University of Oxford

Mathematical Institute

InFoMM
Industrially Focused Mathematical Modelling

MLQMC Methods for Elliptic PDEs Driven by White Noise

M. Croci (Oxford), M. B. Giles (Oxford), M. E. Rognes (Simula), P. E. Farrell (Oxford)

SIAM CSE and ICIAM 2019

EPSRC
Engineering and Physical Sciences Research Council

simula

Overview

Introduction

The motivation of our research is the solution of partial differential equations with random coefficients via the multilevel Quasi-Monte Carlo method.

Today we focus on the ubiquitous model problem,

$$\begin{aligned} -\nabla \cdot (e^{u(x,\omega)} \nabla p(x,\omega)) &= 1, \qquad x \in G \subset \mathbb{R}^d, \qquad \omega \in \Omega, \\ p(x,\omega) &= 0, \qquad x \in \partial G, \qquad \omega \in \Omega. \end{aligned}$$

Here $u(x,\omega) \sim \mathcal{N}(0, \mathcal{C}(x,y))$ is a (Matérn) Gaussian random field and we are interested in computing $\mathbb{E}[P]$, where $P(\omega) = ||p||^2_{L^2(G)}(\omega)$.

Simple approach: standard Monte Carlo (MC) method,

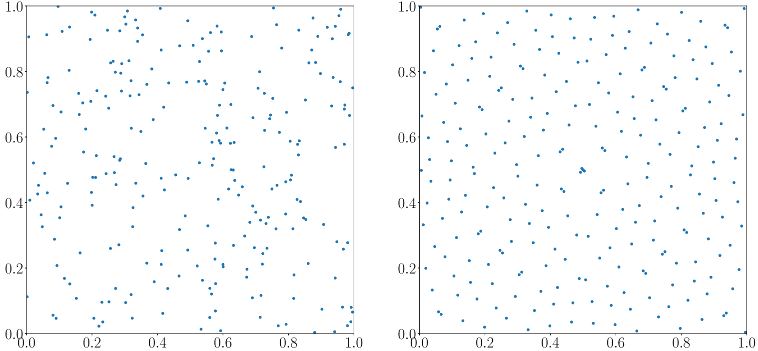
$$\mathbb{E}[P] \approx \frac{1}{N} \sum_{n=1}^{N} P(\omega^n).$$

Convergence rate $O(N^{-1/2})$. $O(\varepsilon^{-q})$ cost per sample $\Rightarrow O(\varepsilon^{-2-q})$ complexity for ε tol.

From standard Monte Carlo to Quasi Monte Carlo

Approximate $\mathbb{E}[P]$ with an s-dimensional integral over $[0,1]^s$,

$$\mathbb{E}[P] \approx \int_{[0,1]^s} Y(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \approx \frac{1}{N}\sum_{n=1}^{N} Y(\boldsymbol{x}_n),$$



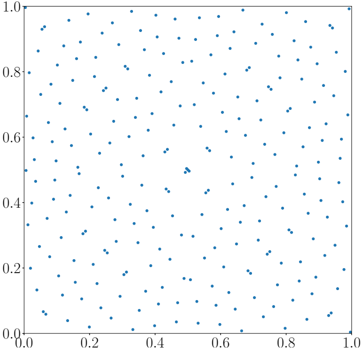
Pseudo-random points. Low-discrepancy point sequence (Sobol').

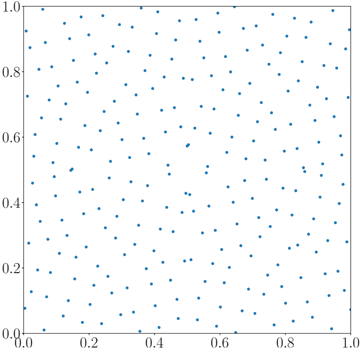
QMC convergence rate up to $O(N^{-1}) \Rightarrow$ up to $O(\varepsilon^{-1-q})$ complexity.

Randomised Quasi Monte Carlo

$$\mathbb{E}[P] \approx \int_{[0,1]^s} Y(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \approx \frac{1}{M}\sum_{m=1}^{M}\left(\frac{1}{N}\sum_{n=1}^{N} Y(\hat{\boldsymbol{x}}_{n,m})\right),$$

where $\{\hat{\boldsymbol{x}}_{n,m}\}_{n=1}^N$ is the m-th randomisation of a low-discrepancy point set $\{\boldsymbol{x}_n\}_{n=1}^N$.





First $N = 256$ Sobol' points before (left) and after (right) scrambling.

QMC convergence rate up to $O(N^{-1}) \Rightarrow$ up to $O(\varepsilon^{-1-q})$ complexity.

Multilevel Monte Carlo [Giles 2008]

Solve for P using a hierarchy of $L+1$ (possibly non-nested) meshes to obtain the approximations P_ℓ of different accuracy for $\ell = 0, \ldots, L$, then

$$\mathbb{E}[P] \approx \mathbb{E}[P_L] = \mathbb{E}[P_0] + \sum_{\ell=0}^{L} \mathbb{E}[P_\ell - P_{\ell-1}].$$

Multilevel Monte Carlo [Giles 2008]

Solve for P using a hierarchy of $L+1$ (possibly non-nested) meshes to obtain the approximations P_ℓ of different accuracy for $\ell = 0, \ldots, L$, then

$$\mathbb{E}[P] \approx \mathbb{E}[P_L] = \mathbb{E}[P_0] + \sum_{\ell=0}^{L} \mathbb{E}[P_\ell - P_{\ell-1}].$$

Apply standard MC to each term on the RHS:

$$\mathbb{E}[P] \approx \frac{1}{N_0}\sum_{n=1}^{N_0} P_0(\omega_0^n) + \sum_{\ell=0}^{L} \frac{1}{N_\ell}\sum_{n=1}^{N_\ell}[P_\ell(\omega_\ell^n) - P_{\ell-1}(\omega_\ell^n)].$$

Under suitable conditions $\implies$ optimal N_ℓ known and $O(\varepsilon^{-2})$ complexity, $O(\varepsilon^{-q})$ better than MC.

Multilevel Quasi Monte Carlo [Giles and Waterhouse 2009]

We now approximate each expectation in the telescoping sum with randomised quasi Monte Carlo (QMC). Rewrite $\mathbb{E}[P_\ell - P_{\ell-1}]$ as the s_ℓ-dimensional integral over $[0,1]^{s_\ell}$,

$$\mathbb{E}[P_\ell - P_{\ell-1}] = \int_{[0,1]^{s_\ell}} Y_\ell(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \approx \frac{1}{M}\sum_{m=1}^{M}\left(\frac{1}{N_\ell}\sum_{n=1}^{N_\ell} Y_\ell(\hat{\boldsymbol{x}}_{n,m}^{\ell})\right) = \hat{Y}_\ell,$$

where $\{\hat{\boldsymbol{x}}_{n,m}^{\ell}\}_{n=1}^{N_\ell}$ is the m-th randomisation of a low-discrepancy point set $\{\boldsymbol{x}_n^{\ell}\}_{n=1}^{N_\ell}$. We use $M = 32$ and random digital shifted Sobol' sequences[1].

QMC integration convergence[2] up to $O(N_\ell^{-1}) \Longrightarrow$ complexity up to $O(\varepsilon^{-1})$.

[1] Python-wrapped Intel® MKL library Sobol' sequence implementation augmented with Joe and Kuo's primitive polynomials and direction numbers ($s_{\text{max}} = 21201$) [Joe and Kuo 2008].

[2] Lots of recent work with randomly shifted lattice rules [Kuo, Schwab, Sloan 2015, Kuo, Scheichl, Schwab, Sloan, Ullmann 2017, Herrmann and Schwab 2017,...].

Gaussian field sampling

Sampling the Gaussian field $u(x,\omega) \sim \mathcal{N}(0,\mathcal{C})$ is hard!

Lots of recent developments related to ML(Q)MC: [Kuo et al. 2015, Kuo et al. 2018, Graham et al. 2018, Drzisga et al. 2017, Herrmann and Schwab 2017, Osborn et al. 2018, C. et al. 2018, ...]

Gaussian field sampling

Sampling the Gaussian field $u(x, \omega) \sim \mathcal{N}(0, \mathcal{C})$ is hard!

Naïve approach

- Discretise u and compute a Cholesky factorization of the covariance matrix.

Better approaches

- Karhunen-Loève.
- FFT + circulant embeddings. [Dietrich and Newsam 1997]
- **PDE approach** [Lindgren et al. 2009] (see next).

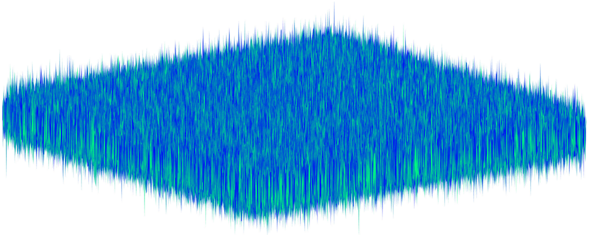
Lots of recent developments related to ML(Q)MC: [Kuo et al. 2015, Kuo et al. 2018, Graham et al. 2018, Drzisga et al. 2017, Herrmann and Schwab 2017, Osborn et al. 2018, C. et al. 2018, ...]

PDE approach to Matérn field sampling [Lindgren et al. 2009]

If the covariance of u is of the Matérn class with smoothness parameter ν, then u approximately satisfies the linear elliptic SPDE,

$$Lu(x,\omega) = \left(\mathcal{I} - \kappa^{-2}\Delta\right)^{k} u(x,\omega) = \eta \dot{W}, \quad x \in D \subset \mathbb{R}^d, \quad \nu = 2k - d/2 > 0,$$

where $\dot{W}$ is spatial white noise, $G \subset D$, $\eta \in \mathbb{R}$ and for today $k \in \mathbb{N}$, $\eta = 1$.



Spatial white noise in $[0,1]^2$.

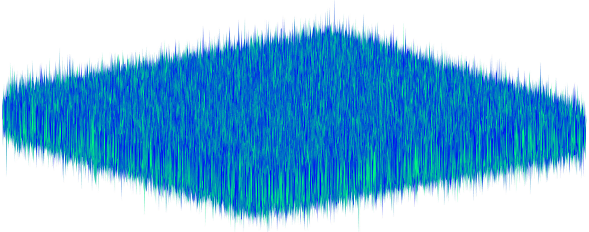
Refs: [Abrahamsen 1997, Scheuerer 2010, Lindgren et al. 2009, Bolin et al. 2017, Khristenko et al. 2018]

PDE approach to Matérn field sampling [Lindgren et al. 2009]

Definition (Spatial White Noise $\dot{W}$ (Hida et al. 1993))

For any $\phi \in L^2(D)$, define $\langle \dot{W}, \phi \rangle := \int_D \phi \, \mathrm{d}\dot{W}$. For any $\phi_i, \ \phi_j \in L^2(D)$, $b_i = \langle \dot{W}, \phi_i \rangle$, $b_j = \langle \dot{W}, \phi_j \rangle$ are zero-mean Gaussian random variables, with,

$$\mathbb{E}[b_i b_j] = \int_D \phi_i \phi_j \ \mathrm{dx} =: M_{ij}, \quad \mathbf{b} \sim \mathcal{N}(0, M). \tag{1}$$



Spatial white noise in $[0,1]^2$.

Refs: [Abrahamsen 1997, Scheuerer 2010, Lindgren et al. 2009, Bolin et al. 2017, Khristenko et al. 2018]

Overview

Efficient white noise sampling for MLMC [C. et al. 2018]

Let $\{V_\ell\}_{\ell=0}^{L}$ be a hierarchy of (possibly non-nested) FEM approximation subspaces with $V_\ell = \text{span}(\{\phi_i^\ell\}_{i=1}^{n_{\text{dofs}}^\ell}) \subseteq H_0^1(D)$. On each MLMC level, we need to solve for u_ℓ and $u_{\ell-1}$,

$$Lu_\ell = \dot{W}, \quad \text{and} \quad Lu_{\ell-1} = \dot{W}, \quad \text{if } \ell > 0,$$

where we use the same white noise sample on both levels to enforce the MLMC coupling.

Efficient white noise sampling for MLMC [C. et al. 2018]

Let $\{V_\ell\}_{\ell=0}^{L}$ be a hierarchy of (possibly non-nested) FEM approximation subspaces with $V_\ell = \text{span}(\{\phi_i^\ell\}_{i=1}^{n_{\text{dofs}}^\ell}) \subseteq H_0^1(D)$. On each MLMC level, we need to solve for u_ℓ and $u_{\ell-1}$,

$$Lu_\ell = \dot{W}, \quad \text{and} \quad Lu_{\ell-1} = \dot{W}, \quad \text{if } \ell > 0,$$

where we use the same white noise sample on both levels to enforce the MLMC coupling.

After discretisation, the above yields the linear system,

$$\left[\begin{array}{c|c} A^\ell & 0 \\ \hline 0 & A^{\ell-1} \end{array}\right] \left[\begin{array}{c} \mathbf{u}^\ell \\ \hline \mathbf{u}^{\ell-1} \end{array}\right] = \left[\begin{array}{c} \mathbf{b}^\ell \\ \hline \mathbf{b}^{\ell-1} \end{array}\right] = \mathbf{b}, \quad \text{where} \quad \boldsymbol{b}_i^\ell = \langle \dot{W}, \phi_i^\ell \rangle.$$

Hence $\boldsymbol{b} \sim \mathcal{N}(0, M)$, where M is the mass matrix over $V_\ell + V_{\ell-1}$, (set $V_{-1} = \emptyset$), i.e.

$$M = \left[\begin{array}{c|c} M^\ell & M^{\ell,\ell-1} \\ \hline (M^{\ell,\ell-1})^T & M^{\ell-1} \end{array}\right], \quad M_{ij}^{\ell,\ell-1} = \int \phi_i^\ell \phi_j^{\ell-1} \, \text{dx}, \quad \text{if } \ell > 0.$$

NOTE: we do not require the FEM approximation subspaces to be nested!

How to sample b?

Sampling b is hard!

How to sample b?

Sampling b is hard!

Naïve approach

- Factorise the covariance M using Cholesky (**cubic complexity!**)
- Works well if M **diagonal**. Previous work under this assumption [Lindgren et al. 2009, Osborn et al. 2017, Drzisga, et al. 2017, Du and Zhang 2002].

How to sample b?

Sampling b is hard!

Naïve approach

- Factorise the covariance M using Cholesky (**cubic complexity!**)
- Works well if M **diagonal**. Previous work under this assumption [Lindgren et al. 2009, Osborn et al. 2017, Drzisga, et al. 2017, Du and Zhang 2002].

Our Work [C. et al. 2018]

- We do not require M to be diagonal.
- We can sample **b** with **linear complexity**.

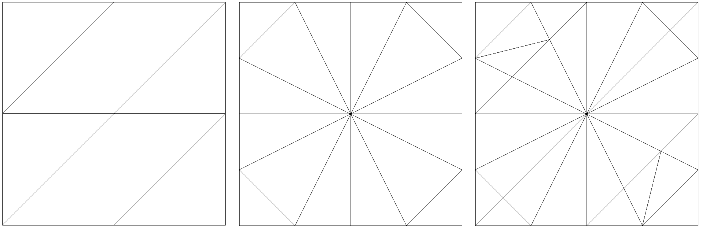
Sampling b in **linear complexity** (Sketch) [C. et al. 2018]

Two ingredients: **supermeshing** [Farrell 2009] and **local factorisation** [Wathen 1987].

Sampling b in **linear complexity** (Sketch) [C. et al. 2018]

Two ingredients: **supermeshing** [Farrell 2009] and **local factorisation** [Wathen 1987].

1. **Supermeshing**: construct a FEM subspace S_ℓ such that V_ℓ and $V_{\ell-1}$ are both nested within S_ℓ. This requires a supermesh construction:

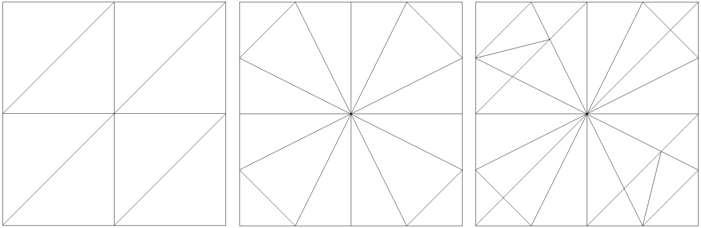


Sample $\boldsymbol{b}_S^\ell \sim \mathcal{N}(0, M_S^\ell)$ where M_S^ℓ is the mass matrix over S_ℓ and get $\boldsymbol{b}_\ell$ and $\boldsymbol{b}_{\ell-1}$ by transferring $\boldsymbol{b}_S^\ell$ onto V_ℓ and $V_{\ell-1}$ using **nested interpolation**.

Sampling b in **linear complexity** (Sketch) [C. et al. 2018]

Two ingredients: **supermeshing** [Farrell 2009] and **local factorisation** [Wathen 1987].

1. **Supermeshing**: construct a FEM subspace S_ℓ such that V_ℓ and $V_{\ell-1}$ are both nested within S_ℓ. This requires a supermesh construction:



   Sample $\boldsymbol{b}_S^\ell \sim \mathcal{N}(0, M_S^\ell)$ where M_S^ℓ is the mass matrix over S_ℓ and get $\boldsymbol{b}_\ell$ and $\boldsymbol{b}_{\ell-1}$ by transferring $\boldsymbol{b}_S^\ell$ onto V_ℓ and $V_{\ell-1}$ using **nested interpolation**.
2. **Local factorisation**: spatially **disjoint pieces of white noise are independent**, we can sample small independent **local white noise vectors** $\boldsymbol{b}_e^\ell \sim \mathcal{N}(0, M_e^\ell)$ on each supermesh cell e and assemble the contributions together to obtain $\boldsymbol{b}_S^\ell$.

RESULT: Linear cost complexity and trivially parallelisable!

Overview

Related work

Herrmann and Schwab (2017 pre-print)

- Same application problem.
- Some aspects of the theory are very general.
- Algorithm is detailed only for 1D, nested, structured grids.
- Truncated QMC.

Our work

- Focussed on algorithm development for 2D and 3D, non-nested, unstructured grids.
- Non-truncated hybrid QMC/MC.

Extending work on MLMC to MLQMC: the challenge

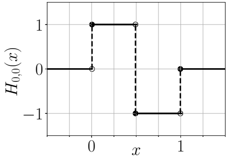
- Low-discrepancy sequences are extremely uniform in the first few dimensions and in low-dimensional projections, but less so across the whole hypercube. Therefore **QMC works best when the integrand has low effective dimension** (adapted from [Joe and Kuo 2008]).

Extending work on MLMC to MLQMC: the challenge

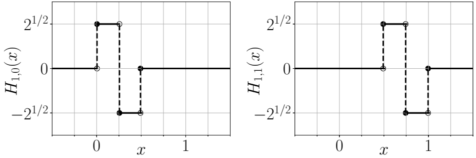
- Low-discrepancy sequences are extremely uniform in the first few dimensions and in low-dimensional projections, but less so across the whole hypercube. Therefore **QMC works best when the integrand has low effective dimension** (adapted from [Joe and Kuo 2008]).
- For good QMC convergence we need to **order the dimensions** in our QMC integrands in order of decaying importance so that the largest error components are on the first dimensions.

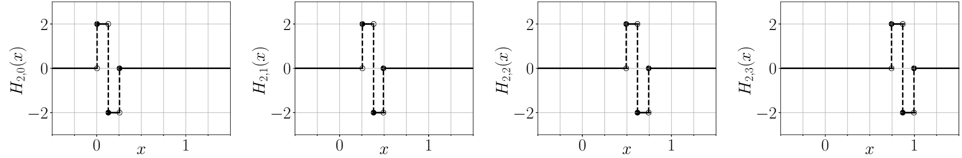
Haar wavelet expansion of white noise

From now on, $D = [0,1]^d$. We expand $\dot{W}$ into a Haar wavelet series. Let $\boldsymbol{l} \in (\{-1\} \cap \mathbb{N})^d$, $\boldsymbol{n} \in \mathbb{N}^d$, $|\boldsymbol{l}| = \max_i(l_i)$, $x^+ = \max(x,0)$, $H_{\boldsymbol{-1},\boldsymbol{0}}(\boldsymbol{x}) = \mathbb{1}_D(\boldsymbol{x})$,

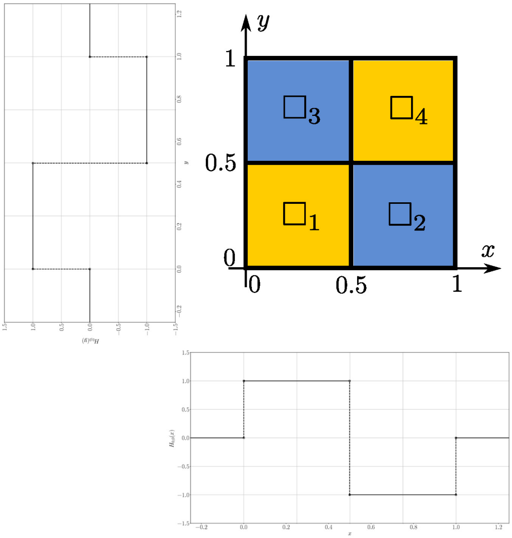


$$\dot{W} = \sum_{|\boldsymbol{l}|=-1}^{|\boldsymbol{l}|=\infty} \sum_{\boldsymbol{n}=\boldsymbol{0}}^{(2^{\boldsymbol{l}}-\boldsymbol{1})^+} z_{\boldsymbol{l},\boldsymbol{n}}(\omega) H_{\boldsymbol{l},\boldsymbol{n}}(x), \quad z_{\boldsymbol{l},\boldsymbol{n}} \sim \mathcal{N}(0,1) \ \text{i.i.d.}$$



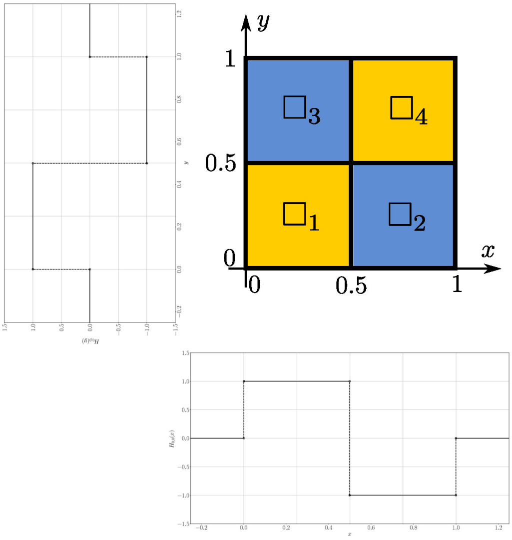


Multilevel Quasi Monte Carlo white noise sampling



The $H_{\mathbf{0},\mathbf{0}} = H_{0,0}(x)H_{0,0}(y)$ *2D Haar wavelet and the level* $\mathscr{L} = |\boldsymbol{l}| = 1$ *Haar mesh.*

Multilevel Quasi Monte Carlo white noise sampling



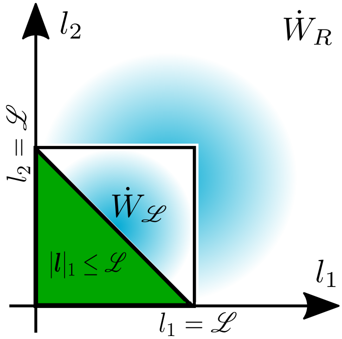
The $H_{\mathbf{0},\mathbf{0}} = H_{0,0}(x)H_{0,0}(y)$ 2D Haar wavelet and the level $\mathscr{L} = |\boldsymbol{l}| = 1$ Haar mesh.

Truncate the series of $\dot{W}$ at Haar level $\mathscr{L}$.
Let $\dot{W}_{\mathscr{L}}$:= truncation, $\dot{W}_R$:= remainder.

$$\dot{W} = \dot{W}_{\mathscr{L}} + \dot{W}_R.$$

1) Order all the $N_{\mathscr{L}} = 2^{d(\mathscr{L}+1)}$ coefficients in $\dot{W}_{\mathscr{L}}$ according to $|\boldsymbol{l}|_1$.

2) Use hybrid MC/QMC sampling (figure).

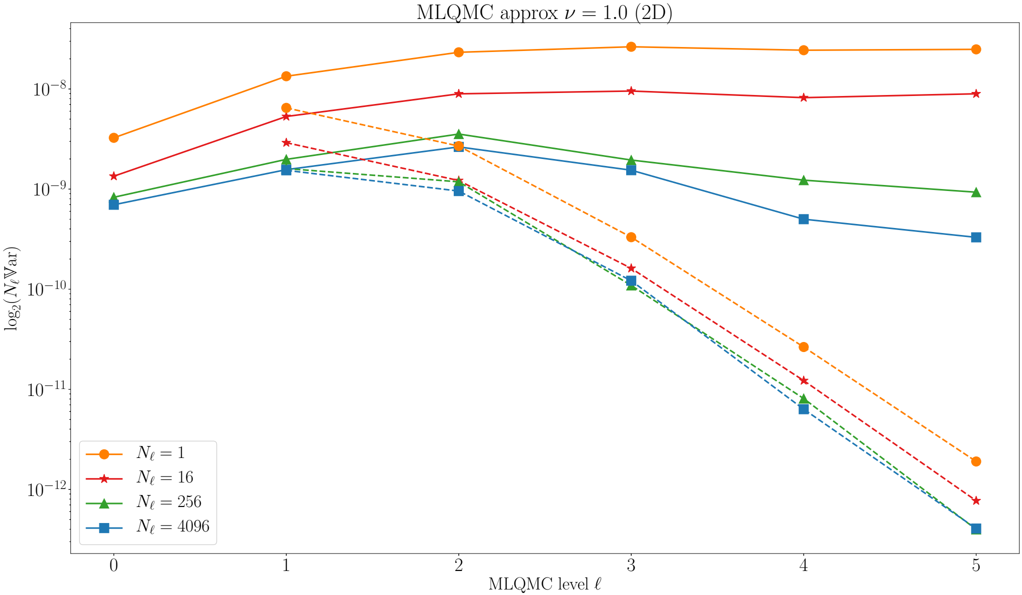


Multilevel Quasi Monte Carlo white noise sampling

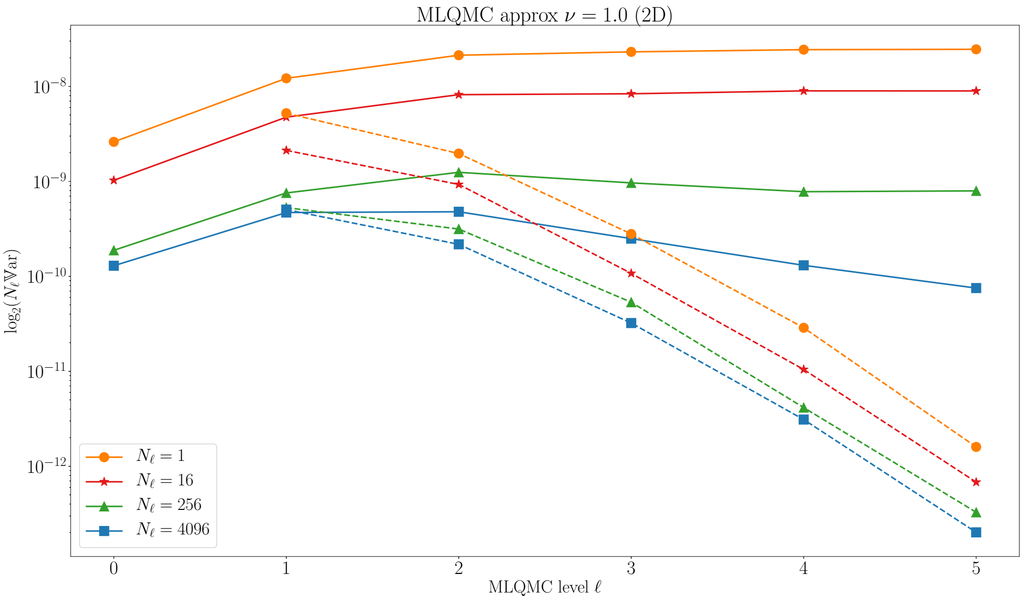
- $\dot{W}_R$ is sampled with pseudo-random points so no ordering needed. The covariance of $\dot{W}_R$ is known and a similar technique as in MLMC can be used for the sampling.

- $\dot{W}_R$ is sampled with pseudo-random points so no ordering needed. The covariance of $\dot{W}_R$ is known and a similar technique as in MLMC can be used for the sampling.
- This time a 3-way supermesh is required. Let $N_S = \#$ supermesh cells. The Haar mesh is "nice" so we expect $N_S \leq C(d)$ ($\#$ cells of the finest parent mesh). Numerical experiments suggest $C(d) < 3$ in 1D and 2D and $C(d) \approx O(50)$ in 3D.
- Overall Matérn field sampling cost is $O(N_{\mathscr{L}} \log(N_{\mathscr{L}}) + N_S)$, where the log term can be dropped if using locally supported Haar wavelets or if $d = 1$. $\mathscr{L} \leq \mathscr{L}_{\max}$.

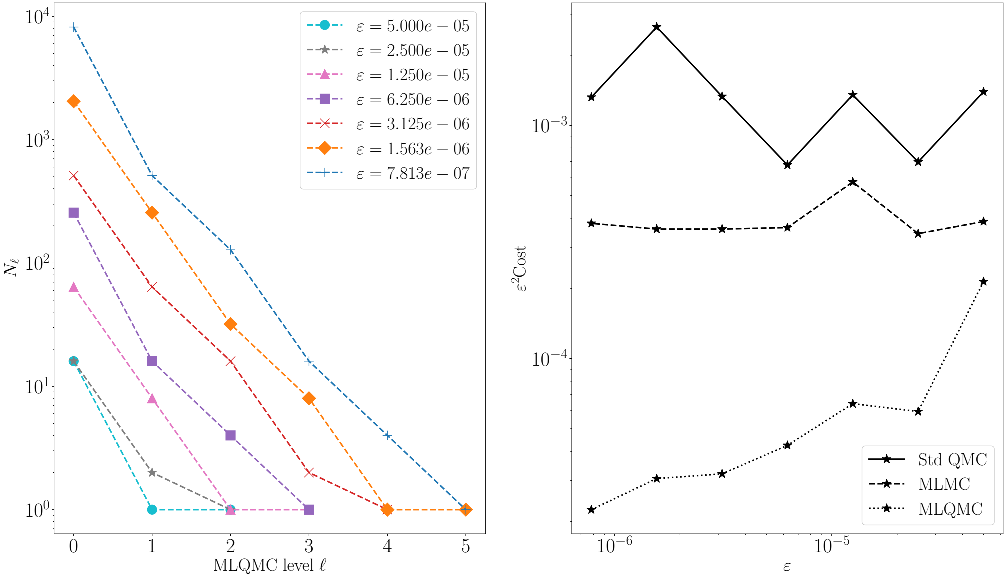
MC vs QMC vs MLQMC (2D, $\nu = 1$, $M = 256$)



MC vs QMC vs MLQMC (2D, $\nu = 1$, $M = 256$)



MLQMC convergence and cost (2D, $\nu = 1$)



Overview

Conclusions

Outlook

- **Ordering** the **variables** is **needed** for good QMC convergence.
- Asymptotic QMC convergence rate is still $O(N^{-1/2})$. However, large gains to be found in the pre-asymptotic regime while keeping the QMC dimension contained.
- We can couple white noise between different FEM approximation subspaces. A supermesh construction is not needed in the nested case. 2-way supermesh needed for MLMC and 3-way supermesh needed for MLQMC.
- The overall order of **complexity is linear** in the number of elements of the supermesh and it can be **trivially parallelised**. Standard techniques usually have cubic complexity.
- Software used: **FEniCS**, libsupermesh, Intel® MKL. Fast FEniCS-based Matérn field and white noise sampling **software to appear on Bitbucket soon**!

JUQ paper (MLMC only) and slides: `https://croci.github.io`

MLQMC paper and PhD thesis to be submitted soon, please read both!

Selected references - Thank you for listening!

JUQ paper [1] and slides: `https://croci.github.io`

[1] M. Croci, M. B. Giles, M. E. Rognes, and P. E. Farrell. Efficient white noise sampling and coupling for multilevel Monte Carlo with non-nested meshes. *SIAM/ASA Journal on Uncertainty Quantification*, 6(4):1630–1655, 2018. doi: 10.1137/18M1175239.

[2] P. E. Farrell, M. D. Piggott, C. C. Pain, G. J. Gorman, and C. R. Wilson. Conservative interpolation between unstructured meshes via supermesh construction. *Computer Methods in Applied Mechanics and Engineering*, 198:2632–2642, 2009.

[3] M. B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008. doi: 10.1287/opre.1070.0496.

[4] M. B. Giles and B. J. Waterhouse. Multilevel quasi-Monte Carlo path simulation. *Advanced Financial Modelling, Radon Series on Computational and Applied Mathematics*, (8):165–181, 2009.

[5] L. Herrmann and C. Schwab. Multilevel quasi-Monte Carlo integration with product weights for elliptic PDEs with lognormal coefficients Multilevel quasi-Monte Carlo integration with product weights for elliptic PDEs with lognormal coefficients. Technical report, ETH Zurich, Zurich, 2017.

[6] F. Y. Kuo, R. Scheichl, C. Schwab, I. H. Sloan, and E. Ullmann. Multilevel Quasi-Monte Carlo Methods for Lognormal Diffusion Problems. *Mathematics of Computation*, 86(308):2827–2860, 2017.

[7] F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2009.

[8] A. J. Wathen. Realistic eigenvalue bounds for the Galerkin mass matrix. *IMA Journal of Numerical Analysis*, 7(August 1985):449–457, 1987. doi: 10.1093/imanum/7.4.449.