha}$. Then, minimizing equivalently $\sqrt{Work(\mathcal{I})}$, the previous optimization problem can be recast into a knapsack problem with profits defined for each multi-index $\boldsymbol{\alpha}$. The corresponding profit is

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

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, $\mathfrak{E}(\tilde{\mathcal{I}}) > \mathfrak{E}(\mathcal{I}(\nu))$.

Defining the optimal index-set for MIMC

In particular, under **Assumptions 1-3**, the optimal index-set can be written as

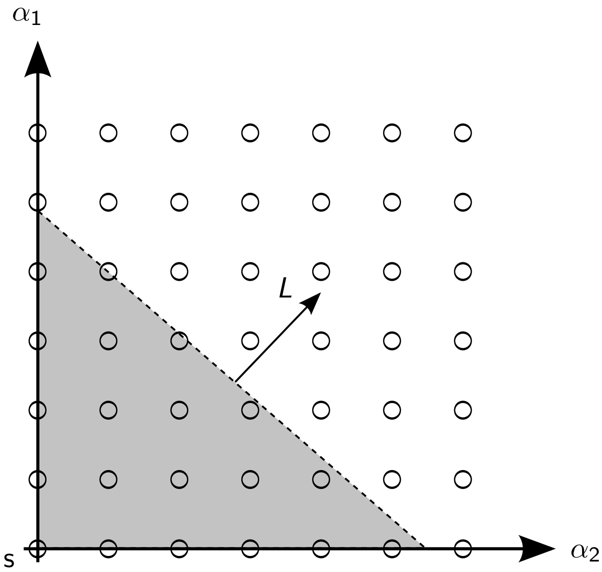
$$\mathcal{I}_{\boldsymbol{\delta}}(L) = \{\boldsymbol{\alpha} \in \mathbb{N}^{d} : \boldsymbol{\alpha} \cdot \boldsymbol{\delta} = \sum_{i=1}^{d} \boldsymbol{\alpha}_i \delta_i \leq L\}. \quad (5)$$

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

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

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

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
L
s
α_2

MIMC work estimate

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

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 $0 \leq \mathfrak{p} \leq 3d + 2(d-1)\zeta$ *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 $\mathfrak{p} = d - 1 + 2(\mathfrak{z} - 1)(\zeta + 1)$.

Fully Isotropic Case: Rough noise case

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

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

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

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

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

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

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

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

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

MIMC: Case with a single worst direction

Recall $\zeta = \max_{i\in\{1\cdots d\}} \frac{\gamma_i - s_i}{2w_i}$ and $\mathfrak{z} = \#\{i \in \{1 \cdots d\} : \frac{\gamma_i - s_i}{2w_i} = \zeta\}$.
In the special case when $\zeta > 0$ and $\mathfrak{z} = 1$, i.e. when the directions are dominated by a single "worst" direction with the maximum difference between the work rate and the rate of variance convergence. In this case, the value of L becomes

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

and MIMC with a TD index-set achieves a better rate for the computational complexity, namely $\mathcal{O}\left(\mathrm{TOL}^{2-2\zeta}\right)$. In other words, the logarithmic term disappears from the computational complexity.

Observe: TD-MIMC with a single worst direction has the same rate of computational complexity as a **one-dimensional** MLMC along that single direction.

Problem description

We test our methods on a three-dimensional, linear elliptic PDE with variable, smooth, stochastic coefficients. The problem is isotropic and we have

$$\gamma_i = 2,$$

$$w_i = 2,$$

and

$$s_i = 4$$

as $\mathrm{TOL} \to 0$.

Problem description

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

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

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

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

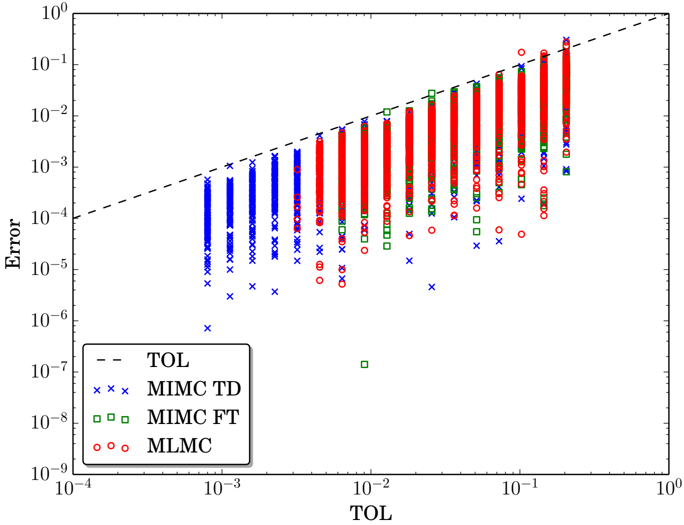
$$
\text{and} \qquad \phi_i(x) = \begin{cases} \cos\left(\frac{i}{2}\pi x\right) & i \text{ is even}, \\ \sin\left(\frac{i+1}{2}\pi x\right) & i \text{ is odd}, \end{cases}
$$

Finally, the quantity of interest, S, is

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

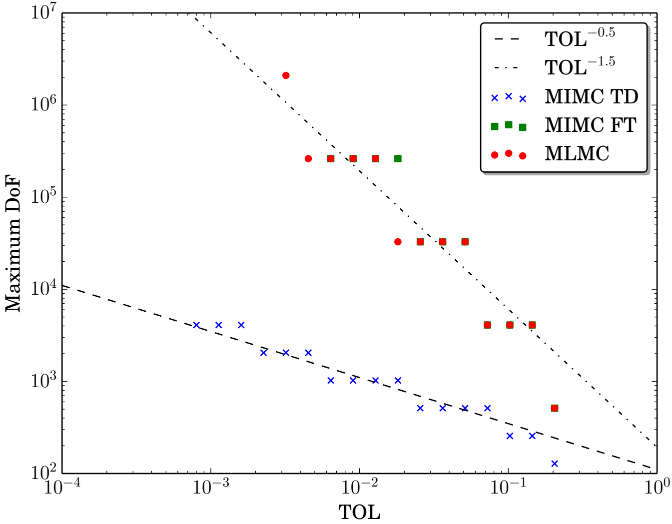
and the selected parameters are $\sigma = 0.04$ and $\boldsymbol{x}_0 = [0.5, 0.2, 0.6]$.

Numerical test: Computational Errors



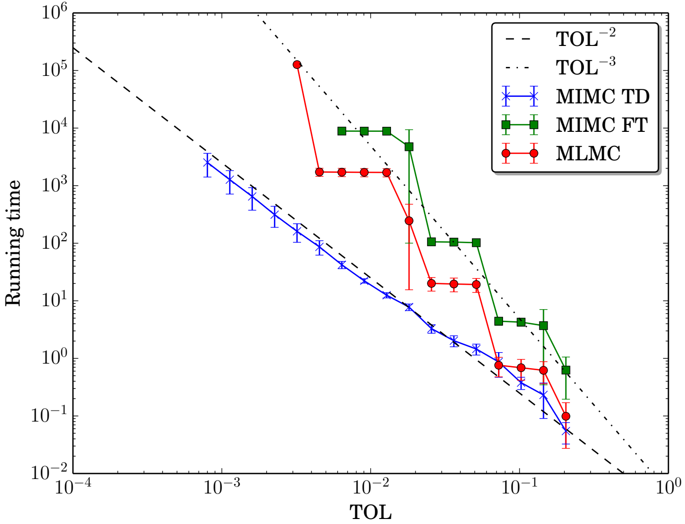
Several runs for different TOL values. Error is satisfied in probability but not over-killed.

Numerical test: Maximum degrees of freedom



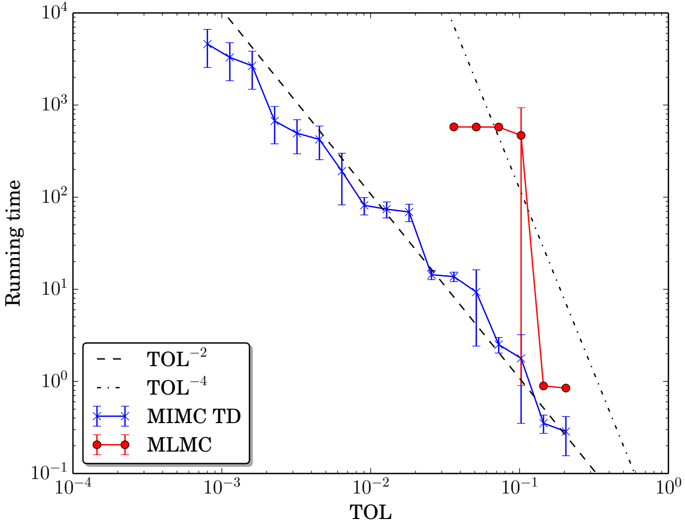
Maximum number of degrees of freedom of a sample PDE solve for different TOL values. This is an indication of required memory.

Numerical test: Running time, 3D problem



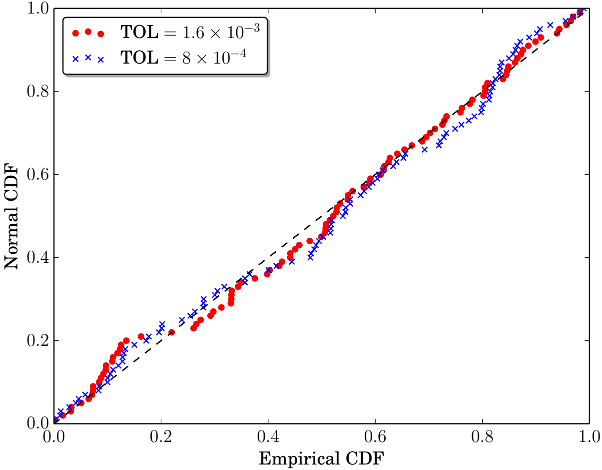
Recall that the work complexity of MC is $\mathcal{O}\left(\mathrm{TOL}^{-5}\right)$

Numerical test: Running time, 4D problem



A similar PDE problem with $d=4$.

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. 2014).

Conclusions and Extra Points

- MIMC is a generalization of MLMC and performs better, especially in higher dimensions.
- For optimal rate of computational complexity, MIMC requires mixed regularity between discretization parameters.
- MIMC may have better complexity rates when applied to non-isotropic problems, for example problems with a single worst direction.
- A different set of regularity assumptions would yield a different optimal index-set and related complexity results.
- A 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 $\mathcal{O}(\mathrm{TOL}^{-5})$ to $\mathcal{O}(\mathrm{TOL}^{-2} \log\left(\mathrm{TOL}^{-1}\right)^2)$

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. Submitted, August 2015.

[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!

Preliminary: Interpolation

Let $\Gamma \subseteq \mathbb{R}$, $\mathbb{P}^q(\Gamma)$ be the space of polynomials of degree q over Γ, and $\mathcal{C}^0(\Gamma)$ the set of real-valued continuous functions over Γ. Given m interpolation points $y_1, y_2 \ldots y_m \in \Gamma$ define the one-dimensional Lagrangian interpolant operator $\mathcal{U}^m : \mathcal{C}^0(\Gamma) \to \mathbb{P}^{m-1}(\Gamma)$ as

$$\mathcal{U}^m[u](y) = \sum_{j=1}^{m} u(y^j)\psi_j(y), \qquad \text{where } \psi_j(y) = \prod_{k \neq j} \frac{y - y_k}{y_j - y_k}.$$

Preliminary: Interpolation

Let $\Gamma \subseteq \mathbb{R}$, $\mathbb{P}^q(\Gamma)$ be the space of polynomials of degree q over Γ, and $\mathcal{C}^0(\Gamma)$ the set of real-valued continuous functions over Γ. Given m interpolation points $y_1, y_2 \dots y_m \in \Gamma$ define the one-dimensional Lagrangian interpolant operator $\mathcal{U}^m : \mathcal{C}^0(\Gamma) \to \mathbb{P}^{m-1}(\Gamma)$ as

$$\mathcal{U}^m[u](y) = \sum_{j=1}^{m} u(y^j)\psi_j(y), \qquad \text{where } \psi_j(y) = \prod_{k \neq j} \frac{y - y_k}{y_j - y_k}.$$

Then, given a tensor grid $\bigotimes_{j=1}^{n}\{y_1^j, y_2^j \dots y_{m_j}^j \in \Gamma_n\}$ with cardinality $\prod_{j=1}^{n} m_j$, the n-variate lagrangian interpolant $\mathcal{U}^{\boldsymbol{m}}[u] : \mathcal{C}^0(\Gamma) \to \mathbb{P}^{\boldsymbol{m}-\boldsymbol{1}}(\Gamma)$ can be written as

$$\begin{aligned}\mathcal{U}^{\boldsymbol{m}}[u](\boldsymbol{y}) &= \left(\mathcal{U}^{m_1} \otimes \cdots \otimes \mathcal{U}^{m_n}\right)[u](\boldsymbol{y}) \\ &= \sum_{i_1=1}^{m_1} \cdots \sum_{i_n=1}^{m_n} u(y_1^{i_1}, \dots y_n^{i_n}) \cdot \left(\psi_{i_1}(y_1) \cdots \psi_{i_n}(y_n)\right).\end{aligned}$$

Preliminary: Stochastic Collocation

It is also straightforward to deduce a n-variate quadrature rule from the lagrangian interpolant. In particular, if $(\Gamma, \mathcal{B}(\Gamma), \rho)$ is a probability space, where $\mathcal{B}(\Gamma)$ is the Borel σ-algebra and $\rho(\boldsymbol{y})d\boldsymbol{y}$ is a probability measure, the expected value of the tensor interpolant can be computed as

$$\mathrm{E}[\mathcal{U}^{\boldsymbol{m}}[u](\boldsymbol{y})] = \sum_{i_1=1}^{m_1} \cdots \sum_{i_1=1}^{m_n} u(y_1^{i_1}, \ldots y_n^{i_n}) \cdot \mathrm{E}[\psi_{i_1}(y_1) \cdots \psi_{i_n}(y_n)].$$

Preliminary: Stochastic Collocation

It is also straightforward to deduce a n-variate quadrature rule from the lagrangian interpolant. In particular, if $(\Gamma, \mathcal{B}(\Gamma), \rho)$ is a probability space, where $\mathcal{B}(\Gamma)$ is the Borel σ-algebra and $\rho(\boldsymbol{y})d\boldsymbol{y}$ is a probability measure, the expected value of the tensor interpolant can be computed as

$$\mathrm{E}[\mathcal{U}^{\boldsymbol{m}}[u](\boldsymbol{y})] = \sum_{i_1=1}^{m_1} \cdots \sum_{i_1=1}^{m_n} u(y_1^{i_1}, \ldots y_n^{i_n}) \cdot \mathrm{E}[\psi_{i_1}(y_1) \cdots \psi_{i_n}(y_n)].$$

Moreover, if $(y_1, \ldots, y_n)$ are jointly independent then the probability density function ρ factorizes, i.e. $\rho(\boldsymbol{y}) = \prod_{n=1}^{N} \rho_n(y_n)$, and there holds

$$\mathrm{E}[\psi_{i_1}(y_1) \cdots \psi_{i_n}(y_n)] = \prod_{n=1}^{N} \mathrm{E}[\psi_{i_n}(y_n)]$$

MISC Main Operator

Assume S is a function of n random variables. Instead of estimating $\mathrm{E}[S_{\boldsymbol{\alpha}}]$ using Monte Carlo we can use Stochastic Collocation with $\boldsymbol{\tau} \in \mathbb{N}^n$ points, as follows

$$\mathrm{E}[S_{\boldsymbol{\alpha}}] = S_{\boldsymbol{\alpha},\boldsymbol{\tau}}(\boldsymbol{Y}) = \mathcal{U}^{(\tau_1,\ldots,\tau_n)}[S_{\boldsymbol{\alpha}}](\boldsymbol{Y}).$$

MISC Main Operator

Assume S is a function of n random variables. Instead of estimating $\mathrm{E}[S_{\boldsymbol{\alpha}}]$ using Monte Carlo we can use Stochastic Collocation with $\boldsymbol{\tau} \in \mathbb{N}^n$ points, as follows

$$\mathrm{E}[S_{\boldsymbol{\alpha}}] = S_{\boldsymbol{\alpha},\boldsymbol{\tau}}(\boldsymbol{Y}) = \mathcal{U}^{(\tau_1,\ldots,\tau_n)}[S_{\boldsymbol{\alpha}}](\boldsymbol{Y}).$$

Then we can define the Delta operators along the stochastic and deterministic dimensions

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

$$\Delta_j^n S_{\boldsymbol{\alpha},\boldsymbol{\tau}} = \begin{cases} S_{\boldsymbol{\alpha},\boldsymbol{\tau}} - S_{\boldsymbol{\alpha},\boldsymbol{\tau}-\boldsymbol{e}_j}, & \text{if } \tau_j > 0, \\ S_{\boldsymbol{\alpha},\boldsymbol{\tau}} & \text{if } \tau_j = 0, \end{cases}$$

MISC Estimator

We use these operator to define the following Multi-index Stochastic Collocation (MISC) estimator of $\mathrm{E}[S]$,

$$\mathcal{A}_{\mathsf{MISC}}(\nu) = \mathrm{E}\left[\sum_{(\boldsymbol{\alpha},\boldsymbol{\tau})\in\mathcal{I}} \boldsymbol{\Delta}^n \left(\boldsymbol{\Delta}^d S_{\boldsymbol{\alpha},\boldsymbol{\tau}}\right)\right] = \sum_{(\boldsymbol{\alpha},\boldsymbol{\tau})\in\mathcal{I}} c_{\boldsymbol{\alpha},\boldsymbol{\tau}} \mathrm{E}[S_{\boldsymbol{\alpha},\boldsymbol{\tau}}],$$

for some index set $\mathcal{I} \in \mathbb{N}^{d+n}$.

MISC Estimator

We use these operator to define the following Multi-index Stochastic Collocation (MISC) estimator of $\mathrm{E}[S]$,

$$\mathcal{A}_{\mathsf{MISC}}(\nu) = \mathrm{E}\left[\sum_{(\boldsymbol{\alpha},\boldsymbol{\tau})\in\mathcal{I}} \boldsymbol{\Delta}^{n}\left(\boldsymbol{\Delta}^{d} S_{\boldsymbol{\alpha},\boldsymbol{\tau}}\right)\right] = \sum_{(\boldsymbol{\alpha},\boldsymbol{\tau})\in\mathcal{I}} c_{\boldsymbol{\alpha},\boldsymbol{\tau}} \mathrm{E}[S_{\boldsymbol{\alpha},\boldsymbol{\tau}}],$$

for some index set $\mathcal{I} \in \mathbb{N}^{d+n}$.

Question: Optimal choice for $\mathcal{I}$?

MISC Estimator

We use these operator to define the following Multi-index Stochastic Collocation (MISC) estimator of $\mathrm{E}[S]$,

$$\mathcal{A}_{\mathsf{MISC}}(\nu) = \mathrm{E}\left[\sum_{(\boldsymbol{\alpha},\boldsymbol{\tau})\in\mathcal{I}} \boldsymbol{\Delta}^{n}\left(\boldsymbol{\Delta}^{d} S_{\boldsymbol{\alpha},\boldsymbol{\tau}}\right)\right] = \sum_{(\boldsymbol{\alpha},\boldsymbol{\tau})\in\mathcal{I}} c_{\boldsymbol{\alpha},\boldsymbol{\tau}} \mathrm{E}[S_{\boldsymbol{\alpha},\boldsymbol{\tau}}],$$

for some index set $\mathcal{I} \in \mathbb{N}^{d+n}$.

Question: Optimal choice for $\mathcal{I}$?
Can be found computationally using the knapsack optimization theory we outlined.

MISC Estimator

We use these operator to define the following Multi-index Stochastic Collocation (MISC) estimator of $\mathrm{E}[S]$,

$$\mathcal{A}_{\mathsf{MISC}}(\nu) = \mathrm{E}\left[\sum_{(\boldsymbol{\alpha},\boldsymbol{\tau})\in\mathcal{I}} \boldsymbol{\Delta}^{n}\left(\boldsymbol{\Delta}^{d} S_{\boldsymbol{\alpha},\boldsymbol{\tau}}\right)\right] = \sum_{(\boldsymbol{\alpha},\boldsymbol{\tau})\in\mathcal{I}} c_{\boldsymbol{\alpha},\boldsymbol{\tau}} \mathrm{E}[S_{\boldsymbol{\alpha},\boldsymbol{\tau}}],$$

for some index set $\mathcal{I} \in \mathbb{N}^{d+n}$.

Question: Optimal choice for $\mathcal{I}$?
Can be found computationally using the knapsack optimization theory we outlined.
Question: Can we say something about the rate of work complexity using the optimal $\mathcal{I}$?

MISC Assumptions

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

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

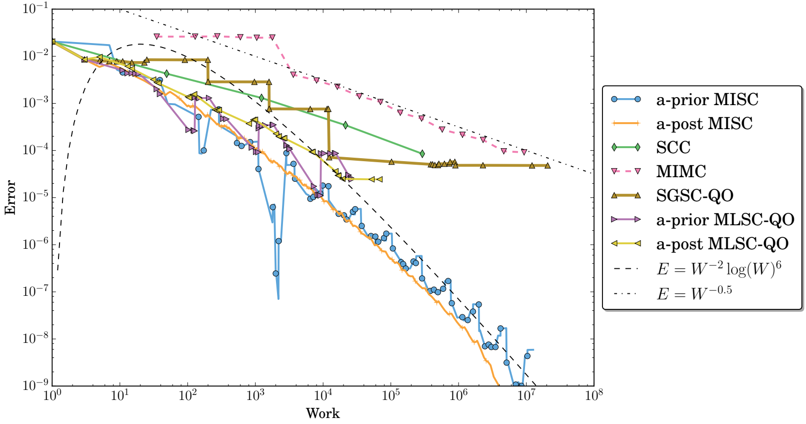
$$\textit{and} \quad \lim_{\mathrm{TOL}\downarrow 0} \frac{\mathrm{Work}[\mathcal{A}_{MISC}(\mathcal{I})]}{\mathrm{TOL}^{-\zeta} \left(\log\left(\mathrm{TOL}^{-1}\right)\right)^{(\mathfrak{z}-1)(\zeta+1)}} = C(n,d) < \infty \tag{7}$$

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 $\mathcal{B} \subset \mathbb{R}^d$

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

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 $\psi_j \in C^\infty(\mathcal{B})$.

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*

$$\left|\mathrm{E}[S] - \mathcal{M}_{\mathcal{I}}[S]\right| \leq \tilde{C}_P \, \mathrm{Work}[\mathcal{M}_{\mathcal{I}}]^{-r_{\mathrm{MISC}}}.$$

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]**

- Given the sequences

$$b_{0,j} = \|\psi_j\|_{L^\infty(\mathcal{B})}, \qquad j \geq 1, \tag{8}$$

$$b_{s,j} = \max_{\boldsymbol{s}\in\mathbb{N}^d : |\boldsymbol{s}|\leq s} \|D^{\boldsymbol{s}}\psi_j\|_{L^\infty(\mathcal{B})}, \qquad j \geq 1, \tag{9}$$

  we assume that there exist $0 < p_0 \leq p_s < \frac{1}{2}$ such that $\{b_{0,j}\}_{j\in\mathbb{N}_+} \in \ell^{p_0}$ and $\{b_{s,j}\}_{j\in\mathbb{N}_+} \in \ell^{p_s}$,
- Shift theorem: From regularity of a and f to regularity of $u \in H^{1+s}(\mathcal{B}) \Rightarrow u \in \mathcal{H}^{1+q}_{mix}(\mathcal{B})$, for $0 < q < s/d$.
- Extend holomorphically $u(\cdot, \boldsymbol{z}) \in H^{1+r}(\mathcal{B})$ 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 $\nu > 0$

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

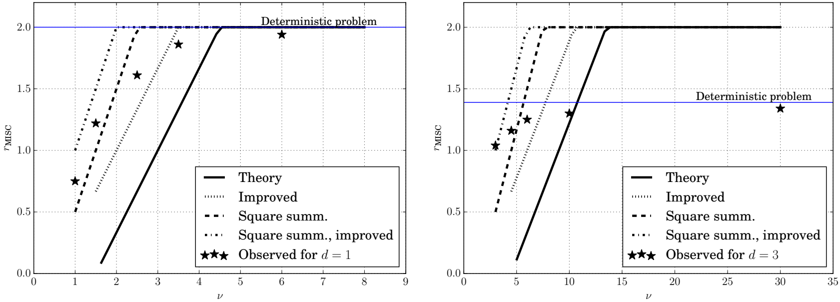
where the coefficients $A_{\boldsymbol{k}}$ are taken as

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

We have

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

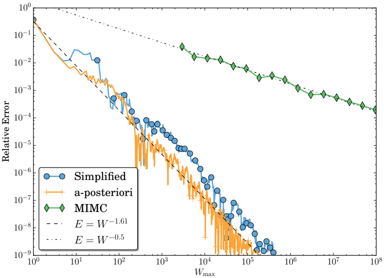
Application of main theorem **[MISC2, 2015]**

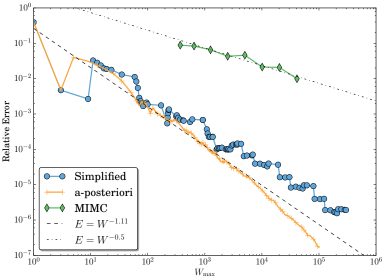


$$Error \propto Work^{-r_{MISC}(\nu, d)}$$

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]**

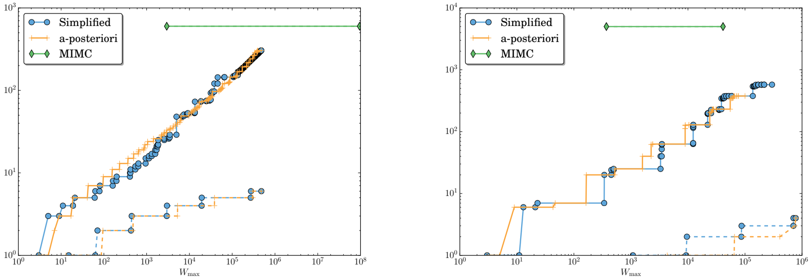




Left: $d = 1, \nu = 2.5$. Right: $d = 3, \nu = 4.5$.

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

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

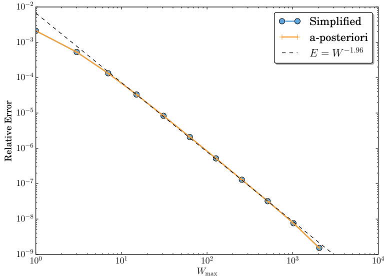


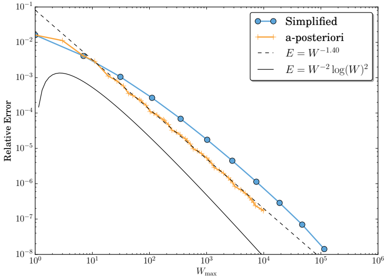
Left: $d = 1, \nu = 2.5$. Right: $d = 3, \nu = 4.5$.

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

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.

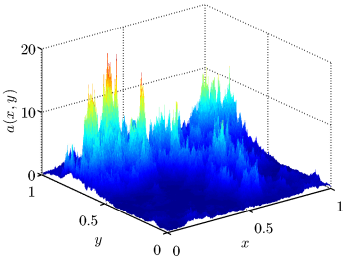


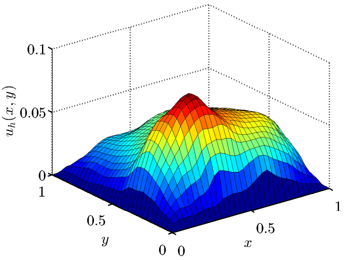


Left: $d = 1$. Right: $d = 3$.

Error Estimation for PDEs with rough stochastic random coefficients

- E. J. Hall, H. Hoel, M. Sandberg, A. Szepessy and R. T. "Computable error estimates for finite element approximations of elliptic partial differential equations with lognormal data", Submitted, 2015.





$$-\nabla \cdot a\nabla u = f, \; a \in C^{1/2-\epsilon}(D)$$

More References: Inverse Probl. and Experimental Design

- H. Hoel, K. J. H. Law, R. T., Multilevel ensemble Kalman filtering. Submitted, 2015.
- F. Bisetti, D. Kim, O. Knio, Q. Long, and R. T. Optimal Bayesian Experimental Design for Priors of Compact Support with Application to Shock-Tube Experiments for Combustion Kinetics. Accepted, Int. J. for Num. Met. in Engineering, 2016.
- C. Bayer, A. Moraes, R. T., and P. Vilanova, An Efficient Forward-Reverse Expectation-Maximization Algorithm for Statistical Inference in Stochastic Reaction Networks. Accepted, Stochastic Analysis and Applications, 2015.
- A. Beskos, Ajay Jasra, K. Law, R. T. and Y. Zhou, Multilevel Sequential Monte Carlo Samplers. Submitted, 2015.
- I. Babuska, Z. Sawlan, M. Scavino, B. Szabo and R. T. Bayesian inference and model comparison for metallic fatigue data. Accepted, Computers and Mathematics with Applications, 2016.
- F. Ruggeri, Z. Sawlan, M. Scavino, R. T. A hierarchical Bayesian setting for an inverse problem in linear parabolic PDEs with noisy boundary conditions. Submitted, 2014.
- Q. Long, M. Motamed and R. T. Fast Bayesian Optimal Experimental Design for Seismic Source Inversion, by CMAME, vol. 291, pp. 123-145, 2015.
- Q. Long, M. Scavino, R. T. and S. Wang. A Laplace method for under determined Bayesian experimental design, CMAME, Vol. 285, pp. 849–876, March 2015.
- Q. Long, M. Scavino, R. T. and S. Wang. Fast Estimation of Expected Information Gains for Bayesian Experimental Designs Based on Laplace Approximations. CMAME, 259(1), pp. 24–39, 2013.

More References: more on forward problems

- F. Nobile, L. Tamellini, and R. T., Convergence of quasi-optimal sparse-grid approximation of Hilbert-space-valued functions: application to random elliptic PDEs, Accepted, Num. Mat., 2015.
- F. Nobile, L. Tamellini, R. T. and F. Tesei, An adaptive sparse grid algorithm for elliptic PDEs with lognormal diffusion coefficient. Accepted, Lecture Notes in Comp. Sci. and Eng. 2015.
- G. Migliorati, F. Nobile and R. T., Convergence estimates in probability and in expectation for discrete least squares with noisy evaluations in random points, Journal of Multivariate Analysis, 2015
- G. Malenova, M. Motamed, O. Runborg, R. T., A Sparse Stochastic Collocation Technique for High Frequency Wave Propagation with Uncertainty, Submitted July 2015.

References

- K.A. Cliffe, M.B. Giles, R. Scheichl, A.L. Teckentrup, "Multilevel Monte Carlo Methods and Applications to Elliptic PDEs with Random Coefficients", Computing and Visualization in Science, 14(1):3-15 (2011).
- A.-L. Haji-Ali, F. Nobile, and R. T. "Multi-Index Monte Carlo: When Sparsity Meets Sampling", arXiv preprint arXiv:1405.3757 (2014).