Adaptive strategies for Multilevel Monte Carlo

Raúl Tempone

http://sri-uq.kaust.edu.sa,
http://stochastic_numerics.kaust.edu.sa

King Abdullah University of Science and Technology **(KAUST)**, Saudi Arabia.

December 18, 2014, **FoCM**, Montevideo, Uruguay

Abstract Setting for MLMC

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

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

Goal: to approximate $\mathrm{E}[g] \in \mathbb{R}$ where $g = \Psi(u)$ for some sufficiently "smooth" u (solution of a random PDE/stochastic differential equation) and a given functional g.
Assumption: We assume we have an approximation of u, say $\bar{u}_h$ (time discretization, FEM, FD, FV, ...) based on discretization parameter h.
Notation: g_h is the approximation of g calculated using a discretization defined by h.

Monte Carlo complexity analysis

Monte Carlo (MC) approximates expectations by sample averages of i.i.d. approximate realizations

$$\mathrm{E}[g] \approx \mathrm{E}[g_h] \approx \frac{1}{M} \sum_{m=1}^{M} g_h(\omega_m).$$

Error splitting: $\mathrm{E}[g] - \frac{1}{M} \sum_{m=1}^{M} g_h(\omega_m) = \mathcal{E}_{Bias}(h) + \mathcal{E}_{Stat}(M)$

$$|\mathcal{E}_{Bias}(h)| = \underbrace{|\mathrm{E}[g] - \mathrm{E}[g_h]|}_{\text{Discretization Error}} \leq C h^w$$

$$|\mathcal{E}_{Stat}(M)| = \underbrace{|\mathrm{E}[g_h] - \frac{1}{M} \sum_{m=1}^{M} g_h(\omega_m)|}_{\text{Statistical Error}} \lesssim c_0 \sqrt{\frac{\mathrm{Var}\,[g_h]}{M}}$$

The last estimate is motivated in probability by a Central Limit Theorem.

Monte Carlo complexity analysis

Let us assume now that the computational work to solve for each sample of g_h is $\mathcal{O}(h^{-d\gamma})$. Thus, we have the following estimates

$$\textbf{Total work}: \quad W \lesssim M\, h^{-d\gamma}$$

$$\textbf{Total error}: \quad |\mathcal{E}_{Bias}(h)| + |\mathcal{E}_{Stat}(M)| \leq C_1 h^w + \frac{C_2}{\sqrt{M}}$$

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

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

The resulting complexity (error versus computational work) is then

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

Multilevel Monte Carlo (MLMC)

Seminal works: [Giles06, Heinrich01].
Construction: Take $\beta > 1$ and for each $\ell = 1, 2, \ldots$ use discretizations with $h_\ell = h_0 \beta^{-\ell}$.
Recall the standard MLMC difference operator

$$\Delta g_\ell = \begin{cases} g_{h_0} & \text{if } \ell = 0, \\ g_{h_\ell} - g_{h_{\ell-1}} & \text{if } \ell > 0. \end{cases}$$

Observe the telescopic identity

$$\mathrm{E}[g] \approx \mathrm{E}[g_{h_L}] = \sum_{\ell=0}^{L} \mathrm{E}[\Delta g_\ell].$$

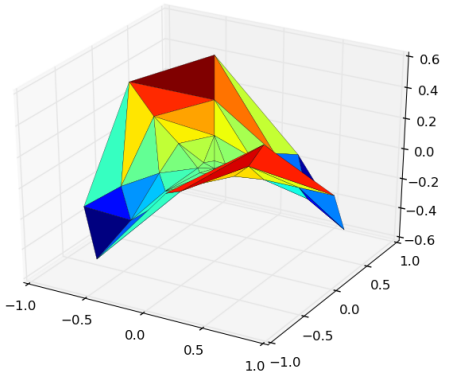
Multilevel Monte Carlo (MLMC)

Seminal works: [Giles06, Heinrich01].

Construction: Take $\beta > 1$ and for each $\ell = 1, 2, \ldots$ use discretizations with $h_\ell = h_0 \beta^{-\ell}$.

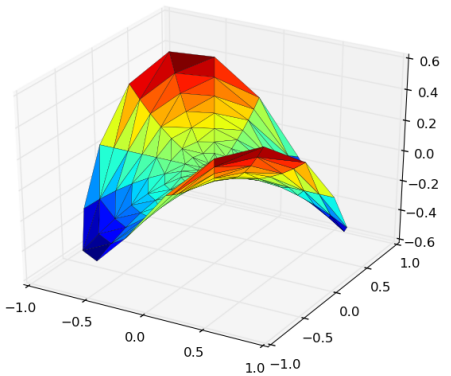
Recall the standard MLMC difference operator

$$\Delta g_\ell = \begin{cases} g_{h_0} & \text{if } \ell = 0, \\ g_{h_\ell} - g_{h_{\ell-1}} & \text{if } \ell > 0. \end{cases}$$



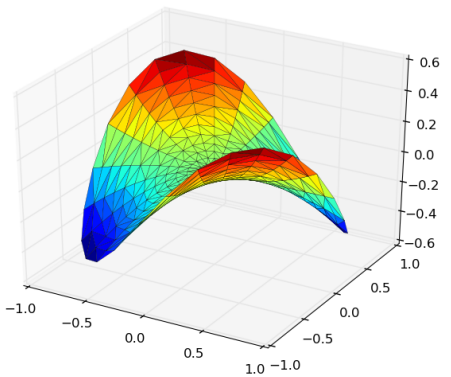
Observe the telescopic identity

$$\mathrm{E}[g] \approx \mathrm{E}[g_{h_L}] = \sum_{\ell=0}^{L} \mathrm{E}[\Delta g_\ell].$$



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

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



Variance reduction: MLMC

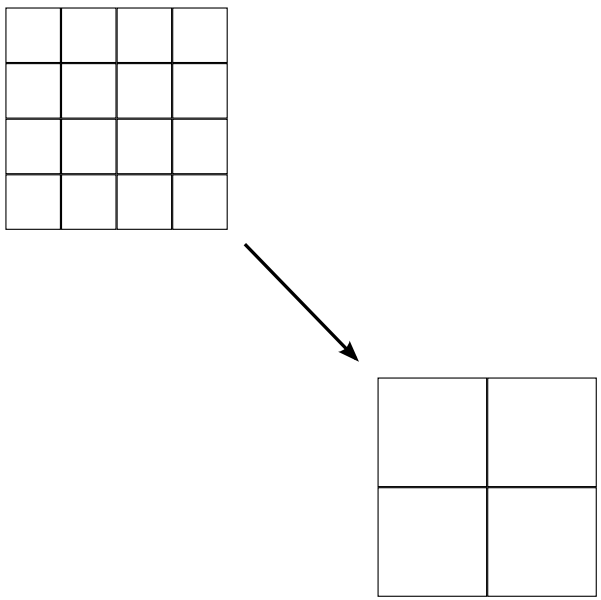
Recall: With Monte Carlo we have to satisfy

$$\mathrm{Var}\left[A_{MC}\right] = \frac{1}{M_L}\mathrm{Var}\left[g_L\right] \approx \frac{1}{M_L}\mathrm{Var}\left[g\right] \leq TOL^2.$$

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

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

(2)



Observe: Level 0 in MLMC is usually determined by *both* stability and accuracy, i.e.

$$\mathrm{Var}\left[\Delta g_1\right] << \mathrm{Var}\left[g_{\mathbf{0}}\right] \approx \mathrm{Var}\left[g\right] < \infty.$$

Assumptions for MLMC

For every level ℓ, assume:

Assumption $\widetilde{1}$ (Bias): $|\mathrm{E}[g - g_\ell]| \leq C\beta^{-w\ell}$,

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

Assumption $\widetilde{3}$ (Work): $\mathrm{Work}(\Delta g_\ell) \leq C\beta^{d\gamma\ell}$,

for positive constants C, γ, w and $s < 2w$.

Assumptions for MLMC

For every level ℓ, assume:

Assumption $\widetilde{1}$ (Bias): $|\mathrm{E}[g - g_\ell]| \leq C\beta^{-w\ell}$,

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

Assumption $\widetilde{3}$ (Work): $\mathrm{Work}(\Delta g_\ell) \leq C\beta^{d\gamma\ell}$,

for positive constants C, γ, w and $s < 2w$.

Example: For the smooth linear elliptic PDE (1) approximated with multilinear piecewise continuous FEM we have:

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

Assumptions for MLMC

For every level ℓ, assume:

Assumption $\widetilde{1}$ (Bias): $|\mathrm{E}[g - g_\ell]| \leq C\beta^{-w\ell}$,

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

Assumption $\widetilde{3}$ (Work): $\mathrm{Work}(\Delta g_\ell) \leq C\beta^{d\gamma\ell}$,

for positive constants C, γ, w and $s < 2w$.

Example: For the smooth linear elliptic PDE (1) approximated with multilinear piecewise continuous FEM we have:

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

Total Work MLMC:

$$\begin{aligned}\mathrm{Work}(\mathrm{MLMC}) &= \sum_{\ell=0}^{L} M_\ell \ \mathrm{Work}(\Delta g_\ell) \\ &\propto TOL^{-2} \left(\sum_{\ell=0}^{L} \sqrt{\mathrm{Work}(\Delta g_\ell)\mathrm{Var}\left[\Delta g_\ell\right]} \right)^2\end{aligned}$$

MLMC Computational Complexity

We choose the number of levels to bound the bias

$$|\mathrm{E}[g - g_L]| \propto \beta^{-Lw} \leq CTOL \quad \Rightarrow \quad L \geq \frac{\log(TOL^{-1}) - \log(C)}{w \log(\beta)},$$

and then the samples $(M_\ell)_{\ell=0}^{L}$ to minimize the total work s.t. (2) [Giles et al., 08,11]:

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

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

Today's questions : How to extend MLMC into non-uniform, adaptive discretization settings? Is it worth it?

Some of the developments on MLMC

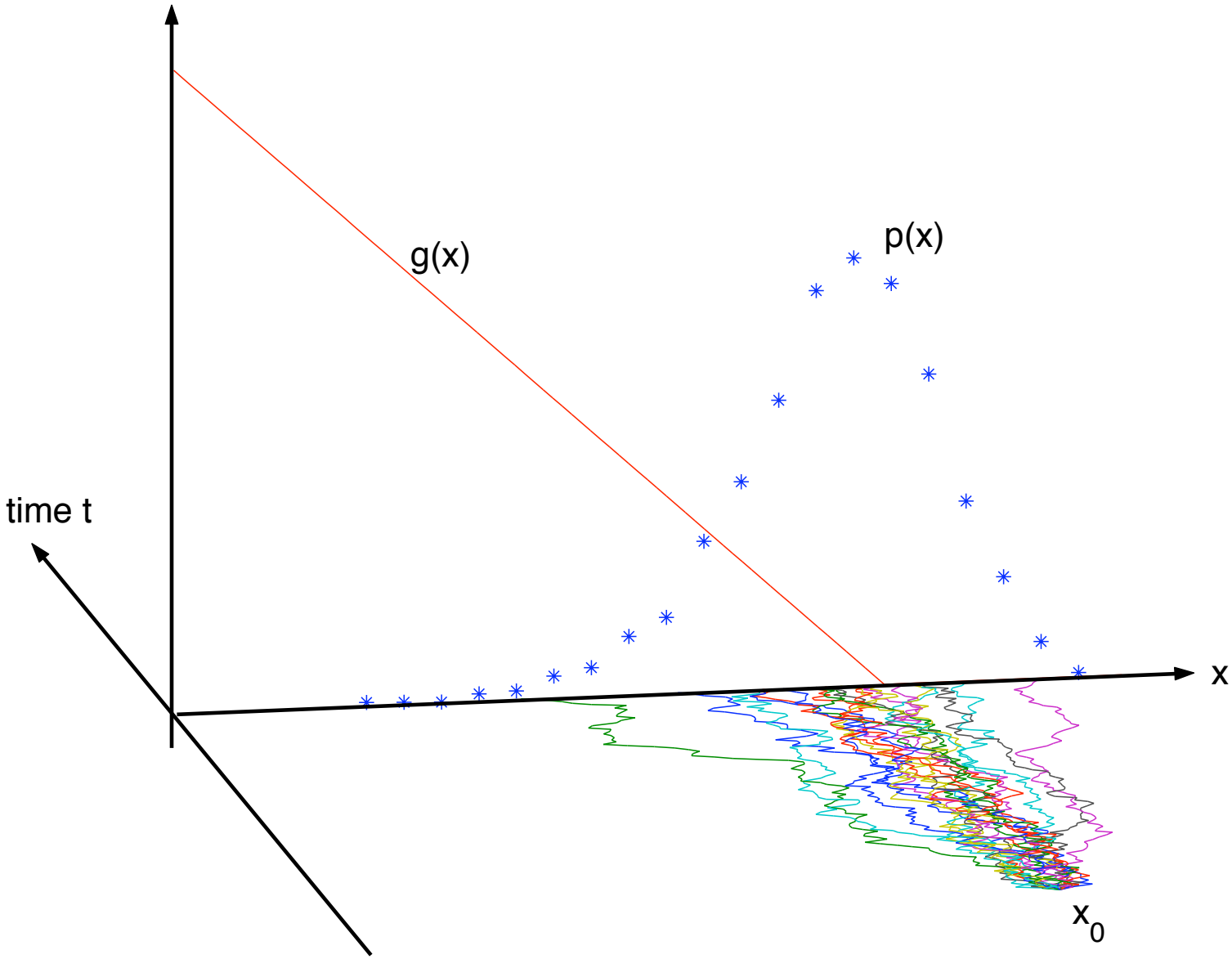
Seminal works: (Giles, Heinrich).

An incomplete list of later developments:

- Antithetic variates for MLMC with the Milstein scheme, (Giles, Szpruch)
- MLMC for SDEs with Rough Observables/Payoffs, (Avikainen, Giles, Mao, Higham)
- MLMC with Tamed Time stepping for stiff SDEs, (Jentzen, Kloeden)
- MLMC complexity, (Creutzig, Dereich, Gronbach, Ritter)
- MLMC for jump diffusion SDEs, (Xia, Giles, Abdulle, Blumenthal, Buckwar)
- MLMC for Levy driven SDEs, (Dereich, Heidenreich)
- MLMC for Tau Leap (Anderson, Higham, Moraes, T, Vilanova)
- MLMC for SPDEs: (Giles, Graham, Scheichl, Teckentrup, Kuo, Sloan, Barth, Gittelson, Schwab, Reisinger, Haji-Ali, . . .)
- Optimized hierarchies for MLMC, Continuation-MLMC, (Haji-Ali, Nobile, von Schwerin, T.)
- Multi-Index-Monte-Carlo (MIMC), (Haji-Ali, Nobile, T.)

https://people.maths.ox.ac.uk/gilesm/mlmc_community.html

Weak approximation of an Itô SDE



References:

- Adaptive weak approximation of stochastic differential equations, by A. Szepessy, R. T. and G. Zouraris. Comm. Pure and Appl. Math., **54**(10), (2001).
- 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. **Preprint**, arXiv:1411.5515v1 (2014).

Problem formulation

For the Itô SDE

$$dX_t = a(X_t, t)\,dt + \sum_{k=1}^{K} b^k(X_t, t)\,dW_t^k, \quad 0 < t < T, \tag{3}$$
$$X_0 = x_0,$$

and $g : \mathbb{R}^d \to \mathbb{R}$, approximate $E[g(X_T)]$ to a given accuracy TOL and with prescribed confidence.
W_t is a K-dimensional Wiener process.

Applications: Diffusion processes, Langevin Dynamics, Computational Finance, Crowd Flows,. . .

Forward Euler Maruyama Method

1. Forward Euler scheme on a grid $t_0 = 0 < t_1 < \ldots < t_N = T$

$$\overline{X}_{n+1} = \overline{X}_n + a(\overline{X}_n, t_n)\Delta t_n + \sum_{k=1}^{K} b^k(\overline{X}_n, t_n)\Delta W_n^k$$

gives approximate realisations $\overline{X}_T(\omega)$.

Here $\Delta t_n = t_{n+1} - t_n$ and independent Wiener increments

$$\Delta W_n^k = W_{n+1}^k - W_n^k \sim N(0, \Delta t_n) \sim \sqrt{\Delta t_n} N(0,1).$$

2. Monte Carlo estimator:

$$E[g(X_T)] \approx \sum_{i=1}^{M} \frac{g(\overline{X}_T(\omega_i; \Delta t))}{M}$$

The error contributions in MC: Bias and Statistical

Total error:

$$\left| E[g(X_T)] - \sum_{i=1}^{M} \frac{g(\overline{X}_T(\omega_i; \Delta t))}{M} \right|$$

$$\leq \left| E[g(X_T) - g(\overline{X}_T)] \right| + \left| E[g(\overline{X}_T)] - \sum_{i=1}^{M} \frac{g(\overline{X}_T(\omega_i; \Delta t))}{M} \right|$$

$$\leq TOL_T + TOL_S = TOL$$

Requirement for the time discretization error:

$$|E[g(X_T) - g(\overline{X}_T)]| \leq TOL_T$$

Requirement for the statistical error (with high probability):

$$\left| E[g(\overline{X}_T)] - \sum_{i=1}^{M} \frac{g(\overline{X}_T(\omega_i; \Delta t))}{M} \right| \leq TOL_S$$

Error Control and Complexity

Weak convergence for smooth drift and diffusion:

$$|E[g(X_T) - g(\overline{X}_T(\cdot; \Delta t))]| = \mathcal{O}(\Delta t).$$

$\Delta t \propto TOL$ needed for $|E[g(X_T) - g(\overline{X}_T)]| \leq \mathcal{O}(TOL_T)$.

By the Central Limit Theorem, as $M \to \infty$,

$$\sqrt{M}\left(\sum_{i=1}^{M} \frac{g(\overline{X}_T(\omega_i; \Delta t)) - E[g(\overline{X}_T)]}{M}\right) \xrightarrow{D} N\left(0, \sqrt{Var[g(\overline{X}_T)]}\right).$$

Thus, $M \propto \frac{1}{TOL^2}$ needed for sufficient probability that $\left|E[g(\overline{X}_T)] - \sum_{i=1}^{M} \frac{g(\overline{X}_T(\omega_i;\Delta t))}{M}\right| \leq \mathcal{O}(TOL_S)$.

Computational complexity $= M\frac{T}{\Delta t} \lesssim TOL^{-3}$.

Adaptivity, weak approximation of SDE

Given TOL_T, use adaptive refinements to generate stochastic grids $t_0 = 0 < t_1(\omega) < \ldots < t_N = T$ to create realizations $\overline{X}_T(\omega; \Delta t(\omega))$.

Why? Non-smooth $a(X_s, s)$ or $b(X_s, s)$ can decrease convergence rates.

How? Use a posteriori SDE weak error density [STZ01], [MSTZ06],[MSTZ08].
For $\max \Delta t(TOL) \to 0$ as $TOL \to 0$ we have

$$E\left[g(X_T) - g(\overline{X}_T)\right] \simeq \int_0^T E[\Delta t(s) \rho_W(\overline{X}_s, s)] ds$$

One also has a strong error density [HHT14] **(see H. Hoel's upcoming talk)**

$$E\left[\left(g(X_T) - g(\overline{X}_T)\right)^2\right] \simeq \int_0^T E[\Delta t(s) \rho_S(\overline{X}_s, s)] ds$$

Single Level, Adaptive Time Step algorithm

Adaptive refinements start from a coarse initial grid, and

(1) compute solution and error indicators r_n for each time step n,

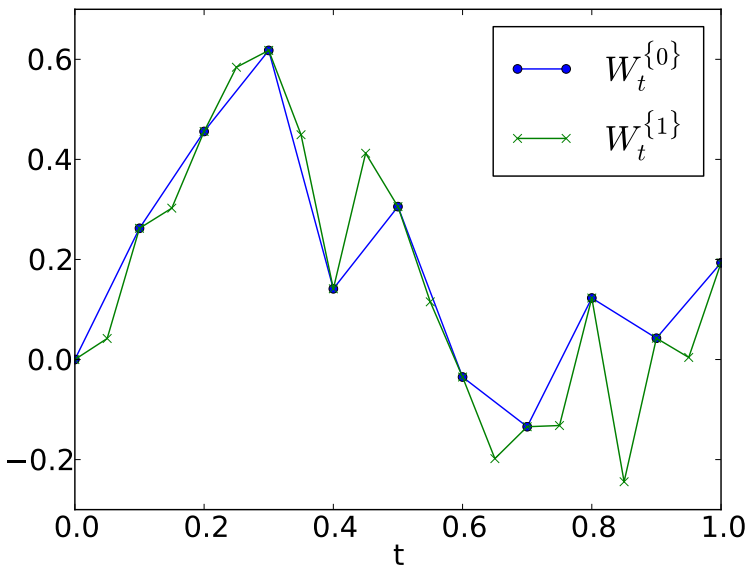
(2) as long as

$$\max_n r_n \geq C_S \frac{TOL_T}{E[N]}, \tag{4}$$

(3) refine all time steps s.t.

$$r_n \geq C_R \frac{TOL_T}{E[N]}, \tag{5}$$

refine sampling of W by Brownian bridges, and go to (1).



$$r_n = \rho_n \Delta t_n^2$$

Idea of Adaptive Multilevel Algorithm [HSST08]

How to apply the multilevel Monte Carlo idea with adaptive time stepping?

Let the tolerance in the adaptive algorithm define the hierarchy!

Grid hierarchy defined by

$$TOL_{T,\ell} = \frac{TOL_{T,0}}{2^{\ell}}.$$

When is adaptivity useful for multilevel Monte Carlo?

- Non-smooth $a(X_s, s)$ or $b(X_s, s)$ can decrease both weak and strong convergence rates for uniform grids, but also affects proportionality constants.
- Jentzen & Kloeden pointed out that stability problems on coarse grids can ruin multilevel Monte Carlo convergence for stiff SDEs.

Our results on adaptive MLMC

Developed the first adaptive MLMC algorithm. Theoretical analysis of our adaptive MLMC algorithm proves

- Stopping: The adaptive MLMC Algorithm stops.
- Asymptotic Normality: Show CLT for MLMC based on Lindeberg-Feller's CLT theorem.
- Asymptotic Accuracy: With prescribed confidence, the Error is bounded by TOL.
- Complexity: Essentially the same as MLMC has with uniform time steps in smooth problems.

All these results hold for our **adaptive** ($\Delta t(t, \omega)$ non uniform), **non-adapted** ($\bar{X}$ depends on the future values of W through Δt) **discretizations**.

Example: Drift singularity

Consider for a constant $\alpha \in (0, T)$, the SDE

$$dX_t = \begin{cases} X_t dW_t, & t \in [0, \alpha] \\ \frac{X_t}{2\sqrt{t-\alpha}} dt + X_t dW_t, & t \in (\alpha, T] \end{cases}$$
$$X_0 = 1, \qquad \text{with } T = 1 \text{ and } \alpha = T/3$$

with the unique solution

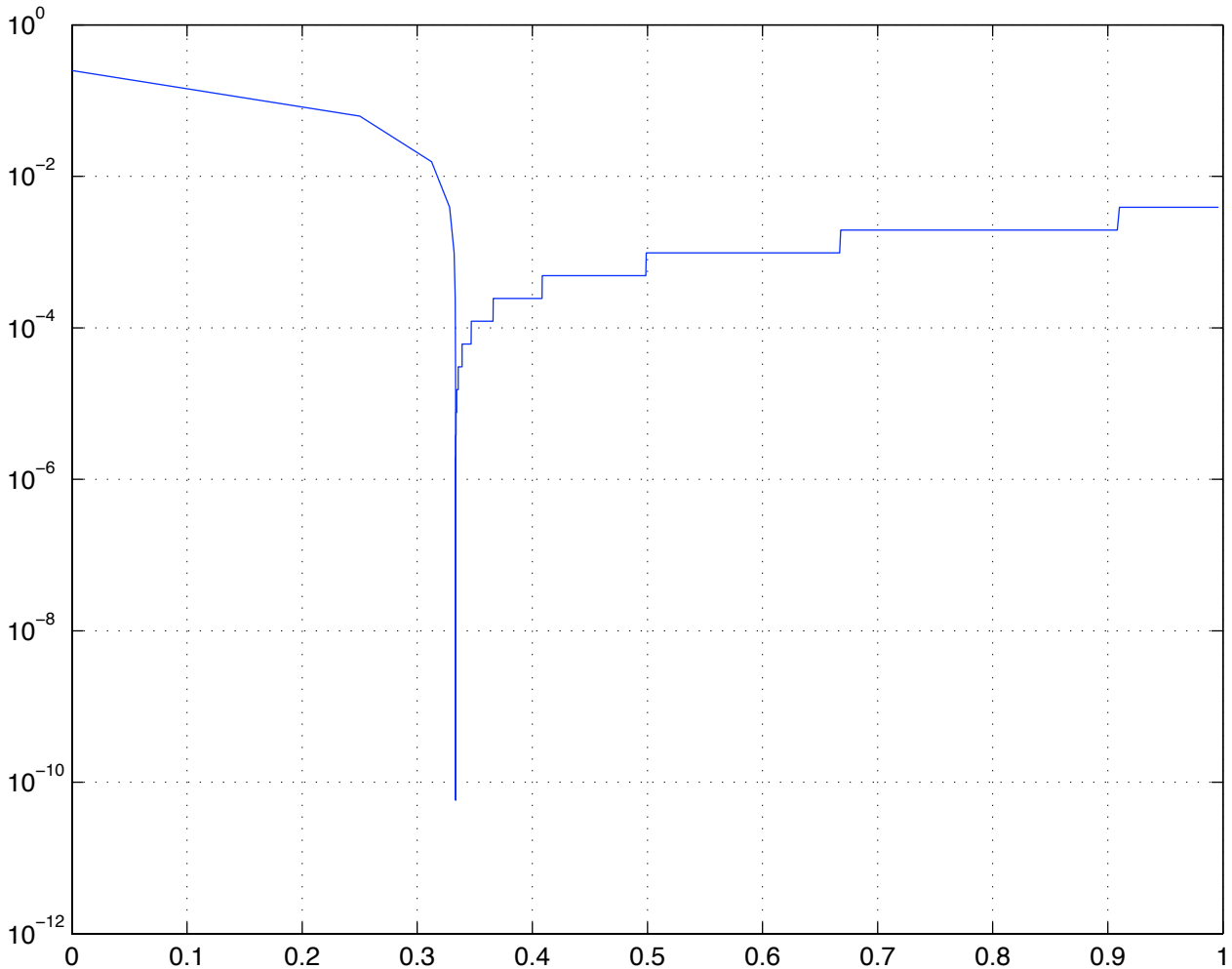
$$X_t = \begin{cases} \exp(W_t - t/2), & t \in [0, \alpha] \\ \exp(W_t - t/2)\exp(\sqrt{t-\alpha}), & t \in (\alpha, T]. \end{cases}$$

Goal: Approximate $E[X_T] = \exp(\sqrt{t-\alpha})$.
Weak convergence **uniform** order $1/2$; **adaptive** order 1.

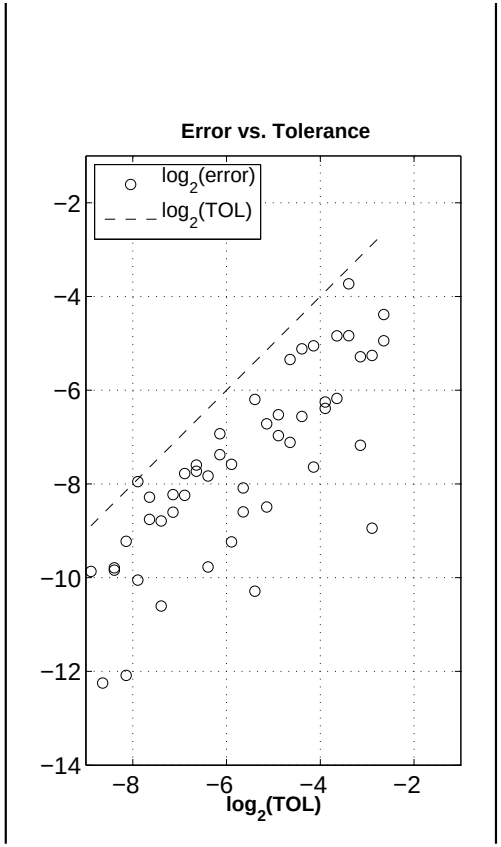
Drift singularity adaptive strategy

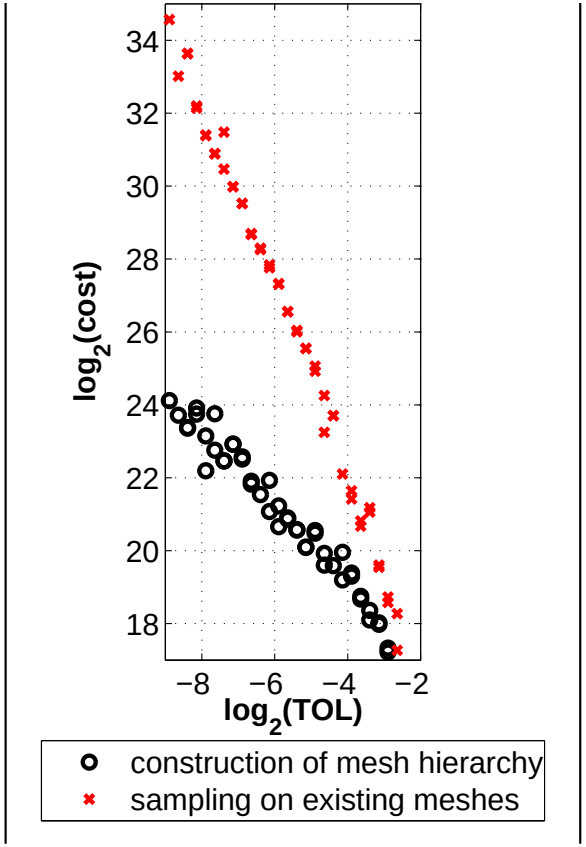
- Drift singularity at a deterministic (but unknown!) time
- Grid generation phase – use sample averaged error indicators to generate the grid hierarchy
- Sampling phase – control statistical error by performing multilevel simulations on the existing grid hierarchy

Experimental Complexity: Adapted time step size

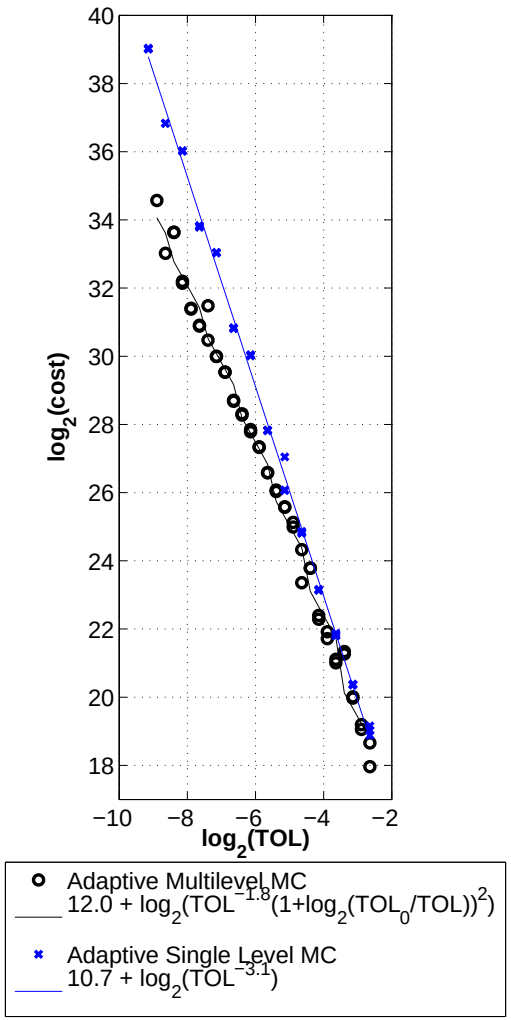


Experimental Complexity: Drift Singularity

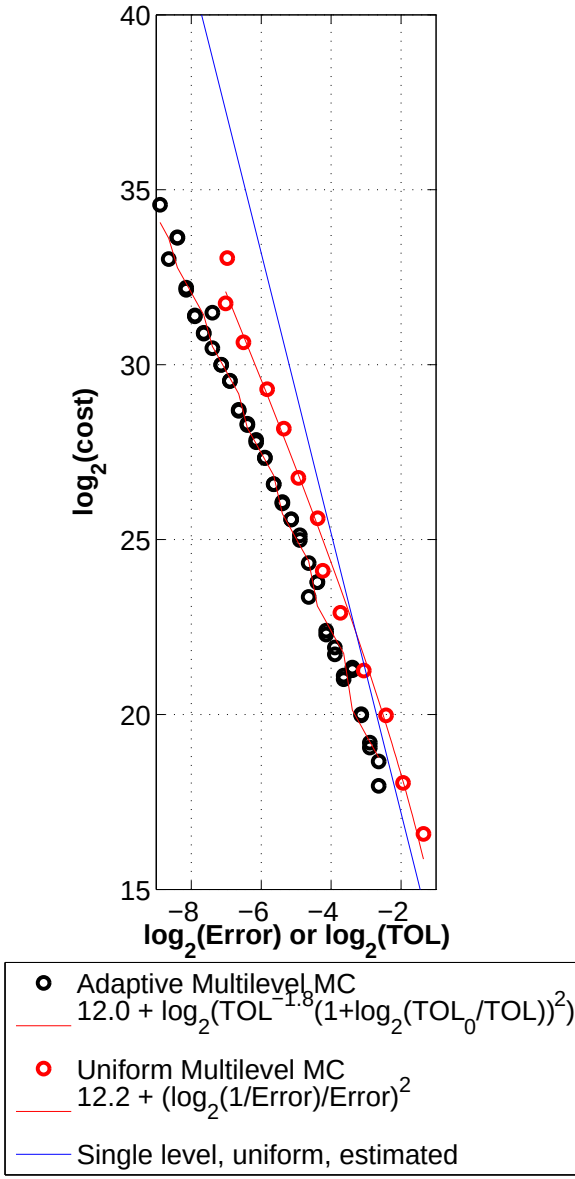




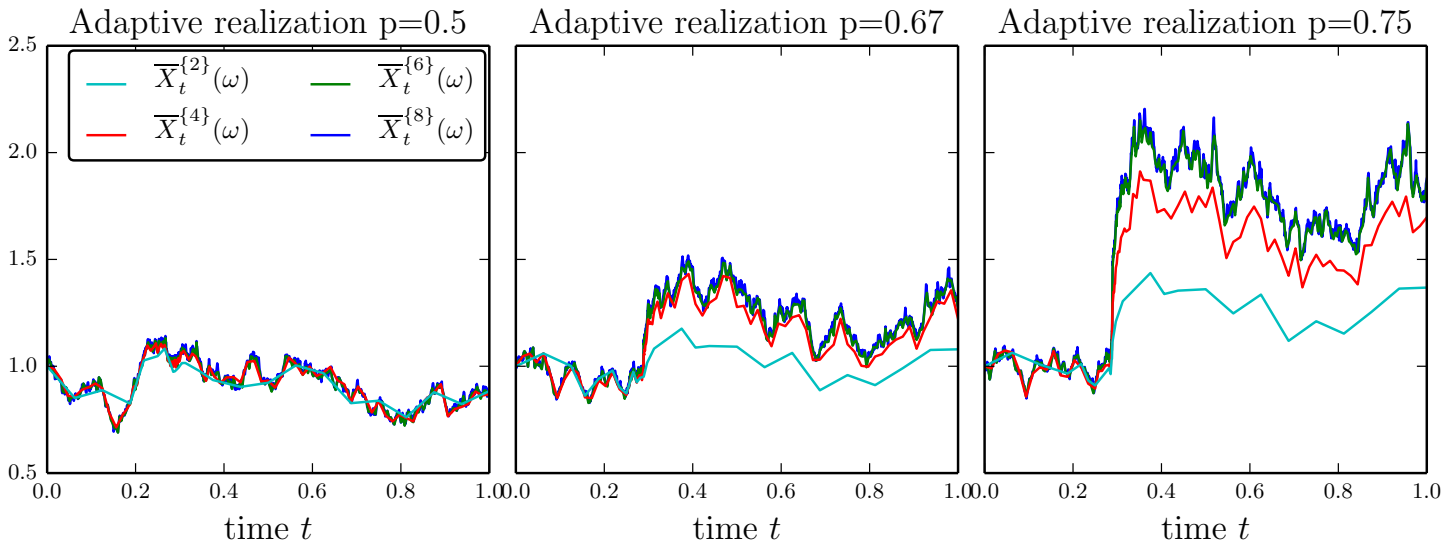
Experimental Complexity: Drift Singularity

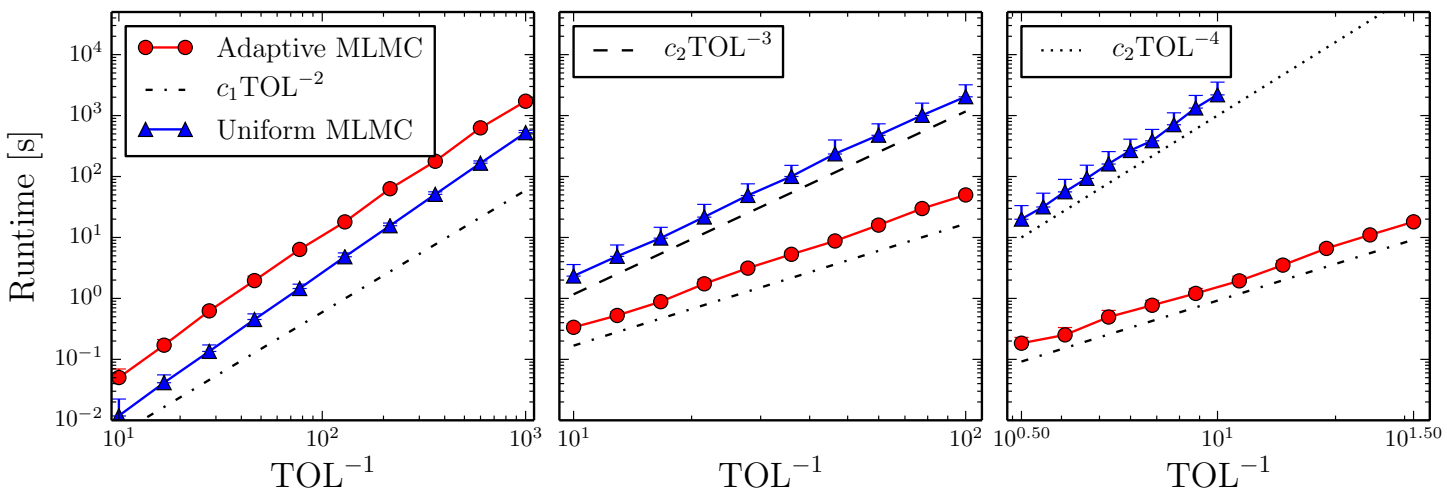


Method	Complexity
uniform SLMC	ϵ^{-4}
adaptive SLMC	$\epsilon^{-3.1}$
uniform MLMC	$\log(1/\epsilon)^2\epsilon^{-2.0}$
adaptive MLMC	$\log(1/\epsilon)^2\epsilon^{-1.8}$



Drift Singularity, gains from adaptive MLMC





$$dX_t = \frac{X_t}{5|t-\alpha|^p}dt + 0.5X_t dW_t, X_0 = 1, \qquad \alpha \sim U(1/4, 3/4)$$

Conclusions

- Extended single level adaptive time stepping algorithms to the MLMC setting.
- Showed a CLT result that allows prescription of confidence level in our computations.
- Asymptotic estimates describe the behavior of the resulting adaptive algorithms, numerical experiments confirm the predicted bounds.
- Extension to stopped [DMSST05] and jump diffusions [MSTZ08] is direct.

Noise driven by Poisson Random Measures

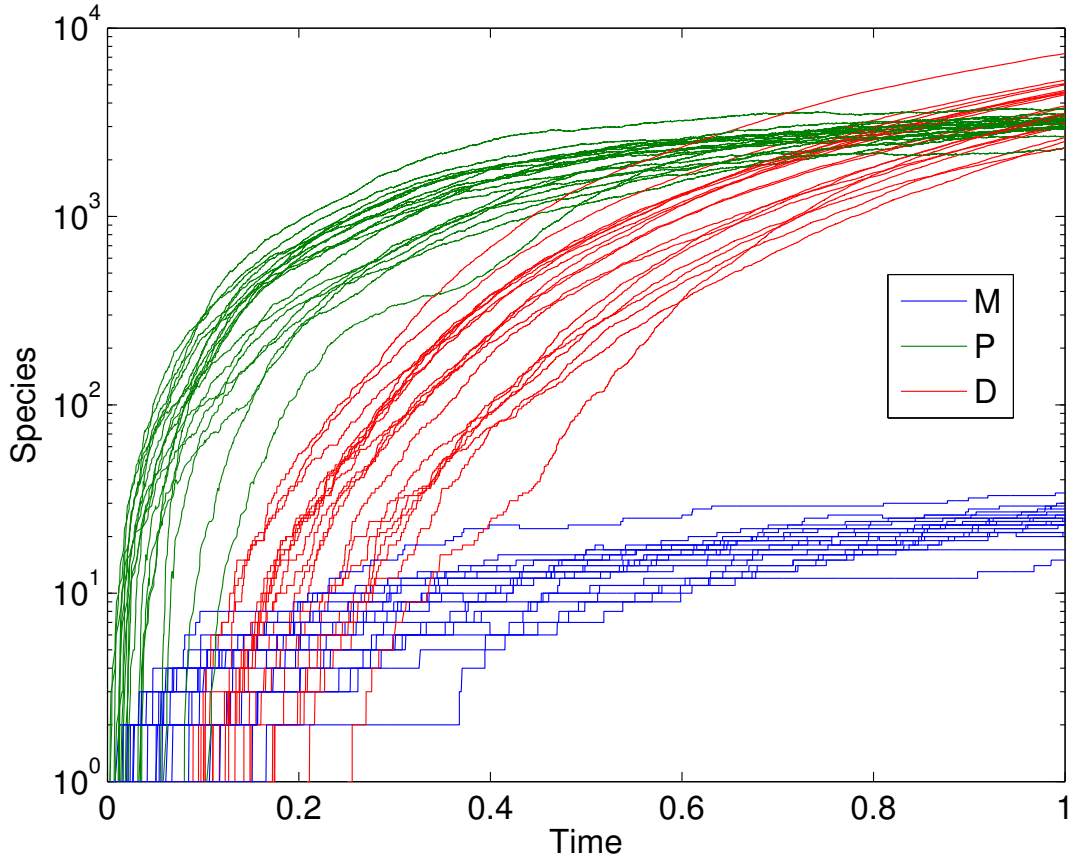


Figure: Semilogarithmic plot showing 20 exact i.i.d. paths of 3 biochemical species in a problem from genomics.

Motivational Example: Gene transcription and translation
In **[Anderson2012]** the following example is proposed:

- $G \xrightarrow{25} G+M$, a single gene is being transcribed into mRNA.
- $M \xrightarrow{1000} M+P$ mRNA is then being translated into proteins.
- $P+P \xrightarrow{0.001} D$, finally the proteins produce stable Dimers.
- $M \xrightarrow{0.1} \emptyset$, $P \xrightarrow{1} \emptyset$ degradation of mRNA and proteins, respectively.

Initial state: $X(0) = (0, 0, 0)$, where X_1, X_2, X_3 give the molecular counts of the mRNA, proteins, and dimers, respectively.

Goal: *We want to estimate the expected number of Dimers at time* $T = 1$ up to certain tolerance, with high probability.

Mathematical statement of the problem

Let X be a Pure Jump Process $X=(X_1,\ldots,X_d):[0,T]\times\Omega\to\mathbb{Z}_+^d$ described by

- Finite number of possible reaction channels $\nu_j\in\mathbb{Z}^d$.

$$x\in\mathbb{Z}_+^d,\ x\to x+\nu_j$$

- and its corresponding propensity functions, $a_j:\mathbb{R}^d\to\mathbb{R}^+$ s.t.

$$\mathrm{P}\left(X(t+\mathrm{d}t)=x+\nu_j\,\middle|\,X(t)=x\right)=a_j(x)\mathrm{d}t+o\left(\mathrm{d}t\right),$$

Motivation: accurately approximate the expected value

$$\mathrm{E}\left[g(X(T))\right],$$

for some *given* observable $g:\mathbb{R}^d\to\mathbb{R}$.
Applications: Systems biology, complex reaction networks, biochemical kinetics, stochastic epidemic spread modeling, ...

References: MLMC for Pure Jump Processes

- J. Karlsson and R. T. "Towards Automatic Global Error Control: Computable Weak Error Expansion for the Tau-Leap Method", Monte Carlo Methods and Applications, 17(3), 233–278, (2011).
- D. Anderson and D. Higham, "Multilevel Monte Carlo for continuous Markov chains, with applications in biochemical kinetics". SIAM Multiscal Model. Simul., 10(1), (2012).
- A. Moraes, P. Vilanova and R. T., "Hybrid Chernoff Tau-Leap", SIAM Multiscale Modeling and Simulation, Vol. 12, Issue 2, (2014).
- A. Moraes, P. Vilanova and R. T., "Multilevel Hybrid Chernoff Tau-Leap", *arXiv:1403.2943*. Submitted, April 2014.
- A. Moraes, P. Vilanova and R. T., "Multilevel adaptive reaction-splitting kinetic simulation method for stochastic reaction networks", *arXiv:1406.1989v1*. Submitted, June 2014.

See Pedro & Alvaro's posters outside for further details

Pathwise Exact and approximate simulation methods

- **Exact algorithms** like the Stochastic Simulation Algorithm (SSA) [Gillespie76] and the Modified Next Reaction Method (MNRM) [Anderson07] sometimes are too expensive for path simulation. Just consider the inter-arrival time between transitions $\tau_{SSA}|X(t)=x \sim$ exponential $\left(\sum_j a_j(x)\right)$.
- **Approximate algorithms** that evolve with fixed time steps, like the Tau-leap, may be faster [Gillespie01] .Two drawbacks: i) time discretization errors ii) may lead to negative population numbers, *i.e.*, non-physical results. Pre-leap: adjust adaptively the time step to control the one-step exit probability.

[Moraes,Vilanova,T.14]:
– single level hybrid algorithm that, at each time step, adaptively switches between the SSA and the Tau-leap to min comp. cost.
– related hybrid adaptive MLMC algorithms for error control
– a variance reduction technique based on Kurtz representation and a deterministic time change.

A Chernoff bound for the tau-leap method I

From Kurtz's random time change representation [Kurtz78]:

$$X(t) = X(0) + \sum_{j=1}^{J} Y_j \left(\int_0^t a_j(X(s))ds \right) \nu_j,$$

where Y_j are independent unit-rate Poisson processes, we obtain the Tau-leap method (forward Euler approximation):

$$\bar{X}(t+\tau) = \bar{X}(t) + \sum_{j=1}^{J} \mathcal{P}_j \left(\underbrace{a_j(\bar{X}(t))\tau}_{=\lambda_j} \right) \nu_j,$$

where $\mathcal{P}_j(\lambda_j)$ are independent Poisson random vars with rate λ_j.
Problem: Given $\delta>0$, find the largest $\tau = \tau(x, \delta)$ s.t.

$$\mathrm{P}\left(\bar{X}(t+\tau) \notin \mathbb{Z}_+^d \,\middle|\, \bar{X}(t) = x\right) \leq \mathit{ChBnd}(x, \tau) < \delta.$$

One-step hybrid switching rule: Take the T-Leap step if the $\min\{\tau, \Delta t\} >> \tau_{SSA}$

An exiting path is a rare event, the role of δ

Let A be the event that a hybrid trajectory arrived to final time T without exiting $\mathbb{Z}_+^d$. We show in [MVT13-14] that

$$\mathrm{P}\left(A^c\right) \leq \delta \mathrm{E}\left[N_{TL}\right] - \frac{\delta^2}{2}\left(\mathrm{E}\left[N_{TL}^2\right] - \mathrm{E}\left[N_{TL}\right]\right) + o(\delta^2).$$

In practice, we use $\delta \mathrm{E}\left[N_{TL}\right]$ as an upper bound of $\mathrm{P}\left(A^c\right)$.

Remark: The role of δ is to turn A^c into a rare event. Direct sampling of hybrid paths to estimate $P(A^c)$ is non feasible, while the estimate of $\mathrm{E}\left[N_{TL}\right]$ is straightforward.

The MLMC estimator I

Consider the following telescopic decomposition with $\Delta t_\ell = 2^{-\ell}\Delta t_0$, $\ell = 1, \ldots, L$:

$$\mathrm{E}\left[g_L \mathbf{1}_{A_L}\right] = \mathrm{E}\left[g_0 \mathbf{1}_{A_0}\right] + \sum_{\ell=1}^{L} \mathrm{E}\left[g_\ell \mathbf{1}_{A_\ell} - g_{\ell-1}\mathbf{1}_{A_{\ell-1}}\right],$$

which motivates the definition of our MLMC estimator of $\mathrm{E}\left[g(X(T))\right]$,

$$\mathcal{M}_L := \frac{1}{M_0}\sum_{m_0=1}^{M_0} g_0 \mathbf{1}_{A_0}(\omega_{m_0}) + \sum_{\ell=1}^{L}\frac{1}{M_\ell}\sum_{m_\ell=1}^{M_\ell}[g_\ell \mathbf{1}_{A_\ell} - g_{\ell-1}\mathbf{1}_{A_{\ell-1}}](\omega_{m_\ell}).$$

We define the computational global error, $\mathcal{E}_L$, as

$$\mathcal{E}_L := \mathrm{E}\left[g(X(T))\right] - \mathcal{M}_L.$$

The MLMC estimator II

Now, consider the following decomposition of $\mathcal{E}_L$

$$\begin{aligned}\mathcal{E}_L &= \mathrm{E}\left[g(X(T))(\mathbf{1}_{A_L} + \mathbf{1}_{A_L^c})\right] \pm \mathrm{E}\left[g_L \mathbf{1}_{A_L}\right] - \mathcal{M}_L \\ &= \underbrace{\mathrm{E}\left[g(X(T))\mathbf{1}_{A_L^c}\right]}_{=:\mathcal{E}_{E,L}\,(\text{exit})} + \underbrace{\mathrm{E}\left[(g(X(T)) - g_L)\,\mathbf{1}_{A_L}\right]}_{=:\mathcal{E}_{I,L}\,(\text{weak})} + \underbrace{\mathrm{E}\left[g_L \mathbf{1}_{A_L}\right] - \mathcal{M}_L}_{=:\mathcal{E}_{S,L}\,(\text{statistical})}.\end{aligned}$$

Problem: Given $TOL > 0$, find the parameters for computing $\mathcal{M}_L$ such that $|\mathcal{E}_L| < TOL$ with high probability, and with nearly optimal computational work.

Issues addressed:

i) Simulated coupled hybrid pairs $(g_\ell, g_{\ell-1})(\omega)$ for $\ell = 1, \ldots, L$,

ii) Estimated accurately and controlled all the global error components.

iii) Showed the resulting complexity $\mathcal{O}(TOL^{-2})$.

Example: Gene transcription and translation

In **[Anderson2012]** the following example is proposed:

- $G \xrightarrow{25} G+M$, a single gene is being transcribed into mRNA.
- $M \xrightarrow{1000} M+P$, mRNA is then being translated into proteins.
- $P+P \xrightarrow{0.001} D$, finally the proteins produce stable Dimers.
- $M \xrightarrow{0.1} \emptyset$, $P \xrightarrow{1} \emptyset$, degradation of mRNA and proteins, respectively.

Initial state: $X(0) = (0, 0, 0)$, where X_1, X_2, X_3 give the molecular counts of the mRNA, proteins, and dimers, respectively.

Goal: *We want to estimate the expected number of Dimers at time* $T = 1$ up to certain tolerance, with high probability.

Example: Gene transcription and translation

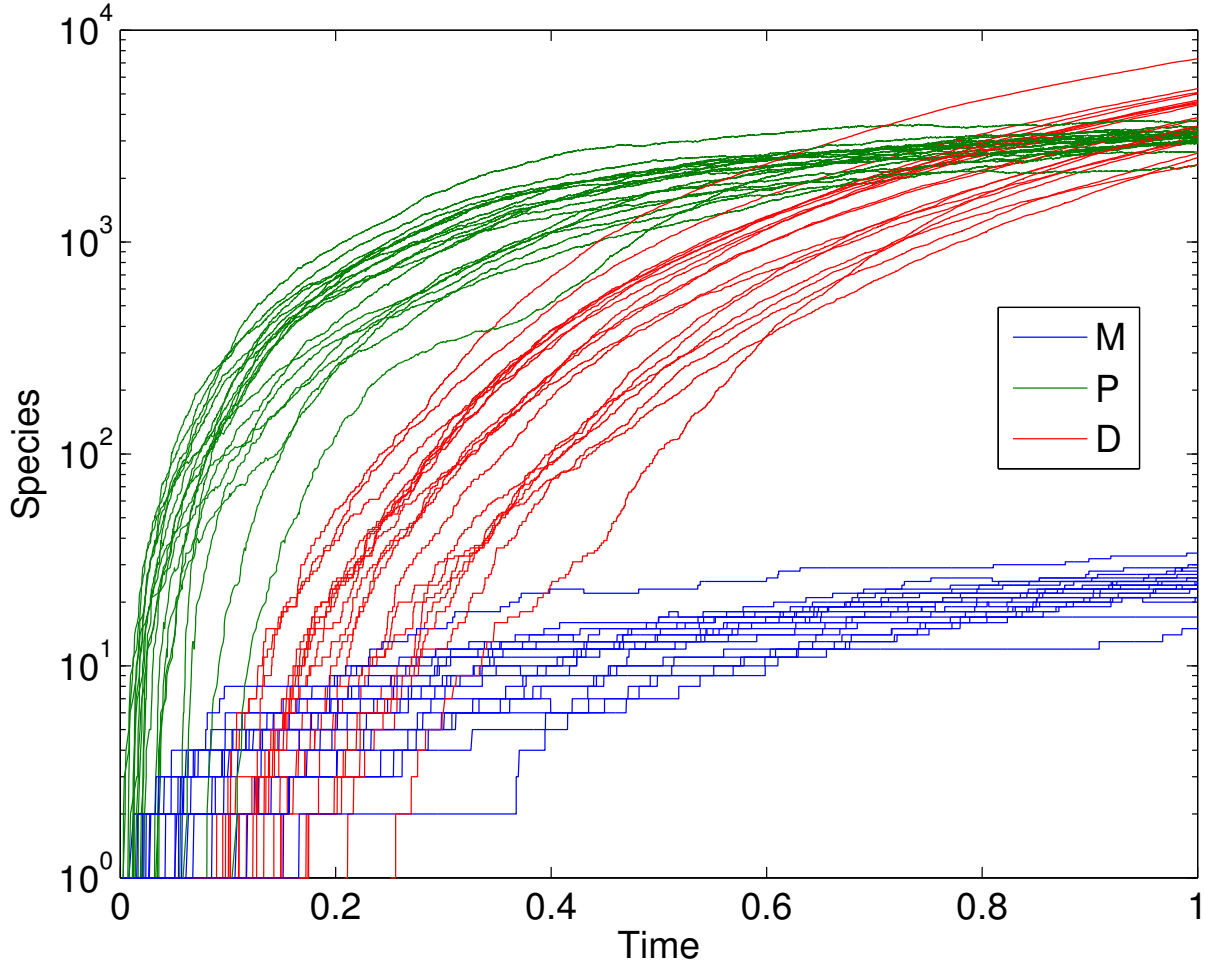


Figure: Semilog scale. Some paths of the time evolution of the species.

Numerical results: global error vs actual work

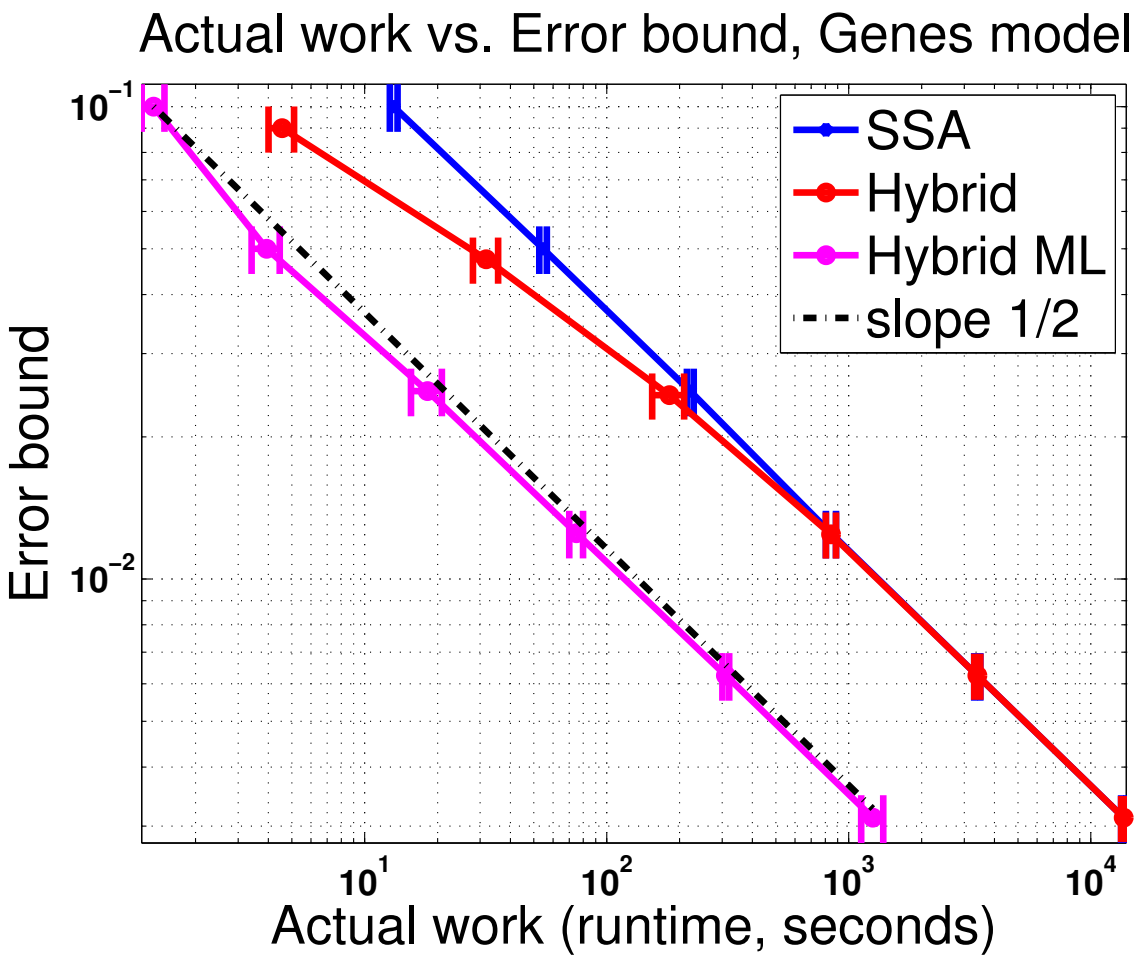


Figure: Actual work for each one of the one hundred adaptive runs. Our hybrid MLMC is 10 times faster than the SSA.

Numerical results: error vs TOL

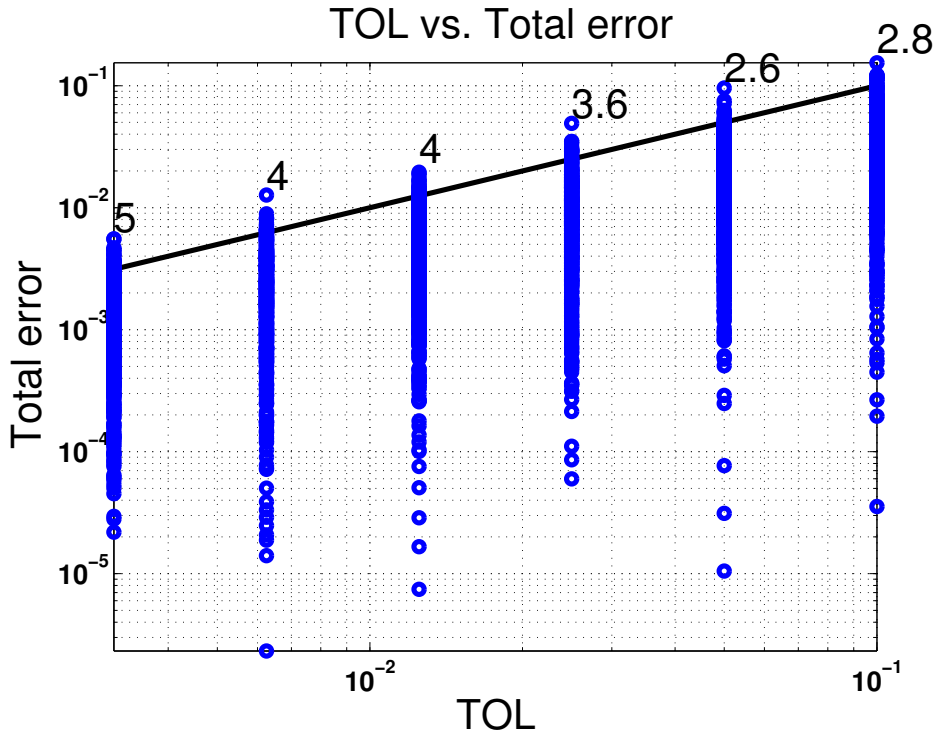


Figure: *TOL* versus the actual computational error. The numbers above the straight line show the percentage of runs that had errors larger than the required tolerance. We observe that in all cases the computational error follows the imposed tolerance closely with the expected confidence of 95%.

Novel MLMC Reaction-Splitting Method: error vs TOL

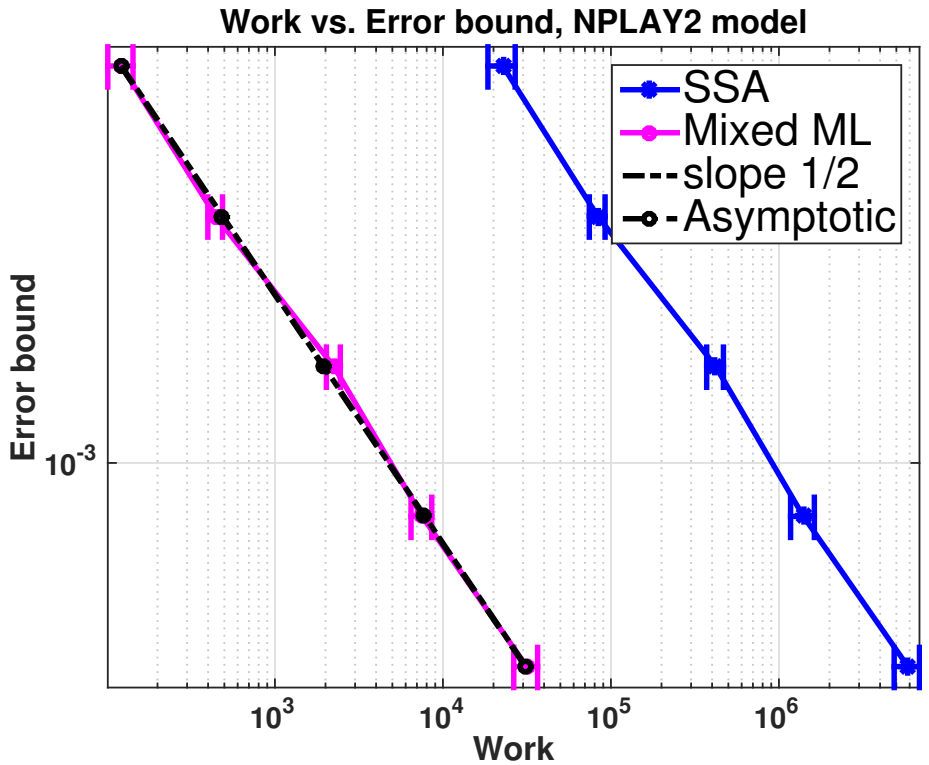


Figure: Speed-ups of order $10^2 - 10^4$ are obtained in stiff problems using our novel reaction-splitting Multilevel Monte Carlo method. We developed also a novel Control Variate based on Kurtz representation and a deterministic-time change approximation.

$$\emptyset \xrightarrow{k} X_n,\ X_n \xrightarrow{c} Y,\ n = 1, \ldots, N = 200$$

$$Y \xrightarrow{a} \emptyset,\ Y \xrightarrow{b} 50\ Y$$

Novel MLMC Reaction-Splitting Method: error vs TOL

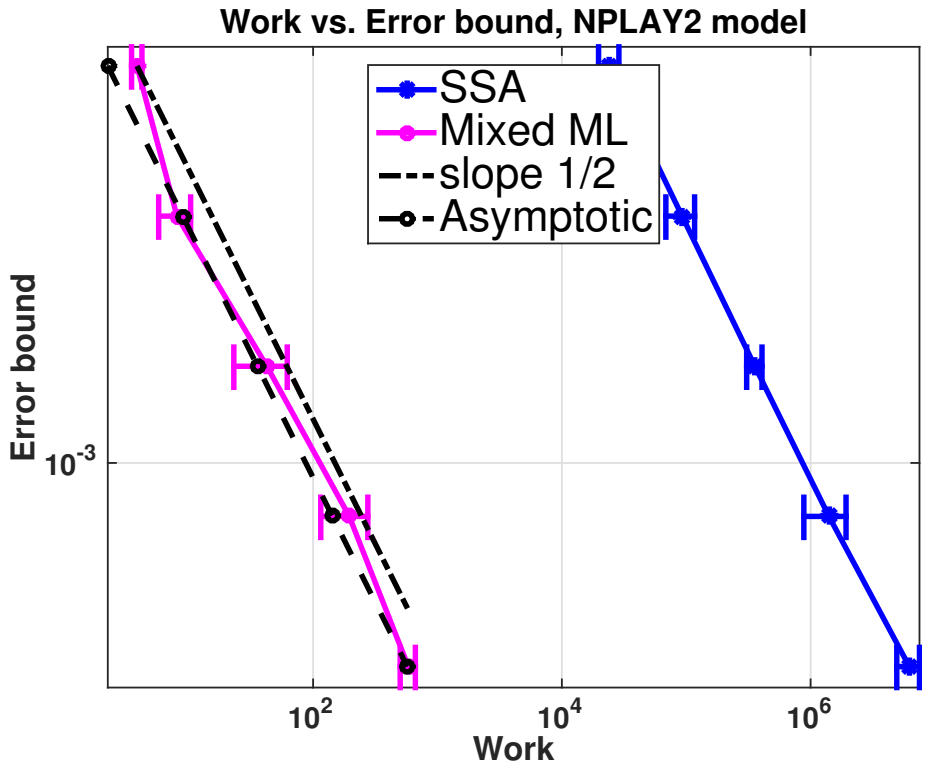


Figure: Speed-ups of order $10^2 - 10^4$ are obtained in stiff problems using our novel reaction-splitting Multilevel Monte Carlo method. We developed also a novel Control Variate based on Kurtz representation and a deterministic-time change approximation.

$$\emptyset \xrightarrow{k} X_n,\ X_n \xrightarrow{c} Y,\ n = 1, \ldots, N = 200$$

$$Y \xrightarrow{a} \emptyset,\ Y \xrightarrow{b} 50\ Y$$

Novel MLMC Reaction-Splitting Method: error vs TOL

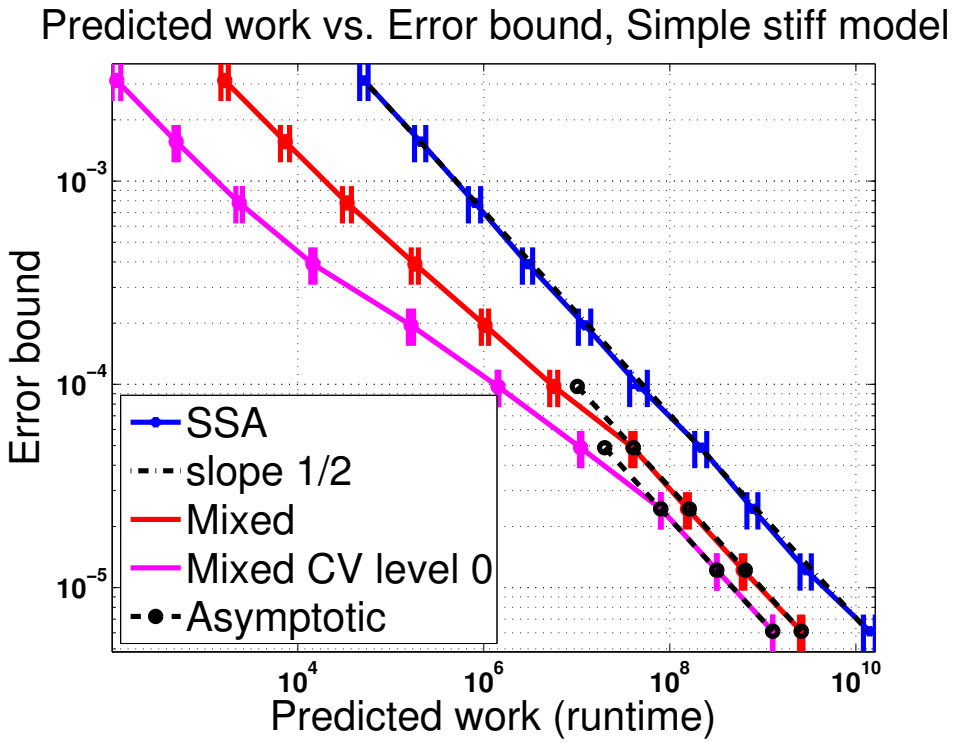


Figure: Speed-ups of order 10^3 are obtained in stiff problems using our novel reaction-splitting Multilevel Monte Carlo method. We developed also a novel Control Variate based on a deterministic-time change approximation. Example adapted from [Cao-Gillespie-Petzold,2005].

$$X_1 \underset{c_2}{\overset{c_1}{\rightleftharpoons}} X_2 \xrightarrow{c_3} X_3 \xrightarrow{c_4} \emptyset, \quad c_2 \gg c_3 > c_4.$$

Conclusions

- The computational complexity of this method is of order $\mathcal{O}\left(TOL^{-2}\right)$, and therefore, it can be seen as a variance reduction of the SSA method, which has the same complexity. This represents an important advantage of the hybrid tau-leap with respect to the pure tau-leap in the MLMC context.
- Our algorithm provides the elements for the simulation setting (*i.e.*, initial time mesh, number of levels, one-step exit probabilities and number of coupled hybrid paths at each level) that optimizes the computational work.
- For reaching this optimality, we derived novel formulas based on dual-weighted residual to estimate the variance of the difference of the observables between two consecutive levels in coupled hybrid paths.

Optimal MLMC hierarchies and CMLMC

Goal: compute $\mathrm{E}[g]$ where $g = g(u)$. Here g is either a bounded linear functional or a Lipschitz functional with respect to u, and u solves a stochastic equation.

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

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

for sufficiently smooth a, f, k.

References (See E. von Schwerin's upcoming talk):

[ANvST00] *"Optimization of mesh hierarchies for Multilevel Monte Carlo"*, by A.-L Haji-Ali, F. Nobile, E. von Schwerin and R. T. Preprint arXiv:1403.2480, 2014.

[NANvST01] *"A Continuation Multilevel Monte Carlo"*, by N. Collier, A.-L Haji-Ali, F. Nobile, E. von Schwerin and R. T. (in BIT Num., 2014).

Optimal MLMC hierarchies and CMLMC

Following the standard MLMC approach, we introduce a hierarchy of $L+1$ meshes defined by decreasing mesh sizes $\{h_\ell\}_{\ell=0}^{L}$ and we denote the approximation of g using mesh size h_ℓ by g_ℓ. We then write the MLMC estimator as

$$\mathcal{A} = \frac{1}{M_0}\sum_{m=1}^{M_0} g_0(\omega_{0,m}) + \sum_{\ell=1}^{L}\frac{1}{M_\ell}\sum_{m=1}^{M_\ell}\left(g_\ell(\omega_{\ell,m}) - g_{\ell-1}(\omega_{\ell,m})\right). \tag{7}$$

We assume positive constants Q_W, Q_S, q_1, q_2, d and γ s.t.

$$\left|\mathrm{E}[g_\ell - g]\right| \approx Q_W h_\ell^{q_1}, \tag{8a}$$

$$\mathrm{Var}\left[g_\ell - g_{\ell-1}\right] := V_\ell \approx Q_S h_{\ell-1}^{q_2}, \tag{8b}$$

$$\text{Work per sample of level } \ell := W_\ell \approx h_\ell^{-d\gamma}. \tag{8c}$$

Goal: Choose $(\{h_\ell\}_{\ell=0}^{L}, \{M_\ell\}_{\ell=0}^{L})$ optimally to minimize work while meeting prescribed accuracy and confidence constraints.

Examples

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 linear or piecewise bilinear continuous 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: $d = 3$ and $\gamma = 3$ for a naive Gaussian elimination implementation. Moreover, Using an *Iterative* solver has $\gamma \approx 1$ while using *Direct* solver has $\gamma \approx 1.5$.

Examples

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 linear or piecewise bilinear continuous 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: $d = 3$ and $\gamma = 3$ for a naive Gaussian elimination implementation. Moreover, Using an *Iterative* solver has $\gamma \approx 1$ while using *Direct* solver has $\gamma \approx 1.5$.

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

In our PDE example: $\chi \approx 1.34$ for iterative solver and $\chi \approx 0.89$ for direct solver.

Problem (Optimization of MLMC 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*

$$\mathrm{WORK}(\mathbf{H}) = \sum_{\ell=0}^{L} \frac{M_\ell}{h_\ell^{d\gamma}}, \tag{9a}$$

is minimized while satisfying the constraints

$$Q_W h_L^{q_1} \leq (1-\theta)\, TOL, \tag{9b}$$

$$\frac{V_0}{M_0} + Q_S \sum_{\ell=1}^{L} \frac{h_{\ell-1}^{q_2}}{M_\ell} \leq \left(\frac{\theta\, TOL}{C_\alpha}\right)^2, \tag{9c}$$

Obs: *We showed a MLMC-CLT result justifying* (9c). *For a confidence parameter,* C_α, *such that* $\Phi(C_\alpha) = 1 - \frac{\alpha}{2}$; *here,* $0 < \alpha \ll 1$ *and* Φ *is the standard normal cdf.*

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

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

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

$$1 \leq \beta = \left\{ \left(\frac{(1-\theta)\, TOL}{Q_W} \right)^{\frac{1}{q_1}} \left(\frac{Q_S}{V_0} \right)^{\frac{1}{q_2}} \right\}^{-\frac{1}{L+1}},$$

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

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

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 1 are given by

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

where the optimal choice of the splitting parameter is

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

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

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

MLMC Optimal hierarchies III

Corollary (On the asymptotic work with optimal hierarchies)

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

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

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

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

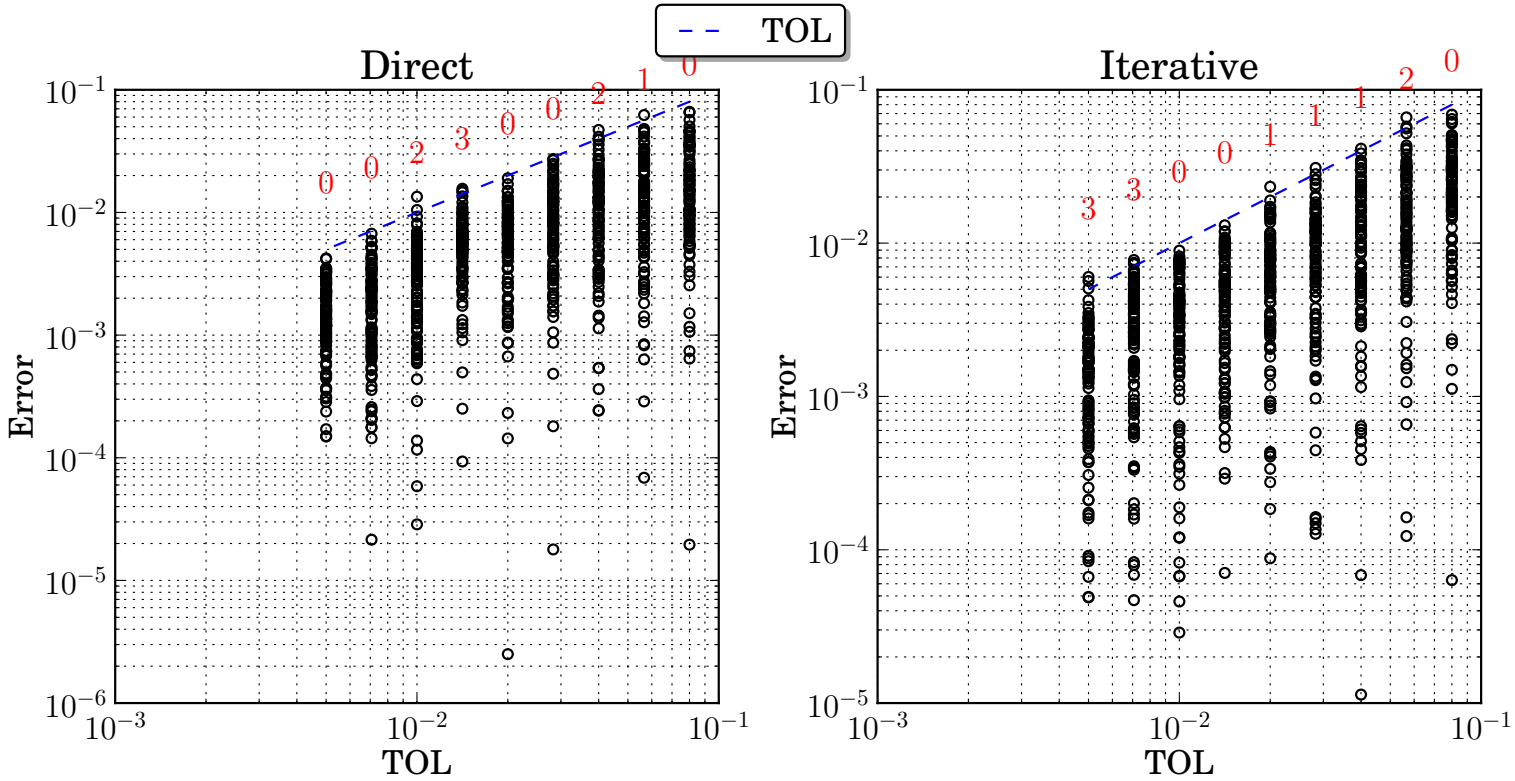
with known constants of proportionality,

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

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

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

Continuation-MLMC: Error plots

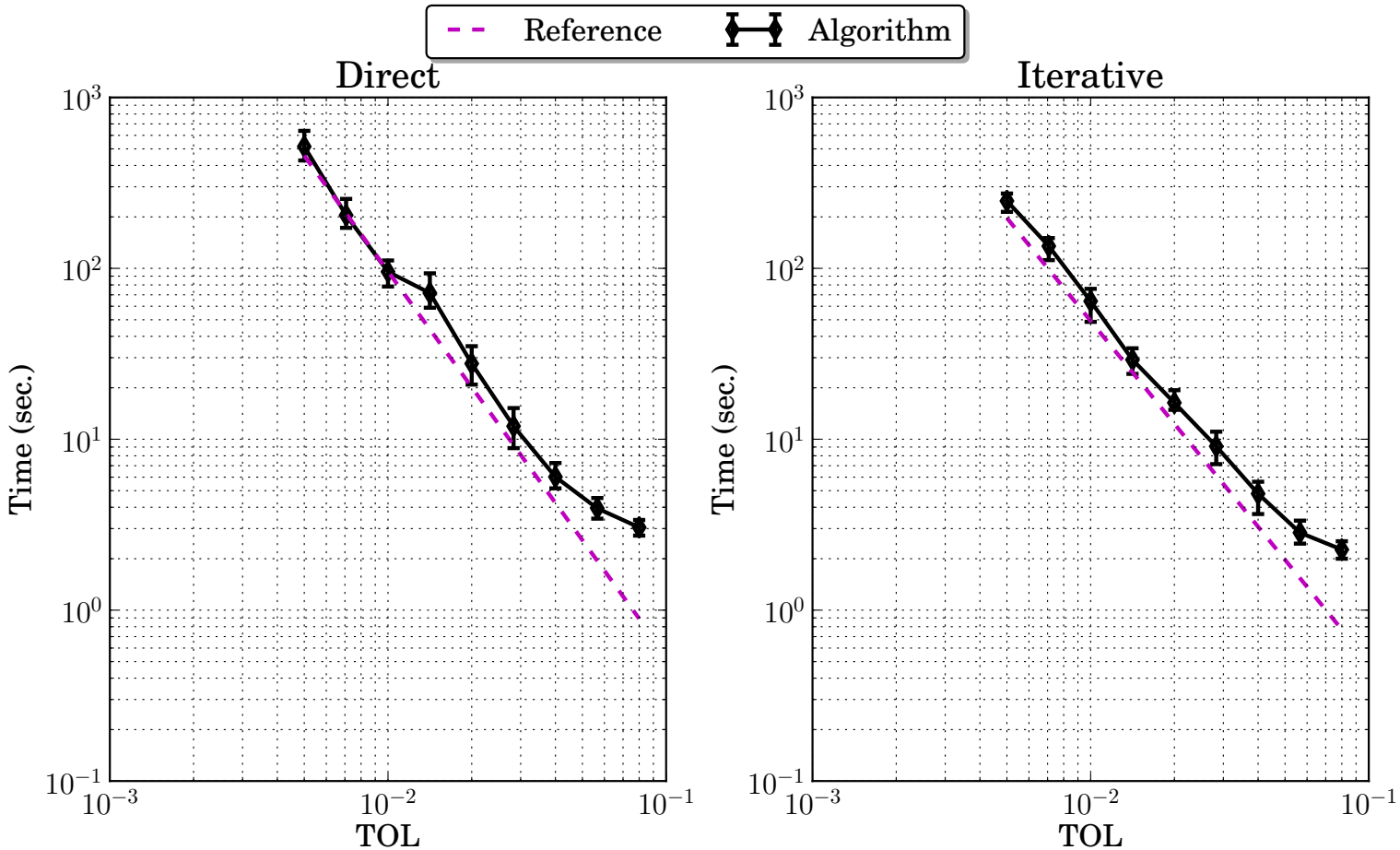


The algorithm was run with $C_\alpha = 2$ so that the bound holds with 95% confidence.

Our contribution:

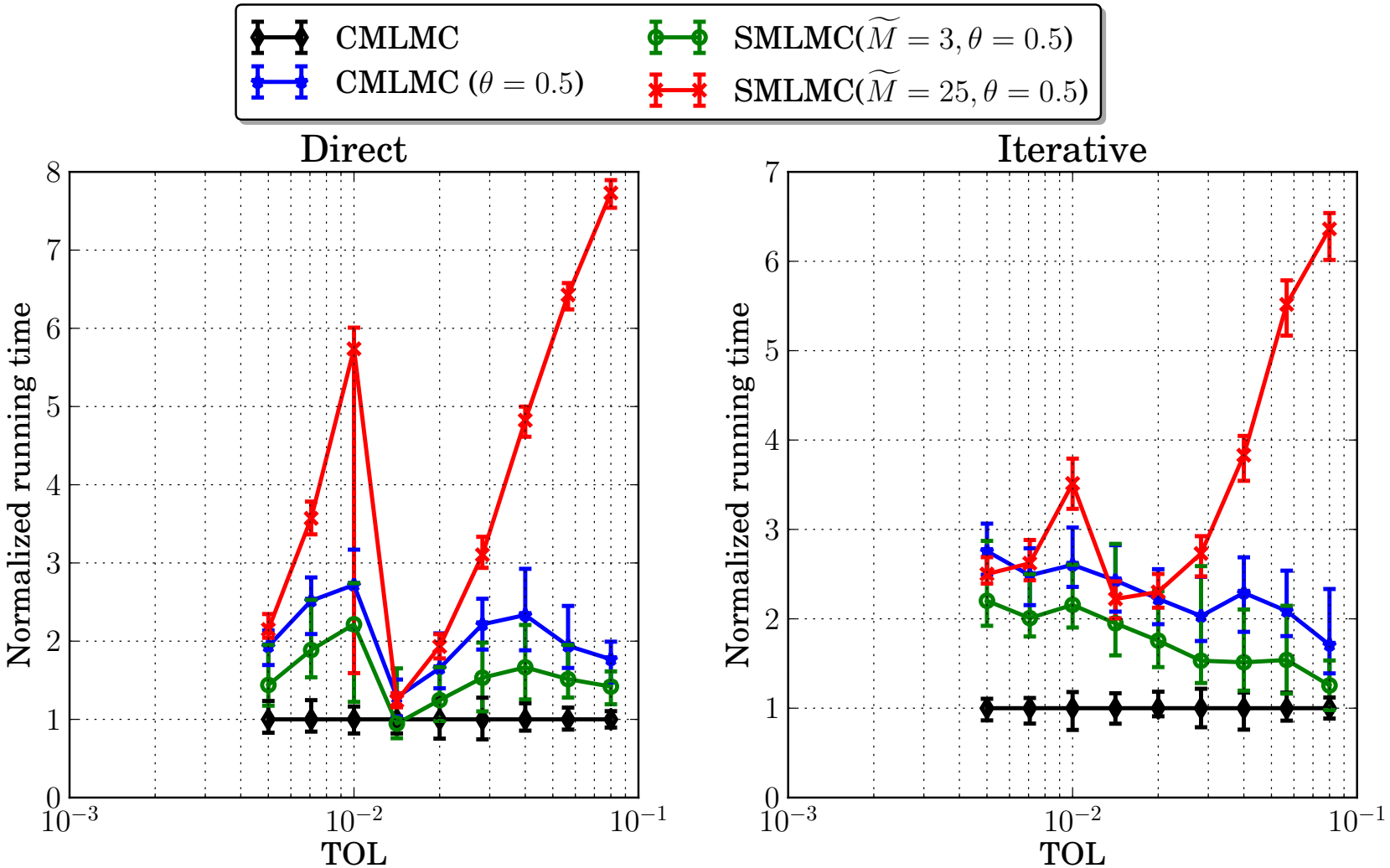
We developed a Continuation MLMC algorithm that, given a hierarchy, solves the given approximation problem for a sequence of decreasing tolerances, ending with the desired one. The sequence is chosen such that the total work is dominated by the last iteration, estimating all necessary parameters on the fly.

Measured total work of CMLMC



Reference lines are $TOL^{-2.25}$ and TOL^{-2}, respectively.

Continuation-MLMC: Running times



Improvement in running time due to better choice of splitting parameter, θ.

Conclusions

- 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.
- We show that geometric hierarchies are near-optimal. Moreover, we derive the computational complexity with known rates and constants.

We want to compute $\mathbb{E}[S] \in \mathbb{R}$. Assume $S_{\boldsymbol{\alpha}}$ is a discretization of S with discretization parameters of the form

$$h_i = \beta_i^{-\alpha_i}$$

for $i \in 1, 2, \ldots, d$. Assume $\mathbb{E}[S_{\boldsymbol{\alpha}}] \to \mathbb{E}[S]$ as $\min_{1 \le i \le d} \alpha_i \to \infty$. Define

$$\Delta_i S_{\boldsymbol{\alpha}} = \begin{cases} S_{\boldsymbol{\alpha}} & \text{if } \alpha_i = 0, \\ S_{\boldsymbol{\alpha}} - S_{\boldsymbol{\alpha} - \boldsymbol{e}_i} & \text{if } \alpha_i > 0, \end{cases} \tag{10}$$

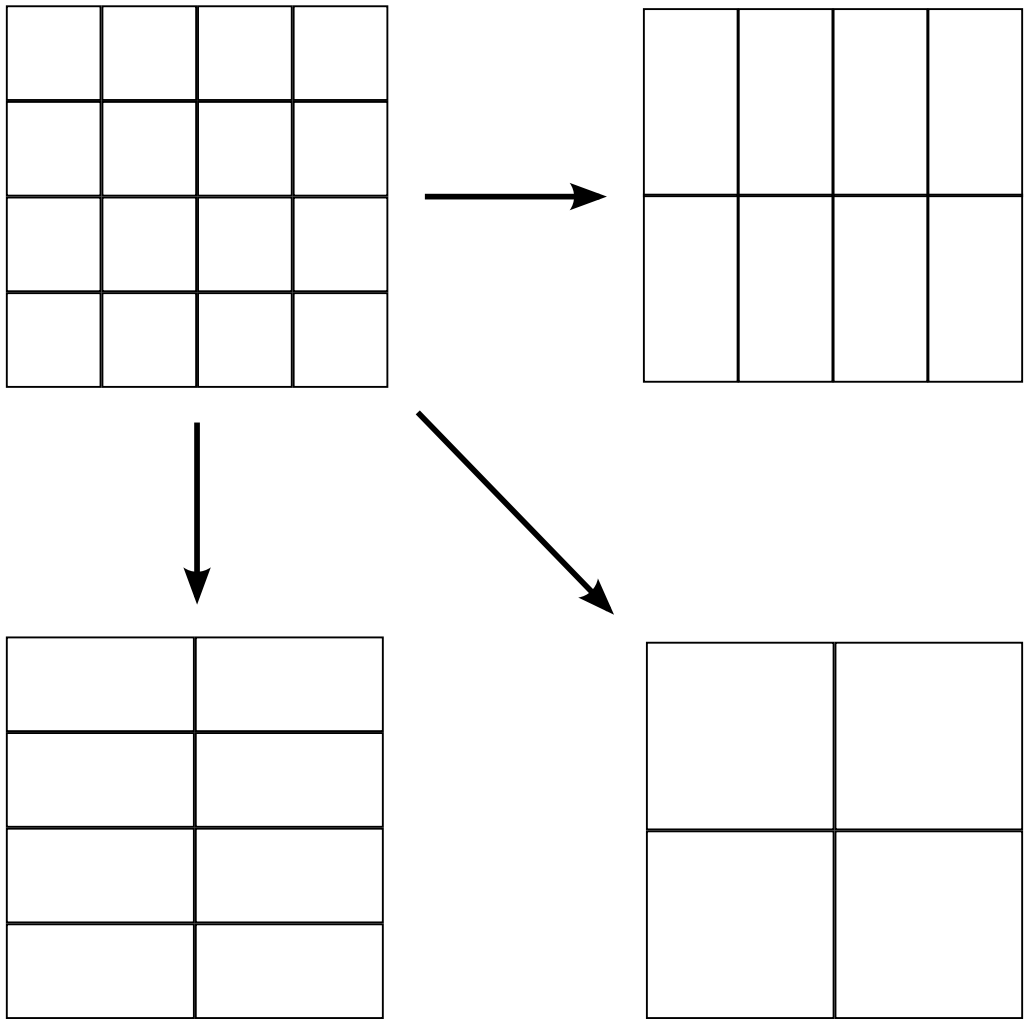
and let $\Delta S_{\boldsymbol{\alpha}} = \left(\otimes_{i=1}^{d} \Delta_i\right) S_{\boldsymbol{\alpha}}$. Then the MIMC estimator can be written as

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

with index set $\mathcal{I} \subset \mathbb{N}^d$ and independent samples.

[H-ANT] A.-L Haji-Ali, F. Nobile, R.T. *Multi Index Monte Carlo: Where Sparsity Meets Sampling.* Preprint, arXiv:1405.3757, May 2014. (See Abdul-Lateef's talk this afternoon)

Variance reduction: Further potential



Assumptions for MIMC

Provided that we satisfy mixed regularity assumptions, MIMC yields (up to log factors) the rates of MLMC with $d = 1$.

Assumption 1 (Bias) : $E_{\boldsymbol{\alpha}} = |\mathrm{E}[\Delta \mathcal{S}_{\boldsymbol{\alpha}}]| \leq Q_W \, \beta^{-w|\alpha|}$

for constants $0 < Q_W$ and $w > 0$

Assumption 2 (Variance) : $V_{\boldsymbol{\alpha}} = \mathrm{Var}\left[\Delta \mathcal{S}_{\boldsymbol{\alpha}}\right] \leq Q_S \, \beta^{-s|\alpha|}$,

for constants $0 < Q_S$ and $0 < s \leq 2w$

Assumption 3 (Work) : $W_{\boldsymbol{\alpha}} = \mathrm{Work}(\Delta \mathcal{S}_{\boldsymbol{\alpha}}) \leq C_{\mathrm{work}} \, \beta^{\gamma|\alpha|}$,

for constants $0 < C_{\mathrm{work}}$ and $0 < \gamma$

Fully Isotropic Case: Comparing MIMC to MLMC

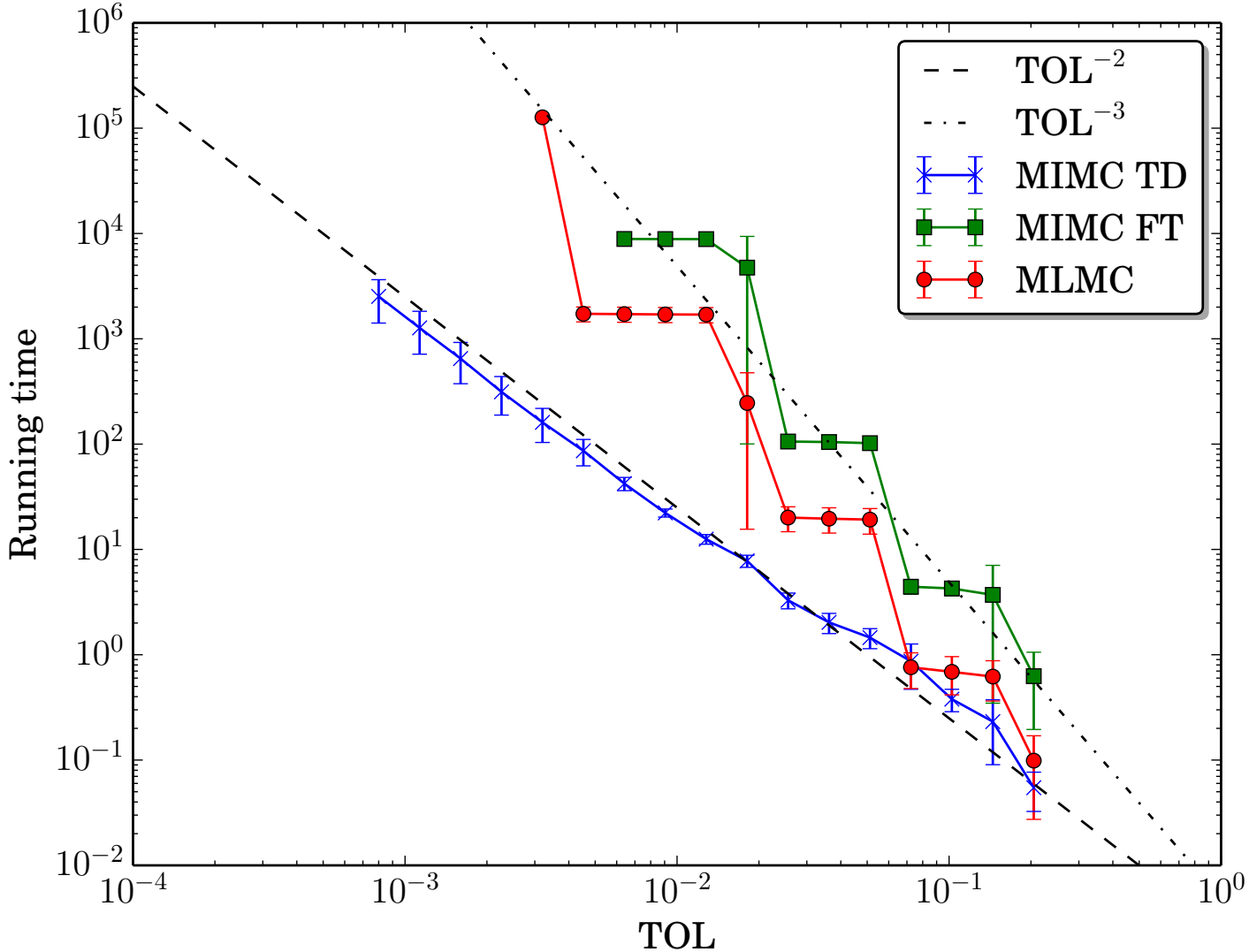
Assume $s_i = s$, $w_i = w$, $\beta_i = \beta$ and $\gamma_i = \gamma$ for all $i = 1, 2, \ldots d$. Then

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

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

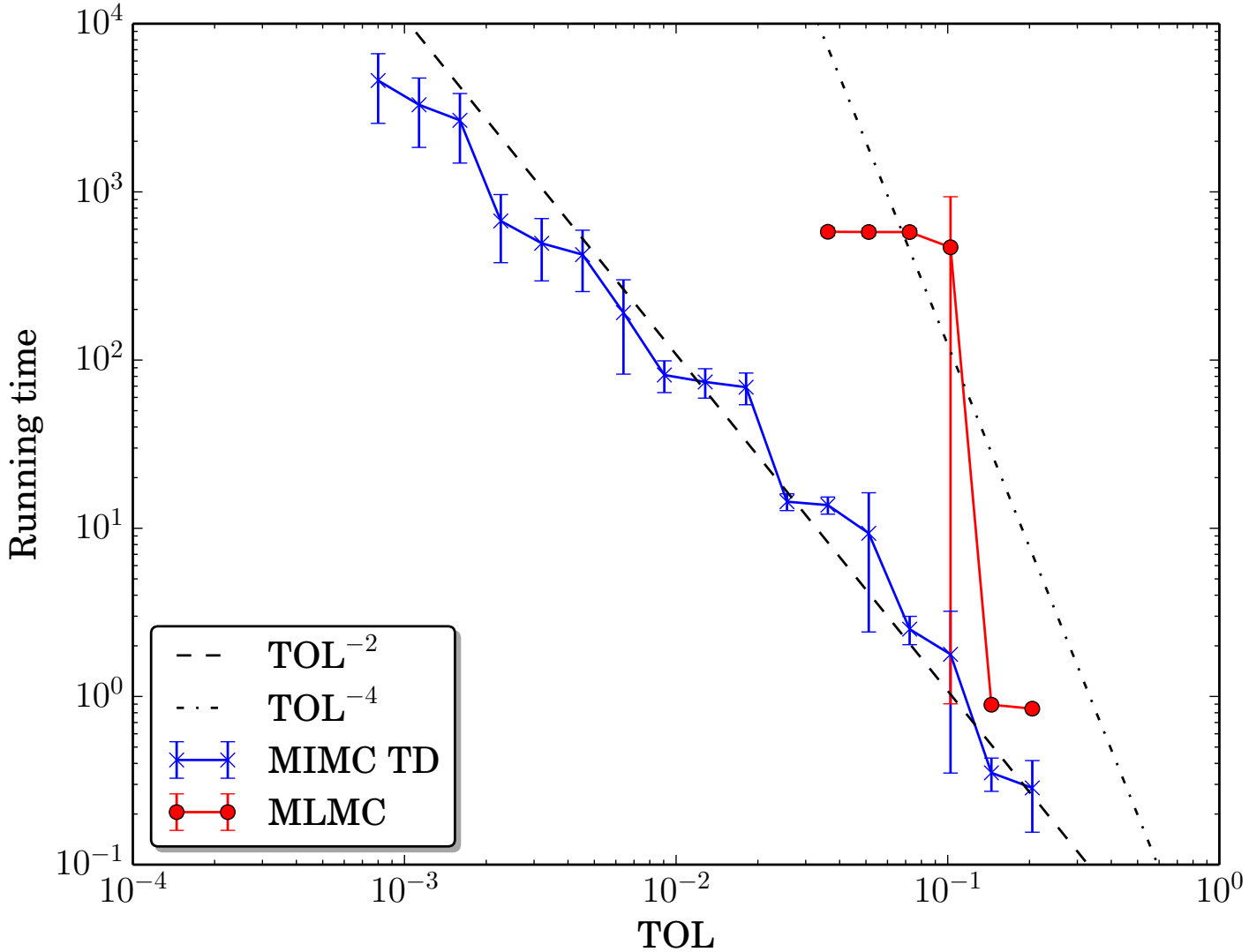
Provided that we satisfy mixed regularity assumptions, MIMC yields (up to log factors) the rates of MLMC with $d = 1$.

Numerical test: Running time, 3D problem



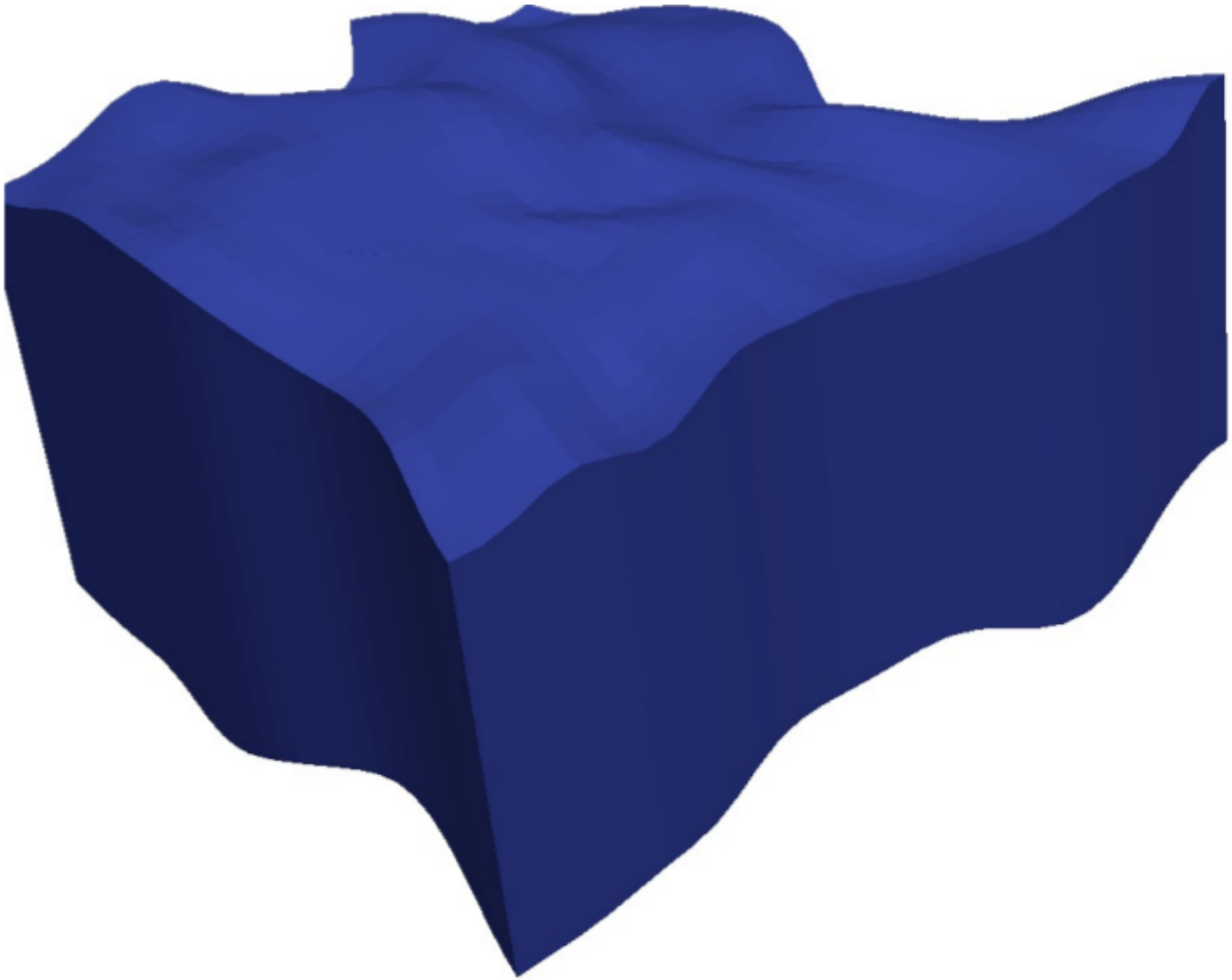
Recall that the work complexity of MC is $\mathcal{O}\left(TOL^{-5}\right)$

Numerical test: Running time, 4D problem

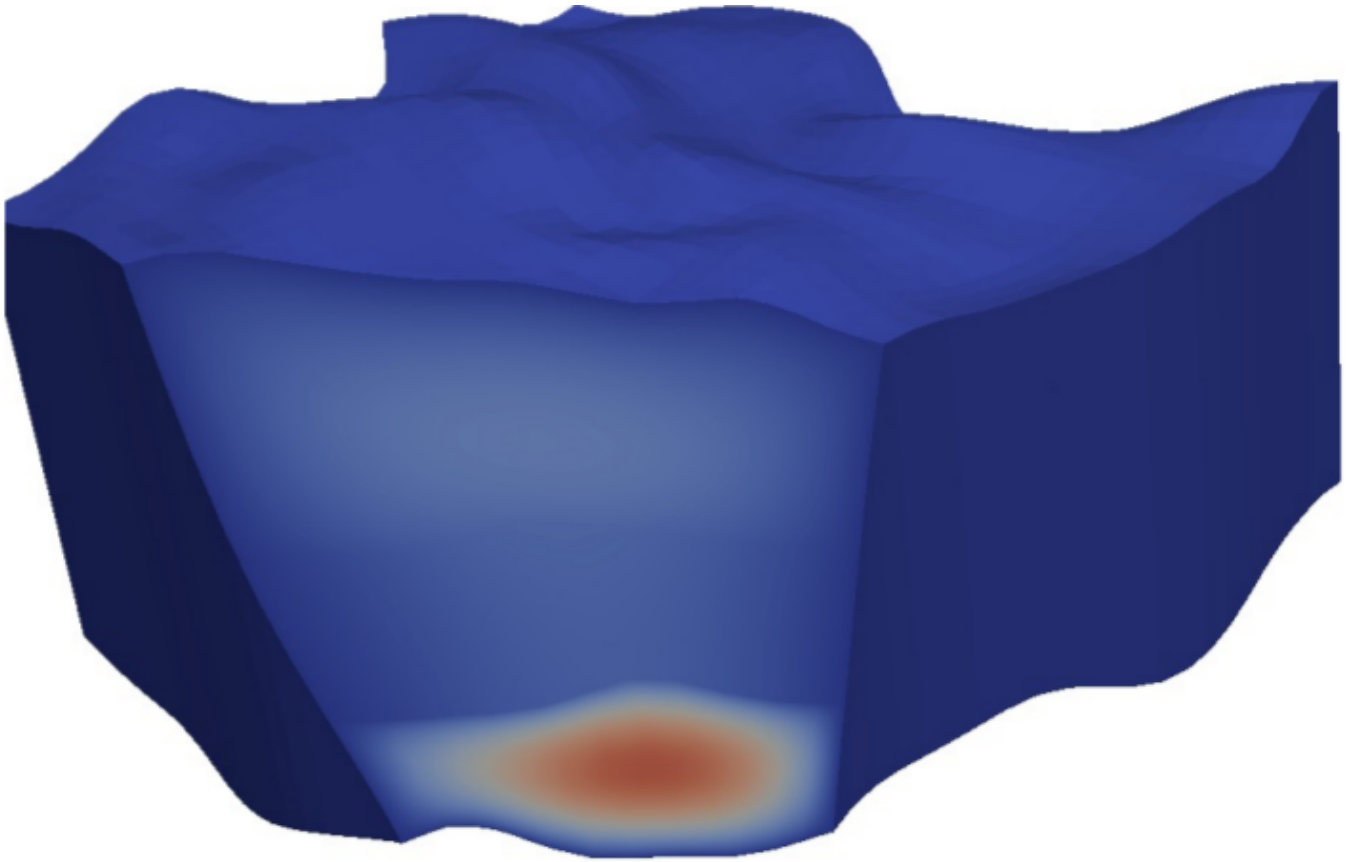


A Similar d=4 problem.

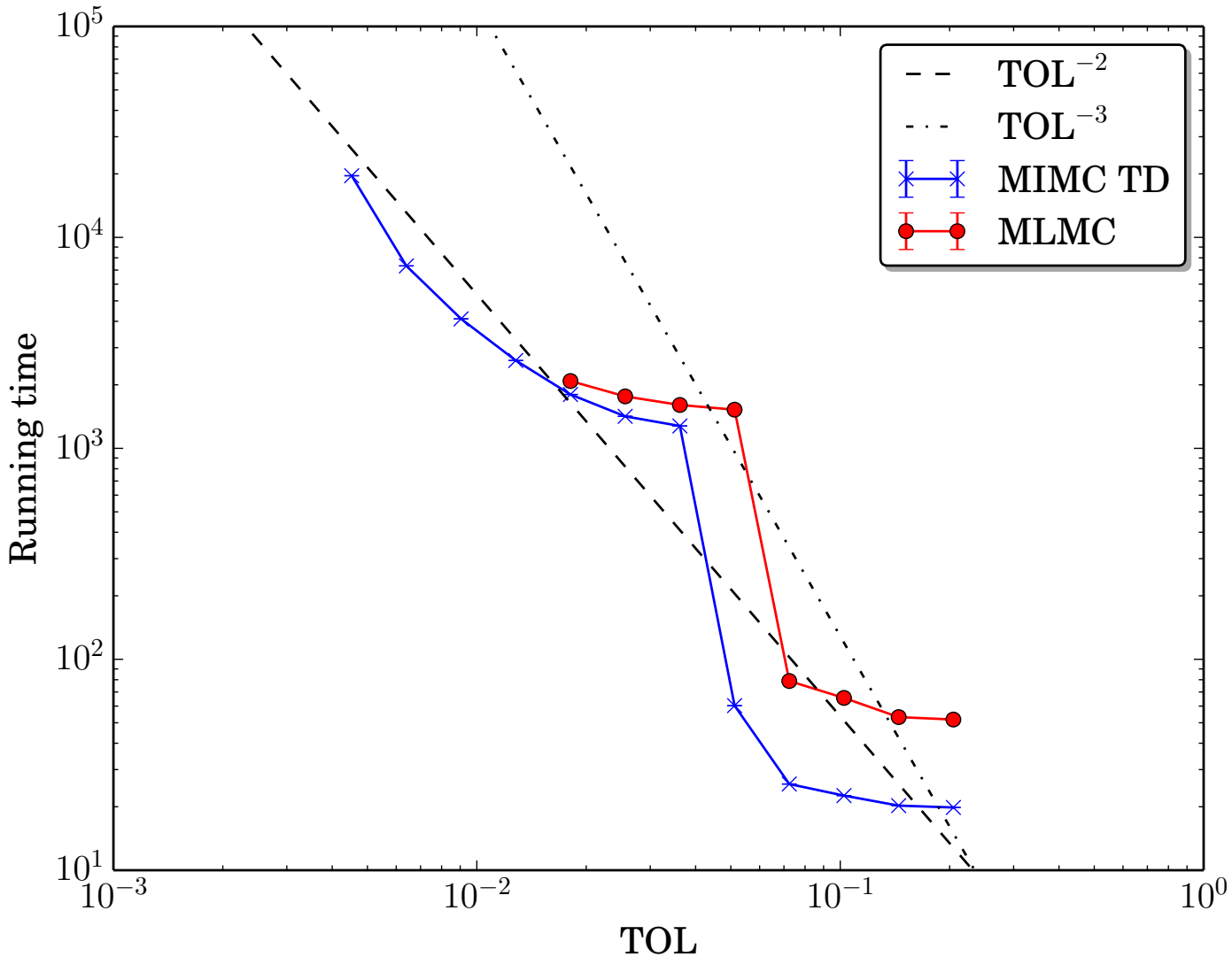
Numerical test: layered media



Numerical test: layered media



Numerical test: layered media



Conclusions and Extra Points

- MIMC may perform better than MLMC. Especially in higher dimensions.
- MIMC requires mixed regularity between discretization parameters.
- Just like MLMC reduces to MC when $L = 0$. MIMC reduces to MLMC when $d = 1$ or we do not have mixed regularity.
- MIMC can take advantage of non-isotropic behavior in different directions.
- A different set of regularity assumptions would yield a different optimal index set $\mathcal{I}$.
- A direction does not have to be a spatial dimension. It can represent any form of discretization parameter (Number of particles, Number of terms in an expansion, etc...).

Itô SDEs: Regularity Assumptions

Assumption

Assume that the following regularity conditions hold:

(i) *The functions $a(t,x)$ and $b(t,x)$ are continuous in (t,x) and are twice continuously differentiable with respect to x.*

(ii) *The partial derivatives of first and second order with respect to x of the functions a and b are uniformly bounded.*

(iii) *The function g is twice continuously differentiable, and together with its partial derivatives of first and second order it is uniformly bounded.*

Regularity on the cost to go function

Lemma (Regularity)

Suppose that a, b, g, X satisfy Assumption 1. Then the cost to go function, defined by

$$u(t,x) = E\big[g(X(T)) \mid X(t) = x\big], \tag{11}$$

satisfies the Kolmogorov equation

$$\partial_t u(t,x) + a_k \partial_k u(t,x) + d_{kn} \partial_{kn} u(t,x) = 0, \tag{12}$$

with the final condition

$$u(T,\cdot) = g.$$

Notation: $d_{jn} := \sum_k bj^k b_n^k$.

Furthermore, if the following regularity conditions are satisfied:

(i) the functions $\partial_\beta a(t,\cdot)$, $\partial_\beta b(t,\cdot)$, and are bounded uniformly in t for $1 \leq |\beta| \leq 8$;

(ii) the functions $a(\cdot,x)$, $b(\cdot,x)$ have continuous and uniformly bounded first order time derivatives;

(iii) the function g has spatial derivatives $\partial_\beta g$, with polynomial growth for $|\beta| \leq 8$;

then the function u has continuous partial derivatives with respect to x up to the order 8, satisfying the following polynomial growth condition: for all $i \in \{0,1,2\}$ and $\alpha \in \mathbb{N}^d$ with $i + |\alpha| \leq 8$ there exists $p_{\alpha,i} \in \mathbb{N}$ and $C_{\alpha,i} > 0$ such that

$$\max_{0\leq t\leq T} \left| \partial_t^i \partial_\alpha u(t,x) \right| \leq C_{\alpha,i} \left(1 + |x|^{p_{\alpha,i}} \right) \quad \forall\, x \in \mathbb{R}^d.$$

A priori [Talay and Tubaro 90],

$$E[g(X_T) - g(\overline{X}_T)] \simeq \int_0^T E[\Delta t(s)\Psi(X_s, s)]ds = \mathcal{O}(\Delta t_{max}).$$

A posteriori SDE error density [STZ01], [MSTZ06],[MSTZ08]. For $TOL \to 0$ we have

$$E[g(X_T) - g(\overline{X}_T)] \simeq \int_0^T E[\Delta t(s)\rho(\overline{X}_s, s)]ds$$

Two adaptive strategies

- Δt stochastic $\Rightarrow$ determined by error density ρ,
- Δt deterministic $\Rightarrow$ determined by error density $E[\rho]$.

Imposed bounds for the a posteriori error density

For technical reasons we impose, for exponents $\bar{\gamma}, \bar{r} > 0$,

$$TOL^{\bar{\gamma}} = \rho_{low}(TOL) \leq |\rho| \leq \rho_{up}(TOL) = TOL^{-\bar{r}} \qquad (13)$$

for instance to ensure that the refinement algorithm stops and that $\Delta t_{\max}(TOL) \to 0$ as $TOL \to 0$.
This in turn implies the a.s. convergence of the error density,

$$\rho \to \hat{\rho}, \text{ as } TOL \to 0.$$

Lemma (strong convergence, STZ01)

Suppose that a, b, g, X satisfy Assumption 1,and $\overline{X}$ is constructed using Euler Maruyama, based on the stochastic time stepping algorithm above. Then, as $TOL \to 0$ and for $p = 2, 4$ we have

$$\sup_{0 \leq t \leq T} E[|X(t) - \overline{X}(t)|^p]^{1/p} = \mathcal{O}\left(\sqrt{\frac{TOL}{\rho_{low}(TOL)}}\right) \to 0$$

A posteriori SDE weak error density

$$E[g(X_T)-g(\overline{X}_T)] = E\left[\sum_{n=0}^{N-1}\rho(t_n,\omega)(\Delta t_n)^2\right] + \mathcal{O}\left(\left(\frac{TOL}{\rho_{low}(TOL)}\right)^{1/2}\left(\frac{\rho_{up}(TOL)}{\rho_{low}(TOL)}\right)^{\epsilon}\right) E\left[\sum_{n=0}^{N-1}(\Delta t_n)^2\right], \tag{14}$$

The weak error density $\rho = \frac{1}{2}\partial_t a \cdot \varphi + \ldots$ is based on computable adjoints, i.e. $\overline{X}_{n+1} = \hat{A}(\overline{X}_n)$,

$$\begin{aligned} \varphi_n &= \partial_x \hat{A}(\overline{X}_n)\varphi_{n+1}, \\ \varphi_T &= \partial_x g(\overline{X}_T), \\ \varphi'_n &= \ldots \\ \varphi''_n &= \ldots \end{aligned}$$

New theoretical results

New results inspired by the treatment by Chow and Robbins on the accuracy and complexity of sequential stopping rules for sampling i.i.d. random variables.

" On the asymptotic theory of fixed-width sequential confidence intervals for the mean." *The Annals of Mathematical Statistics*, 36(2):pp. 457–462, 1965.

Now we can let the number of realizations in the coarsest level, $M_0(\omega)$, to be random in our analysis.

We also have a sharper computational complexity result that improves our previous analysis.

Accuracy

Lemma (strong convergence)

There exists a constant $C_G > 0$ *such that, for* $TOL_\ell = TOL_0 2^{-\ell}$ *we have*

$$\limsup_{\ell \to +\infty} Var(g_\ell - g_{\ell-1}) \frac{\rho_{low}(TOL_\ell)}{TOL_\ell} = C_G. \tag{15}$$

Theorem (Multi level accuracy)

Suppose that the modeling assumptions of Lemma 5 hold and that $\mathrm{TOL_T} \leq \mathrm{TOL_S}$. *Then the adaptive MLMC algorithm with confidence parameter* $C_C > 0$ *cf.* (**??**) *and stochastic time steps* (4) *and* (5) *satisfies*

$$\liminf_{TOL \downarrow 0} P\left(|\mathrm{E}[g(X(T))] - A(M_0)| \leq TOL\right) \geq \int_{-C_C}^{C_C} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx. \tag{16}$$

Theorem (Multi level computational complexity)

Suppose the assumptions of Lemma 5 and (6) *hold and that the lower bound for the error density is as in* (13)*. Then the work for the MLMC algorithm using stochastic time steps, as defined by*

$$\mathrm{WORK}(TOL) = \sum_{\ell=0}^{L} \mathrm{E}[M_\ell]\mathrm{E}[N_\ell], \tag{17}$$

satisfies

$$\begin{aligned} &\limsup_{TOL\downarrow 0} \frac{\mathrm{WORK}(TOL)\,TOL^2}{L^2}\, C(\bar{\gamma}\, L) \\ &\qquad \leq \frac{16\, C_C^2\, C_G}{\mathrm{TOL}_{\mathrm{T}\,Max}\, C_R} \left(\mathrm{E}\left[\int_0^T \sqrt{|\widehat{\rho}(\tau)|}\, d\tau \right] \right)^2 . \end{aligned} \tag{18}$$

(A) If $\rho_{low}(\mathrm{TOL_T}) = \rho_{\min} \in \mathbb{R}_+$ (i.e. $\bar{\gamma} = 0$) and

$$\min_{\tau \in [0,T]} |\widehat{\rho}(\tau)| \geq \rho_{\min} \text{ a.s.} \tag{19}$$

then

$$C(\bar{\gamma}\, L) = 1. \tag{20}$$

(B) If $\bar{\gamma} \to 0$ and $L\bar{\gamma} \to \infty$ as $TOL \downarrow 0$, then

$$C(\bar{\gamma}\, L) = \frac{\bar{\gamma}\, L\, \log(2)}{2^{\bar{\gamma} L}} \to 0. \tag{21}$$

Here, the number of levels is $L = \mathcal{O}\left(\log(TOL^{-1})\right)$, C_C is the confidence parameter from (**??**) , C_R and C_S are refinement parameters described by (4) and (5), C_G is the constant in the variance estimate (15), where $\mathrm{TOL_{TMax}}$ is the upper bound of the time discretization tolerance at level $\ell = 0$, and $\bar{\gamma}$ is the lower bound error density exponent; $\rho_{low}(\mathrm{TOL_T}) = \mathrm{TOL_T}^{\bar{\gamma}}$, cf. (13).

Remark (Complexity example)

Case **(B)** *of Theorem 9 implies that if the exponent of the lower error density* ρ_{low} *is given by* $\bar{\gamma}(TOL) = \log_2(\log_2(L))/L \to 0$, *then*

$$C(\bar{\gamma}\, L) = \frac{\bar{\gamma}\, L\, \log(2)}{2^{\bar{\gamma} L}} = \frac{\log_2(\log_2(L))\, \log(2)}{\log_2(L)}.$$

This is still very close to case **(A)** *and close to the standard complexity in the setting of uniform time steps.*

Remark (Uniform time steps)

$$\limsup_{TOL \downarrow 0} \frac{\mathrm{WORK}(TOL)\, TOL^2}{L^2} \leq \frac{8\, C_C^2\, C_G}{\mathrm{TOL}_{\mathrm{T}\, Max}} \mathrm{E}\left[\int_0^T \widehat{\rho}(\tau) d\tau\right]. \quad (22)$$

Lemma (M_0 asymptotic estimate)

For a given confidence interval parameter $C_C > 0$, the stopping criterion implies

$$\limsup_{TOL \to 0} \frac{E[M_0]\,\mathrm{TOL_S}^2}{L} \leq 2\,(C_C)^2\,C_G\,\frac{TOL_0}{\rho_{low}(TOL_0)} \tag{23}$$

Lemma (CLT approximation)

Assume that $Var(g_0) > 0$. *Then the multilevel estimator* $A = \mathcal{E}_{\{\mathcal{S}_\ell\}_{\ell=0}^{L}}\left(g(\overline{X}_L(T))\right)$, *satisfies the following weak convergence*

$$\frac{A - E[A]}{\sqrt{Var(A)}} \rightharpoonup N(0,1), \quad as\ TOL \to 0 \tag{24}$$

Proof: verify that Lindeberg's CLT conditions are satisfied.

MC Stopping: Problem statement

- For a sequence of i.i.d. random variables $X_1, X_2, \ldots$ and *fixed* $\mathrm{TOL} > 0$ and $\delta > 0$, determine the number of samples M that is required to ensure that

$$\mathrm{P}\left(\left|\frac{1}{M}\sum_{j=1}^{M} X_j - \mathrm{E}[X]\right| \geq \mathrm{TOL}\right) \leq \delta. \qquad (25)$$

- Assumptions: $X_1, X_2, \ldots$ have a continuous probability distribution function and $\mathrm{Var}(X) < \infty$.
- To determine $M(\mathrm{TOL}, \delta)$ by a sequential stopping rule that only uses first and second sample moments gives an unreliable result.
- In **[BHvST]** we propose a more reliable higher moments based sequential stopping rule for determining $M(\mathrm{TOL}, \delta)$.

[BHvST] Bayer, Hoel, von Schwerin, T. *On non-asymptotic optimal stopping criteria in Monte Carlo simulations.* To appear in SIAM Journal on Scientific Computing.

General Monte Carlo sequential stopping rule

General stopping rule

1. Generate a batch of M i.i.d. samples $X_1, X_2, \ldots, X_M$.
2. Estimate the probability of failure through sample moment functions $V(X_1, X_2, \ldots, X_M)$, e.g., the sample variance.
3. **If** the estimate indicate that (26) is violated, increase the number of samples M and return to step 1.
   Else; return M and **Break**.

$$\underbrace{\mathrm{P}\left(\left|\frac{1}{M}\sum_{j=1}^{M} X_j - \mathrm{E}[X]\right| \geq \mathrm{TOL}\right)}_{\text{the probability of failure}} \leq \delta. \tag{26}$$

A second moment based stopping rule

The central limit theorem motivates replacing the goal

$$\mathrm{P}\left(\left|\frac{1}{M}\sum_{j=1}^{M} X_j - \mu\right| \geq \mathrm{TOL}\right) \leq \delta \quad \textbf{by} \quad 2\left(1 - \Phi\left(\frac{\sqrt{M}\mathrm{TOL}}{\sigma}\right)\right) \leq \delta.$$

This leads to the "intuitive" stopping rule:

Algorithm 1

0. Initialize $M = M_0$.
1. Sample M realizations X_j and compute the mean and sample variance

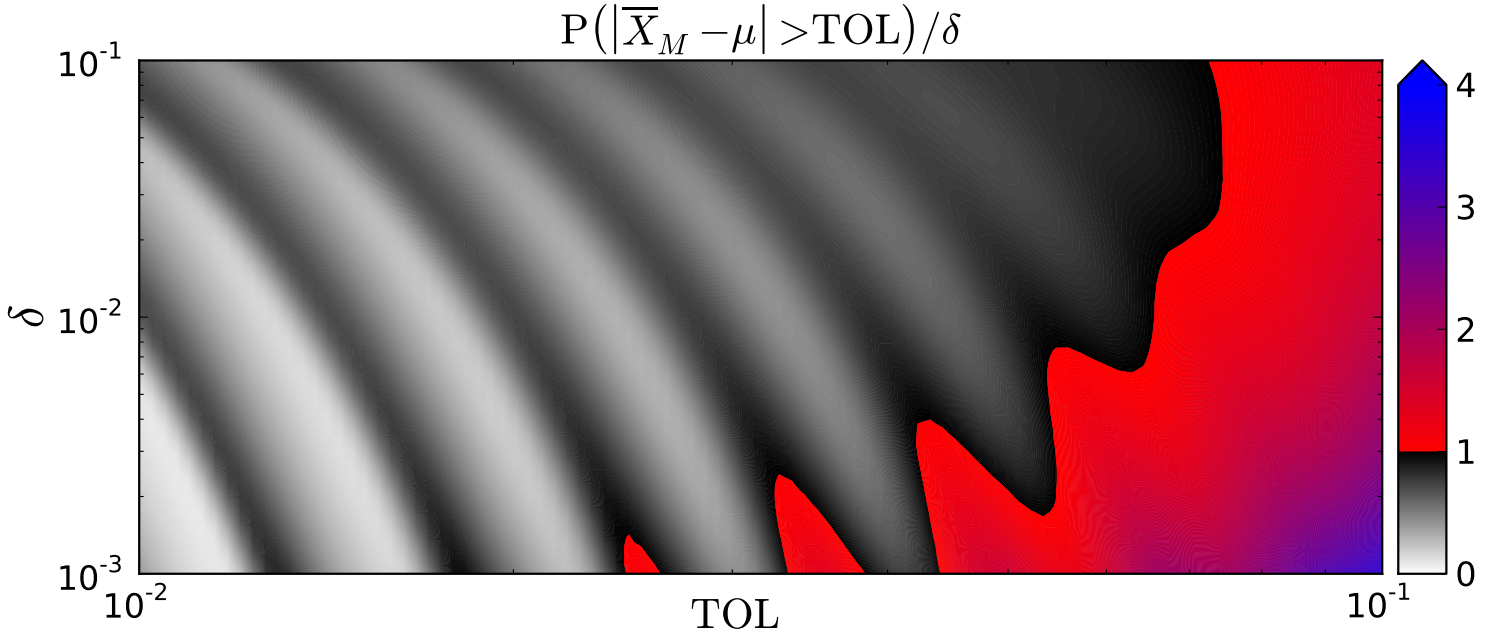
$$\overline{X}_M := \frac{1}{M}\sum_{j=1}^{M} X_j, \quad \overline{\sigma}_M^2 := \frac{1}{M-1}\sum_{j=1}^{M}(X_j - \overline{X}_M)^2.$$

2. **IF** $2\left(1 - \Phi\left(\frac{\sqrt{M}\mathrm{TOL}}{\overline{\sigma}_M}\right)\right) > \delta$; set $M = 2M$ and return to point 1. **ELSE**; output M and **BREAK**.

Numerical test of Algorithm 1

The Pareto distribution with PDF

$$p(x) = Cx^{-4.1}, \quad \text{for } x \geq 1. \qquad \text{With } \sigma = 1 \text{ and } E[X^4] = \infty.$$



Our goal: $\mathrm{P}\left(\left|\overline{X}_M - \mu\right| > \mathrm{TOL}\right) \leq \delta.$

Algorithm 1 is unreliable on heavy-tailed distributions!

A more reliable stopping rule

- Idea for a more reliable stopping rule: double number of samples M until
  $2\left(1-\Phi\left(\frac{\sqrt{M}\mathrm{TOL}}{\overline{\sigma}_M}\right)\right)+$ Penalty(higher moments) $\leq \delta$.
- Bounds involving higher moments:

Theorem (Berry-Esseen, uniform and non-uniform)

Suppose $X_1, X_2, \ldots$ *are i.i.d. r.v. with* $\mathrm{E}[X]=\mu$, $\sigma^2=\mathrm{Var}(X)$ *and* $\beta=\frac{\mathrm{E}\left[|X|^3\right]}{\sigma^3}<\infty$. *Then*

$$\left|P\left(\sum_{i=1}^{n}\frac{X_i-\mu}{\sigma\sqrt{n}}<x\right)-\Phi(x)\right|\leq\frac{C_{\mathrm{BE}}(x,\beta)}{\sqrt{n}}.$$

Bounds involving higher moments

Theorem (Edgeworth expansion)

Suppose $X_1, X_2, \ldots$ are i.i.d. random variables with a non-lattice distribution, $\mathrm{E}[X] = \mu$, $\sigma^2 = \mathrm{Var}(X)$ and $\mathrm{E}\left[X^3\right] < \infty$. Then

$$P\left(\sum_{i=1}^{n} \frac{X_i - \mu}{\sigma\sqrt{n}} < x\right) = \Phi(x) + \frac{(x^2-1)e^{-x^2/2}\mathrm{E}\left[(X-\mu)^3\right]}{6\sqrt{2\pi n}\sigma^3} + o\left(n^{-1/2}\right),$$

uniformly for $x \in \mathbb{R}$.

Bound for the variance of the sample variance

$$\mathrm{P}\left(|\overline{\sigma}^2_M - \sigma^2| \geq \sigma^2 \middle| M\right) \leq \min\left(\left(\frac{2}{M-1} + \frac{\kappa}{M}\right), 1\right),$$

with the *kurtosis*

$$\kappa = \frac{\mathrm{E}\left[|X-\mu|^4\right]}{\sigma^4} - 3.$$

Algorithm 2 – a higher moments based algorithm

0 Initialize $M = M_0$.

1 Sample M realizations X_j, compute the mean and higher order sample moments

$$\overline{X}_M := \sum_{i=1}^{M} \frac{X_i}{M}, \quad \overline{\sigma}_M := \sqrt{\sum_{i=1}^{M} \frac{(X_i - \overline{X}_M)^2}{M}}, \quad \overline{\beta}_M := \sum_{i=1}^{M} \frac{|X_i - \overline{X}_M|^3}{M\overline{\sigma}_M^3},$$

$$\widehat{\beta}_M := \sum_{i=1}^{M} \frac{(X_i - \overline{X}_M)^3}{M\overline{\sigma}_M^3}, \quad \text{and} \quad \overline{\kappa}_M := \sum_{i=1}^{M} \frac{(X_i - \overline{X}_M)^4}{M\overline{\sigma}_M^4} - 3.$$

2 **IF**

$$2\left(1 - \Phi\left(\frac{\sqrt{M}\mathrm{TOL}}{\overline{\sigma}_M}\right)\right) + 2\min\left(4\left(\frac{2}{M-1} + \frac{\overline{\kappa}_M}{M}\right), 1\right) \frac{C_{\mathrm{BE}}\left(\frac{\sqrt{M}\mathrm{TOL}}{\overline{\sigma}_M}, \overline{\beta}_M\right)}{\sqrt{M}}$$

$$+ \left(1 - \min\left(4\left(\frac{2}{M-1} + \frac{\overline{\kappa}_M}{M}\right), 1\right)\right) \frac{\left|\frac{M\mathrm{TOL}^2}{\overline{\sigma}_M^2} - 1\right| \exp\left(-\frac{M\mathrm{TOL}^2}{\overline{\sigma}_M^2}\right) |\widehat{\beta}_M|}{3\sqrt{2\pi M}} > \delta,$$

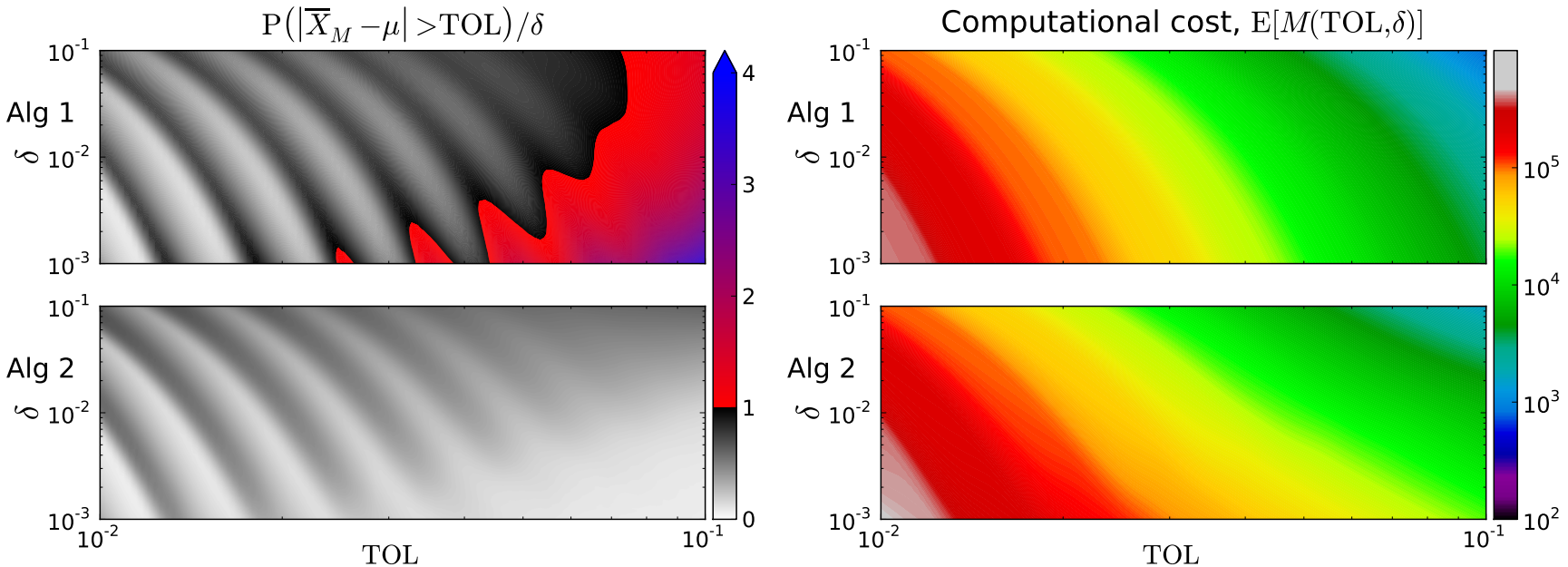
then set $M = 2M$ and go to point 1.
ELSE output M and **BREAK**.

Numerical examples

The Pareto distribution with PDF

$$p(x) = Cx^{-4.1}, \quad \text{for } x \geq 1,$$

and moments $\sigma = 1$ and $\kappa = \infty$ (since $\mathrm{E}\left[X^4\right] = \infty$).



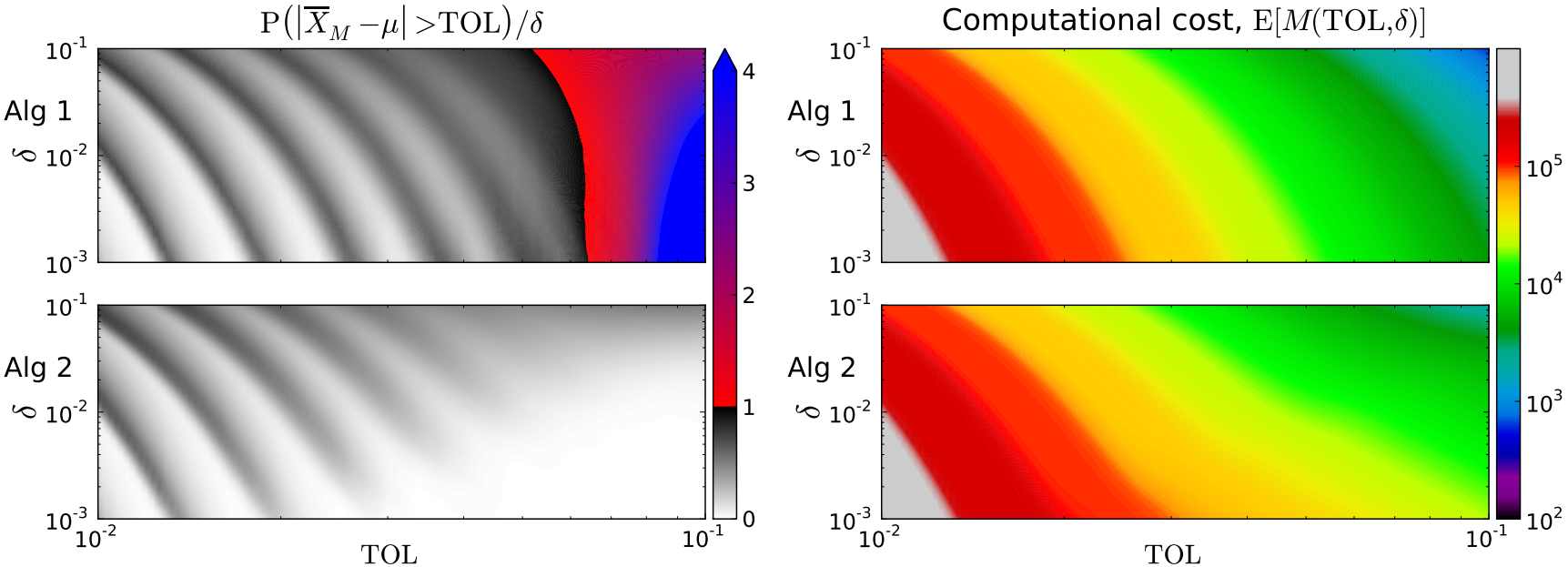
Our goal: $\mathrm{P}\left(\left|\overline{X}_M - \mu\right| > \mathrm{TOL}\right) \leq \delta.$

Numerical examples

The Normal-inverse Gaussian with PDF

$$p(x) = \frac{\alpha_1 \alpha_4 K_1(\alpha_1 \sqrt{\alpha_4^2 + (x-\alpha_5)^2})}{\pi \sqrt{\alpha_4^2 + (x-\alpha_5)^2}} e^{\alpha_3 \alpha_4 + \alpha_2 (x-\alpha_5)},$$

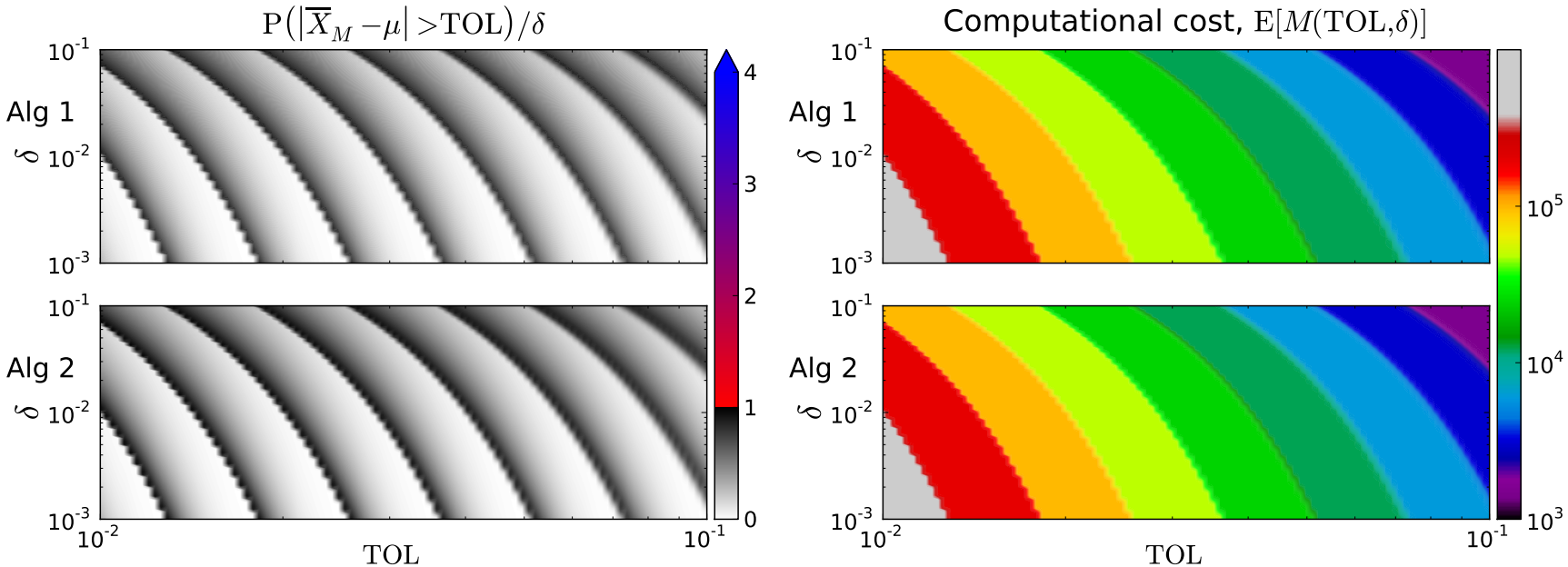
with the parameters α_i chosen so that $\sigma = 1$ and $\kappa = 123$.



Our goal: $\mathrm{P}\left(\left|\overline{X}_M - \mu\right| > \mathrm{TOL}\right) \leq \delta.$

Numerical examples

The uniform distribution $X_i \sim U(-\sqrt{3}, \sqrt{3})$ with $\sigma = 1$ and $\kappa = -6/5$.



Our goal: $\mathrm{P}\left(\left|\overline{X}_M - \mu\right| > \mathrm{TOL}\right) \leq \delta.$

Summary

- For a given sequence of i.i.d. r.v. $X_1, X_2, \ldots$, we considered the problem of determining $M(\mathrm{TOL}, \delta)$ by a sequential stopping rule such that

$$\mathrm{P}\left(\left|\frac{1}{M}\sum_{j=1}^{M} X_j - \mathrm{E}[X]\right| \geq \mathrm{TOL}\right) \leq \delta$$

  for fixed $\mathrm{TOL},\ \delta > 0$.
- Showed that the "intuitive" sample variance based stopping rule (Algorithm 1) performs unreliably on heavy-tailed distributions.
- Proposed a higher moments based stopping rule (Algorithm 2) which performs more reliably than Algorithm 1, but which also requires slightly more computational cost.
- For more, see:

[BHvST] Bayer, Hoel, von Schwerin, Tempone. *On non-asymptotic optimal stopping criteria in Monte Carlo simulations.* SIAM Journal on Scientific Computing, Vol. 36, Issue 2, 2014.