The Maxwell Institute Graduate School in Mathematical Modelling, Analysis and Computation, Edinburgh

Multilevel Monte Carlo Methods with Smoothing

Uncertainty Quantification for High-Dimensional Problems Workshop
Amsterdam, The Netherlands

Anastasia Istratuca, Aretha Teckentrup

arXiv: 2306.13493
(to appear in IMA Journal of Numerical Analysis)

November 11, 2024

Outline

Motivation

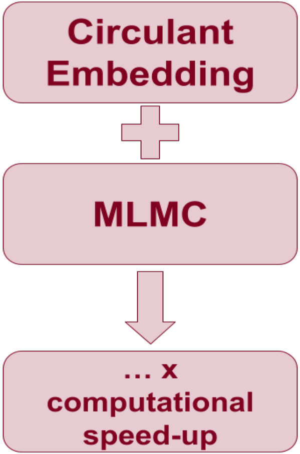
Circulant Embedding

Multilevel Monte Carlo Methods

Smoothing

Numerical Experiments

Conclusions and Outlook



Motivation

Application - Modelling and simulation of groundwater flow

$$-\nabla \cdot (k(\mathbf{x},\omega)\nabla u(\mathbf{x},\omega)) = f(\mathbf{x}), \quad \mathbf{x} \in D, \quad D = (0,1) \times (0,1)$$

$$u|_{x_1=0} = 1, \qquad u|_{x_1=1} = 0,$$

$$\left.\frac{\partial u}{\partial \mathbf{n}}\right|_{x_2=0} = 0, \quad \left.\frac{\partial u}{\partial \mathbf{n}}\right|_{x_2=1} = 0.$$

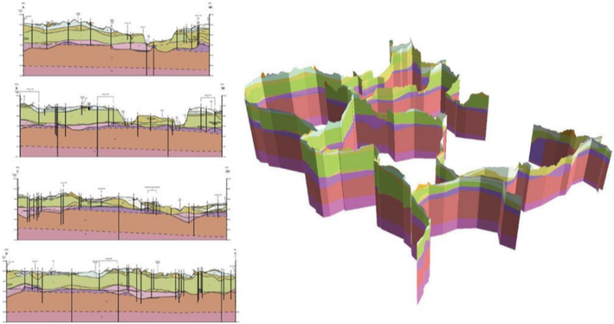


Figure: Cross-section of subsurface example (DeMeritt, 2012).

Suppose we are interested in finding $\mathbb{E}[Q]$, e.g. $Q = u(\mathbf{x}^*, \cdot)$. Then:

$$-\nabla \cdot (k(\mathbf{x}, \omega^{(i)}) \nabla u(\mathbf{x}, \omega^{(i)})) = f(\mathbf{x})$$

for **one sample** $k\left(\mathbf{x}, \omega^{(i)}\right)$.

Suppose we are interested in finding $\mathbb{E}[Q]$, e.g. $Q = u(\mathbf{x}^*, \cdot)$. Then:

$$-\nabla \cdot (k(\mathbf{x}, \omega^{(i)}) \nabla u(\mathbf{x}, \omega^{(i)})) = f(\mathbf{x})$$

$$\downarrow$$

$$u_h(\mathbf{x}, \omega^{(i)}) \approx u(\mathbf{x}, \omega^{(i)})$$

for **one sample** $k\left(\mathbf{x}, \omega^{(i)}\right)$.

Suppose we are interested in finding $\mathbb{E}[Q]$, e.g. $Q = u(\mathbf{x}^*, \cdot)$. Then:

$$-\nabla \cdot (k(\mathbf{x}, \omega^{(i)}) \nabla u(\mathbf{x}, \omega^{(i)})) = f(\mathbf{x})$$

$$\downarrow$$

$$u_h(\mathbf{x}, \omega^{(i)}) \approx u(\mathbf{x}, \omega^{(i)})$$

$$\downarrow$$

$$Q_h^{(i)} \approx Q^{(i)}$$

for **one sample** $k\left(\mathbf{x}, \omega^{(i)}\right)$.

Suppose $k(\mathbf{x}, \cdot)$ is a log-normal random field, so that:

$$k(\mathbf{x}, \omega) = \exp(Z(\mathbf{x}, \omega)),$$

where $Z(\mathbf{x}, \cdot)$ is a Gaussian random field with:

$$\begin{aligned} \mathbb{E}[Z(\mathbf{x}, \cdot)] &\equiv 0 \\ \mathbb{E}[Z(\mathbf{x}, \cdot)Z(\mathbf{y}, \cdot)] &= r(\mathbf{x}, \mathbf{y}) = C(\mathbf{x} - \mathbf{y}). \end{aligned}$$

Suppose $k(\mathbf{x}, \cdot)$ is a log-normal random field, so that:

$$k(\mathbf{x}, \omega) = \exp(Z(\mathbf{x}, \omega)),$$

where $Z(\mathbf{x}, \cdot)$ is a Gaussian random field with:

$$\mathbb{E}[Z(\mathbf{x}, \cdot)] \equiv 0$$
$$\mathbb{E}[Z(\mathbf{x}, \cdot)Z(\mathbf{y}, \cdot)] = r(\mathbf{x}, \mathbf{y}) = C(\mathbf{x} - \mathbf{y}).$$

The covariance function C selected for this application is given by (Hoeksema and Kitanidis, 1985):

$$C(\mathbf{r}) := \sigma^2 \exp\left(-\frac{\|\mathbf{r}\|_1}{\lambda}\right), \quad \sigma, \lambda > 0, \lambda < \mathrm{diam}(D).$$

We also use the Matérn covariance, given by:

$$C(\mathbf{r}) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}\|\mathbf{r}\|_2}{\lambda}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}\|\mathbf{r}\|_2}{\lambda}\right), \quad \sigma, \lambda, \nu > 0.$$

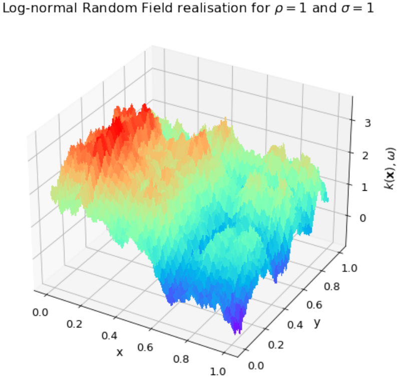


Figure: $\lambda = 1, \sigma = 1$

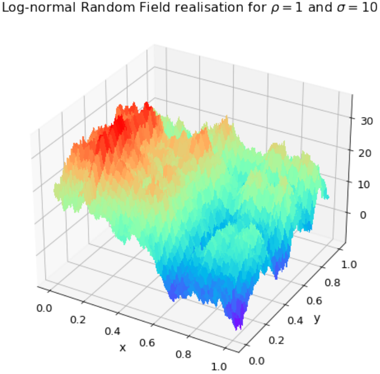


Figure: $\lambda = 1, \sigma = 10$

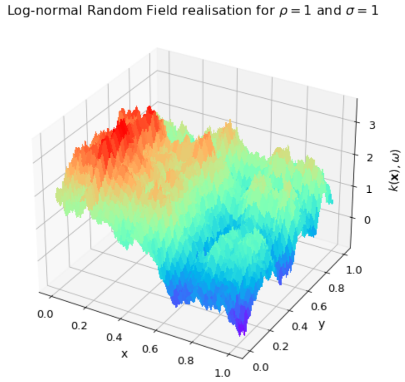


Figure: $\lambda = 1, \sigma = 1$

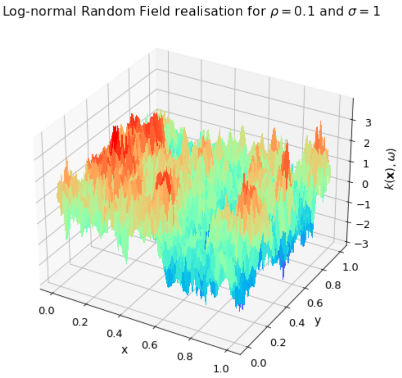


Figure: $\lambda = 0.1, \sigma = 1$

How do we obtain samples $k(\mathbf{x}, \omega)$ on mesh $\mathcal{T}$?

Issue: many classical factorisation methods, such as Cholesky, have cubic cost in the number of mesh points!

How do we obtain samples $k(\mathbf{x}, \omega)$ on mesh $\mathcal{T}$?

For any factorisation of the covariance matrix R of $Z_{\mathcal{T}}(\mathbf{x}, \cdot)$:

$$R = \Theta\Theta^T,$$

and any vector ξ such that:

$$\xi \sim \mathcal{N}(\mathbf{0}, I),$$

we can take:

$$\mathbf{Z} := \Theta\xi,$$

to obtain $\mathbf{Z} \sim Z_{\mathcal{T}}(\mathbf{x}, \cdot)$.

Issue: many classical factorisation methods, such as Cholesky, have cubic cost in the number of mesh points!

Circulant Embedding
Overview [Dietrich and Newsam (1993)]

How do we obtain samples $k(\mathbf{x}, \omega)$ on mesh $\mathcal{T}$?

How do we obtain samples $k(\mathbf{x}, \omega)$ on mesh $\mathcal{T}$?

$\mathcal{T}$

where:

- $\mathcal{T}$ - uniform two-dimensional discretisation mesh;
- R - covariance matrix;
- S - circulant embedding matrix;
- $G = \Re(F) + \Im(F)$, F - two-dimensional Fourier matrix;
- $\Lambda = \sqrt{4 m_1 m_2} F \mathbf{s}$ - diagonal matrix of eigenvalues of S with $\mathbf{s}$ - first column of S;
- $\mathbf{Z}$ - sample from the Gaussian field.

How do we obtain samples $k(\mathbf{x}, \omega)$ on mesh $\mathcal{T}$?

$$\mathcal{T} \xrightarrow{C(\mathbf{x}_i - \mathbf{y}_j)} R$$

where:

- $\mathcal{T}$ - uniform two-dimensional discretisation mesh;
- R - covariance matrix;
- S - circulant embedding matrix;
- $G = \Re(F) + \Im(F)$, F - two-dimensional Fourier matrix;
- $\Lambda = \sqrt{4 m_1 m_2} F \mathbf{s}$ - diagonal matrix of eigenvalues of S with $\mathbf{s}$ - first column of S;
- $\mathbf{Z}$ - sample from the Gaussian field.

How do we obtain samples $k(\mathbf{x}, \omega)$ on mesh $\mathcal{T}$?

$$\mathcal{T} \xrightarrow{C(\mathbf{x}_i - \mathbf{y}_j)} R \xrightarrow{\text{embedding}} S$$

where:

- $\mathcal{T}$ - uniform two-dimensional discretisation mesh;
- R - covariance matrix;
- S - circulant embedding matrix;
- $G = \Re(F) + \Im(F)$, F - two-dimensional Fourier matrix;
- $\Lambda = \sqrt{4 m_1 m_2} F \mathbf{s}$ - diagonal matrix of eigenvalues of S with $\mathbf{s}$ - first column of S;
- $\mathbf{Z}$ - sample from the Gaussian field.

How do we obtain samples $k(\mathbf{x}, \omega)$ on mesh $\mathcal{T}$?

$$\mathcal{T} \xrightarrow{C(\mathbf{x}_i - \mathbf{y}_j)} R \xrightarrow{\text{embedding}} S \xrightarrow[\text{transform}]{\text{Fourier}} S = G\Lambda G^T$$

where:

- $\mathcal{T}$ - uniform two-dimensional discretisation mesh;
- R - covariance matrix;
- S - circulant embedding matrix;
- $G = \Re(F) + \Im(F)$, F - two-dimensional Fourier matrix;
- $\Lambda = \sqrt{4m_1 m_2} F\mathbf{s}$ - diagonal matrix of eigenvalues of S with $\mathbf{s}$ - first column of S;
- $\mathbf{Z}$ - sample from the Gaussian field.

Circulant Embedding
Overview [Dietrich and Newsam (1993)]

How do we obtain samples $k(\mathbf{x}, \omega)$ on mesh $\mathcal{T}$?

$$\mathcal{T} \xrightarrow{C(\mathbf{x}_i - \mathbf{y}_j)} R \xrightarrow{\text{embedding}} S \xrightarrow[\text{transform}]{\text{Fourier}} S = G \Lambda G^T \xrightarrow{\xi \sim \mathcal{N}(\mathbf{0}, I)} \mathbf{Z} = G \Lambda^{1/2} \xi,$$

where:

- $\mathcal{T}$ - uniform two-dimensional discretisation mesh;
- R - covariance matrix;
- S - circulant embedding matrix;
- $G = \Re(F) + \Im(F)$, F - two-dimensional Fourier matrix;
- $\Lambda = \sqrt{4 m_1 m_2} F \mathbf{s}$ - diagonal matrix of eigenvalues of S with $\mathbf{s}$ - first column of S;
- $\mathbf{Z}$ - sample from the Gaussian field.

Circulant Embedding

How? (1D)

$$R = \begin{bmatrix} C_0 & C_1 & \cdots & C_m \\ C_1 & C_0 & \cdots & C_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ C_m & C_{m-1} & \cdots & C_0 \end{bmatrix}, \quad C_i = C\left(\frac{i}{m}\right)$$

Circulant Embedding

How? (1D)

$$R = \begin{bmatrix} C_0 & C_1 & \dots & C_m \\ C_1 & C_0 & \dots & C_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ C_m & C_{m-1} & \dots & C_0 \end{bmatrix}, \quad C_i = C\left(\frac{i}{m}\right)$$

$$\downarrow$$

$$S = \begin{bmatrix} C_0 & C_1 & \dots & C_m & & \\ C_1 & C_0 & \dots & C_{m-1} & & \\ \vdots & \vdots & \ddots & \vdots & & \\ C_m & C_{m-1} & \dots & C_0 & & \\ & & & & & \\ & & & & & \end{bmatrix}$$

Circulant Embedding

How? (1D)

$$R = \begin{bmatrix} C_0 & C_1 & \cdots & C_m \\ C_1 & C_0 & \cdots & C_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ C_m & C_{m-1} & \cdots & C_0 \end{bmatrix}, \quad C_i = C\left(\frac{i}{m}\right)$$

$$\downarrow$$

$$S = \begin{bmatrix} C_0 & C_1 & \cdots & C_m & C_{m-1} & \cdots & C_1 \\ C_1 & C_0 & \cdots & C_{m-1} & C_m & \cdots & C_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ C_m & C_{m-1} & \cdots & C_0 & C_1 & \cdots & C_{m-1} \\ & & & & & & \\ & & & & & & \end{bmatrix}$$

$$R = \begin{bmatrix} C_0 & C_1 & \cdots & C_m \\ C_1 & C_0 & \cdots & C_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ C_m & C_{m-1} & \cdots & C_0 \end{bmatrix}, \quad C_i = C\left(\frac{i}{m}\right)$$

$$\downarrow$$

$$S = \begin{bmatrix} C_0 & C_1 & \cdots & C_m & C_{m-1} & \cdots & C_1 \\ C_1 & C_0 & \cdots & C_{m-1} & C_m & \cdots & C_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ C_m & C_{m-1} & \cdots & C_0 & C_1 & \cdots & C_{m-1} \\ C_{m-1} & C_{m-2} & \cdots & C_1 & & & \\ \vdots & \vdots & \ddots & \vdots & & & \\ C_1 & C_2 & \cdots & C_{m-1} & & & \end{bmatrix}$$

Circulant Embedding

How? (1D)

$$R = \begin{bmatrix} C_0 & C_1 & \dots & C_m \\ C_1 & C_0 & \dots & C_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ C_m & C_{m-1} & \dots & C_0 \end{bmatrix}, \quad C_i = C\left(\frac{i}{m}\right)$$

$$\downarrow$$

$$S = \begin{bmatrix} C_0 & C_1 & \dots & C_m & C_{m-1} & \dots & C_1 \\ C_1 & C_0 & \dots & C_{m-1} & C_m & \dots & C_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ C_m & C_{m-1} & \dots & C_0 & C_1 & \dots & C_{m-1} \\ C_{m-1} & C_{m-2} & \dots & C_1 & C_0 & \dots & C_m \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ C_1 & C_2 & \dots & C_{m-1} & C_m & \dots & C_0 \end{bmatrix}$$

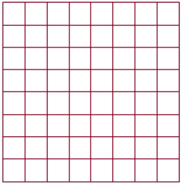
Multilevel Monte Carlo Methods
Standard Monte Carlo

For N i.i.d. samples of $k(\mathbf{x}, \cdot)$:

$$\mathbb{E}[Q_h] \approx \widehat{Q}^{\mathrm{MC}}_{h,N} := \frac{1}{N}\sum_{i=1}^{N} Q_h^{(i)}.$$

Problem: N is typically very large and h is very small:

$$e\left(\widehat{Q}^{\mathrm{MC}}_{h,N}\right)^2 := \mathbb{E}\left[\left(\widehat{Q}^{\mathrm{MC}}_{h,N} - \mathbb{E}[Q]\right)^2\right] = \frac{1}{N}\mathbb{V}[Q_h] + \left(\mathbb{E}[Q_h - Q]\right)^2.$$



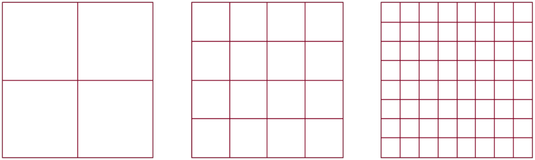
Solution: spread the approximation cost over multiple "levels":

$$\mathbb{E}[Q_{h_L}] = \mathbb{E}[Q_{h_0}] + \sum_{\ell=1}^{L} \mathbb{E}[Q_{h_\ell} - Q_{h_{\ell-1}}]$$

$$\approx \frac{1}{N_0}\sum_{i=1}^{N_0} Q_{h_0}^{(i)} + \sum_{\ell=1}^{L}\left(\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\left(Q_{h_\ell}^{(i)} - Q_{h_{\ell-1}}^{(i)}\right)\right) =:\approx \widehat{Q}_L^{\text{MLMC}},$$

where $h_\ell = 2^{-\ell} h_0$ and $N_0 > N_1 > \ldots > N_L$. This gives:

$$e\left(\widehat{Q}_L^{\text{MLMC}}\right)^2 = \sum_{\ell=0}^{L}\frac{1}{N_\ell}\mathbb{V}[Q_{h_\ell} - Q_{h_{\ell-1}}] + \left(\mathbb{E}[Q_{h_L} - Q]\right)^2.$$



Assume that:

(A1) $|\mathbb{E}[Q_h - Q]| \leq C_\alpha h^\alpha$ *(bias decay)*

(A2) $\mathbb{V}[Q_{h_\ell} - Q_{h_{\ell-1}}] \leq C_\beta h_\ell^\beta$ *(variance decay)*

(A3) $\text{Cost}(Q_h^{(i)}) \leq C_\gamma h_\ell^\gamma,$ *(cost of one sample)*

for some constants $C_\alpha, C_\beta, C_\gamma, \alpha, \beta, \gamma > 0$ with $2\alpha \geq \min(\beta, \gamma)$.

Then, there exist L and $\{N_\ell\}_{\ell=0}^L$ such that $e\left(\widehat{Q}_L^{\text{MLMC}}\right)^2 < \varepsilon^2$ and

$$\mathcal{C}\left(\widehat{Q}_L^{\text{MLMC}}\right) \lesssim \begin{cases} \mathcal{O}(\varepsilon^{-2}), & \text{if } \beta > \gamma, \\ \mathcal{O}(\varepsilon^{-2}(\log \varepsilon)^2), & \text{if } \beta = \gamma, \\ \mathcal{O}(\varepsilon^{-2-(\gamma-\beta)/\alpha}), & \text{if } \beta < \gamma. \end{cases}$$

The Monte Carlo estimator cost is $\mathcal{C}(\widehat{Q}_{h,N}^{\text{MC}}) = \mathcal{O}(\varepsilon^{-2-\frac{\gamma}{\alpha}})$.

Issue: If the random field is extremely oscillatory (small λ and ν), these fluctuations cannot be resolved on a very coarse grid.

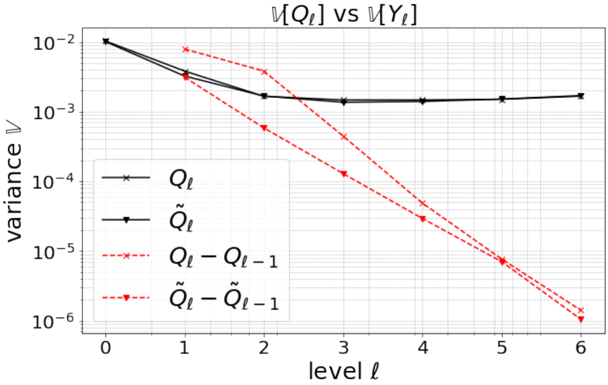


Figure: $Q = u(\mathbf{x}^*)$, $\nu = 1.5$ and $\lambda = 0.03$.

Heuristically: $h_0 \leq \lambda$ for exponential or $h_0 \leq \sqrt{8\nu}\lambda$ for Matérn.

Solution: "Smooth" samples of $k(\mathbf{x}, \omega)$ so that bulk behaviour is captured correctly, and variations are resolved more easily.

How: Drop the τ smallest eigenvalues in a given sample $\mathbf{Z} = G\Lambda^{1/2}\xi$, which correspond to the sharpest oscillations.

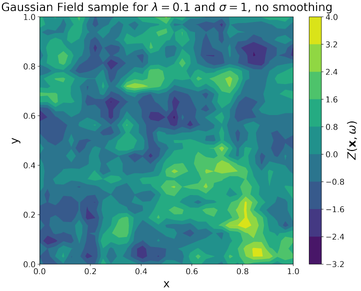


Figure: Without smoothing

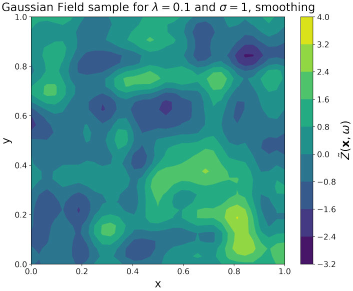


Figure: With smoothing

Smoothing
Example

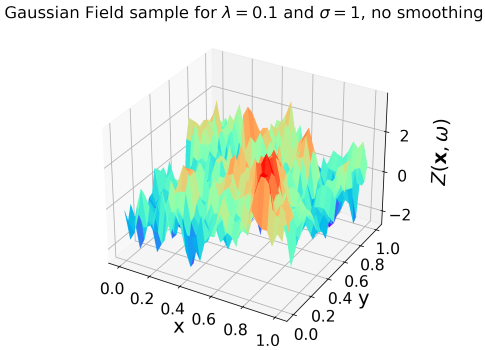


Figure: Without smoothing

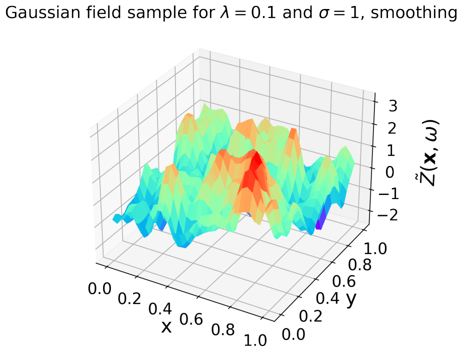


Figure: With smoothing

Smoothing

Error [I., Teckentrup (to appear in IMAJNA)]

Theorem

Let τ be the truncation index and $\tilde{\mathbf{Z}}$ be the resulting smoothed sample. Then, for any $p \in \mathbb{N}$:

$$\mathbb{E}\left[\|\mathbf{Z} - \tilde{\mathbf{Z}}\|_\infty^p\right] \lesssim s^{-\frac{p}{2}} \left(\max_{j=s-\tau+1,\dots,s} \sqrt{\lambda_j}\right)^p \tau^p,$$

where $s = \prod_{i=1}^d 2(m_i + J_i)$.

- Here, s is the dimension of the circulant matrix S.
- We have m_i mesh points in $\mathcal{T}$ in dimension i.
- J_i are "padding" values that might be necessary to ensure S is symmetric positive definite. (Not needed for separable exponential covariance.)

The previous theorem can be used to obtain convergence rates in τ of $Q_h - \tilde{Q}_h$, given convergence rates of the eigenvalues.

Theorem

Let τ be the truncation index and $\tilde{Q}_h$ be the resulting smoothed quantity of interest. Then, for the separable exponential covariance function and any $p \in [1, \infty)$:

$$\mathbb{E}\left[|Q_h - \tilde{Q}_h|^p\right] \lesssim (s - \tau + 1)^{-p}\, \tau^p,$$

where $s = \prod_{i=1}^{d} 2m_i$.

(Similar results are obtained for Matérn covariance kernels.)

- Using smoothing in multilevel Monte Carlo, we introduce a level-dependent truncation index τ_ℓ.

- Using smoothing in multilevel Monte Carlo, we introduce a level-dependent truncation index τ_ℓ.
- We choose $\tau_L = 0$, so that this strategy does not introduce additional bias in the final result.

- Using smoothing in multilevel Monte Carlo, we introduce a level-dependent truncation index τ_ℓ.
- We choose $\tau_L = 0$, so that this strategy does not introduce additional bias in the final result.
- Combining the previous theorem with an error bound on $Q - Q_h$, i.e. the finite element error, we obtain:

$$\mathbb{E}[|Q - \tilde{Q}_{h_\ell}|] \leq Ch_\ell^\alpha + C'(s_\ell - \tau_\ell + 1)^{-1}\tau_\ell.$$

- We choose τ_ℓ as a function of h_ℓ to balance the two error contributions.

- Using smoothing in multilevel Monte Carlo, we introduce a level-dependent truncation index τ_ℓ.
- We choose $\tau_L = 0$, so that this strategy does not introduce additional bias in the final result.
- Combining the previous theorem with an error bound on $Q - Q_h$, i.e. the finite element error, we obtain:

$$\mathbb{E}[|Q - \tilde{Q}_{h_\ell}|] \leq Ch_\ell^\alpha + C'(s_\ell - \tau_\ell + 1)^{-1}\tau_\ell.$$

- We choose τ_ℓ as a function of h_ℓ to balance the two error contributions.
- Note that this means that the convergence rates α and β in the multilevel Monte Carlo complexity theorem do not change.

Numerical Experiments
Computational complexity - $\lambda = 0.3$ for $Q = u(\mathbf{x}^*)$

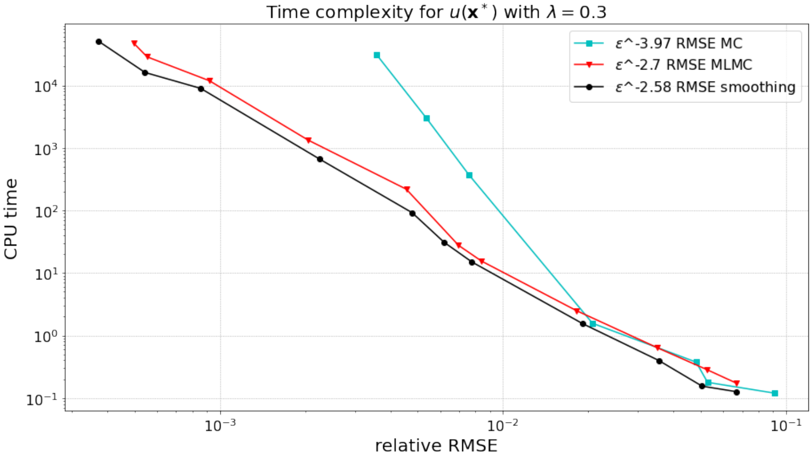


Figure: MC vs MLMC vs MLMC with smoothing for $\lambda = 0.3$.

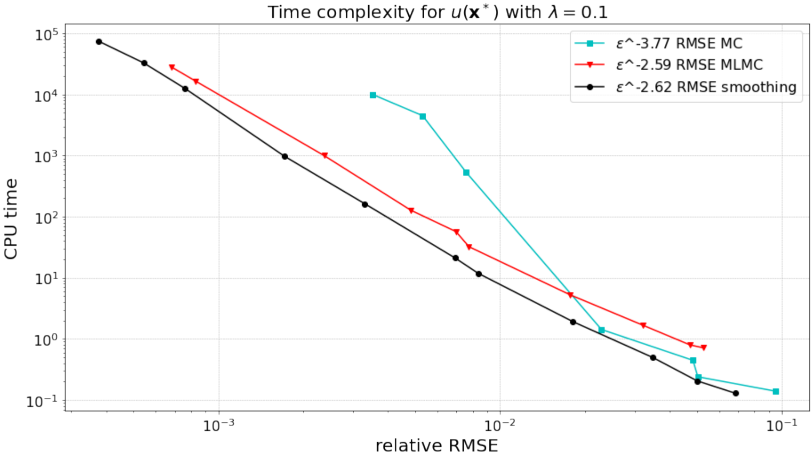


Figure: MC vs MLMC vs MLMC with smoothing for $\lambda = 0.1$.

Conclusions and Outlook

- Circulant Embedding is an efficient technique for sampling from a random field on a discrete mesh.

Conclusions and Outlook

- Circulant Embedding is an efficient technique for sampling from a random field on a discrete mesh.

$$\Downarrow$$

- Explore different approaches for sampling from $k(\mathbf{x}, \omega)$, such as SPDE-based methods.

Conclusions and Outlook

- Circulant Embedding is an efficient technique for sampling from a random field on a discrete mesh.

$$\Downarrow$$

- Explore different approaches for sampling from $k(\mathbf{x}, \omega)$, such as SPDE-based methods.

- Multilevel Monte Carlo with smoothed Circulant Embedding works well for linear elliptic PDE.

Conclusions and Outlook

- Circulant Embedding is an efficient technique for sampling from a random field on a discrete mesh.

$$\Downarrow$$

- Explore different approaches for sampling from $k(\mathbf{x}, \omega)$, such as SPDE-based methods.

- Multilevel Monte Carlo with smoothed Circulant Embedding works well for linear elliptic PDE.

$$\Downarrow$$

- Try different types of PDEs with less regularity, where correlation between levels is challenging to establish.

References

[1] Anastasia Istratuca and Aretha Teckentrup. *Smoothed Circulant Embedding with Applications to Multilevel Monte Carlo Methods for PDEs with Random Coefficients*. 2023. arXiv: 2306.13493.

[2] Markus Bachmayr et al. “Unified Analysis of Periodization-Based Sampling Methods for Matérn Covariances”. In: *SIAM Journal on Numerical Analysis* 58 (2020), pp. 2953–2980.

[3] C. R. Dietrich and G. N. Newsam. “A fast and exact method for multidimensional gaussian stochastic simulations”. en. In: *Water Resources Research* 29 (1993), pp. 2861–2869.

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

[5] I. G. Graham et al. “Quasi-Monte Carlo finite element methods for elliptic PDEs with lognormal random coefficients”. In: *Numerische Mathematik* 131 (2015), pp. 329–368.

[6] Julia Charrier. “Strong and Weak Error Estimates for Elliptic Partial Differential Equations with Random Coefficients”. en. In: *SIAM J. Numer. Anal.* 50 (2012), pp. 216–246. (Visited on 12/12/2022).

Thank you!

The eigenvalues $\{\Lambda_j\}_{j=1}^{s}$ of the matrix S are intricately linked with the eigenvalues of the covariance operator of the periodised covariance function:

$$\lambda_{\mathbf{j}}^{\text{ext}} = \int_{[-\ell,\ell]^d} C^{\text{ext}}(\mathbf{x}) \exp\left(\frac{\pi \mathrm{i} \mathbf{j} \mathbf{x}}{\ell}\right) \mathrm{d}\mathbf{x}, \quad \mathbf{j} \in \mathbf{Z}^d.$$

For the exponential covariance, we have [Charrier (2012)]:

$$\lambda_j^{\text{ext}} \lesssim j^{-2},$$

while for the Matérn covariance [Graham et al. (2015)]:

$$\lambda_j^{\text{ext}} \lesssim j^{-1-\frac{2\nu}{d}}, \quad \forall j.$$

Let the extended covariance matrix R^{ext} be given by:

$$R^{\text{ext}} = \text{Toeplitz}(C_0, C_1, \ldots, C_m, C_{m+1}, \ldots, C_{m+J})$$

where $C_m, \ldots, C_{m+J}$ are given by the periodic extension:

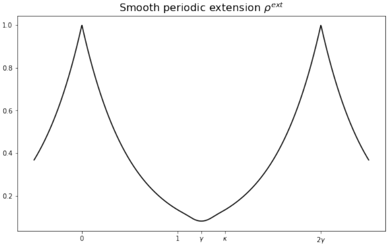
$$C^{\text{ext}}(x) = \sum_{n \in \mathbb{Z}^d} (C\varphi_\kappa)(x + 2\ell n), \quad x \in \mathbb{R}.$$

For example:

$$\varphi_\kappa(t) = \frac{\eta\left(\frac{\kappa - |t|}{\kappa - 1}\right)}{\eta\left(\frac{\kappa - |t|}{\kappa - 1}\right) + \eta\left(\frac{|t| - 1}{\kappa - 1}\right)},$$

where:

$$\eta(x) = \begin{cases} \exp(-x^{-1}), & x > 0, \\ 0, & x \leq 0. \end{cases}$$



Numerical Experiments

Smoothing error - Exponential covariance for $Q = u(\mathbf{x}^*)$

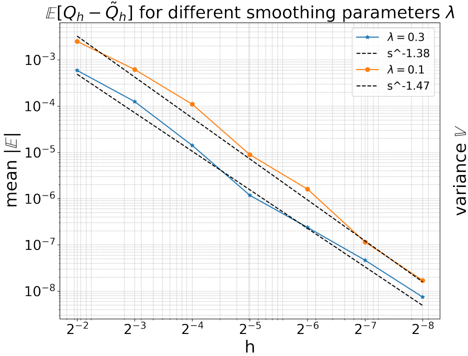


Figure: $|\mathbb{E}[Q_h - \tilde{Q}_h]|$.

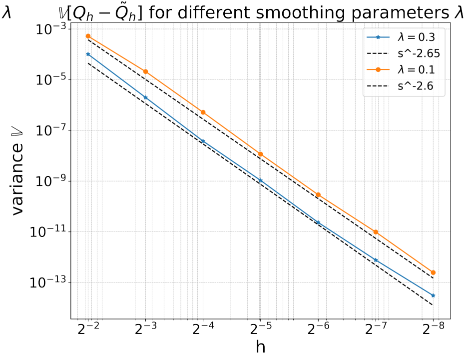


Figure: $\mathbb{V}[Q_h - \tilde{Q}_h]$.

Sketch of proof of theorem 2

$$
\begin{aligned}
\mathbb{E}\left[\left|Q_h(\cdot,\omega)-\tilde{Q}_h(\cdot,\omega))\right|^p\right] &\lesssim \mathbb{E}\left[\|u_h(\cdot,\omega)-\tilde{u}_h(\cdot,\omega)\|_{H^1(D)}^p\right] \\
&\lesssim \mathbb{E}\left[\frac{1}{\kappa(\omega)^p\tilde{\kappa}(\omega)^p}\|k(\cdot,\omega)-\tilde{k}(\cdot,\omega)\|_{C^0(\bar{D})}^p\right] \\
&\leq \mathbb{E}\left[\frac{1}{\kappa(\omega)^{pq_1}\tilde{\kappa}(\omega)^{pq_1}}\right]^{\frac{1}{q_1}} \mathbb{E}[\|k(\cdot,\omega)-\tilde{k}(\cdot,\omega)\|_{C^0(\bar{D})}^{pr_1}]^{\frac{1}{r_1}}.
\end{aligned}
$$

To show that $\mathbb{E}\left[\frac{1}{\kappa(\omega)^{pq_1}\tilde{\kappa}(\omega)^{pq_1}}\right]^{\frac{1}{q_1}}$ is bounded independent of h:

$$
\begin{aligned}
\mathbb{E}[\|\tilde{Z}_{\mathcal{T}}(\cdot,\omega)\|_{C^0(\bar{D})}^p] &\leq \mathbb{E}[(\|Z_{\mathcal{T}}(\cdot,\omega)\|_{C^0(\bar{D})}+\|Z_{\mathcal{T}}(\cdot,\omega)-\tilde{Z}_{\mathcal{T}}(\cdot,\omega)\|_{C^0(\bar{D})})^p] \\
&\leq 2^{p-1}\mathbb{E}[\|Z_{\mathcal{T}}(\cdot,\omega)\|_{C^0(\bar{D})}^p]+2^{p-1}\mathbb{E}[\|\mathbf{Z}(\omega)-\tilde{\mathbf{Z}}(\omega)\|_\infty^p].
\end{aligned}
$$

Finally:

$$
\begin{aligned}
&\mathbb{E}\left[\|\exp(Z_{\mathcal{T}}(\cdot,\omega))+\exp(\tilde{Z}_{\mathcal{T}}(\cdot,\omega))\|_{C^0(\bar{D})}^{pr_1q_2}\right]^{\frac{1}{r_1q_2}} \\
&\leq 2^{\frac{pr_1q_2-1}{r1q_2}}\left(\mathbb{E}\left[\exp(\|Z_{\mathcal{T}}(\cdot,\omega)\|)_{C^0(\bar{D})}^{pr_1q_2}\right]^{\frac{1}{r_1q_2}}+\mathbb{E}\left[\exp(\|\tilde{Z}_{\mathcal{T}}(\cdot,\omega)\|)_{C^0(\bar{D})}^{pr_1q_2}\right]^{\frac{1}{r_1q_2}}\right).
\end{aligned}
$$

Numerical Experiments
KL vs CE comparison I - $\lambda = 0.1$ for $Q = u(\mathbf{x}^*)$

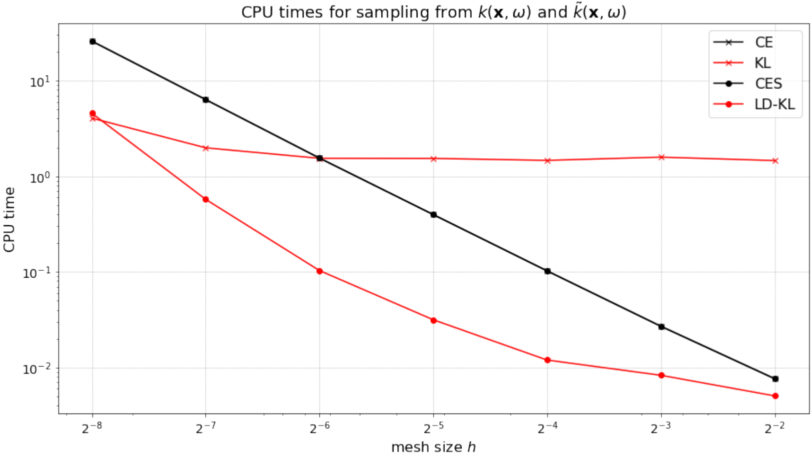


Figure: Cost per sample (no PDE solve).

Numerical Experiments

KL vs CE comparison II - $\lambda = 0.1$ for $Q = u(\mathbf{x}^*)$

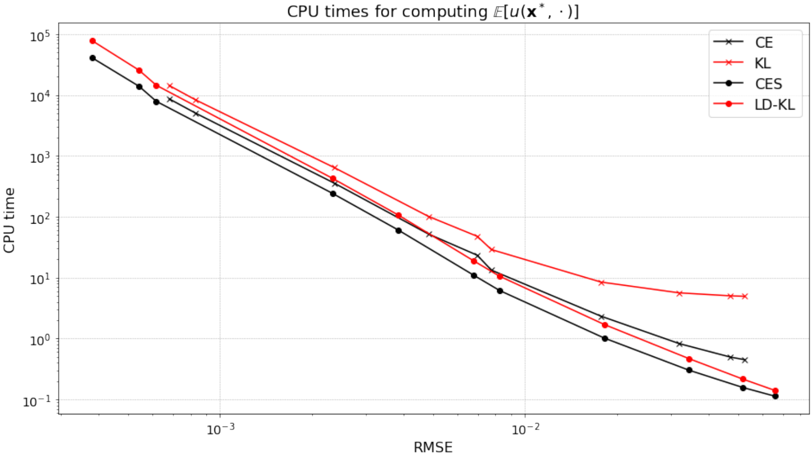


Figure: MLMC cost.