MLMC Combined with Numerical Smoothing: Applications in Probabilities/Densities Computation and Option Pricing

Chiheb Ben Hammouda

Christian Bayer

Weierstrass Institute for Applied Analysis and Stochastics

Raúl Tempone

King Abdullah University of Science and Technology

MCM23, mini-symposuim: MLMC for discontinuous functionals
Paris, France, June 28, 2023

Outline

Framework

- **Goal**: Approximate efficiently $\mathrm{E}\left[g(\mathbf{X}(T))\right]$
- **Setting**:
  - Given a (smooth) $\phi : \mathbb{R}^d \to \mathbb{R}$, the function $g : \mathbb{R}^d \to \mathbb{R}$:
    - ★ Indicator functions: $g(\mathbf{x}) = \mathbf{1}_{(\phi(\mathbf{x}) \geq 0)}$ (probabilities, pricing digital/barrier options, ...)
    - ★ Dirac Delta functions: $g(\mathbf{x}) = \delta_{(\phi(\mathbf{x}) = 0)}$ (densities, ...)
  - $\mathbf{X}$: solution process of a d-dimensional system of SDEs, approximated by $\overline{\mathbf{X}}$ (via a discretization scheme with N time steps), E.g., stochastic volatility model: E.g., the Heston model

$$
\begin{aligned}
dX_t &= \mu X_t dt + \sqrt{v_t} X_t dW_t^X \\
dv_t &= \kappa(\theta - v_t) dt + \xi \sqrt{v_t} dW_t^v,
\end{aligned}
$$

$\left(W_t^X, W_t^v\right)$: correlated Wiener processes with correlation ρ.

- **Challenge**: High-dimensional, non-smooth integration problem

$$
\mathrm{E}\left[g(\overline{\mathbf{X}}^{\Delta t}(T))\right] = \int_{\mathbb{R}^{d \times N}} G(\mathbf{z}) \rho_{d \times N}(\mathbf{z}) dz_1^{(1)} \dots dz_N^{(1)} \dots dz_1^{(d)} \dots dz_N^{(d)},
$$

with $G(\cdot)$ maps $N \times d$ random inputs to $g(\overline{\mathbf{X}}^{\Delta t}(T))$; and $\rho_{d \times N}(\mathbf{z})$: joint density function of $\mathbf{z}$.

Framework

- **Goal**: Approximate efficiently $\mathrm{E}\left[g(\mathbf{X}(T))\right]$
- **Setting**:
  - Given a (smooth) $\phi : \mathbb{R}^d \to \mathbb{R}$, the function $g : \mathbb{R}^d \to \mathbb{R}$:
    - ★ Indicator functions: $g(\mathbf{x}) = \mathbf{1}_{(\phi(\mathbf{x}) \geq 0)}$ (probabilities, pricing digital/barrier options, ...)
    - ★ Dirac Delta functions: $g(\mathbf{x}) = \delta_{(\phi(\mathbf{x})=0)}$ (densities, ...)
  - $\mathbf{X}$: solution process of a d-dimensional system of SDEs, approximated by $\overline{\mathbf{X}}$ (via a discretization scheme with N time steps), E.g., stochastic volatility model: E.g., the Heston model

$$
\begin{aligned}
dX_t &= \mu X_t dt + \sqrt{v_t} X_t dW_t^X \\
dv_t &= \kappa(\theta - v_t) dt + \xi \sqrt{v_t} dW_t^v,
\end{aligned}
$$

$\left(W_t^X, W_t^v\right)$: correlated Wiener processes with correlation ρ.

- **Challenge**: High-dimensional, non-smooth integration problem

$$
\mathrm{E}\left[g(\overline{\mathbf{X}}^{\Delta t}(T))\right] = \int_{\mathbb{R}^{d\times N}} G(\mathbf{z}) \rho_{d\times N}(\mathbf{z}) dz_1^{(1)} \dots dz_N^{(1)} \dots dz_1^{(d)} \dots dz_N^{(d)},
$$

with $G(\cdot)$ maps $N \times d$ random inputs to $g(\overline{\mathbf{X}}^{\Delta t}(T))$; and $\rho_{d\times N}(\mathbf{z})$: joint density function of $\mathbf{z}$.

Motivation

Table 1: Complexity comparison of the different methods for approximating $\mathrm{E}[g(\mathbf{X}(T))]$ within a pre-selected error tolerance, TOL. Given the same initial problem, and using a weak order one scheme, E.g, the Euler-Maruyama scheme.

Method	General Complexity	Optimal Complexity
MC	$\mathcal{O}\left(\mathrm{TOL}^{-3}\right)$	$\mathcal{O}\left(\mathrm{TOL}^{-3}\right)$
MLMC	$\mathcal{O}\left(\mathrm{TOL}^{-3+\beta}\right)$, $\frac{1}{2} \leq \beta \leq 1$	$\mathcal{O}\left(\mathrm{TOL}^{-2}\right)$
Quasi-MC (QMC)	$\mathcal{O}\left(\mathrm{TOL}^{-1-\frac{2}{1+2\delta}}\right)$, $0 \leq \delta \leq \frac{1}{2}$	$\mathcal{O}\left(\mathrm{TOL}^{-2}\right)$
Adaptive sparse grids quad (ASGQ)	$\mathcal{O}\left(\mathrm{TOL}^{-1-\frac{2}{p}}\right)$, $p > 0$	$\mathcal{O}\left(\mathrm{TOL}^{-1}\right)$

- **Sufficient Regularity Conditions for Optimal Complexity:**
  - **MLMC** (Cliffe et al. 2011; Giles 2015):
    g is Lipschtiz $\Rightarrow$ (sub) canonical complexity: $\mathcal{O}\left(\mathrm{TOL}^{-2}\right)$ up to log terms.
  - **QMC** (Dick, Kuo, and Sloan 2013):
    1. g belongs to the d-dimensional weighted Sobolev space of functions with square-integrable mixed (partial) first derivatives.
    2. High anisotropy between the different dimensions.
  - **ASGQ** (Chen 2018; Ernst, Sprungk, and Tamellini 2018):
    p is related to the order of bounded weighted mixed (partial) derivatives of g and the anisotropy between the different dimensions.
    $\Rightarrow$ ASGQ Complexity: $\mathcal{O}\left(\mathrm{TOL}^{-1-\frac{2}{p}}\right)$ ($\mathcal{O}\left(\mathrm{TOL}^{-1}\right)$ when $p \gg 1$).

Our Proposed Strategy to Recover Optimal Complexities

1. For QMC/ASGQ:
   Christian Bayer, Chiheb Ben Hammouda, and Raúl Tempone. "Numerical smoothing with hierarchical adaptive sparse grids and quasi-Monte Carlo methods for efficient option pricing". In: *Quantitative Finance* 23.2 (2023), pp. 209–227.
2. For MLMC (**Topic of the Talk**)
   Christian Bayer, Chiheb Ben Hammouda, and Raúl Tempone. "Multilevel Monte Carlo with Numerical Smoothing for Robust and Efficient Computation of Probabilities and Densities". In: *arXiv preprint arXiv:2003.05708* (2022).

⚠ The numerical smoothing idea in (Bayer, Ben Hammouda, and Tempone 2023) and (Bayer, Ben Hammouda, and Tempone 2022) is similar. However, the analysis is different.

⚠ For a survey on the different smoothing/adaptivity techniques for MLMC: see (Giles 2023).

Numerical Smoothing Steps

Motivating Example:

$$E[g(\overline{\mathbf{X}}_T^{\Delta t})] = ?$$

- $g : \mathbb{R}^d \to \mathbb{R}$ nonsmooth function: (E.g., $g(\mathbf{x}) = \mathbf{1}_{(\phi(\mathbf{x}) \geq 0)}$)
- $\overline{\mathbf{X}}_T^{\Delta t}$ ($\Delta t = \frac{T}{N}$) Euler discretization of d-dimensional SDE , E.g., $dX_t^{(i)} = a_i(\mathbf{X}_t)dt + \sum_{j=1}^d b_{ij}(\mathbf{X}_t)dW_t^{(j)}$, where $\{W^{(j)}\}_{j=1}^d$ are standard Brownian motions.
- $$\overline{\mathbf{X}}_T^{\Delta t} = \overline{\mathbf{X}}_T^{\Delta t}(\underbrace{\Delta W_1^{(1)}, \dots, \Delta W_N^{(1)}, \dots, \Delta W_1^{(d)}, \dots, \Delta W_N^{(d)}}_{:=\mathbf{\Delta W}})$$
  $$\equiv \overline{\mathbf{X}}_T^{\Delta t}(\mathbf{Z}), \quad \mathbf{Z} = (Z_i)_{i=1}^{dN} \sim \mathcal{N}(0, I_{dN}).$$
- The discontinuity is in $(N \times d)$-dimensional space characterised by
  $$\phi(\overline{\mathbf{X}}_T^{\Delta t}(\mathbf{Z})) = 0.$$

Numerical Smoothing Steps

1. Identify hierarchical representation of integration variables $\Rightarrow$ locate the discontinuity in a smaller dimensional space
   (a) $\overline{\mathbf{X}}_T^{\Delta t}(\mathbf{\Delta W}) \equiv \overline{\mathbf{X}}_T^{\Delta t}(\mathbf{Z})$, $\mathbf{Z} = (Z_i)_{i=1}^{dN} \sim \mathcal{N}(0, I_{dN})$:
   s.t. "$\mathbf{Z}_1 := (Z_1^{(1)}, \dots, Z_1^{(d)})$ (coarse rdvs) substantially contribute even for $\Delta t \to 0$", through hierarchical path generation (Brownian bridges / Haar wavelet construction)
   $\Rightarrow$ Discontinuity in d-dimensional space instead of $(N \times d)$-dimensions.

Haar wavelet construction in one dimension

For i.i.d. standard normal rdvs Z_1, $Z_{n,k}$, $n \in \mathbb{N}_0$, $k = 0, \dots, 2^n - 1$, we define the (truncated) standard Brownian motion

$$W_t^N := Z_1 \Psi_{-1}(t) + \sum_{n=0}^{N} \sum_{k=0}^{2^n-1} Z_{n,k} \Psi_{n,k}(t).$$

with $\Psi_{-1}(\cdot)$ and $\Psi_{n,k}(\cdot)$ are the antiderivatives of the Haar basis functions.

⚠ Our approach is different from previous MLMC techniques which uses conditional expectation at the final step w.r.t ΔW $\Rightarrow$ smoothing effect vanishes as $\Delta t \to 0$.

Numerical Smoothing Steps

1. Identify hierarchical representation of integration variables $\Rightarrow$ locate the discontinuity in a smaller dimensional space
   (b) If $d > 1$, introduce a linear mapping using $\mathcal{A}$: rotation matrix whose structure depends on the function g.

$$\mathbf{Y} = \mathcal{A}\mathbf{Z}_1.$$

E.g., for an observable $g(\mathbf{x}) = \mathbf{1}_{\{(\sum_{i=1}^{d} c_i x_i(T) - K) \geq 0\}}$, a suitable $\mathcal{A}$ is a rotation matrix, with the first row leading to $Y_1 = \sum_{i=1}^{d} Z_1^{(i)}$ up to rescaling without any constraint for the remaining rows (Gram-Schmidt procedure).

$\Rightarrow$ Discontinuity in 1-dimensional space instead of d-dimensions.

$\Rightarrow$ $y_1^*(\mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \ldots, \mathbf{z}_{-1}^{(d)})$: the exact discontinuity location s.t

$$\phi(\overline{\mathbf{X}}_T^{\Delta t}) = \phi(\overline{\mathbf{X}}_T^{\Delta t}(y_1^*; \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \ldots, \mathbf{z}_{-1}^{(d)})) = 0. \quad (1)$$

Notation

- $\mathbf{x}_{-j}$: vector of length $N-1$ denoting all the variables other than x_j in $\mathbf{x} \in \mathbb{R}^N$.

Numerical Smoothing Steps

2.

$$
\begin{aligned}
\mathrm{E}\left[g(\mathbf{X}(T))\right] &\approx \mathrm{E}\left[g(\overline{\mathbf{X}}^{\Delta t}(T))\right] \\
&= \int_{\mathbb{R}^{d\times N}} G(\mathbf{z})\rho_{d\times N}(\mathbf{z}) dz_1^{(1)}\dots dz_N^{(1)}\dots dz_1^{(d)}\dots dz_N^{(d)} \\
&= \int_{\mathbb{R}^{dN-1}} I(\mathbf{y}_{-1},\mathbf{z}_{-1}^{(1)},\dots,\mathbf{z}_{-1}^{(d)})\rho_{d-1}(\mathbf{y}_{-1})d\mathbf{y}_{-1}\rho_{dN-d}(\mathbf{z}_{-1}^{(1)},\dots,\mathbf{z}_{-1}^{(d)})d\mathbf{z}_{-1}^{(1)}\dots d\mathbf{z}_{-1}^{(d)} \\
&= \mathrm{E}\left[I(\mathbf{Y}_{-1},\mathbf{Z}_{-1}^{(1)},\dots,\mathbf{Z}_{-1}^{(d)})\right] \approx \mathrm{E}\left[\overline{I}(\mathbf{Y}_{-1},\mathbf{Z}_{-1}^{(1)},\dots,\mathbf{Z}_{-1}^{(d)})\right], \qquad (2)
\end{aligned}
$$

3.

$$
\begin{aligned}
&I(\mathbf{y}_{-1},\mathbf{z}_{-1}^{(1)},\dots,\mathbf{z}_{-1}^{(d)}) = \int_{\mathbb{R}} G(y_1,\mathbf{y}_{-1},\mathbf{z}_{-1}^{(1)},\dots,\mathbf{z}_{-1}^{(d)})\rho_1(y_1)dy_1 \\
&= \int_{-\infty}^{y_1^*} G(y_1,\mathbf{y}_{-1},\mathbf{z}_{-1}^{(1)},\dots,\mathbf{z}_{-1}^{(d)})\rho_1(y_1)dy_1 + \int_{y_1^*}^{+\infty} G(y_1,\mathbf{y}_{-1},\mathbf{z}_{-1}^{(1)},\dots,\mathbf{z}_{-1}^{(d)})\rho_1(y_1)dy_1 \\
&\approx \overline{I}(\mathbf{y}_{-1},\mathbf{z}_{-1}^{(1)},\dots,\mathbf{z}_{-1}^{(d)}) := \sum_{k=0}^{M_{\text{lag}}} \eta_k G\left(\zeta_k\left(\overline{y}_1^*\right),\mathbf{y}_{-1},\mathbf{z}_{-1}^{(1)},\dots,\mathbf{z}_{-1}^{(d)}\right),
\end{aligned}
$$

4. Compute the remaining $(dN-1)$-integral (expectation) in (2) by MLMC.

Notation

- G maps $N\times d$ Gaussian random inputs to $g(\overline{\mathbf{X}}^{\Delta t}(T))$;
- $y_1^*(\mathbf{y}_{-1},\mathbf{z}_{-1}^{(1)},\dots,\mathbf{z}_{-1}^{(d)})$: the exact discontinuity location (see (1))
- $\overline{y}_1^*(\mathbf{y}_{-1},\mathbf{z}_{-1}^{(1)},\dots,\mathbf{z}_{-1}^{(d)})$: the approximated discontinuity location via root finding.
- M_{Lag}: number of Laguerre quadrature points $\zeta_k\in\mathbb{R}$, and weights η_k;
- $\rho_{d\times N}(\mathbf{z}) = \frac{1}{(2\pi)^{d\times N/2}}e^{-\frac{1}{2}\mathbf{z}^T\mathbf{z}}$.

Some Remarks

- ⚠ In (Bayer, Ben Hammouda, and Tempone 2023), we show that $\overline{I}(\cdot)$ in (2) is C^{∞} $\Rightarrow$ optimal complexity for ASGQ and QMC.
- ⚠ Here, for MLMC we need different analysis/arguments to show that we get the optimal complexity of MLMC (**see next slides**).
- The numerical smoothing can be extended to the case of finitely many roots.

Extending Numerical Smoothing for Density Estimation

- **Goal:** Approximate the density ρ_X at u, for a stochastic process X

$$\rho_X(u) = \mathrm{E}\left[\delta(X-u)\right], \quad \delta \text{ is the Dirac delta function.}$$

⚠ Without any smoothing techniques (regularization, KDE,...) MC/MLMC fail due to the infinite variance caused by the Dirac distribution function, $\delta(\cdot)$.

- **Strategy** in (Bayer, Ben Hammouda, and Tempone 2022): Conditioning with respect to $\mathbf{Z}_{-1}$ (randomness related to the Brownian bridge)

$$\begin{aligned}\rho_X(u) &= \frac{1}{\sqrt{2\pi}} E\left[\exp\left(-\left(Y_1^*(u)\right)^2/2\right)\left|\frac{dY_1^*}{dx}(u)\right|\right]\\ &\approx \frac{1}{\sqrt{2\pi}} E\left[\exp\left(-\left(\overline{Y}_1^*(u)\right)^2/2\right)\left|\frac{d\overline{Y}_1^*}{dx}(u)\right|\right],\end{aligned}$$

$Y_1^*(x;\mathbf{Z}_{-1})$: the exact singularity; $\overline{Y}_1^*(x;\mathbf{Z}_{-1})$: the approximated singularity obtained by solving $\overline{X}^{\Delta t}(T;\overline{Y}^*(x),\mathbf{Z}_{-1}) = x$.

Why not Kernel Density Estimator (KDE) in Multiple Dimensions?

- Similar to approaches based on MLMC with parametric regularization (Giles, Nagapetyan, and Ritter 2015) or QMC with KDE techniques (Ben Abdellah et al. 2021).
- This class of approaches has a pointwise error that increases exponentially with respect to the dimension of the state vector $\mathbf{X}$.
- For a d-dimensional problem, a KDE with a bandwidth matrix, $\mathcal{H} = \mathrm{diag}(h, \ldots, h)$

$$\mathbf{MSE} \approx c_1 M^{-1} h^{-d} + c_2 h^4. \tag{3}$$

  M is the number of samples, and c_1 and c_2 are constants.
- **Our approach in high dimension**: For $\mathbf{u} \in \mathbb{R}^d$

$$\begin{aligned}\rho_{\mathbf{X}}(\mathbf{u}) = \mathrm{E}\left[\delta(\mathbf{X} - \mathbf{u})\right] &= \mathrm{E}\left[\rho_d\left(\mathbf{Y}^*(\mathbf{u})\right)\left|\det\left(\mathbf{J}\left(\mathbf{u}\right)\right)\right|\right] \\ &\approx \mathrm{E}\left[\rho_d\left(\overline{\mathbf{Y}}^*(\mathbf{u})\right)\left|\det\left(\overline{\mathbf{J}}\left(\mathbf{u}\right)\right)\right|\right],\end{aligned} \tag{4}$$

  - $\mathbf{Y}^*(\mathbf{u};\cdot)$: the exact discontinuity; $\overline{\mathbf{Y}}^*(\mathbf{u};\cdot)$: the approximated discontinuity.
  - $\mathbf{J}$ is the Jacobian matrix, with $\mathbf{J}_{ij} = \frac{\partial y_i^*}{\partial u_j}$; $\rho_d(\cdot)$ is the multivariate Gaussian density.
- Exact conditioning with respect to the remaining Brownian bridge noise $\Rightarrow$ the smoothing error in our approach is insensitive to the dimension of the problem.

Multilevel Monte Carlo (MLMC) (Heinrich 2001; Kebaier 2005; Giles 2008)

- **Setting**
  - A hierarchy of nested meshes of $[0,T]$ (sequence of finer discretizations).
  - $\Delta t_\ell := K^{-\ell}\Delta t_0$: the time steps size for levels $\ell \geq 0$; $K{>}1$, $K \in \mathbb{N}$. $(\Delta t_0 > \ldots > \Delta t_L)$
  - $\overline{\mathbf{X}}_\ell := \overline{\mathbf{X}}^{\Delta t_\ell}$: The approximate process generated using a step size of Δt_ℓ.
- **MLMC idea**

$$\mathrm{E}\left[g(\mathbf{X}(T))\right] \approx \mathrm{E}\left[g(\overline{\mathbf{X}}_L(T))\right] = \underbrace{\mathrm{E}\left[g(\overline{\mathbf{X}}_0(T))\right]}_{} + \underbrace{\sum_{\ell=1}^{L} \mathrm{E}\left[g(\overline{\mathbf{X}}_\ell(T)) - g(\overline{\mathbf{X}}_{\ell-1}(T))\right]}_{} \tag{5}$$

$\mathrm{Var}[g(\overline{\mathbf{X}}_0(T))] \gg \mathrm{Var}[g(\overline{\mathbf{X}}_\ell(T)) - g(\overline{\mathbf{X}}_{\ell-1}(T))] \searrow$ as $\ell \nearrow$

$M_0 \gg M_\ell \searrow$ as $\ell \nearrow$

- **MLMC estimator**: $\widehat{Q}^{\mathrm{MLMC}} := \sum_{\ell=0}^{L} \widehat{Q}_\ell$, (sample independently each term of (5) with MC)

$$\widehat{Q}_0 := \frac{1}{M_0}\sum_{m_0=1}^{M_0} g(\overline{\mathbf{X}}_0(T;\omega_{m_0})); \quad \widehat{Q}_\ell := \frac{1}{M_\ell}\sum_{m_\ell=1}^{M_\ell} \left(g(\overline{\mathbf{X}}_\ell(T;\omega_{m_\ell})) - g(\overline{\mathbf{X}}_{\ell-1}(T;\omega_{m_\ell}))\right),\ 1 \leq \ell \leq L$$

- **Compared to MC**: MLMC reduces the variance of the deepest level using samples on coarser (less expensive) levels.

Multilevel Monte Carlo with Numerical Smoothing: Estimator and Notation

- Recall
  - $\mathrm{E}\left[g(\mathbf{X}(T))\right] \approx \mathrm{E}\left[g(\overline{\mathbf{X}}^{\Delta t}(T))\right] = \mathrm{E}\left[I(\mathbf{Y}_{-1}, \mathbf{Z}_{-1}^{(1)}, \dots, \mathbf{Z}_{-1}^{(d)})\right] \approx \mathrm{E}\left[\overline{I}(\mathbf{Y}_{-1}, \mathbf{Z}_{-1}^{(1)}, \dots, \mathbf{Z}_{-1}^{(d)})\right]$
  -
$$
\begin{aligned}
I(\mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) &= \int_{\mathbb{R}} G(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) \rho_1(y_1) dy_1 \\
&= \int_{-\infty}^{y_1^*} G(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) \rho_1(y_1) dy_1 + \int_{y_1^*}^{+\infty} G(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) \rho_1(y_1) dy_1 \\
&\approx \overline{I}(\mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) := \sum_{k=0}^{M_{\text{lag}}} \eta_k G\left(\zeta_k\left(\overline{y}_1^*\right), \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}\right),
\end{aligned}
$$

    where $\overline{y}_1^*(\mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)})$: the approximated discontinuity location via root finding.

- $\overline{I}_\ell := \overline{I}_\ell(\mathbf{y}_{-1}^{\ell}, \mathbf{z}_{-1}^{(1),\ell}, \dots, \mathbf{z}_{-1}^{(d),\ell})$: level ℓ approximation of $\overline{I}$ in $\widehat{Q}^{\text{MLMC}}$, computed with step size Δt_ℓ; $M_{\text{Lag},\ell}$ Laguerre points; $\text{TOL}_{\text{Newton},\ell}$ as the Newton tolerance at level ℓ.
-
$$
\widehat{Q}^{\text{MLMC}} := \sum_{\ell=L_0}^{L} \widehat{Q}_\ell,
$$

  with

$$
\widehat{Q}_{L_0} := \frac{1}{M_{L_0}} \sum_{m_{L_0}=1}^{M_{L_0}} \overline{I}_{L_0,[m_{L_0}]}; \quad \widehat{Q}_\ell := \frac{1}{M_\ell} \sum_{m_\ell=1}^{M_\ell} \left(\overline{I}_{\ell,[m_\ell]} - \overline{I}_{\ell-1,[m_\ell]}\right), \; L_0 + 1 \leq \ell \leq L, \tag{6}
$$

MLMC with Numerical Smoothing: Analysis

Let $g(\mathbf{x}) = \mathbf{1}_{(\phi(\mathbf{x})\geq 0)}$ or $\delta(\phi(\mathbf{x}) = 0)$

Theorem 3.1 (Variance Decay (Bayer, Ben Hammouda, and Tempone 2022))

Under some regularity assumptions for the drift and diffusion, using Euler–Maruyama, $V_\ell := Var\left[\overline{I}_\ell - \overline{I}_{\ell-1}\right] = \mathcal{O}\left(\Delta t_\ell^1\right)$, compared with $\mathcal{O}\left(\Delta t_\ell^{1/2}\right)$ for MLMC without smoothing.

⚠ **General MLMC Complexity**: $\mathcal{O}\left(\mathrm{TOL}^{-2-\max\left(0,\frac{\gamma-\beta}{\alpha}\right)} \log\left(\mathrm{TOL}\right)^{2\times \mathbf{1}_{\{\beta=\gamma\}}}\right)$,
where α: weak rate; β: variance decay rate; γ : work growth rate.

Corollary 3.2 (Complexity (Bayer, Ben Hammouda, and Tempone 2022))

Under some regularity assumptions for the drift and diffusion, the complexity of MLMC combined with numerical smoothing is $\mathcal{O}\left(TOL^{-2}\right)$ up to log terms, compared with $\mathcal{O}\left(TOL^{-2.5}\right)$ for MLMC without smoothing.

⚠ Milstein scheme: we show that we obtain the canonical complexity ($\mathcal{O}\left(\mathrm{TOL}^{-2}\right)$).

Corollary 3.3 (Robustness (Bayer, Ben Hammouda, and Tempone 2022))

Let κ_ℓ be the kurtosis of the r.d.v $\overline{I}_\ell - \overline{I}_{\ell-1}$, then under some regularity assumptions of the drift & diffusion, we get $\kappa_\ell = \mathcal{O}(1)$ compared to $\mathcal{O}\left(\Delta t_\ell^{-1/2}\right)$ for MLMC without smoothing.

⚠ The assumptions in Theorem 3.1 and Corollary 3.2 are sufficient but not necessary.

Sketch of the Proof of Theorem 3.1: Goal and Notations

Goal: We want to show $V_\ell := \mathrm{Var}\left[\overline{I}_\ell - \overline{I}_{\ell-1}\right] \leq \mathrm{E}\left[\left(\overline{I}_\ell - \overline{I}_{\ell-1}\right)^2\right] = \mathcal{O}\left(\Delta t_\ell\right)$.

Notations

- $\overline{X}_\ell, \overline{X}_{\ell-1}$: the coupled paths of the approximate process $\overline{X}$, simulated with time step sizes Δt_ℓ and $\Delta t_{\ell-1}$, respectively.
- W_ℓ and B_ℓ: coupling Wiener and related Brownian bridge processes at levels ℓ and $\ell-1$, respectively.
- For $t \in [0,T]$, $e_\ell(t; Y, B_\ell)$ is defined as

$$
\begin{aligned}
(\overline{X}_\ell - \overline{X}_{\ell-1})(t) &= \int_0^t \left(\overline{a}(\overline{X}_\ell(s)) - \overline{a}(\overline{X}_{\ell-1}(s))\right) ds + \int_0^t \left(\overline{b}(\overline{X}_\ell(s)) - \overline{b}(\overline{X}_{\ell-1}(s))\right) dW_\ell(s) \\
&= \int_0^t \left(\overline{a}(\overline{X}_\ell(s)) - \overline{a}(\overline{X}_{\ell-1}(s))\right) ds + \int_0^t \left(\overline{b}(\overline{X}_\ell(s)) - \overline{b}(\overline{X}_{\ell-1}(s))\right) \frac{Y}{\sqrt{T}} ds \\
&\quad + \int_0^t \left(\overline{b}(\overline{X}_\ell(s)) - \overline{b}(\overline{X}_{\ell-1}(s))\right) dB_\ell(s) \\
&=: e_\ell(t; Y, B_\ell),
\end{aligned}
$$

where $\overline{a}(\overline{X}(s)) = a(\overline{X}(t_n))$, $\overline{b}(\overline{X}(s)) = b(\overline{X}(t_n))$, for $t_n \leq s < t_{n+1}$, on the time grid $0 = t_0 < t_1 < \ldots < t_N = T$.

Sketch of the Proof of Theorem 3.1: Step 1

For Euler–Maruyama scheme and $p \geq 1$,

- Under global Lipschitzity of drift and diffusion coefficients Assumption, we have (Kloeden and Platen 1992)

$$\mathrm{E}\left[e_\ell^{2p}(T)\right] = \mathcal{O}\left(\Delta t_\ell^p\right). \tag{7}$$

- In (Bayer, Ben Hammouda, and Tempone 2022), assuming further regularity assumptions of the drift and diffusion, we prove that

$$\mathrm{E}\left[(\partial_y e_\ell)^{2p}(T)\right] = \mathcal{O}\left(\Delta t_\ell^p\right). \tag{8}$$

⚠ The proof is based on the Grönwall, Hölder, Jensen and Burkholder-Davis-Gundy inequalities.

Sketch of the Proof of Theorem 3.1: Step 2

Using (i) integration by parts, and (ii) the mean value, Fubini, and dominated convergence theorems, we show that

$$\begin{aligned}\Delta I_\ell(B_\ell) &:= (\overline{I}_\ell - \overline{I}_{\ell-1})(B_\ell) := \int_{\mathbb{R}} \left(g(\overline{X}_\ell(T;y,B_\ell)) - g(\overline{X}_{\ell-1}(T;y,B_\ell))\right)\rho_1(y)dy \\ &= \underbrace{-\int_0^1 \left[\int_{\mathbb{R}} e_\ell(T;y,B_\ell) g(z(\theta;y,B_\ell)) \Big(\partial_y \left((\partial_y z(\theta;y,B_\ell))^{-1}\right) - y\,(\partial_y z(\theta;y,B_\ell))^{-1}\Big)\rho_1(y)dy\right] d\theta}_{(I)} \\ &\quad \underbrace{- \int_0^1 \left[\int_{\mathbb{R}} \partial_y e_\ell(T;y,B_\ell) g(z(\theta;y,B_\ell)) (\partial_y z(\theta;y,B_\ell))^{-1}\rho_1(y)dy\right] d\theta}_{(II)}, \end{aligned} \tag{9}$$

with

$$\begin{aligned} z(\theta;y,B_\ell) &:= \overline{X}_{\ell-1}(T;y,B_\ell) + \theta e_\ell(T;y,B_\ell),\ \theta \in (0,1) \\ &= (1-\theta)\overline{X}_{\ell-1}(T;y,B_\ell) + \theta \overline{X}_\ell(T;y,B_\ell) \end{aligned}$$

Sketch of the Proof of Theorem 3.1: Step 3

- For term (I), taking expectation w.r.t the Brownian bridge and using Hölder's inequality ($p, q, p_1, q_1 \in (1, +\infty)$, $\frac{1}{p} + \frac{1}{q} = 1$ and $\frac{1}{p_1} + \frac{1}{q_1} = 1$), result in

$$
\begin{aligned}
E\left[(I)^2\right] &\leq \left(E\left[\left\|g(z(\cdot;\cdot,B_\ell))\left(\partial_y\left((\partial_y z(\cdot;\cdot,B_\ell))^{-1}\right) - Y\left(\partial_y z(\cdot;\cdot,B_\ell)\right)^{-1}\right)\right\|^{2q_1}_{L^q_{\rho_1}([0,1]\times\mathbb{R})}\right]\right)^{1/q_1} \\
&\quad \times \left(E\left[\|e_\ell(T;\cdot,B_\ell)\|^{2p_1}_{L^p_{\rho_1}(\mathbb{R})}\right]\right)^{1/p_1} \\
&= \mathcal{O}\left(\Delta t_\ell\right). \qquad (10)
\end{aligned}
$$

- Choosing p and p_1 such that $\frac{2p_1}{p} \leq 1$, and applying Jensen's inequality:

$$
\begin{aligned}
\left(E\left[\|e_\ell(T;\cdot,B_\ell)\|^{2p_1}_{L^p_{\rho_1}(\mathbb{R})}\right]\right)^{1/p_1} &= \left(E\left[\left(\int_{\mathbb{R}}\left|e^p_\ell(T;y,B_\ell)\right|\rho_1 dy\right)^{\frac{2p_1}{p}}\right]\right)^{1/p_1} \\
&\leq \left(E\left[\int_{\mathbb{R}}\left|e^p_\ell(T;y,B_\ell)\right|\rho_1 dy\right]\right)^{\frac{2}{p}} \\
&= \mathcal{O}\left(\Delta t_\ell\right) \text{ (using Fubini's theorem and (7))}.
\end{aligned}
$$

- We show that

$$
\left(E\left[\left\|g(z(\cdot;\cdot,B_\ell))\left(\partial_y\left((\partial_y z(\cdot;\cdot,B_\ell))^{-1}\right) - Y\left(\partial_y z(\cdot;\cdot,B_\ell)\right)^{-1}\right)\right\|^{2q_1}_{L^q_{\rho_1}([0,1]\times\mathbb{R})}\right]\right)^{1/q_1} < \infty,
$$

Sketch of the Proof of Theorem 3.1: Step 4

- For the term (II) in (9), we redo same steps as for term (I)

$$
\begin{aligned}
E\left[(II)^2\right] &\leq \left(E\left[\left\|g(z(\cdot;\cdot,B_\ell))\left(\partial_y z(\cdot;\cdot,B_\ell)\right)^{-1}\right\|^{2q_1}_{L^q_{\rho_1}([0,1]\times\mathbb{R})}\right]\right)^{1/q_1} \\
&\times \left(E\left[\|\partial_y e_\ell(T;\cdot,B_\ell)\|^{2p_1}_{L^p_{\rho_1}(\mathbb{R})}\right]\right)^{1/p_1} \\
&= \mathcal{O}\left(\Delta t_\ell\right)
\end{aligned}
\tag{11}
$$

- Using (8), we show $\left(E_{B_\ell}\left[\|\partial_y e_\ell(T;\cdot,B_\ell)\|^{2p_1}_{L^p_{\rho_1}(\mathbb{R})}\right]\right)^{1/p_1} = \mathcal{O}\left(\Delta t_\ell\right)$.
- We show that

$$
\left(E\left[\left\|g(z(\cdot;\cdot,B_\ell))\left(\partial_y z(\cdot;\cdot,B_\ell)\right)^{-1}\right\|^{2q_1}_{L^q_{\rho_1}([0,1]\times\mathbb{R})}\right]\right)^{1/q_1} < \infty,
$$

Error Discussion for MLMC

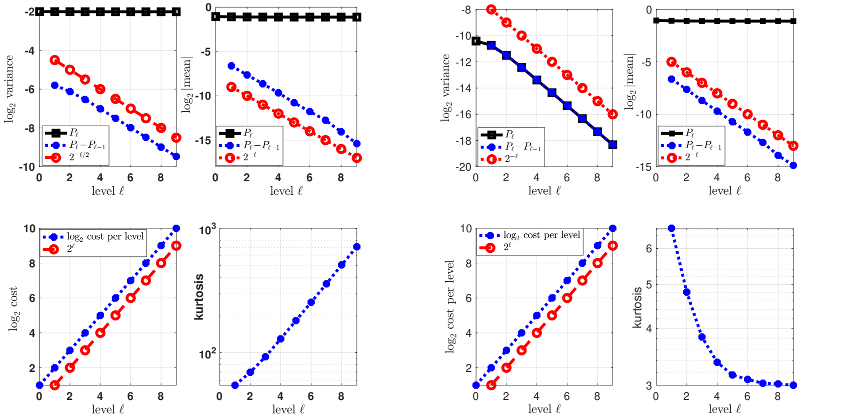
$\widehat{Q}^{\mathrm{MLMC}}$: the MLMC estimator

$$
\begin{aligned}
\mathrm{E}\left[g(X(T)\right] - \widehat{Q}^{\mathrm{MLMC}} &= \underbrace{\mathrm{E}\left[g(X(T))\right] - \mathrm{E}\left[g(\overline{\mathbf{X}}^{\Delta t_L}(T))\right]}_{\text{Error I: bias or weak error of } \mathcal{O}(\Delta t_L)} \\
&+ \underbrace{\mathrm{E}\left[I_L\left(\mathbf{Y}_{-1}, \mathbf{Z}_{-1}^{(1)}, \dots, \mathbf{Z}_{-1}^{(d)}\right)\right] - \mathrm{E}\left[\overline{I}_L\left(\mathbf{Y}_{-1}, \mathbf{Z}_{-1}^{(1)}, \dots, \mathbf{Z}_{-1}^{(d)}\right)\right]}_{\text{Error II: numerical smoothing error of } \mathcal{O}\left(M_{\mathrm{Lag},L}^{-s/2}\right) + \mathcal{O}(\mathrm{TOL}_{\mathrm{Newton},L})} \\
&+ \underbrace{\mathrm{E}\left[\overline{I}_L\left(\mathbf{Y}_{-1}, \mathbf{Z}_{-1}^{(1)}, \dots, \mathbf{Z}_{-1}^{(d)}\right)\right] - \widehat{Q}^{\mathrm{MLMC}}}_{\text{Error III: MLMC statistical error of } \mathcal{O}\left(\sqrt{\sum_{\ell=L_0}^{L} \sqrt{M_{\mathrm{Lag},\ell} + \log\left(\mathrm{TOL}_{\mathrm{Newton},\ell}^{-1}\right)}}\right)}
\end{aligned}
$$

Notations

- $\overline{y}_1^*$: the approximated location of the non smoothness obtained by Newton iteration $\Rightarrow |y_1^* - \overline{y}_1^*| = \mathrm{TOL}_{\mathrm{Newton}}$
- M_{Lag} is the number of points used by the Laguerre quadrature for the one dimensional pre-integration step.
- $s > 0$: For the parts of the domain separated by the discontinuity location, derivatives of G with respect to y_1 are bounded up to order s.

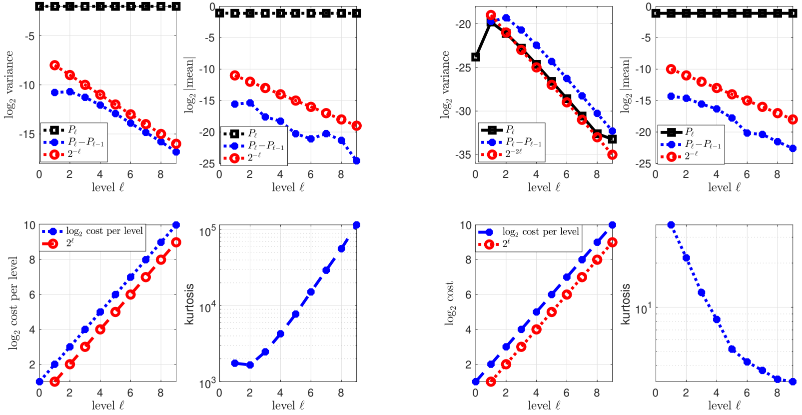
MLMC for Probability in the GBM model: Euler–Maruyama



(a) without numerical smoothing. (b) With numerical smoothing.

Figure 4.1: MLMC for probability computation under the geometric Brownian motion (GBM): Variance, cost, L^1-distance and kurtosis per level. P_ℓ: the numerical approximation of the QoI at level ℓ.

MLMC for Probability in the GBM Model: Milstein



(a) without numerical smoothing.

(b) With numerical smoothing.

Figure 4.2: MLMC with Milstein scheme for probability computation under the geometric Brownian motion (GBM): Variance, cost, L^1-distance and kurtosis per level. P_ℓ: the numerical approximation of the QoI at level ℓ.

Probability Computation under the GBM Model: Numerical Complexity Comparison

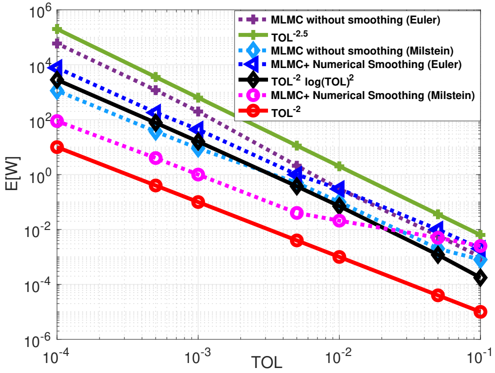
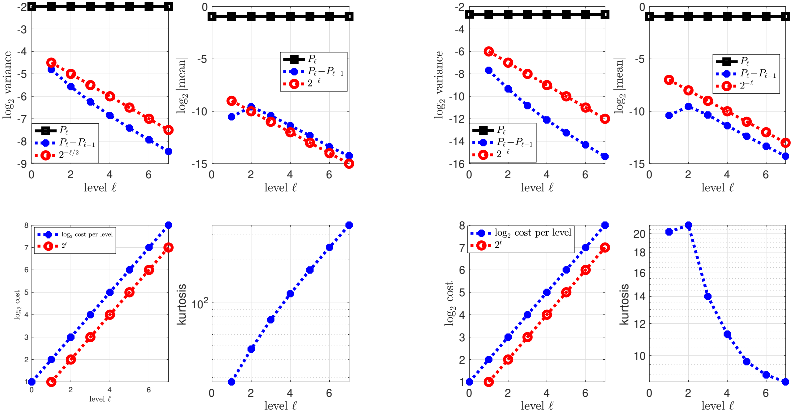


Figure 4.3: Probability Computation under GBM: Comparison of the numerical complexity of the different MLMC estimators.

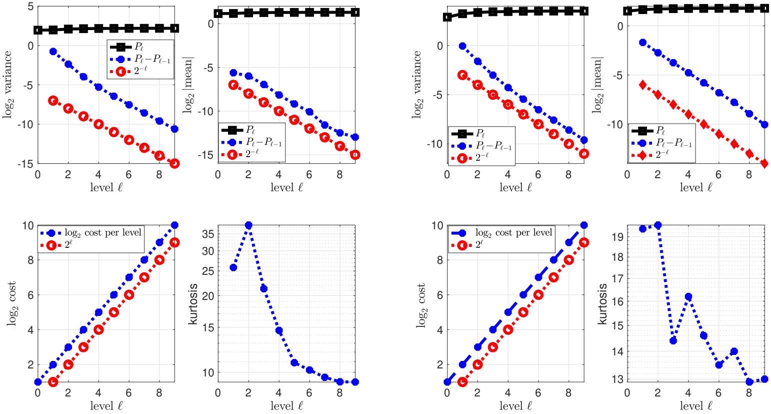
MLMC for Probability under the Heston Model



(a) without numerical smoothing. (b) With numerical smoothing.

Figure 4.4: MLMC with FT Euler–Maruyama scheme for probability computation under the Heston model: Variance, cost, L^1-distance and kurtosis per level. P_ℓ: the numerical approximation of the QoI at level ℓ.

Density Estimation under the Heston Model



(a) Asset price density (b) Joint density

Figure 4.5: Density of Heston: Convergence plots for MLMC with numerical smoothing combined with the FT Euler scheme, for computing the asset price density $\rho_{X(T)}$ at $u = 1$ and the joint density $\rho_{X(T),v(T)}$ at $u = 1$ and $v = 0.04$.

Conclusions

1. The numerical smoothing approach is adapted to the MLMC context for efficient probability computation, univariate/multivariate density estimation, and option pricing.
2. Compared to the case without smoothing
   - We significantly reduce the kurtosis at the deep levels of MLMC (becomes bounded instead of blow-up) which improves the robustness of the estimator.
   - We improve the MLMC strong convergence (variance decay) rate $\Rightarrow$ improvement of MLMC complexity from $\mathcal{O}\left(\mathrm{TOL}^{-2.5}\right)$ to $\mathcal{O}\left(\mathrm{TOL}^{-2}\right)$ (we recover the MLMC complexities obtained for Lipschitz functionals).
3. When estimating densities: Compared to the smoothing strategies based on MLMC with parametric regularization as in (Giles, Nagapetyan, and Ritter 2015) or QMC with kernel density techniques as in (Ben Abdellah et al. 2021), the error of our approach does not increase exponentially with respect to the dimension of state vector

Extensions

1. Extend our techniques to efficiently compute
   - Sensitivities (Financial Greeks): $\frac{\partial}{\partial \alpha} E[f(\omega, \alpha)]$.
   - Risk quantities $\Rightarrow$ nested expectations problems $E\left[g\left(E\left[f(X,Y)|X\right]\right)\right]$.
   - Computing nonsmooth quantities (such as probabilities) of a functional of a solution arising from a random PDE.
2. Combine the numerical smoothing technique with multilevel QMC to profit from the good features of QMC and MLMC.
3. Combine the numerical smoothing technique with antithetic MLMC (Giles and Szpruch 2014) for multi-dimensional SDEs to recover the optimal complexity.

Related References

Thank you for your attention!

[1] C. Bayer, C. Ben Hammouda, R. Tempone. *Numerical Smoothing with Hierarchical Adaptive Sparse Grids and Quasi-Monte Carlo Methods for Efficient Option Pricing*. Quantitative Finance 23, no. 2 (2023): 209-227.

[2] C. Bayer, C. Ben Hammouda, R. Tempone. *Multilevel Monte Carlo with Numerical Smoothing for Robust and Efficient Computation of Probabilities and Densities*, arXiv:2003.05708 (2022).

Assumptions about $\phi(\cdot)$

$$\frac{\partial \phi}{\partial x_j}(\mathbf{x}) > 0,\ \forall \mathbf{x} \in \mathbb{R}^d \ \textbf{(Monotonicity condition)} \tag{12}$$

$$\lim_{x_j \to +\infty} \phi(\mathbf{x}) = \lim_{x_j \to +\infty} \phi(x_j, \mathbf{x}_{-j}) = +\infty,\ \forall \mathbf{x}_{-j} \in \mathbb{R}^{d-1} \text{ or } \frac{\partial^2 \phi}{\partial x_j^2}(\mathbf{x}) \geq 0,\ \forall \mathbf{x} \in \mathbb{R}^d, \tag{13}$$

$$\textbf{(Growth condition)}.$$

(1) and (2) $\Rightarrow$ the function $\phi(x_j, \mathbf{x}_{-j})$ either has a simple root or is positive for all $x_j \in \mathbb{R}$

⚠ **Notation**: $\mathbf{x}_{-j}$ denotes the vector of length $d-1$ denoting all the variables other than x_j in $\mathbf{x}$.

How Does Regularity Affect MLMC Complexity?

- Complexity analysis for MLMC

MLMC Complexity (Cliffe et al. 2011)

$$\mathcal{O}\left(\mathrm{TOL}^{-2-\max\left(0,\frac{\gamma-\beta}{\alpha}\right)}\log\left(\mathrm{TOL}\right)^{2\times\mathbf{1}_{\{\beta=\gamma\}}}\right) \tag{14}$$

i) Weak rate:
$\left|\mathrm{E}\left[g\left(\overline{\mathbf{X}}_\ell(T)\right)-g\left(\mathbf{X}(T)\right)\right]\right| \le c_1 2^{-\alpha\ell}$

ii) Variance decay rate:
$\underbrace{\mathrm{Var}\left[g\left(\overline{\mathbf{X}}_\ell(T)\right)-g\left(\overline{\mathbf{X}}_{\ell-1}(T)\right)\right]}_{:=V_\ell} \le c_2 2^{-\beta\ell}$

iii) Work growth rate: $W_\ell \le c_3 2^{\gamma\ell}$ (W_ℓ: expected cost)

- For Euler-Maruyama ($\gamma = 1$):
  - If g is Lipschitz $\Rightarrow V_\ell \simeq \Delta t_\ell$ due to strong rate 1/2, that is $\beta = \gamma$ and MLMC complexity $\mathcal{O}\left(\mathrm{TOL}^{-2}\right)$ (up to log terms);
  - **Otherwise (without any smoothing or adaptivity techniques):** $\beta < \gamma \Rightarrow$ worst-case complexity, $\mathcal{O}\left(\mathrm{TOL}^{-\frac{5}{2}}\right)$.
- Higher order schemes, E.g., the Milstein scheme, may lead to better complexities even for non-Lipschitz observables (Giles, Debrabant, and Rößler 2013; Giles 2015). **However,**
  - For moderate/high-dimensional SDEs, the scheme becomes computationally expensive.
  - Deterioration of the robustness of the MLMC estimator because the kurtosis explodes as Δt_ℓ decreases: $\mathcal{O}\left(\Delta t_\ell^{-1}\right)$ compared with $\mathcal{O}\left(\Delta t_\ell^{-1/2}\right)$ for Euler-Maruyama without smoothing (Giles, Nagapetyan, and Ritter 2015).

How Does Regularity Affect MLMC Robustness?

- ☹ For non-lipschitz payoffs (without any smoothing or adaptivity techniques):
  The Kurtosis, $\kappa_\ell := \frac{\mathrm{E}\left[(Y_\ell - \mathrm{E}[Y_\ell])^4\right]}{(\mathrm{Var}[Y_\ell])^2}$ is of $\mathcal{O}(\Delta t_\ell^{-1/2})$ for Euler-Maruyama.
- Large kurtosis problem: discussed previously in (Ben Hammouda, Moraes, and Tempone 2017; Ben Hammouda, Ben Rached, and Tempone 2020) $\Rightarrow$ ☹ Expensive cost for reliable/robust estimates of sample statistics.
- Why is large kurtosis bad?

$$\sigma_{\mathcal{S}^2(Y_\ell)} = \frac{\mathrm{Var}[Y_\ell]}{\sqrt{M_\ell}} \sqrt{(\kappa_\ell - 1) + \frac{2}{M_\ell - 1}}; \; ⚠ \; M_\ell \gg \kappa_\ell.$$

- Why are accurate variance estimates, $V_\ell = \mathrm{Var}[Y_\ell]$, important?

$$M_\ell^* \propto \sqrt{V_\ell W_\ell^{-1}} \sum_{\ell=0}^{L} \sqrt{V_\ell W_\ell}.$$

Notation

- $Y_\ell := g(\overline{X}_\ell(T)) - g(\overline{X}_{\ell-1}(T))$
- $\sigma_{\mathcal{S}^2(Y_\ell)}$: Standard deviation of the sample variance of Y_ℓ;
- M_ℓ^*: Optimal number of samples per level; W_ℓ: Cost per sample path.

Extending Numerical Smoothing for Multiple Discontinuities

- Multiple Discontinuities: Due to the payoff structure/use of Richardson extrapolation.
- R different ordered multiple roots, e.g., $\{y_i^*\}_{i=1}^R$, the smoothed integrand is

$$I\left(\mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}\right) = \int_{-\infty}^{y_1^*} G\left(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}\right) \rho_1(y_1) dy_1 + \int_{y_R^*}^{+\infty} G\left(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}\right) \rho_1(y_1) dy_1 + \sum_{i=1}^{R-1} \int_{y_i^*}^{y_{i+1}^*} G\left(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}\right) \rho_1(y_1) dy_1,$$

and its approximation $\overline{I}$ is given by

$$\begin{aligned}\overline{I}(\mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) &:= \sum_{k=0}^{M_{\text{Lag},1}} \eta_k^{\text{Lag}}\, G\left(\zeta_{k,1}^{\text{Lag}}\left(\overline{y}_1^*\right), \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}\right) \\ &+ \sum_{k=0}^{M_{\text{Lag},R}} \eta_k^{\text{Lag}}\, G\left(\zeta_{k,R}^{\text{Lag}}\left(\overline{y}_R^*\right), \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}\right) \\ &+ \sum_{i=1}^{R-1} \left(\sum_{k=0}^{M_{\text{Leg},i}} \eta_k^{\text{Leg}}\, G\left(\zeta_{k,i}^{\text{Leg}}\left(\overline{y}_i^*, \overline{y}_{i+1}^*\right), \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}\right)\right),\end{aligned}$$

$\{\overline{y}_i^*\}_{i=1}^R$: the approximated discontinuities locations; $M_{\text{Lag},1}$ and $M_{\text{Lag},R}$: the number of Laguerre quadrature points $\zeta_{.,.}^{\text{Lag}} \in \mathbb{R}$ with corresponding weights $\eta_.^{\text{Lag}}$; $\{M_{\text{Leg},i}\}_{i=1}^{R-1}$: the number of Legendre quadrature points $\zeta_{.,.}^{\text{Leg}}$ with corresponding weights $\eta_.^{\text{Leg}}$.

- $\overline{I}$ can be approximated further depending on (i) the decay of $G \times \rho_1$ in the semi-infinite domains and (ii) how close the roots are to each other.

Notations and Assumptions

Notation

- $\overline{X}_\ell(T) \coloneqq \overline{X}_\ell(T;(Z_1^\ell, \mathbf{Z}_{-1}^\ell))$.
- We denote $\overline{X}_\ell(T)$ by $\overline{X}_T^{N_\ell}$.
- $\overline{X}_k^{N_\ell}$ are the Euler–Maruyama increments of $\overline{X}_T^{N_\ell}$ for $0 \leq k \leq N_\ell$ with $\overline{X}_T^{N_\ell} = \overline{X}_{N_\ell}^{N_\ell}$.

Assumption 5.1

For $p \in \mathbb{N}$ s.t. $1 \leq p \leq 4$, there are positive rdvs C_p with finite moments of all orders such that

$$\forall N_\ell \in \mathbb{N},\ \ \forall k_1, \ldots, k_p \in \{0, \ldots, N_\ell - 1\} : \left| \frac{\partial^p \overline{X}_T^{N_\ell}}{\partial \overline{X}_{k_1}^{N_\ell} \cdots \partial \overline{X}_{k_p}^{N_\ell}} \right| \leq C_p \ \ a.s.$$

Assumption 5.1 is fulfilled if the drift and diffusion coefficients are smooth.

Notations and Assumptions

Assumption 5.2

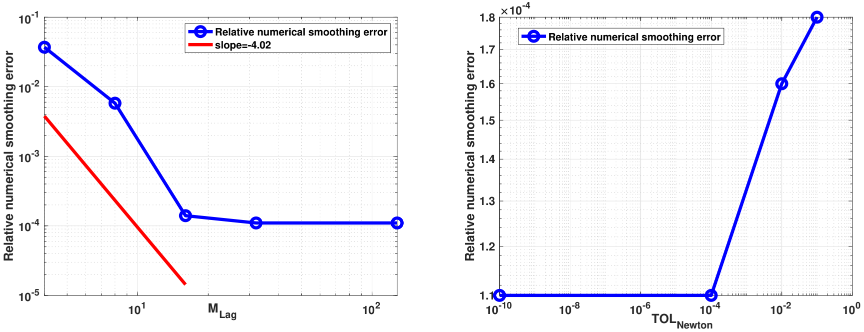
For $p \in \mathbb{N}$ s.t. $1 \le p \le 4$, there are positive rdvs D_p with finite moments of all orders such that [a]

$$\left(\frac{\partial \overline{X}_T^{N_\ell}}{\partial y} \left(Z_1^\ell, \boldsymbol{Z}_{-1}^\ell \right) \right)^{-p} \le C_p \;\; a.s.$$

[a] $y := z_{-1}$

- In (Bayer, Ben Hammouda, and Tempone 2023), we show sufficient conditions where this assumption is valid.
- For instance, Assumption 5.2 is valid for
  - one-dimensional SDEs with a linear or constant diffusion.
  - multivariate SDEs with a linear drift and constant diffusion, including the multivariate lognormal model (see (Bayer, Siebenmorgen, and Tempone 2018)).

Errors in the Numerical Smoothing



(a) (b)

Figure 5.1: Call option under GBM with $N = 4$: The relative numerical smoothing error for a fixed number of ASGQ points $M_{\text{ASGQ}} = 10^3$ plotted against (a) different values of M_{Lag} with a fixed Newton tolerance $\text{TOL}_{\text{Newton}} = 10^{-10}$, (b) different values of $\text{TOL}_{\text{Newton}}$ with a fixed number of Laguerre quadrature points $M_{\text{Lag}} = 128$.

References I

[1] Christian Bayer, Chiheb Ben Hammouda, and Raúl Tempone. "Multilevel Monte Carlo with Numerical Smoothing for Robust and Efficient Computation of Probabilities and Densities". In: *arXiv preprint arXiv:2003.05708* (2022).

[2] Christian Bayer, Chiheb Ben Hammouda, and Raúl Tempone. "Numerical smoothing with hierarchical adaptive sparse grids and quasi-Monte Carlo methods for efficient option pricing". In: *Quantitative Finance* 23.2 (2023), pp. 209–227.

[3] Christian Bayer, Markus Siebenmorgen, and Rául Tempone. "Smoothing the payoff for efficient computation of basket option pricing.". In: *Quantitative Finance* 18.3 (2018), pp. 491–505.

[4] Amal Ben Abdellah et al. "Density estimation by randomized quasi-Monte Carlo". In: *SIAM/ASA Journal on Uncertainty Quantification* 9.1 (2021), pp. 280–301.

References II

[5] Chiheb Ben Hammouda, Nadhir Ben Rached, and Raúl Tempone. "Importance sampling for a robust and efficient multilevel Monte Carlo estimator for stochastic reaction networks". In: *Statistics and Computing* 30.6 (2020), pp. 1665–1689.

[6] Chiheb Ben Hammouda, Alvaro Moraes, and Raúl Tempone. "Multilevel hybrid split-step implicit tau-leap". In: *Numerical Algorithms* 74.2 (2017), pp. 527–560.

[7] Peng Chen. "Sparse quadrature for high-dimensional integration with Gaussian measure". In: *ESAIM: Mathematical Modelling and Numerical Analysis* 52.2 (2018), pp. 631–657.

[8] K Andrew Cliffe et al. "Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients". In: *Computing and Visualization in Science* 14.1 (2011), p. 3.

References III

[9] Josef Dick, Frances Y Kuo, and Ian H Sloan. "High-dimensional integration: the quasi-Monte Carlo way". In: *Acta Numerica* 22 (2013), pp. 133–288.

[10] Oliver G Ernst, Bjorn Sprungk, and Lorenzo Tamellini. "Convergence of sparse collocation for functions of countably many Gaussian random variables (with application to elliptic PDEs)". In: *SIAM Journal on Numerical Analysis* 56.2 (2018), pp. 877–905.

[11] Michael B Giles. "MLMC techniques for discontinuous functions". In: *arXiv preprint arXiv:2301.02882* (2023).

[12] Michael B Giles. "Multilevel Monte Carlo methods". In: *Acta Numerica* 24 (2015), pp. 259–328.

[13] Michael B Giles. "Multilevel Monte Carlo path simulation". In: *Operations Research* 56.3 (2008), pp. 607–617.

References IV

[14] Michael B Giles, Kristian Debrabant, and Andreas Rößler. "Numerical analysis of multilevel Monte Carlo path simulation using the Milstein discretisation". In: *arXiv preprint arXiv:1302.4676* (2013).

[15] Michael B Giles, Tigran Nagapetyan, and Klaus Ritter. "Multilevel Monte Carlo approximation of distribution functions and densities". In: *SIAM/ASA Journal on Uncertainty Quantification* 3.1 (2015), pp. 267–295.

[16] Michael B Giles and Lukasz Szpruch. "Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation". In: (2014).

[17] Stefan Heinrich. "Multilevel monte carlo methods". In: *International Conference on Large-Scale Scientific Computing*. Springer. 2001, pp. 58–67.

References V

[18] Ahmed Kebaier. "Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing". In: *The Annals of Applied Probability* 15.4 (2005), pp. 2681–2705.

[19] Peter E Kloeden and Eckhard Platen. "Stochastic differential equations". In: *Numerical solution of stochastic differential equations*. Springer, 1992, pp. 103–160.

[20] Pierre L'Ecuyer, Florian Puchhammer, and Amal Ben Abdellah. "Monte Carlo and Quasi–Monte Carlo Density Estimation via Conditioning". In: *INFORMS Journal on Computing* (2022).