Adaptive Multilevel Monte Carlo Simulation of Stochastic Ordinary Differential Equations

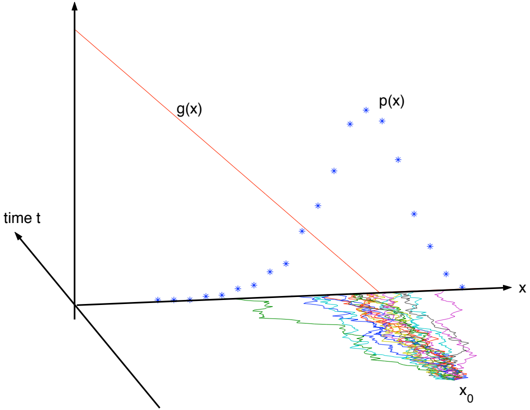
H. Hoel[1] E. von Schwerin[2] A. Szepessy[1] R. Tempone[2].

[1]Department of Numerical Analysis (CSC),
Royal Institute of Technology

[2]Division of Applied Mathematics and Computational Science,
King Abdullah University of Science and Technology

BIRS Meeting, Stochastic Multiscale Methods, Banff,
2011-04-01

Weak approximation of SDE



Outline

1. Formulation of SDE approximation
2. Single level Monte Carlo
3. Multilevel Monte Carlo
4. Adaptive multilevel Monte Carlo

Problem formulation

For the Ito SDE

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

$$X_0 = x_0, \tag{2}$$

and $g : \mathbb{R}^d \to \mathbb{R}$, approximate $E[g(X_T)]$ to a given accuracy TOL.

W_t is a K-dimensional Wiener process.

Euler Maruyama Method

1. Forward Euler (Euler Maruyama) scheme

$$\overline{X}_{n+1} = \overline{X}_n + a(\overline{X}_n, t_n)\Delta t_n + \sum_{k=1}^{K} b^k(\overline{X}_n, t_n)\Delta W_n^k \qquad (3)$$

   gives approximate realisations $\overline{X}_T(\omega)$ on a grid
   $t_0 = 0 < t_1 < \ldots < t_N = T$.
   $\Delta t_n = t_{n+1} - t_n$, $\Delta W_n^k = W_{n+1}^k - W_n^k$

2. Monte Carlo estimate

$$E[g(X_T)] \approx \sum_{i=1}^{M} \frac{g(\overline{X}_T(\omega_i; \Delta t))}{M} \qquad (4)$$

The error contributions

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:

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

Error Control and Complexity

Weak convergence for smooth drift and diffusion:

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

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

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

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

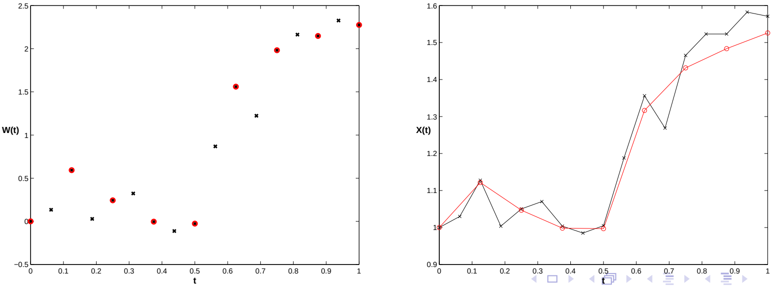
$M \propto \frac{1}{TOL^2}$ needed for sufficient probability that $\left| E[g(\overline{X}_T)] - \sum_{i=1}^{M} \frac{g(\overline{X}_T(\omega_i; \Delta t))}{M} \right| \leq O(TOL_S)$.

Computational complexity $= M\frac{T}{\Delta t} \propto 1/TOL^3$.

Variance reduction

Control variate: $E[Z]$ unkown, $E[Y]$ known,
$E[Z] = E[Z - Y] + E[Y] \approx \frac{1}{M}\sum_{i=1}^{M}(Z(\omega_i) - Y(\omega_i)) + E[Y]$

Use $g(\overline{X}_T(\cdot;\Delta t))$ and $g(\overline{X}_T(\cdot;2\Delta t))$ for Y and Z
Order 1/2 strong convergence of X_T. Assume e.g. uniform Lipschitz g. Then
$Var\left(g(X_T(\cdot;\Delta t)) - g(X_T(\cdot;2\Delta t))\right) = O(\Delta t)$



Giles' multilevel idea 2006

On a hierarchy of uniform grids $\Delta t_\ell = \Delta t_0/2^\ell$, $\ell = 0, \ldots, L$, let $g_\ell = g(\overline{X}_T(\cdot; \Delta t_\ell))$.

Step 1 Write the telescopic sum

$$E[g_L] = E[g_0] + \sum_{\ell=1}^{L} E[g_\ell - g_{\ell-1}].$$

Step 2 Now use $L+1$ batches, each with M_ℓ independent realizations, $\ell = 0, \ldots, L$ to create the estimator

$$A(M_0) = \sum_{i_0=1}^{M_0} \frac{g_0(\omega_{i_0})}{M_0} + \sum_{\ell=1}^{L} \sum_{i_\ell=1}^{M_\ell} \frac{(g_\ell - g_{\ell-1})(\omega_{i_\ell})}{M_\ell}.$$

Giles' multilevel idea 2006

With $Var(g_\ell - g_{\ell-1}) = O(\Delta t_\ell) = O(\Delta t_0/2^\ell)$ choose $M_\ell = M_0/2^\ell$ so that

$$Var(A) = \frac{Var(g_0)}{M_0} + \sum_{\ell=1}^{L} \frac{Var(g_\ell - g_{\ell-1})}{M_\ell} = \frac{O(1)}{M_0}(1 + \Delta t_0 L)$$

To achieve $Var(A) = O(TOL^2)$ take

$$M_0 \propto (TOL^{-2}(1 + \Delta t_0 L))$$

giving the total work to achieve accuracy TOL

$$Work = \sum_{\ell=0}^{L} \frac{M_\ell}{\Delta t_\ell} = O\left((1+L)\frac{M_0}{\Delta t_0}\right)$$

$$= O\left(\left(\log_2(\Delta t_0/TOL)\, TOL^{-1}\right)^2 \right)$$

Adaptivity

Given TOL_T, use adaptive refinements to generate 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.

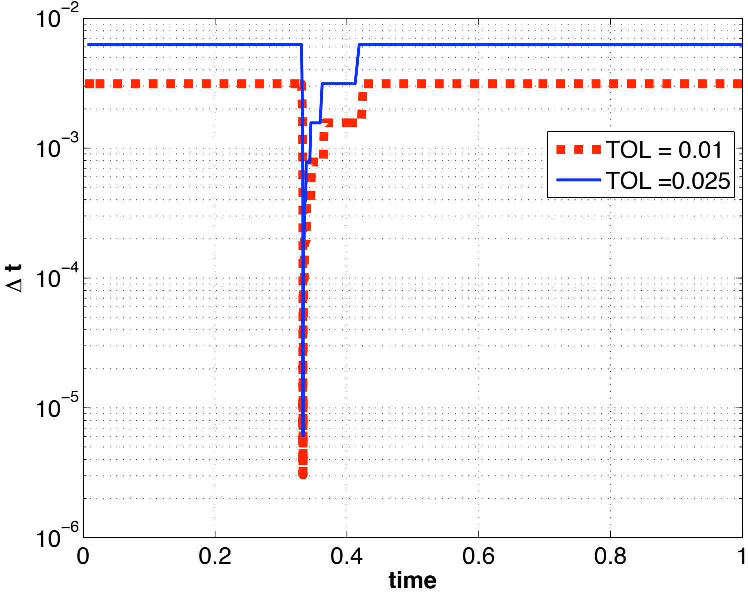
How? Adaptive refinements start from a coarse initial grid, and

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

(2) as long as $\max_n r_n \geq C_S \frac{TOL_T}{E[N]}$,

(3) refine all time steps s.t. $r_n \geq C_R \frac{TOL_T}{E[N]}$, refine sampling by Brownian bridges, and go to (1)

Multilevel: Grid hierarchy defined by $TOL_{T,\ell} = TOL_{T,L} 2^{L-\ell}$.



10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
Δ t
0
0.2
0.4
0.6
0.8
1
time
TOL = 0.01
TOL =0.025

Time discretization, weak approximation of SDE

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]

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

Two adaptive strategies

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

A posteriori SDE error density

$$E[g(X_T)-g(\overline{X}_T)] = E\left[\sum_{n=0}^{N-1} \tilde{\rho}(t_n,\omega)(\Delta t_n)^2\right] + \mathcal{O}\left(\left(\frac{TOL}{\rho_{low}(TOL)}\right)^{1/2}\left(\frac{\rho_{up}(TOL)}{\rho_{low}(TOL)}\right)^{\epsilon}\right) E\left[\sum_{n=0}^{N-1}(\Delta t_n)^2\right], \tag{5}$$

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

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

Bounds for the error density

For technical reasons we impose

$$\rho_{low}(TOL) \leq |\rho| \leq \rho_{up}(TOL)$$

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

$$\rho \to \hat{\rho},$$

as $TOL \to 0$.

Idea adaptive multilevel algorithm

Given tolerance TOL_T, TOL_S, inital grid Δt, M_0 and $L = O(-\log_2(TOL))$

1. Set $M_\ell = 2^{-\ell} M_0$.

Idea adaptive multilevel algorithm

Given tolerance TOL_T, TOL_S, inital grid Δt, M_0 and $L = O(-\log_2(TOL))$

1. Set $M_\ell = 2^{-\ell} M_0$.
2. Compute M_0 realizations of $g_0(\omega)$ on **adaptive** grids. Compute M_ℓ realizations of $(g_\ell - g_{\ell-1})(\omega)$ by succesively halving the tolerance from $TOL_{T,0}$ to $TOL_{T,\ell-1}$ and $TOL_{T,\ell}$ on **adaptive** grids s.t. $E[g(X_T) - g_\ell] \leq TOL_T 2^{L-\ell}$.

Idea adaptive multilevel algorithm

Given tolerance TOL_T, TOL_S, inital grid Δt, M_0 and $L = O(-\log_2(TOL))$

1. Set $M_\ell = 2^{-\ell} M_0$.
2. Compute M_0 realizations of $g_0(\omega)$ on **adaptive** grids. Compute M_ℓ realizations of $(g_\ell - g_{\ell-1})(\omega)$ by succesively halving the tolerance from $TOL_{T,0}$ to $TOL_{T,\ell-1}$ and $TOL_{T,\ell}$ on **adaptive** grids s.t. $E[g(X_T) - g_\ell] \leq TOL_T 2^{L-\ell}$.
3. Compute $A(M_0) = \sum_{i_0=1}^{M_0} \frac{g_0(\omega_{i_0})}{M_0} + \sum_{\ell=1}^{L} \sum_{i_\ell=1}^{M_\ell} \frac{(g_\ell - g_{\ell-1})(\omega_{i_\ell})}{M_\ell}$. and its "sample variance" $V(A(M_0)) := \frac{V_{M_0}(g_0)}{M_0} + \sum_{\ell=1}^{L} \frac{V_{M_\ell}(g_\ell - g_{\ell-1})}{M_\ell}$.

Idea adaptive multilevel algorithm

Given tolerance TOL_T, TOL_S, inital grid Δt, M_0 and $L = O(-\log_2(TOL))$

1. Set $M_\ell = 2^{-\ell} M_0$.
2. Compute M_0 realizations of $g_0(\omega)$ on **adaptive** grids. Compute M_ℓ realizations of $(g_\ell - g_{\ell-1})(\omega)$ by succesively halving the tolerance from $TOL_{T,0}$ to $TOL_{T,\ell-1}$ and $TOL_{T,\ell}$ on **adaptive** grids s.t. $E[g(X_T) - g_\ell] \leq TOL_T 2^{L-\ell}$.
3. Compute $A(M_0) = \sum_{i_0=1}^{M_0} \frac{g_0(\omega_{i_0})}{M_0} + \sum_{\ell=1}^{L} \sum_{i_\ell=1}^{M_\ell} \frac{(g_\ell - g_{\ell-1})(\omega_{i_\ell})}{M_\ell}$. and its "sample variance" $V(A(M_0)) := \frac{V_{M_0}(g_0)}{M_0} + \sum_{\ell=1}^{L} \frac{V_{M_\ell}(g_\ell - g_{\ell-1})}{M_\ell}$.
4. If $V(A(M_0)) > \frac{TOL_S^2}{C_C}$, statistical error is too large: Set $M_0 = 2M_0$ and go to (1).

Drift singularity

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

$$dX_t = \begin{cases} X_t dW_t, & t \in [0, \alpha] \\ \frac{X_t}{2\sqrt{t-\alpha}} dt + X_t dW_t, & t \in (\alpha, T] \end{cases}$$

$$X_0 = 1,$$

with the unique solution

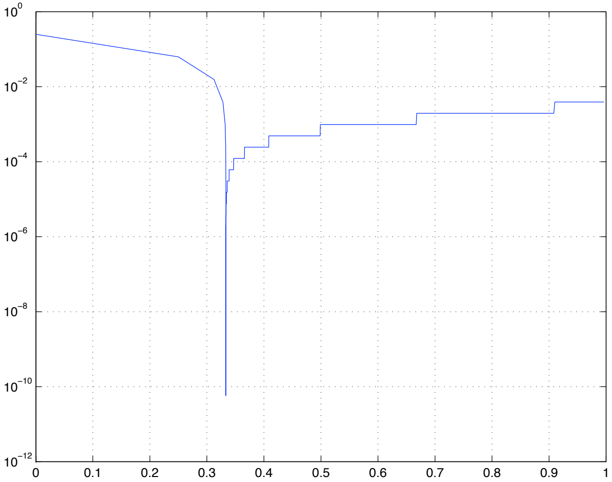
$$X_t = \begin{cases} \exp(W_t - t/2), & t \in [0, \alpha] \\ \exp(W_t - t/2) \exp(\sqrt{t-\alpha}), & t \in (\alpha, T]. \end{cases}$$

Goal: Approximate $E[X_T] = \exp(\sqrt{t-\alpha})$ with $T = 1$ and $\alpha = T/3$.

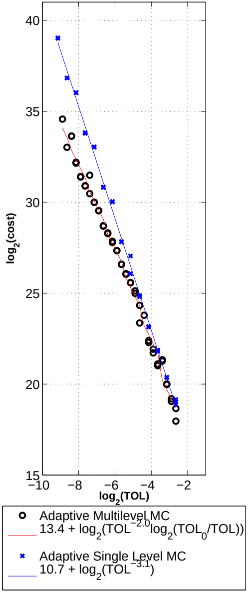
Drift singularity adaptive strategy

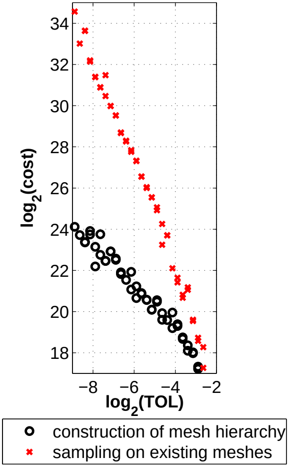
- Drift singularity at a deterministic time
- Grid generation phase – use sample averaged error indicators to generate the grid hierarchy
- Production phase – control statistical error by performing multilevel simulations on the existing grid hierarchy

Experimental Complexity: Adapted time step size

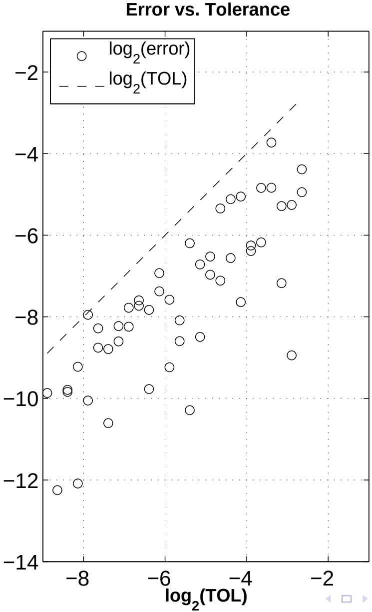


Experimental Complexity: Drift Singularity





Experimental Complexity: Drift Singularity



Stopped diffusion:

Example Stopped diffusion:

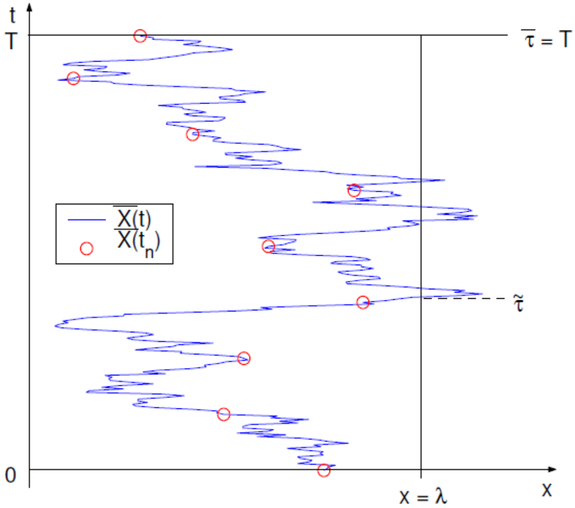
$$dX_t = \begin{cases} \frac{11X_t}{36}dt + \frac{X_t}{6}dW_t, & \text{for } t \in [0,2] \text{ and } X_t \in (-\infty, 2) \\ 0 \quad (\text{Stopped!}), & \text{if } X_t = 2, \end{cases}$$
$$X_0 = 1.6,$$
$$g(x,t) = x^3 e^{-t}$$

compute

$E[g(X_\tau, \tau)]$ τ stopping time

- Weak convergence uniform grid: $O(1/\sqrt{N})$.
- Adaptive grid: $O(1/N)$.

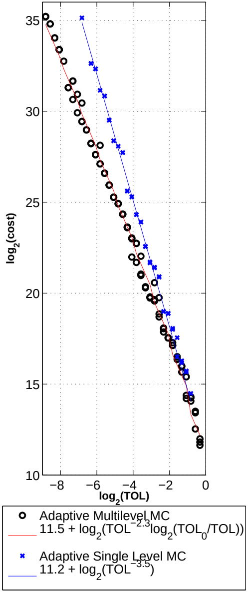
Hitting error

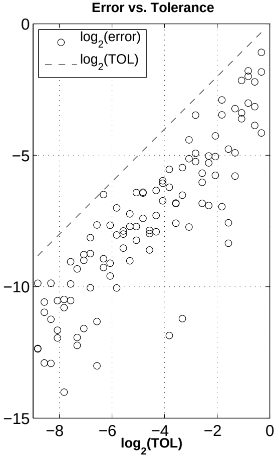


Stopped diffusion adaptive strategy

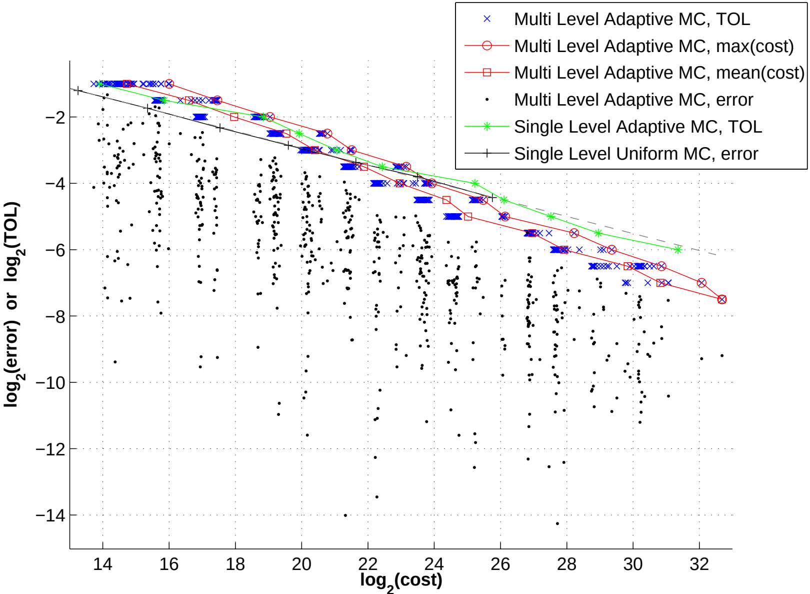
- Take small steps when the path is close to the barrier
- Generate a new adaptive mesh pair for each realization

Experimental Complexity: Barrier





Experimental Complexity and Accuracy



Lemma (Stopping)

Suppose the adaptive algorithm applies the mesh refinement strategy described before on a set of realizations having the same uniform initial mesh of step size Δt_0*. Then, given a prescribed accuracy parameter* $\mathrm{TOL_T} > 0$*, the adaptive refinement algorithm stops after a finite number of iterations.*

Proof: Uses the imposed upper bound on the approximate error density.

Lemma (strong convergence)

*Suppose that a, b, g, X satisfy the assumptions in Lemma **??**,that $\overline{X}$ is constructed by the forward Euler method, based on the stochastic time stepping algorithm above. Then*

$$\sup_{0 \le t \le T} E[|X(t) - \overline{X}(t)|^2] = \mathcal{O}\left(\frac{TOL}{\rho_{low}(TOL)}\right) \to 0$$

Lemma (strong convergence)

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

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

Lemma (Variance Estimate)

Choose the number of realizations on each level, M_ℓ, as follows

$$M_\ell = \left\lceil M_0 \frac{\rho_{low}(TOL_0)\, TOL_\ell}{\rho_{low}(TOL_\ell)\, TOL_0} \right\rceil . \tag{6}$$

Then the variance of the multilevel estimator
$A = \mathcal{E}_{\{\mathcal{S}_\ell\}_{\ell=0}^{L}}\left(g(\overline{X}_L(T))\right)$ *satisfies*

$$\limsup_{TOL \to 0} Var(A) \frac{M_0}{L(\mathrm{TOL_T})} \leq \frac{C_G\, TOL_0}{\rho_{low}(TOL_0)} . \tag{7}$$

Let $M_0(TOL)$ be such that

$$Var(A(M_0)) \leq \frac{\mathrm{TOL_S}^2}{C_C^2}.$$

Lemma (M_0 asymptotic estimate)

For a given confidence interval parameter $C_C > 0$, the stopping criterion and the bound (7) imply

$$\limsup_{TOL \to 0} \frac{E[M_0]\,\mathrm{TOL_S}^2}{L} \leq 2\,(C_C)^2\,C_G\,\frac{TOL_0}{\rho_{low}(TOL_0)} \qquad (8)$$

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*

$$\frac{A - E[A]}{\sqrt{Var(A)}} \rightharpoonup N(0,1), \quad as \ TOL \to 0 \tag{9}$$

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

Accuracy

Choose M_0 deterministically, for instance by using an upper bound on the variance and imposing

$$\begin{aligned} Var(A) &\leq C\frac{L}{M_0} \\ &\leq \left(\frac{\mathrm{TOL_S}}{C_C}\right)^2. \end{aligned} \tag{10}$$

Then, by the CLT result, for any given $y > 0$

$$P\left(\frac{|E[A]-A|}{\mathrm{TOL_S}} \leq y\right) \geq P\left(\frac{|E[A]-A|}{\sqrt{Var(A)}} \leq C_C y\right) \to \frac{1}{\sqrt{2\pi}}\int_{-C_C y}^{C_C y} e^{-\frac{x^2}{2}}dx \text{ as } TOL \to 0.$$

Theorem (Accuracy)

Suppose that the assumptions of Lemma **??** *hold. Then, for any confidence interval parameter $C_C > 0$ in (10) and refinement stopping parameters C_R, C_S the adaptive algorithm with stochastic time steps satisfies*

$$\liminf_{TOL \to 0+} P\left(\frac{\left| E\left[g(X(T))\right] - \mathcal{E}_{\{\mathcal{S}_\ell\}_{\ell=0}^{L}}\left(g(\overline{X}_L(T))\right)\right|}{TOL} \leq \frac{C_S}{2} + \frac{1}{2} \right) \geq \int_{-C_C}^{C_C} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx. \tag{11}$$

Here $\mathrm{TOL_S} = \mathrm{TOL_T} = TOL/2$.

Theorem (Multi level efficiency)

Suppose that the regularity assumptions of Lemma **??** *hold. Choose the number of realizations on each level according to* (6)*. Then the expected value of the final computational work,*

$$E[Work] = E[M_0]E[N_0] + \sum_{\ell=1}^{L} E[M_\ell]\{E[N_\ell] + E[N_{\ell-1}]\}$$

corresponding to the adaptive steps satisfies asymptotically

$$\limsup_{TOL\to 0+} \frac{\mathrm{TOL_S}\, E[Work]}{E[N_{opt}]\, L \sum_{\ell=1}^{L} \rho_{low}^{-1}(TOL_\ell)} \leq C_G (C_C)^2 \frac{28}{C_R} \qquad (12)$$

Corollary

Assume that in our adaptive algorithms we impose a lower bound for the error density of the form $\rho_{low}(TOL) = TOL^{\gamma}$ *e.g.* $\gamma = 1/9$. *Then we have the following estimate for the computational work,*

$$E[Work(TOL)] = \mathcal{O}\left(TOL^{-(2+\gamma)} \log\left(\frac{TOL_0}{TOL}\right)\right) \qquad (13)$$

If, in addition, the exact error density is bounded away from zero on $[0, T]$, *then*

$$E[Work(TOL)] = \mathcal{O}\left(\left(TOL^{-1} \log\left(\frac{TOL_0}{TOL}\right)\right)^2\right). \qquad (14)$$

Conclusions

- Extended adaptive, non adapted, algorithms to the Multi level Monte Carlo setting,
- Asymptotic estimates describe the behavior of the resulting adaptive algorithms, numerical experiments confirm the predicted bounds.
- Extension to jump diffusions as in [MSTZ08] is direct.

Future

- SPDEs
- Processes with jumps, reflections, ...
- Extensions to less regularity in payoff functions.

THANK YOU!