Optimal mesh hierarchies in Multilevel Monte Carlo methods

Nathan Collier[1] Abdul–Lateef Haji–Ali[2] Fabio Nobile[4]
Erik von Schwerin[5] Raúl Tempone[2,3]

[1] Oak Ridge National Lab, Climate Change Science Institute (CCSI), Environmental Sciences Division.
[2] Applied Mathematics and Computational Sciences, KAUST, Thuwal, Saudi Arabia.
[3] Research Center on Uncertainty Quantification in Computational Science and Engineering at KAUST.
[4] MATHICSE-CSQI, EPF de Lausanne, Switzerland.
[5] University of Delaware, Newark, USA.

UQAW '16, January 5 – 10, KAUST, Thuwal, KSA

Goal: compute $\mathrm{E}[g]$ to an accuracy TOL with high probability, where $g = g(u)$:

- g is either a bounded linear functional or a Lipschitz functional with respect to u,
- u solves a stochastic equation,

and

- u is approximated by a discretization based numerical method, characterized by a single mesh parameter, h.

Example:

$$-\nabla \cdot (a(\mathbf{x};\omega)\nabla u(\mathbf{x};\omega)) = f(\mathbf{x};\omega) \qquad \text{, for } \mathbf{x} \in \mathcal{D} := [0,1]^d,$$
$$u(\mathbf{x};\omega) = 0 \qquad \text{, for } \mathbf{x} \in \partial\mathcal{D},$$

and

$$g(u) = \int_{\mathcal{D}} \kappa(\mathbf{x}) u(\mathbf{x}) d\mathbf{x},$$

for sufficiently smooth a, f, κ.

MLMC: use a hierarchy of $L+1$ meshes defined by decreasing mesh sizes $\{h_\ell\}_{\ell=0}^{L}$.
Denote the approximation of g using mesh size h_ℓ by g_ℓ.

MLMC estimator:

$$\mathcal{A} = \frac{1}{M_0}\sum_{m=1}^{M_0} g_0(\omega_{0,m}) + \sum_{\ell=1}^{L}\frac{1}{M_\ell}\sum_{m=1}^{M_\ell}\left(g_\ell(\omega_{\ell,m}) - g_{\ell-1}(\omega_{\ell,m})\right).$$

Questions:

- Standard MLMC methods use geometric sequences of step sizes, $h_\ell = h_0\beta^{-\ell}$. Optimal?
- Is it ideal to make the bias and statistical error equally small?

MLMC estimator:

$$\mathcal{A} = \frac{1}{M_0} \sum_{m=1}^{M_0} g_0(\omega_{0,m}) + \sum_{\ell=1}^{L} \frac{1}{M_\ell} \sum_{m=1}^{M_\ell} \left(g_\ell(\omega_{\ell,m}) - g_{\ell-1}(\omega_{\ell,m}) \right).$$

We assume the following models:

$$\left| \mathrm{E}[g_\ell - g] \right| \approx Q_W h_\ell^{q_1}, \tag{1a}$$

$$\mathrm{Var}[g_\ell - g_{\ell-1}] := V_\ell \approx Q_S h_{\ell-1}^{q_2}, \tag{1b}$$

$$\text{Work per sample of level } \ell := W_\ell \approx h_\ell^{-d\gamma}. \tag{1c}$$

for some positive constants Q_W, Q_S, q_1, q_2, d, and γ.

Examples for q_1, q_2:

- $q_1 = q_2 = 1$ for an SDE with Euler-Maruyama approximation.
- In our example: a PDE with smooth random coefficients and for piecewise trilinear finite element approximations we have $q_1 = 2$ and $q_2 = 4$.

Examples for γ:

- $\gamma = 1$ for an SDE with Euler-Maruyama approximation.
- In our PDE example: $d = 3$.
  - The used *Iterative* solver, GMRES, had $\gamma \approx 1$.
  - The used *Direct* solver, MUMPS, had $\gamma \approx 1.5$.
  - A *naive* Gaussian elimination implementation would give $\gamma \approx 3$.

$$\text{Define:} \quad \chi = \frac{q_2}{d\gamma} \quad \text{and} \quad \eta = \frac{q_1}{d\gamma}.$$

PDE example: $\chi \approx 1.33$ (GMRES) and $\chi \approx 0.89$ (MUMPS).

Problem (Optimization of computational work)

Given $L \in \mathbb{N}$ *and* $\theta \in (0, 1)$*, find* $\mathbf{H} = (\{h_\ell\}_{\ell=0}^{L}, \{M_\ell\}_{\ell=0}^{L}) \in \mathbb{R}_+^{L+1} \times \mathbb{R}_+^{L+1}$ *such that*

$$W(\mathbf{H}) = \sum_{\ell=0}^{L} \frac{M_\ell}{h_\ell^{d\gamma}},$$

is minimized while satisfying the constraints

$$\textit{Bias} \leq (1-\theta)\mathrm{TOL},$$

$$\textit{Statistical error} \leq \theta\, \mathrm{TOL}, \qquad \textit{with high probability.}$$

Problem (Optimization of computational work)

Given $L \in \mathbb{N}$ *and* $\theta \in (0,1)$, *find* $\mathbf{H} = (\{h_\ell\}_{\ell=0}^{L}, \{M_\ell\}_{\ell=0}^{L}) \in \mathbb{R}_+^{L+1} \times \mathbb{R}_+^{L+1}$ *such that*

$$W(\mathbf{H}) = \sum_{\ell=0}^{L} \frac{M_\ell}{h_\ell^{d\gamma}},$$

is minimized while satisfying the constraints

$$Q_W h_L^{q_1} \leq (1-\theta)\mathrm{TOL},$$

$$\frac{V_0}{M_0} + Q_S \sum_{\ell=1}^{L} \frac{h_{\ell-1}^{q_2}}{M_\ell} \leq \left(\frac{\theta \mathrm{TOL}}{C_\alpha}\right)^2,$$

for a confidence parameter, C_α, *such that* $\Phi(C_\alpha) = 1 - \frac{\alpha}{2}$*;*
here, $0 < \alpha \ll 1$ *and* Φ *is the CDF of an* $N(0,1)$ *random variable.*

Lemma (On normality of MLMC estimator)

Assume, as usual, i.i.d. samples of the random variable $G_\ell(\omega_{\ell,m})$ *in the MLMC estimator* $\mathcal{A} = \sum_{\ell=0}^{L} \sum_{m=1}^{M_\ell} \frac{G_\ell(\omega_{\ell,m})}{M_\ell}$.
Moreover, assume the family $\{G_\ell\}_{\ell \geq 0}$ *independent.*
Denoting $Y_\ell = |G_\ell - \mathrm{E}[G_\ell]|$ *assume*

$$C_1 \beta^{-q_3 \ell} \leq \mathrm{E}\left[Y_\ell^2\right] \qquad \text{for all } \ell \geq 0,$$

$$\mathrm{E}\left[Y_\ell^{2+\delta}\right] \leq C_2 \beta^{-\tau \ell} \qquad \text{for all } \ell \geq 0,$$

for $\beta > 1$ *and strictly positive constants* C_1, C_2, q_3, δ, *and* τ.
Choose the number of samples on each level M_ℓ *to satisfy, for* $q_2 > 0$ *and a strictly positive sequence* $\{H_\ell\}_{\ell \geq 0}$

$$M_\ell \geq \beta^{-q_2 \ell} \mathrm{TOL}^{-2} H_\ell^{-1} \left(\sum_{\ell=0}^{L} H_\ell \right) \qquad \text{for all } \ell \geq 0.$$

Lemma (On Normality of MLMC estimator – continues)

If, in addition to the above, we choose the number of levels L to satisfy

$$L \leq \max\left(0, \frac{c \log\left(\mathrm{TOL}^{-1}\right)}{\log \beta} + C\right)$$

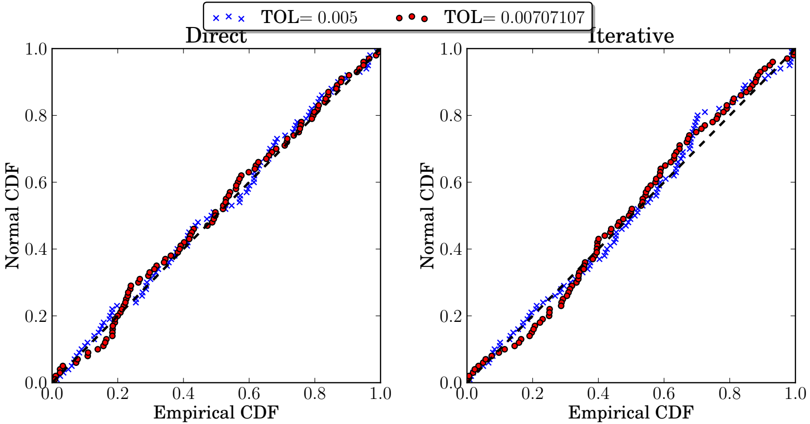
for some constants C, and $c > 0$. Finally, denoting

$$p = (1 + \delta/2)q_3 + (\delta/2)q_2 - \tau,$$

if we have that either $p > 0$ or $c < \delta/p$, then

$$\lim_{\mathrm{TOL}\to 0} \mathrm{P}\left[\frac{\mathcal{A} - \mathrm{E}[\mathcal{A}]}{\sqrt{\mathrm{Var}[\mathcal{A}]}} \leq z\right] = \Phi\left(z\right).$$

Experimental verification of normality of MLMC estimator



This plot shows that the density of the normalized statistical error is well approximated by a standard normal density.

Resulting Optimal Hierarchies

Optimal number of samples

$$M_\ell = \left(\frac{C_\alpha}{\theta \mathrm{TOL}} \right)^2 \sqrt{\frac{V_\ell}{W_\ell}} \left(\sum_{\ell=0}^{L} \sqrt{V_\ell W_\ell} \right). \qquad (2)$$

Given:

- splitting parameter $\theta \in (0, 1)$,
- number of levels $L + 1$,
- mesh sizes h_ℓ defining the variance and work contributions, V_ℓ and W_ℓ, according to the models in (1).

Optimal mesh sizes depend on $\chi \ldots$

Theorem (On the optimal hierarchies when $\chi = 1$)

For any fixed $L \in \mathbb{N}$, with $\chi = 1$, the optimal sequences $\{h_\ell\}_{\ell=0}^{L}$ in Problem 2 are given by

$$h_\ell = \beta^{(L-\ell)} \left(\frac{(1-\theta)\mathrm{TOL}}{Q_W} \right)^{\frac{1}{q_1}}, \qquad \text{for } l = 0, 1, 2, \ldots, L,$$

where the level separation $\beta \in (1, \infty)$ is independent of ℓ (but not TOL), and the optimal choice of the splitting parameter is

$$\theta(1, \eta, L) = \left(1 + \frac{1}{2\eta} \frac{1}{L+1} \right)^{-1} \to 1 \qquad \text{as } L \to \infty.$$

For this case the optimal number of levels, L, satisfies asymptotically

$$\lim_{\mathrm{TOL} \to 0} \frac{L+1}{\log \mathrm{TOL}^{-1}} = \frac{1}{2\eta}.$$

Theorem (On the optimal hierarchies when $\chi \neq 1$)
For any fixed $L \in \mathbb{N}$, with $\chi \neq 1$, the optimal sequences, $\{h_\ell\}_{\ell=0}^{L}$ in Problem 2 are given by

$$h_\ell(\theta, L) = \left(\frac{(1-\theta)\,\mathrm{TOL}}{Q_W}\right)^{\frac{1}{q_1}\frac{1-\chi^{\ell+1}}{1-\chi^{L+1}}} \left(\frac{V_0}{Q_S}\right)^{\frac{1}{d\gamma}\frac{\chi^\ell-\chi^L}{1-\chi^{L+1}}} \cdot \chi^{-\frac{1}{d\gamma}\frac{2}{1-\chi}\left(\frac{\chi^{L+1}-\chi^{\ell+1}}{1-\chi^{L+1}}+\frac{L\left(1-\chi^{\ell+1}\right)-\ell\left(1-\chi^{L+1}\right)}{1-\chi^{L+1}}\right)},$$

where the optimal choice of the splitting parameter is

$$\theta(\chi,\eta,L) = \left(1+\frac{1}{2\eta}\frac{1-\chi}{1-\chi^{L+1}}\right)^{-1} \to \left(1+\frac{1-\min(\chi,1)}{2\eta}\right)^{-1} \qquad \textit{as } L\to\infty.$$

For this case the optimal number of levels, L, satisfies asymptotically

$$\frac{1}{2\eta}\frac{\chi-1}{\log\chi} \le \liminf_{\mathrm{TOL}\to 0}\frac{L+1}{\log\left(\mathrm{TOL}^{-1}\right)} \le \limsup_{\mathrm{TOL}\to 0}\frac{L+1}{\log\left(\mathrm{TOL}^{-1}\right)} \le \frac{\max\{1,\chi\}}{2\eta}\frac{\chi-1}{\log\chi}.$$

Corollary (On the asymptotic work with optimal hierarchies)

For the these optimal hierarchies and using asymptotic upper bounds on L, the total computational complexity

$$\frac{W(\mathbf{H})}{\mathrm{TOL}^{-2\left(1+\frac{1-\chi}{2\eta}\right)}} \to C_0, \qquad \text{as } \mathrm{TOL} \searrow 0 \text{ for } 0 < \chi < 1$$

$$\frac{W(\mathbf{H})}{\mathrm{TOL}^{-2}(\log(\mathrm{TOL}))^2} \to C_1, \qquad \text{as } \mathrm{TOL} \searrow 0 \text{ for } \chi = 1,$$

$$\frac{W(\mathbf{H})}{\mathrm{TOL}^{-2}} \to C_2, \qquad \text{as } \mathrm{TOL} \searrow 0 \text{ for } \chi > 1,$$

with known constants of proportionality,

$$C_0 = C_\alpha^2 Q_S Q_W^{\left\{\frac{1-\chi}{\eta}\right\}} \chi^{\left\{-\frac{2\chi}{1-\chi}\right\}} \left(\frac{1}{2\eta}\right)^2 \left(1+\frac{2\eta}{1-\chi}\right)^{2\left(1+\frac{1-\chi}{2\eta}\right)},$$

$$C_1 = C_\alpha^2 Q_S \exp(2) \left(\frac{1}{2\eta}\right)^2,$$

$$C_2 = C_\alpha^2 V_0^{\left\{\frac{\chi-1}{\chi}\right\}} Q_S^{\left\{\frac{1}{\chi}\right\}} \chi^{2\left\{\frac{\chi}{\chi-1}\right\}} (\chi-1)^{-2}.$$

- For $\chi \neq 1$, the hierarchies are not geometric in general

$$\frac{h_\ell}{h_{\ell+1}} = \chi^{\frac{2}{d\gamma}\left(\frac{1}{\chi-1}+\frac{(L+1)\chi^{\ell+1}}{1-\chi^{L+1}}\right)} \left(\frac{V_0}{Q_S}\right)^{\frac{(1-\chi)\chi^\ell}{d\gamma(1-\chi^{1+L})}} \left(\frac{Q_W}{(1-\theta)\mathrm{TOL}}\right)^{\frac{(1-\chi)\chi^{\ell+1}}{q_1(1-\chi^{L+1})}}.$$

- However, asymptotically as $\mathrm{TOL} \to 0$

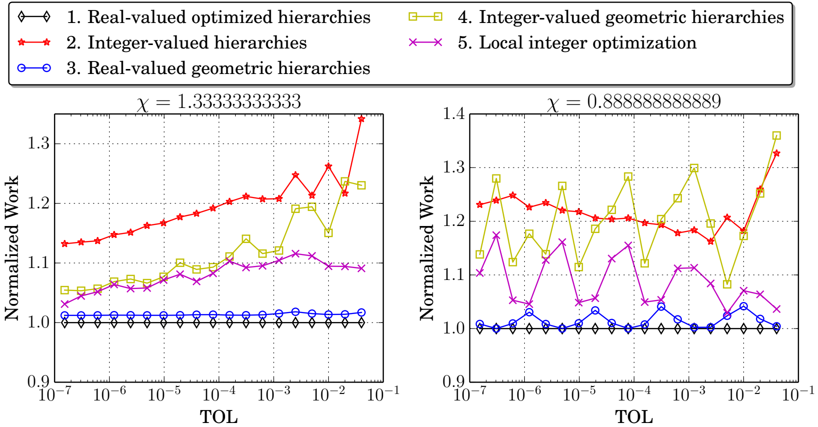
$$\frac{h_\ell}{h_{\ell+1}} \to \chi^{\frac{2}{d\gamma(\chi-1)}} \qquad \text{for most levels } \ell.$$

- Moreover, for geometric hierarchies $h_\ell = h_0\beta^{-\ell}$ (not necessarily nested) with

$$\beta = \begin{cases} \chi^{\frac{2}{d\gamma(\chi-1)}}, & \text{if } \chi \in \mathbb{R}_+ \setminus \{1\}, \\ \exp\left(\frac{2}{q_2}\right), & \text{if } \chi = 1, \end{cases}$$

and $h_0 = \left(\frac{V_0}{Q_S}\right)^{\frac{1}{q_2}} \chi^{\frac{1}{d\gamma(1-\chi)}}$, if $\chi > 1$, the asymptotic computational complexity is the same as the computational complexity of the optimized hierarchies.

Work model – estimated parameters



The number of levels in each hierarchy was chosen using brute-force optimization.

In actual computations the parameters in the work and error models must be estimated.

We propose a Continuation MLMC algorithm that, given a hierarchy, solves the given approximation problem for a sequence of decreasing tolerances, ending with the desired one.

The sequence is chosen such that the total work is dominated by the last iteration.

See also poster

function CMLMC

  Compute with an initial hierarchy.

  Estimate problem parameters $\{V_\ell\}_{\ell=0}^{L}$, Q_S, Q_W, q_1, and q_2.

  Set $i = 0$.

  repeat

    Find optimal integer L.

    Generate hierarchy $\{h_\ell\}_{\ell=0}^{L}$ for TOL_i.

    Using TOL_i and the variance estimates $\{V_\ell\}_{\ell=0}^{L}$ and optimal θ, compute $\{M_\ell\}_{\ell=0}^{L}$ according to (2).

    Compute with the resulting MLMC hierarchy.

    Estimate the parameters, $\{V_\ell\}_{\ell=0}^{L}$, Q_S, Q_W, q_1, and q_2.

    Estimate the total error.

    Set $i = i + 1$

  until Total error estimate is less than TOL

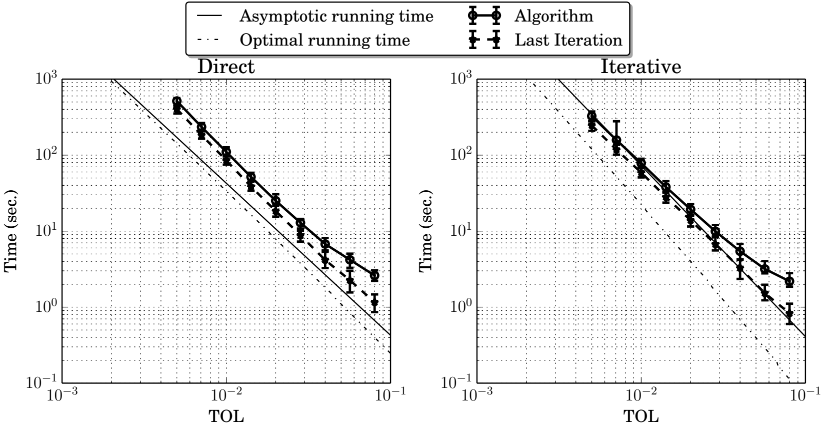
end function

- Estimating q_1, q_2, Q_S, and Q_W is done using a Bayesian approach that allows assuming a prior on q_1 and q_2. Moreover, samples from multiple levels are used to estimate these parameters.
- Estimating V_ℓ for $\ell > 0$ is also done using a Bayesian approach with the models (1) as priors.
- For $r_1 \geq r_2 > 1$ (e.g. $r_1 = 2$ and $r_2 = 1.1$) we choose

$$\mathrm{TOL}_i = \begin{cases} r_1^{i_E - i} r_2^{-1} \mathrm{TOL} & i < i_E, \\ r_2^{i_E - i} r_2^{-1} \mathrm{TOL} & i \geq i_E, \end{cases}$$

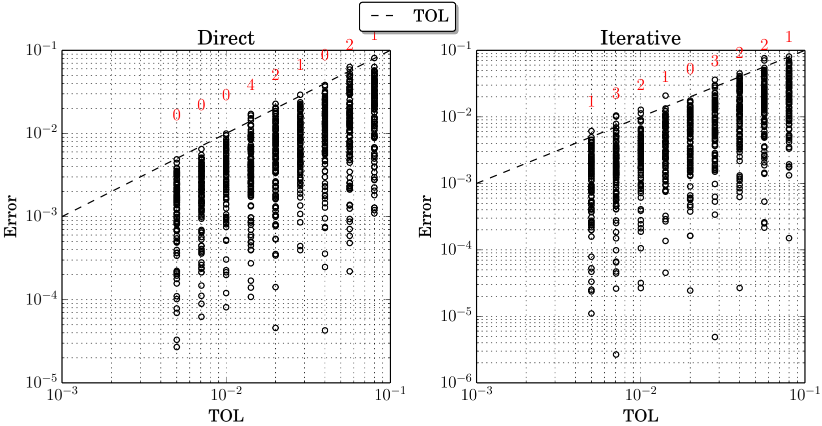
Here $i_E = \left\lfloor \frac{-\log(\mathrm{TOL}) + \log(r_2) + \log(\mathrm{TOL}_{\mathrm{max}})}{\log(r_1)} \right\rfloor$, imposes $\mathrm{TOL}_0 = \mathrm{TOL}_{\mathrm{max}}$ for some maximum tolerance. Moreover $\mathrm{TOL}_{i_E} = r_2^{-1}\mathrm{TOL}$.

Total work of CMLMC – Observed run time



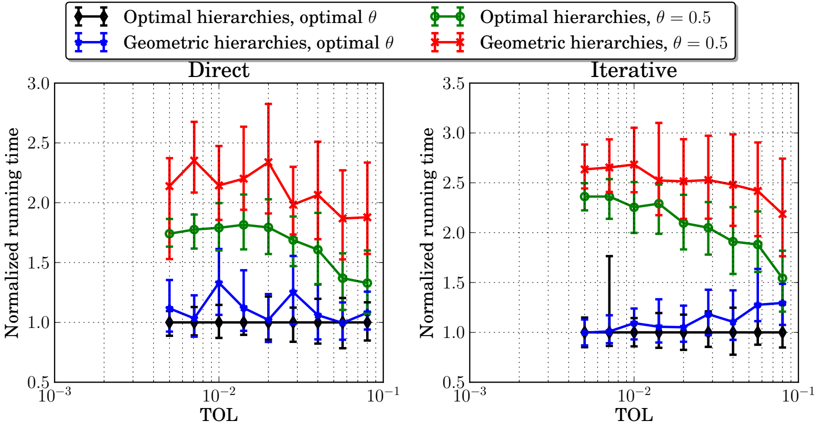
Asymptotic complexities are proportional to $\mathrm{TOL}^{-2.25}$ and TOL^{-2}, respectively.

Accuracy

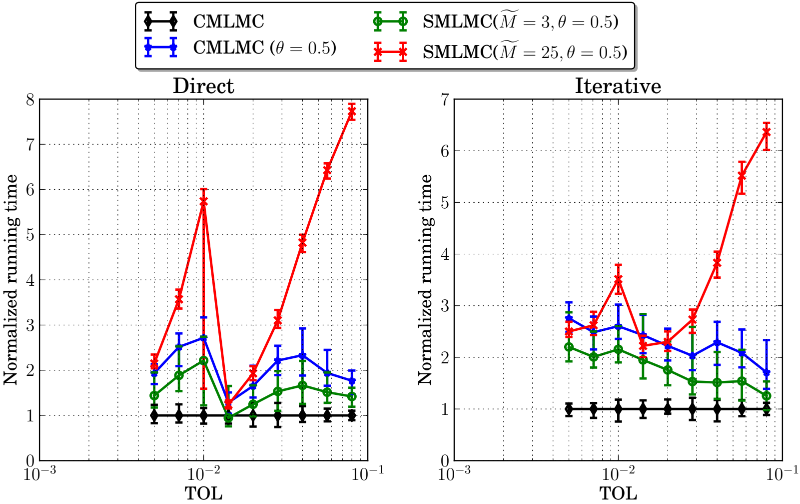


Errors using optimal hierarchies with $C_\alpha = 2$, corresponding to 95% confidence in the error bound.

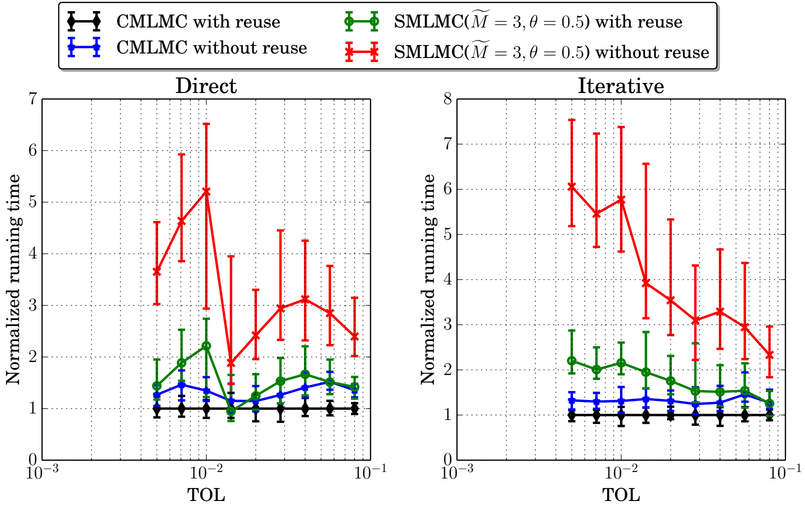
Run time – importance of splitting parameter



Actual run time of the CMLMC algorithm using different tolerance splittings.



Improvement in run time due to better choice of splitting parameter, θ. Normalized by the median run time of CMLMC.



Reusing samples in CMLMC does not significantly improve run time, since the work is dominated by the work of the last iteration.

Conclusions

- We show that geometric hierarchies are near-optimal. Moreover, we derive the computational complexity with known rates and constants.
- Showed normality of MLMC estimator under certain conditions through the use of Lindeberg central limit theorem. We use this in the formulation of our MLMC algorithm and the work optimization problem.
- Computational saving through better tolerance splitting between bias and statistical error contributions.
- A more stable continuation MLMC algorithm with a small overhead. In CMLMC, reusing samples does not introduce significant computational savings.

More information in two posters!

References

N. Collier, A.-L. Haji-Ali, F. Nobile, E. von Schwerin, and R. Tempone, *A continuation multilevel monte carlo algorithm*, BIT Numerical Mathematics, (2014), pp. 1–34.

A.-L. Haji-Ali, F. Nobile, E. von Schwerin, and R. Tempone, *Optimization of mesh hierarchies in multilevel Monte Carlo samplers*, Stochastic Partial Differential Equations: Analysis and Computations, (electronic access 2015 – DOI 10.1007/s40072-015-0049-7).