Hierarchical Monte Carlo Methods for SPDEs

Abdul-Lateef Haji-Ali
Antithetic work: joint with A. Stein (ETH)
MIMC work: joint with H. Hoel (Oslo) & A. Petersson (Linnaeus)

Heriot-Watt University

Overview

Overview

1. Background: SPDEs
2. Antithetic Milstein scheme for SPDEs
3. MIMC for SPDEs
4. Summary and comparison

Stochastic Partial Differential Equations

Consider the H-valued Itô SDE (SPDE)

$$\mathrm{d}X(t) + \big(AX(t) - F(X(t))\big)\,\mathrm{d}t = G(X(t))\,\mathrm{d}W(t), \quad X(0) = X_0,$$

on a separable Hilbert space H.

Stochastic Partial Differential Equations

Consider the H-valued Itô SDE (SPDE)

$$\mathrm{d}X(t) + \big(AX(t) - F(X(t))\big)\,\mathrm{d}t = G(X(t))\,\mathrm{d}W(t), \quad X(0) = X_0,$$

on a separable Hilbert space H.

- $A : D(A) \subset H \to H$ is a densely defined, self-adjoint, positive definite linear operator with orthonormal eigenbasis $\big((e_j, \lambda_j)\big)_{j=1}^{\infty}$ (note that $\lambda_j \to \infty$). We will assume that $\lambda_j \eqsim j^{\nu}$ for some $\nu > 0$.
- A generates an analytic semigroup $(e^{-At}, t \geq 0) \subset \mathcal{L}(H)$ and is boundedly invertible. Here, $e^{-At}e_j = e^{-\lambda_j t}e_j$.
- $W : \Omega \times [0, T] \to H$ is a Q-Wiener process with trace class covariance operator $Q \in \mathcal{L}_1^+(H)$.
- $F : H \to H$ and $G : H \to \mathcal{L}_{HS}(Q^{1/2}H; H)$, where $Q^{1/2}H$ is the RKHS associated to Q.
- $X_0 \in L^2(\Omega; H)$.

Goal: approximate $\mathrm{E}[\,\Psi(X(T))\,]$ for a given QoI $\Psi : H \to \mathbb{R}$.

Regularity of mild solutions

There is a unique mild solution $X : \Omega \times [0, T] \to H$ to the SPDE, given by

$$X(t) = e^{-tA}X_0 + \int_0^t e^{-(t-s)A}F(X(s))ds + \int_0^t e^{-(t-s)A}G(X(s))dW(s),$$

for $t \in [0, T]$. Under some assumptions: $X(t) \in L^p(\Omega; \dot{H}^\kappa)$ for some $\kappa > 0$, $t \in [0, T]$ and $\dot{H}^\kappa := D(A^{\kappa/2})$. (Convergence rates of approximations will be limited by this κ)

Fractional operators $\quad A^r v := \sum_{j=1}^{\infty} \lambda_j^r \langle e_j, v\rangle e_j \quad \text{for} \quad r \in \mathbb{R}$

Extension of norm: $\| \cdot \|_{\dot{H}^r} := \|A^{r/2} \cdot \|$ and associated Hilbert space

$$\dot{H}^r = \begin{cases} \mathrm{D}(A^{r/2}) & r \geq 0 \\ \overline{H}^{\| \cdot \|_{\dot{H}^r}} & r < 0, \end{cases}$$

Canonical example: Stochastic heat equation

$$\mathrm{d}X(t) + \big(-\triangle X(t) - F(X(t))\big)\,\mathrm{d}t = G(X(t))\,\mathrm{d}W(t), \qquad \text{in } [0,T]\times\mathcal{D}$$
$$X(t,\cdot)|_{\partial\mathcal{D}} = 0, \qquad X(0,\cdot) = X_0\,.$$

- Here $X \in L^2(\mathcal{D})$ for a bounded domain $\mathcal{D} \subset \mathbb{R}^d$, $d = 1, 2, 3$, convex or with $\mathcal{C}^2$ boundary $\partial\mathcal{D}$.
- $-\triangle : D(-\triangle) = W^{2,2}(\mathcal{D}) \subset L^2(\mathcal{D}) \to L^2(\mathcal{D})$ is densely defined, positive definite, with eigenbasis $\lambda_j \eqsim j^{2/d}$ (Weyl's law). E.g., $\mathcal{D} = (0,1)$: $e_j(x) = \sqrt{2}\sin(j\pi x)$, $\lambda_j = \pi^2 j^2$.
- **Fractional spaces:** $\dot{H}^\alpha = D((-\triangle)^{\alpha/2}) \cong H_0^\alpha(\mathcal{D})$ (Sobolev of order α, zero BC):

$$\dot{H}^0 = L^2(\mathcal{D}), \qquad \dot{H}^1 = H_0^1(\mathcal{D}), \qquad \dot{H}^2 = H^2(\mathcal{D}) \cap H_0^1(\mathcal{D}).$$

  So $X(t) \in L^p(\Omega; \dot{H}^\alpha)$ means X has α square-integrable weak derivatives a.s.
- A natural example of F: $(F(u))(x) = f(u(x))$, $f : \mathbb{R} \to \mathbb{R}$ smooth with bounded derivatives.

Approximations: Space

Replace $V = D(A^{1/2})$ by a finite dimensional-subspace V_N, $N = \dim(V_N)$. We can then define a discrete operator $A_N : V_N \to V_N$ by

$$(A_N v_N, w_N)_H = (A^{1/2} v_N, A^{1/2} w_N)_H, \quad v_N, w_N \in V_N. \tag{1}$$

Let the associated H-orthogonal projection operation be $P_N : H \to V_N$, the semi-discrete mild problem is then to find X_N

$$\begin{aligned} X_N(t) = e^{-tA_N} P_N X_0 &+ \int_0^t e^{-(t-s)A_N} P_N F(X_N(s)) ds \\ &+ \int_0^t e^{-(t-s)A_N} P_N G(X_N(s)) dW(s). \end{aligned}$$

Approximations: Time

Let $\Delta t = T/M$ and $t_m = m\Delta t$, and we approximate the semigroup with S_N^M,

$$X_N^M(t_{m+1}) = S_N^M(\Delta t) P_N X_N^M(t_m) + \int_{t_m}^{t_{m+1}} S_N^M(t_{m+1} - s) P_N F(X_N^M(t_m)) ds \\ + \int_{t_m}^{t_{m+1}} S_N^M(t_{m+1} - s) P_N G(X_N^M(t_m)) dW(s).$$

Given $(\widetilde{e}_j)_j$ as the H-orthonormal eigen-basis of A_N with corresponding eigenvalues, $\widetilde{\lambda}_j$. We consider two methods for approximating the time and stochastic integrals:

- Rational Approximation: $S_N^M(t_{m+1} - s) =: r(\Delta t A_N) \approx e^{-\Delta t A_N}$ defined as $r(\Delta t A_N)\widetilde{e}_j = r(\Delta t \widetilde{\lambda}_j)\widetilde{e}_j$. E.g., implicit Euler: $r(z) = (1+z)^{-1}$; Crank-Nicolson: $r(z) = (1 - z/2)/(1 + z/2)$.
- Exponential integrator (Jentzen and Kloeden 2009): Set $S_N^M(t_{m+1} - s) = e^{-(t_{m+1}-s)A_N}$ and evaluate both integrals *exactly* via the eigendecomposition of A_N.

When G is non-linear/non-diagonal, or $Q \not\parallel A_N$ is, we also require approximating the Wiener process.

Challenge

- Each sample costs $\mathcal{O}(MN)$, but accuracy requires $M, N \to \infty$.
- A **Monte Carlo** with $\mathcal{N}$ samples of the level-L approximation has cost $\mathcal{O}(\mathcal{N} MN)$, bias $\mathcal{O}(M^{-w_1} + N^{-w_2})$, stat. error $\mathcal{O}(\mathcal{N}^{-1/2})$, giving complexity $\mathcal{O}(\varepsilon^{-2-\frac{1}{w_1}-\frac{1}{w_2}})$.
- **Example**: Assume Q is smooth enough to have $X(t) \in \dot{H}^{\kappa}$. Then $w_1 = \min(1, \kappa)$ (Euler weak temporal) while $w_2 = \kappa\nu$ (spatial weak, e.g., for *spectral Galerkin*). When G is diagonal and $\kappa \geq 1$, this gives complexity $\mathcal{O}(\varepsilon^{-3-1/\kappa\nu})$.

Challenge

- Each sample costs $\mathcal{O}(MN)$, but accuracy requires $M, N \to \infty$.
- A **Monte Carlo** with $\mathcal{N}$ samples of the level-L approximation has cost $\mathcal{O}(\mathcal{N} MN)$, bias $\mathcal{O}(M^{-w_1} + N^{-w_2})$, stat. error $\mathcal{O}(\mathcal{N}^{-1/2})$, giving complexity $\mathcal{O}(\varepsilon^{-2-\frac{1}{w_1}-\frac{1}{w_2}})$.
- **Example**: Assume Q is smooth enough to have $X(t) \in \dot{H}^{\kappa}$. Then $w_1 = \min(1, \kappa)$ (Euler weak temporal) while $w_2 = \kappa\nu$ (spatial weak, e.g., for *spectral Galerkin*). When G is diagonal and $\kappa \geq 1$, this gives complexity $\mathcal{O}(\varepsilon^{-3-1/\kappa\nu})$.

- Extend hierarchical methods from (finite dimensional) SDEs to (infinite dimensional) SPDEs?

- **Goal of this presentation**: Present two examples of such extensions, illustrating main difficulties and when they do not work.
- **Style of this presentation**: hopefully intuitive. Leaving proofs and rigour to the papers.

Overview

Multilevel Monte Carlo[1]

MLMC is based on a telescoping sum:

$$\mathrm{E}[\,\Psi(X_{N_L}^{M_L})\,] = \sum_{\ell=0}^{L} \underbrace{\mathrm{E}[\,\Psi(X_{N_\ell}^{M_\ell}) - \Psi(X_{N_{\ell-1}}^{M_{\ell-1}})\,]}_{\text{correction } \Delta_\ell \Psi} \approx \sum_{\ell=0}^{L} E_{m_\ell}[\Delta_\ell \Psi]$$

The bound on the $\mathrm{Var}[\,\Delta_\ell \Psi\,] = \mathcal{O}(M^{-\beta_1} + N^{-\beta_2})$ become relevant.

For an Euler method in the previous example, $\beta_1 = \min(1, \kappa)$, while $\beta_2 = \kappa\nu$. For $\kappa\nu > 1$, time discretization is the bottleneck. For $\kappa\nu$ small, the spatial discretization becomes the bottleneck.

For $\kappa \geq 1$, an MLMC method would give a complexity of $\mathcal{O}(\varepsilon^{-2-1/\kappa\nu})$ ("β" $= 1$, cost-per-level rate "γ" $= 1 + 1/(\kappa\nu)$).

[1] Giles, M. B. (2008). "Multilevel Monte Carlo path simulation". In: *Oper. Res.* 56.3, pp. 607–617.

SDEs: Euler-Maruyama vs. Milstein

For the finite-dimensional SDE $\mathrm{d}X = a(X)\,\mathrm{d}t + \sum_{i=1}^{d'} b_i(X)\,\mathrm{d}W^i$, the Milstein scheme

$$\begin{aligned}\hat{X}_{t_{m+1}} &= \hat{X}_{t_m} + a\left(\hat{X}_{t_m}\right)\Delta t + \sum_{i=1}^{d'} b_i\left(\hat{X}_{t_m}\right)\Delta_m W^i \\ &\quad + \frac{1}{2}\sum_{i=1}^{d'}\sum_{j=1}^{d'} J_{b_j}\left(\hat{X}_{t_m}\right) b_i(\hat{X}_{t_m})\left(\Delta_m W^i \Delta_m W^j - \delta_{i,j}\Delta t - A_m^{ij}\right)\end{aligned}$$

where A_m^{ij} is Lévy area, which involves an iterated integral in $\mathrm{d}W^i\,\mathrm{d}W^j$.

Milstein MLMC complexity: $\mathcal{O}(\varepsilon^{-2})$ (optimal) whenever Lévy area is available (infeasible for large d') or when noise is "commutative" (note $A_m^{ij} = -A_m^{ji}$)

$$J_{b_j}(x)b_i(x) = J_{b_i}(x)b_j(x)$$

Antithetic Milstein for SDEs[2]

Key idea: avoid Lévy area simulation via **antithetic coupling** of two fine-grid paths.

Set $A_m^{ij} = 0$. Split each coarse step $[t_m, t_{m+1}]$ into two half-steps with increments $\delta_m W$ and $\delta_{m+1/2} W$.

- Fine path X_f: uses $\delta_m W$ first, $\delta_{m+1/2} W$ second.
- Antithetic path X_a: **swaps** the order.
- MLMC correction uses $\overline{\Psi} = \frac{1}{2}(\Psi(X_f) + \Psi(X_a))$.

Result: for $\Psi \in C_b^2$:

$$\mathrm{Var}[\,\overline{\Psi} - \Psi(X_c)\,] = \mathcal{O}(\Delta t^2)$$

without simulating Lévy areas. **Complexity:** $\mathcal{O}(\varepsilon^{-2})$ (optimal).

[2] Giles, M. B. and Szpruch, L. (2014). "Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation". In: *Ann. Appl. Probab.* 24.4, pp. 1585–1620. DOI: 10.1214/13-AAP957.

Truncated Milstein scheme for SPDEs

Recall the mild solution ($F = 0$ for presentation)

$$X_N^M(t_{m+1}) = S_N(\Delta t) P_N X_N^M(t_m) + \int_{t_m}^{t_{m+1}} S_N(t_{m+1} - s) P_N G(X_N^M(t_m)) dW(s).$$

Using a rational time- approximation, and a truncated Wiener process, W_K, the truncated Milstein scheme becomes:

$$X_N^M(t_{m+1}) = r(\Delta t A_N) P_N \Big[X_N^M(t_m) + G(X_m^{N,M}(t_m))\, \Delta_m W_K \\ + \frac{1}{2} \sum_{k,l=1}^{K} G'(X_N^M(t_m)) \big(P_N G(X_N^M(t_m)) \sqrt{\mu_l}\, e_l \big) \sqrt{\mu_k}\, e_k (\Delta_m w_k\, \Delta_m w_l - \delta_{kl} \Delta t) \Big].$$

Compact notation:

$$X_{N,m+1}^K = r(\Delta t A_N) P_N \Big[X_{N,m}^K + G(X_{N,m}^K)\, \Delta_m W_K + \mathcal{G}(X_{N,m}^K)\, \Delta_m \mathcal{W}_{m,K} \Big]$$

for a well-defined $\mathcal{L}_2(H)$-process $\mathcal{W}_{m,K}$ and mapping $\mathcal{G}$, and where $\Delta_m \mathcal{W}_{m,K} = \Delta_m W_K \otimes \Delta_m W_K - \Delta t \sum_{k=1}^{K} \mu_k\, e_k \otimes e_k$.

Antithetic coupling for SPDEs

Coarse approximation (M steps, N, K): standard truncated Milstein ($Y_m^c = X_N^M(m\Delta t)$).

Fine approximation ($2M$ steps, $N_f \geq N$, $K_f \geq K$, step $\delta t = \Delta t/2$): ($Y_m^f = X_N^{M,f}(m\Delta t), Y_m^a = X_N^{M,a}(m\Delta t)$)

$$
\begin{aligned}
Y_{m+1/2}^f &= r(\delta t A_{N_f}) P_{N_f}\Big[Y_m^f + G(Y_m^f)\,\delta_m W_{K_f} + \mathcal{G}(Y_m^f)\,\delta_m \mathcal{W}_{m,K_f} \Big],\\
Y_{m+1}^f &= r(\delta t A_{N_f}) P_{N_f}\Big[Y_{m+1/2}^f + G(Y_{m+1/2}^f)\,\delta_{m+1/2} W_{K_f} + \mathcal{G}(Y_{m+1/2}^f)\,\delta_{m+1/2} \mathcal{W}_{m,K_f} \Big].
\end{aligned}
$$

Antithetic approximation: **swap** the order of increments:

$$
\begin{aligned}
Y_{m+1/2}^a &= r(\delta t A_{N_f}) P_{N_f}\Big[Y_m^a + G(Y_m^a)\,\delta_{m+1/2} W_{K_f} + \mathcal{G}(Y_m^a)\,\delta_{m+1/2} \mathcal{W}_{m,K_f} \Big],\\
Y_{m+1}^a &= r(\delta t A_{N_f}) P_{N_f}\Big[Y_{m+1/2}^a + G(Y_{m+1/2}^a)\,\delta_m W_{K_f} + \mathcal{G}(Y_{m+1/2}^a)\,\delta_m \mathcal{W}_{m,K_f} \Big].
\end{aligned}
$$

MLMC correction: $\frac{1}{2}(\Psi(Y_M^f) + \Psi(Y_M^a)) - \Psi(Y_M^c)$.

Improved variance decay[3]

Let $\overline{Y}_M = \frac{1}{2}(Y_M^f + Y_M^a)$ and Y_M^c the coarse approximation. Recall that for Euler/truncated Milstein **without** antithetic correction:

$$\mathrm{E}\left[\, \| Y_M^f - Y_M^c \|_H^2 \,\right] \lesssim M^{-\min(1,\kappa)} + N^{-\kappa\nu}.$$

Theorem (Haji-Ali & Stein, 2023)

Suppose $X \in L^8(\Omega; \dot{H}^\kappa)$ for some $\kappa \geq 1$, and F, G are twice differentiable with bounded derivatives. Then

$$\mathrm{E}\left[\, \| \overline{Y}_M - Y_M^c \|_H^2 \,\right] \lesssim M^{-\min(2,\kappa)} + N^{-\kappa\nu}.$$

Proof. Exact similar steps/cancellations as antithetic SDE proof.

For $\kappa \geq 2$, MLMC cost is now $\mathcal{O}'(\varepsilon^{-2-\max(0, 2/\kappa\nu - 1)})$.

[3] **H-A** and Stein, A. (2023). *An Antithetic Multilevel Monte Carlo-Milstein Scheme for Stochastic Partial Differential Equations*. arXiv: 2307.14169 [math.NA].

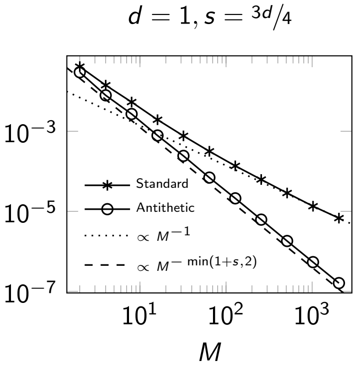
Generality: FEM and arbitrary diffusion

The antithetic approach is broadly applicable:

Spatial discretization:	Any sequence (V_N) with approximation property including finite elements on irregular domains.
Time integrator:	Any stable, consistent rational approximation r: implicit Euler, Crank-Nicolson, etc.
Diffusion G:	Any twice differentiable G with bounded derivatives and linear growth (might increase cost per sample).
Noise Q:	Eigenbasis of Q need not coincide with the eigenbasis of A_N (though approximating W becomes difficult of course).

Numerical results: antithetic variance improvement

Stochastic heat equation, $A = -\triangle$, $\mathcal{D} = [0,1]^d$, spectral Galerkin, $Q = (-\triangle)^{-s}$,$\kappa \in [1, \min(1+s, 2))$:



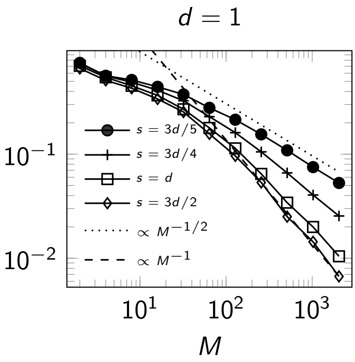


Figure 1: *(left)* Variance of antithetic vs. standard estimator ($d = 1$, $s = {}^3/_4$). *(right)* Relative variance decay for several smoothness parameters s, $d = 1$. Variance estimates from ≥ 4000 MC samples.

Overview

Motivation: two independent parameters

In MLMC we chain M_ℓ and N_ℓ together on each level. But the convergence rates in M and N can differ significantly – either one may dominate the bias or variance bounds.

Multi-Index Monte Carlo (MIMC)[4] treats M and N as **independent** indices (ℓ_1, ℓ_2) and uses a double telescoping sum, potentially gaining a factor in the complexity.

Key challenge: MIMC requires a **multiplicative** bound on the second-order difference — not all schemes satisfy this.

[4] **H-A**, Nobile, F., and Tempone, R. (2016). "Multi-index Monte Carlo: when sparsity meets sampling". In: *Numer. Math.* 132, pp. 767–806.

MIMC estimator[5]

The **second-order difference** is

$$\Delta_{\boldsymbol\ell}\Psi(X) := \Psi(X_{N_{\ell_2}}^{M_{\ell_1}}) - \Psi(X_{N_{\ell_2-1}}^{M_{\ell_1}}) - \Psi(X_{N_{\ell_2}}^{M_{\ell_1-1}}) + \Psi(X_{N_{\ell_2-1}}^{M_{\ell_1-1}}).$$

The **MIMC estimator** is then $\quad \mu_{\mathrm{MI}} := \sum_{\boldsymbol\ell\in\mathcal{I}} \widehat{E}_{m_{\boldsymbol\ell}}[\Delta_{\boldsymbol\ell}\Psi(X)],$

with a **total-degree index set** $\mathcal{I} = \{\boldsymbol\ell \in \mathbb{N}_0^2 : \eta_1\ell_1 + \eta_2\ell_2 \le L(\varepsilon)\}$, for some weights η_1, η_2.

Theorem

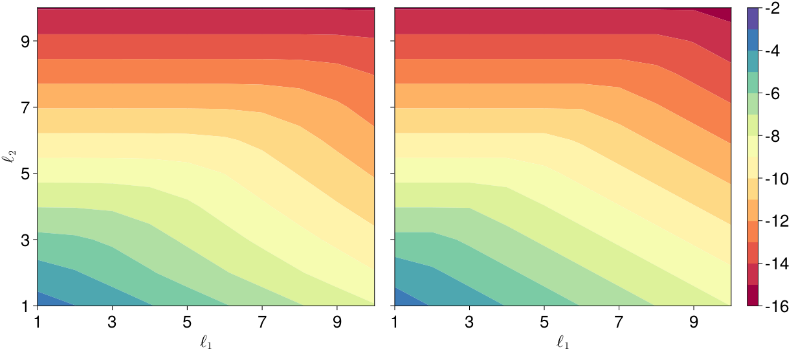
If $\|\Delta_{\boldsymbol\ell}\Psi(X)\|_{L^2}^2 \lesssim M_{\ell_1}^{-\beta_1} N_{\ell_2}^{-\beta_2}$ (multiplicative!) for $\beta_1, \beta_2 > 0$, then optimal MIMC achieves $\mathcal{O}'(\varepsilon^{-2-2u})$ complexity, with $2u = 2\max(0, 1/\beta_1 - 1, 1/\beta_2 - 1)$.

If true, then for $\kappa > 1$, $\mathcal{O}'(\varepsilon^{-2-\max(0,1/\kappa\nu-1)})$. For $\kappa < 1$, the savings are more significant. Also, better weak rates improve the complexity.

[5] **H-A**, Nobile, F., and Tempone, R. (2016). "Multi-index Monte Carlo: when sparsity meets sampling". In: *Numer. Math.* 132, pp. 767–806.

Numerical verification

SPDE with $A = 0.2(-\triangle)^{2/3}$ on $\mathcal{D} = (0,1)$, $\kappa \approx 1$, $\nu = 4/3$. Test: verify the multiplicative structure of $e(\ell_1, \ell_2) := \mathrm{E}[\,\|\Delta_{\boldsymbol{\ell}} X\|^2\,]$.



Left: measured $e(\ell_1, \ell_2)$. **Right:** fitted $p(\ell_1, \ell_2)$ with $\beta_1 = 0.98 \approx 1$ and $\beta_2 = 1.62 \approx 4/3$. Confirms the multiplicative bound.

MIMC second-order difference bound[6]

Theorem

Suppose F, G and Q are "sufficiently smooth" such that $X \in L^{10}(\Omega; \dot{H}^{\kappa})$ for some $\kappa \in (0,2)$, i.e, $\|A^{\frac{\kappa-1}{2}} Q^{1/2}\|_{\mathcal{L}_2(H)} < \infty$, and suppose that (a partial list)

- *$Q \parallel A$ (Can be relaxed, but need to deal with costly simulation of stochastic integrals or introduce a bias.).*
- *F is twice differentiable with bounded derivatives.*
- *Exponential integrator with spectral Galerkin (no FEM nor rational approximation).*
- *$(\mathrm{Id} - P_N)G(u) = 0$ for $u \in P_N(H)$.*

Then

$$\|\Delta_{\ell}\Psi\|^2_{L^2(\Omega;H)} \lesssim 2^{-\min(1,\kappa)\ell_1 - \kappa\nu\ell_2},$$

We also have sharper bounds when $\kappa \in (0,1)$, or when Ψ is linear.

[6] **H-A**, Hoel, H., and Petersson, A. (2025). *The multi-index Monte Carlo method for semilinear stochastic partial differential equations.* arXiv: 2502.00393 [math.NA].

Proof sketch: Decomposition

Recall

$$X_N^M(t_{m+1}) = S_N^M(\Delta t)P_N X_N^M(t_m) + \int_{t_m}^{t_{m+1}} S_N^M(t_{m+1}-s)P_N F(X_N^M(t_m))ds + \int_{t_m}^{t_{m+1}} S_N^M(t_{m+1}-s)P_N G(X_N^M(t_m))dW(s).$$

Let $F_N^M = F(X_N^M(t_m))$, then we have the double difference (suppressing arguments)

$$\begin{aligned}
&\underbrace{S_{\bar N}^{\bar M}P_{\bar N}F_{\bar N}^{\bar M} - S_{\bar N}^{\bar M}P_N F_N^{\bar M} - S_N^M P_{\bar N}F_{\bar N}^M + S_N^M P_N F_N^M}_{=:A} \\
&= \underbrace{(S_{\bar N}^{\bar M} - S_N^M)(P_{\bar N}F_{\bar N}^M - P_N F_N^M)}_{\mathrm{I}_F} \\
&\quad + \underbrace{S_{\bar N}^{\bar M}(P_{\bar N} - P_N)(F_N^{\bar M} - F_N^M)}_{\mathrm{II}_F} + \underbrace{S_{\bar N}^{\bar M}P_{\bar N}[\Delta_\ell F_N^M]}_{\text{Gronwall (feeds back } \Delta X)}
\end{aligned}$$

The G-integrands follow *identically* (with Itô isometry in place of Bochner norm). **Gronwall** will be absorbed (Lipschitz F or G; requires $S_N^M P_N$ uniformly stable).

Proof sketch: Decomposition

Only II_F plays nicely! We "assume" our way out of the others.

- Term $\mathrm{I}_{F/G}$ Product of semi-group approximation (in space/time) and F/G-approximation (in space only).

  $$\mathrm{I}_F = (S_{\bar{N}}^{\bar{M}} - S_N^M)(P_{\bar{N}} F_{\bar{N}}^M - P_N F_N^M)$$

  Using $S_N^M = S = e^{-At}$ (exact semigroup, independent of M), then $I_F = I_G = 0$.

- $\mathrm{II}_G = 0$, using the assumption that $(\mathrm{Id} - P_N) G(u) = 0$.

  $$\mathrm{II}_G = S_{\bar{N}}^{\bar{M}} (P_{\bar{N}} - P_N)(G_N^{\bar{M}} - G_N^M)$$

Proof sketch: the II_F-term

$$\mathrm{II}_F = \int_0^t e^{-A(t-s)} \underbrace{P_{\bar{N}}(\mathrm{Id} - P_N)}_{\text{spatial diff}} \underbrace{(F(\cdots^{\bar{M}}) - F(\cdots^{M}))}_{\text{temporal diff}} \,\mathrm{d}s.$$

Inserting $A^{-\kappa/2} \cdot A^{\kappa/2}$ and using $\|(I - P_N)A^{-\kappa/2}\| = \lambda_{N+1}^{-\kappa/2}$:

$$\lesssim \int_0^t \underbrace{\|A^{\kappa/2} e^{-A(t-s)}\|}_{\lesssim (t-s)^{-\kappa/2}} \lambda_{N+1}^{-\kappa/2} \underbrace{\|F(\cdots^{\bar{M}}) - F(\cdots^{M})\|}_{\lesssim M^{-\min(\kappa,1)/2}} \,\mathrm{d}s.$$

The $(t-s)^{-\kappa/2}$ singularity is integrable for $\kappa < 2$, yielding

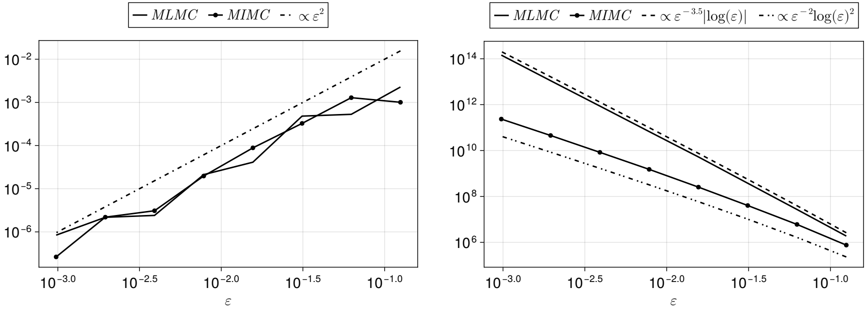
$$\lesssim \lambda_{N+1}^{-\kappa/2} M^{-\min(\kappa,1)/2} \eqsim N^{-\kappa\nu/2} M^{-\min(\kappa,1)/2}.$$

This is multiplicative: one factor from space, one from time.

The ceiling. Using more regularity $r+1 > \kappa$ would give a sharper spatial factor $\lambda_{N+1}^{-(r+1)/2}$, but the singularity $(t-s)^{-(r+1)/2}$ diverges for $r \geq 1$. So $\mathrm{tr}(A^{\kappa-1}Q) < \infty$ is tight: this is the exact level of Q smoothness at which the singularity becomes integrable.

Numerical performance: MLMC vs. MIMC

$A = (-\Delta)^{2/3}$, $\mathcal{D} = (0,1)$, $\kappa \approx 1$, nonlinear $F(x) = \sin(\pi x)$:



Left: RMSE vs. ε. **Right:** cost vs. ε. MIMC achieves $\mathcal{O}(\varepsilon^{-2}|\log \varepsilon|^2)$ vs. MLMC's $\mathcal{O}(\varepsilon^{-2-(2/(\kappa\nu)-1)})$.

General G?

$$\mathrm{II}_G = \int_0^t e^{-A(t-s)} \underbrace{P_{\bar{N}}(\mathrm{Id} - P_N)}_{\text{spatial diff}} \underbrace{\left(G(\cdots^{\bar{M}}) - G(\cdots^{M})\right)}_{\text{temporal diff}} \,\mathrm{d}W(s).$$

- The F-term used $(t-s)^{-\kappa/2}$ (integrable for $\kappa < 2$). Itô isometry *squares* the integrand for the G-term: $(t-s)^{-\kappa/2} \longrightarrow (t-s)^{-\kappa}$ — integrable only for $\kappa < 1$. Again, more Q smoothness cannot fix this.
- Even for $\kappa < 1$, we need a Hilbert-Schmidt Lipschitz bound on G, which requires L^∞ (or its $\dot{H}^{\frac{d^+}{2}}$ embedding) bounds on $X_N^{\bar{M}} - X_N^M$, reducing the temporal rate and restricting the bound to $d = 1$.
- Can also be seen, using the MVT, by considering

$$\sum_{j=N}^{\bar{N}} \int_{t_{m-1}}^{t_m} \left\| e^{-(t_m - s)A} \, G'(\cdot)\left(X_N^{\bar{M}} - X_N^M, \, \mu_j^{1/2} e_j\right) \right\|^2 \mathrm{d}s,$$

which requires a $\dot{H}^{-\alpha}$ norm on $\mu_j^{1/2} e_j$ to get the spatial convergence factor $\lambda_N^{-\alpha}$, and hence a $\dot{H}^{\alpha}$ norm on $X_N^{\bar{M}} - X_N^M$, reducing the temporal rate.

Why implicit Euler *fails* for MIMC[7]

Recall

$$\mathrm{I}_G = \int_0^T (S_{\bar N}^{\bar M} - S_N^M)(P_{\bar N} - P_N) G(\cdot)\,\mathrm{d}W(s).$$

The difference $S^{\bar M} - S^M$ is a *perturbation* to bound, not an analytic semigroup like $S(t)$ to exploit.
Only regularity available is the Q-budget $\mathrm{tr}(A^{\kappa-1}Q) < \infty$ ($\kappa - 1$ degrees), split $\alpha + r' = \kappa - 1$:

$$\mathrm{E}[\,\|\mathrm{I}_G\|^2\,]^{1/2} \lesssim \underbrace{\lambda_{N+1}^{-\alpha/2}}_{\alpha \text{ deg. (spatial)}} \cdot \underbrace{M^{-(1+r')/2}}_{r'=\kappa-1-\alpha \text{ deg. (semigroup error)}}.$$

[7] Reisinger, C. and Wang, Z. (2018). "Analysis of multi-index Monte Carlo estimators for a Zakai SPDE". In: *J. Comput. Math.* 36.2, pp. 202–236.

FEM: structural failure

Recall

$$\mathrm{II}_F = \int_0^T e^{-(T-s)A_N}\,(P_{\bar N} - P_N)(F_N^{\bar M} - F_N^M)\,\mathrm{d}s$$

.
FEM ($h \sim N^{-1/d}$): e^{-tA_N} has no spectral structure. Use stability $\|e^{-tA_N}\| \le 1$ and FEM approximation property

$$\|\mathrm{II}_F^h\| \;\lesssim\; h^\kappa \cdot \|F_N^{\bar M} - F_N^M\|_{\dot H^\kappa}.$$

Since $F_N^{\bar M} - F_N^M \in V_N^h$, a inverse estimate applies: FEM basis functions have support on scale h, so $\|v_h\|_{\dot H^\kappa} \lesssim h^{-\kappa}\|v_h\|_H$. Hence No spatial rate.

Overview

Conclusions and future work

Summary:

- Extend hierarchical methods from SDEs to SPDEs, exploiting smoothness.
- **Antithetic MLMC:** broad applicability (FEM, general G); avoids Lévy area; improves variance from $\mathcal{O}(M^{-\min(1,\kappa)})$ to $\mathcal{O}(M^{-\min(2,\kappa)})$.
- **MIMC:** restricted in setting but achieves optimal multiplicative rates.

Future work:

- Antithetic with MIMC?
- Sharp temporal rates for the additive noise case via stochastic sewing.
- SPDEs with Lévy noise; first-order hyperbolic SPDEs.

Comparison of the two approaches

	Antithetic MLMC	**MIMC**
Spatial disc.	FEM or spectral	Spectral only
Diffusion G	Twice differentiable	Diagonal/shift form on spectrum
Time integrator	Rational approx. (stable)	Exponential integrator
Parameters	M, N, K chained via MLMC	M, N treated independently
Key insight	Antithetic swap kills $O(\Delta t^{1/2})$ terms	Semigroup factorization kills stochastic cross-term

References

- **H-A** and Stein, A. (2023). *An Antithetic Multilevel Monte Carlo-Milstein Scheme for Stochastic Partial Differential Equations*. arXiv: 2307.14169 [math.NA]
- **H-A**, Hoel, H., and Petersson, A. (2025). *The multi-index Monte Carlo method for semilinear stochastic partial differential equations*. arXiv: 2502.00393 [math.NA]
- Giles, M. B. and Szpruch, L. (2014). "Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation". In: *Ann. Appl. Probab.* 24.4, pp. 1585–1620. DOI: 10.1214/13-AAP957
- **H-A**, Nobile, F., and Tempone, R. (2016). "Multi-index Monte Carlo: when sparsity meets sampling". In: *Numer. Math.* 132, pp. 767–806
- Jentzen, A. and Kloeden, P. E. (2009). "Overcoming the order barrier in the numerical approximation of stochastic partial differential equations with additive space-time noise". In: *Proc. R. Soc. Lond., Ser. A, Math. Phys. Eng. Sci.* 465.2102, pp. 649–667. DOI: 10.1098/rspa.2008.0325

References

- Reisinger, C. and Wang, Z. (2018). "Analysis of multi-index Monte Carlo estimators for a Zakai SPDE". In: *J. Comput. Math.* 36.2, pp. 202–236
- Giles, M. B. (2008). "Multilevel Monte Carlo path simulation". In: *Oper. Res.* 56.3, pp. 607–617

Q-Wiener process on Hilbert Space

W is a $Q-$Wiener process, meaning it has independent increments

$$W(t+\Delta t) - W(t) \sim \mathcal{N}(0, \Delta t Q).$$

Here, for any $\phi, \psi \in H, t \geq 0$,

$$\mathrm{E}[\,\langle W(t), \phi\rangle \,\langle W(t), \psi\rangle\,] = t\langle Q\phi, \psi\rangle,$$

for $Q \in \mathcal{L}_1^+(H)$ being a covariance operator (non-negative, trace-class, self-adjoint) with eigenpairs (e_k, μ_k) (same eigenbasis as operator A!).

Representation: $$W(t) = \sum_{j=1}^{\infty} \sqrt{\mu_j} B_j(t) e_j$$

with $B_j(t)$ independent, scalar-valued Brownian motions, defined on a complete probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$.
When approximating the X, we truncate the representation to some number of terms.

Q-Wiener process on Hilbert Space

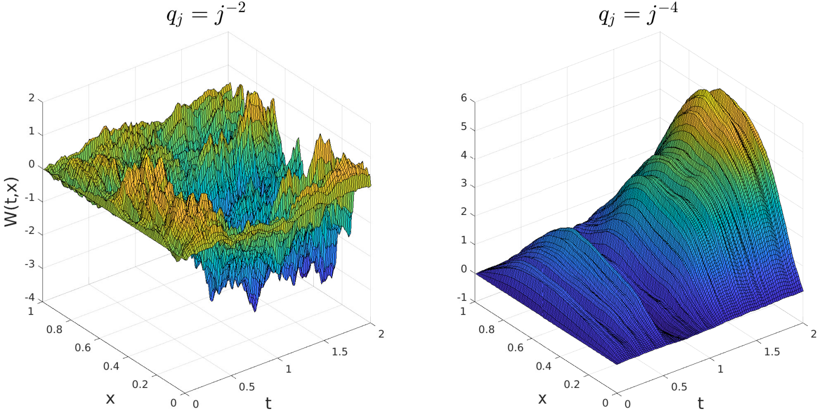


Figure 2: Regularity of W depends on decay rate of $(\mu_j = q_j)$

Prior MIMC for SPDEs[8]

MIMC applied to the Zakai SPDE

$$\mathrm{d}X + \left(\mu \frac{\partial}{\partial x} - \frac{1}{2}\frac{\partial^2}{\partial x^2}\right) X \, \mathrm{d}t = -\sqrt{\rho}\, \frac{\partial X}{\partial x} \, \mathrm{d}B(t)$$

using finite differences in space and Milstein in time.

- Explicit scheme: **unstable** for MIMC (stability condition $\Delta t / \Delta x^2 \lesssim 1$ couples M and N — defeats the purpose).
- Implicit scheme: unconditionally stable, but MIMC complexity is $\mathcal{O}(\varepsilon^{-2} |\log \varepsilon|^z)$ with $z \in \{1, 3\}$ — slightly worse than MLMC.

We also tested implicit Euler MIMC on our SPDE and found no improvement over MLMC. Why? The implicit Euler stochastic cross-term is non-zero and, though it admits a product bound, its temporal variance rate is < 1 — insufficient for MIMC optimality.

$\Rightarrow$ A different time-stepping scheme is required.

[8] Reisinger, C. and Wang, Z. (2018). "Analysis of multi-index Monte Carlo estimators for a Zakai SPDE". In: *J. Comput. Math.* 36.2, pp. 202–236.

Proof sketch: antithetic cancellations

A Taylor expansion of G around $\overline{Y}_m$ gives

$$\overline{Y}_{m+1} = r(\delta t A_{N_f})^2 P_{N_f}\Big[\overline{Y}_m + G(\overline{Y}_m)\,\Delta_m W_{K_f} + \mathcal{G}(\overline{Y}_m)\,\Delta_m \mathcal{W}_{m,K_f}\Big] + \overline{\Xi}_m + \overline{O}_m,$$

where $\overline{O}_m$ is a martingale increment. The **key cancellation**:

- In $Y^f_{m+1} - Y^c_{m+1}$ the leading $\mathcal{O}(\delta t^{1/2})$ noise is $\propto G(\overline{Y}_m)(\delta_{m+1/2} W - \frac{1}{2}\Delta_m W)$.
- In $Y^a_{m+1} - Y^c_{m+1}$ it is $\propto G(\overline{Y}_m)(\delta_m W - \frac{1}{2}\Delta_m W)$ — opposite sign.
- Their **average cancels** the $\mathcal{O}(\delta t^{1/2})$ term; what remains in $\overline{\Xi}_m$ are cross-products $\delta_m W \otimes \delta_{m+1/2} W = \mathcal{O}(\delta t)$.

Combined with semigroup analyticity $\|r(\delta t A)^2 - r(\Delta t A)\|_{\mathcal{L}(\dot{H}^\alpha, H)} \lesssim \delta t^{\alpha/2}$, Grönwall's inequality gives $\mathrm{E}[\,\|\overline{Y}_M - Y^c_M\|^2\,] = \mathcal{O}(M^{-\min(\kappa,2)})$.